

Sensitivity Analysis in Population Shares*

Filippo Palomba[§]
Princeton University

October 21, 2026

Latest version available [here](#).

Empirical work often relies on assumptions on the distribution of random variables such as potential outcomes or latent variables to identify some parameter of interest. We introduce a unified framework for assessing the sensitivity of a target estimand when these distributional assumptions are relaxed for a share of the population, which serves as our sensitivity parameter. We derive bounds on the values the estimand can take as this share varies between 0 and 1 and compute them by means of tractable finite-dimensional programs. We propose plug-in estimators for the bounds and characterize their large sample properties. We also show that our bounds recover exact nonparametric sharp bounds at no additional computational cost, and introduce summary measures that separate misspecification uncertainty from sampling uncertainty and rank studies by their robustness. We apply our framework to assess the robustness of (i) counterfactual predictions in structural models, (ii) average treatment effects in randomized experiments with attrition, and (iii) formula instruments. We provide **R** and **Python** packages implementing all our methods.

Keywords: sensitivity analysis, robustness, partial identification, convex duality.

*I am thankful to my committee, Matias Cattaneo, Kevin Dano, David Lee, and Ulrich Müller, for their guidance, patience, and encouragement throughout this project. I also thank Thomas Bearpark, Francesco Bilotta, Giampaolo Bonomi, Andrea Cerrato, Federico Crippa, Carlo Cusumano, Pietro Dedola, Nicolas Hommel, Bo Honoré, Michal Kolesár, Elena Manresa, Matt Masten, Geert Mesters, Mikkel Plagborg-Møller, Chiara Motta, Frank Schorfheide, and Bartolomeo Stellato for comments that made the paper better, and Claude for giving me back a lot of hours. All remaining mistakes are obviously mine. All methods are implemented in the companion **R** and **Python** packages **tvbounds** (<https://github.com/filippopalomba/tvbounds>).

[§]Department of Economics. fpalomba@princeton.edu.

1 Introduction

A long tradition has stressed the need for sensitivity analyses to assess the fragility of a study's conclusions (Leamer, 1983; Rosenbaum and Rubin, 1983; Leamer, 1985; Tamer, 2015). Consider a policymaker deciding which of several structural models or natural experiments should inform policy, a manager allocating a finite budget across pilot studies, or a drug regulator weighing the evidence from competing trials. These alternative empirical analyses report point estimates and confidence intervals for a common target parameter, but all rest on untestable, sometimes not even falsifiable, distributional assumptions: that attrition in a randomized experiment is random, that unobserved preference shocks in an indirect utility follow an extreme value distribution, or that the exogenous shocks in a formula instrument are uniformly distributed. When the studies disagree, the decision maker needs to know how much of the disagreement reflects misspecification rather than sampling uncertainty, and two questions arise. Can we quantify how sensitive each conclusion is to distributional assumptions, so that studies can be ranked? And can we do so in units interpretable enough to communicate the ranking to a broad audience of stakeholders?

This paper answers both questions with a framework that assesses the sensitivity of a target estimand to violations of distributional assumptions and measures the extent of a violation by the share of the population that does not satisfy such assumptions. The framework gauges, for instance, how much an average treatment effect would change if nonresponse were nonrandom for 60% of the experimental treated units, how much a counterfactual welfare prediction from a structural model would change if 30% of the agents drew shocks that were not extreme value distributed, or how much an estimated elasticity would change if 40% of the time the exogenous shocks were assigned by a mechanism other than the one the researcher postulated.

Formally, the object of sensitivity analysis is an expectation $\mathbb{E}_{P_*}[g(Z; \theta)]$, where g is a known function, Z is a random vector, θ is a structural parameter disciplined by moment equalities or inequalities, and P_* is the *baseline distribution*, the model the econometrician would use absent any concern for misspecification. In this scenario, our goal is to obtain the smallest and largest values the estimand can take over a nonparametric neighborhood of candidate distributions that allow up to a fraction δ of the population to violate the distributional assumption imposed by P_* . We refer to these two values as the *sensitivity bounds* at δ , to the neighborhood as the *robustness set*, and to δ as the *sensitivity parameter*. In a randomized experiment with attrition, δ bounds the share of treated respondents for whom attrition is not random. In a discrete choice model, it bounds the share of agents whose taste shocks do not follow the postulated distribution. With formula instruments, it bounds the probability that the shocks were assigned by a different mechanism.

We show that the sensitivity bounds are the optimal values of two symmetric infinite-dimensional programs, which in general admit no closed form solution. Using tools from convex analysis (Rockafellar, 1970; Rockafellar and Wets, 2009; Kuhn, Shafiee and Wiesemann, 2025), we recast them as finite-dimensional convex programs, so that the entire path of bounds can be traced as δ ranges between 0 and 1. At $\delta = 0$, the bounds reduce to the value of the estimand under the baseline specification P_* . At $\delta = 1$, we prove that they coincide with the sharp nonparametric bounds,

namely the tightest bounds on the estimand implied by the model’s remaining restrictions once the distributional assumption is dropped entirely. As a consequence, our procedure can also be used to compute nonparametric sharp bounds directly from a tractable program whose dimensionality grows only with the number of moment conditions imposed. Put differently, δ trades off informativeness against credibility, in the spirit of *the law of decreasing credibility* of Manski (2003). At $\delta = 0$, the estimand is point identified, but only under the full distributional assumption. At $\delta = 1$, the bounds rest on the model’s remaining restrictions alone, but they are at their widest.

Regarding estimation and inference, we propose simple plug-in estimators for the bounds. We allow for both a known baseline, as when P_* is the distribution of taste shocks in a discrete choice model, and an estimated one, as when P_* is the outcome distribution of the treated respondents in a randomized experiment. We establish the large sample properties of the estimators by showing their consistency, deriving their limiting distribution, and proposing a bootstrap procedure to construct valid confidence intervals for the bounds.

We also propose summary measures that separate misspecification uncertainty from sampling uncertainty and make studies comparable. Together with the bounds, they answer the two questions raised in our opening paragraphs. Finally, we apply the framework to the online tutoring experiment of Carlana and La Ferrara (2025), in which attrition threatens identification, to the marital college premium application of Christensen and Connault (2023), and to the formula instruments of Borusyak and Hull (2023), who estimate the employment effects of market access growth from Chinese high-speed rail construction.

Related literature. Some recent work studies the sensitivity of counterfactuals in structural models. In Christensen and Connault (2023), the values a counterfactual takes are bounded as the distribution of latent variables ranges over a ϕ -divergence neighborhood of a parametric specification while the structural model’s moment restrictions are maintained. Gu and Russell (2024) bound scalar counterfactuals in partially identified structural models over Wasserstein neighborhoods of the latent distribution using random set theory, and Lieber, Tebaldi and Torgovitsky (2026) bound counterfactuals in nonparametric discrete choice models when the latent density lies in a pointwise band around a known reference density, using tools from graph theory and mechanism design. This paper differs from these works mainly in that our sensitivity parameter δ admits a population share interpretation and our framework applies beyond the structural models these papers consider.

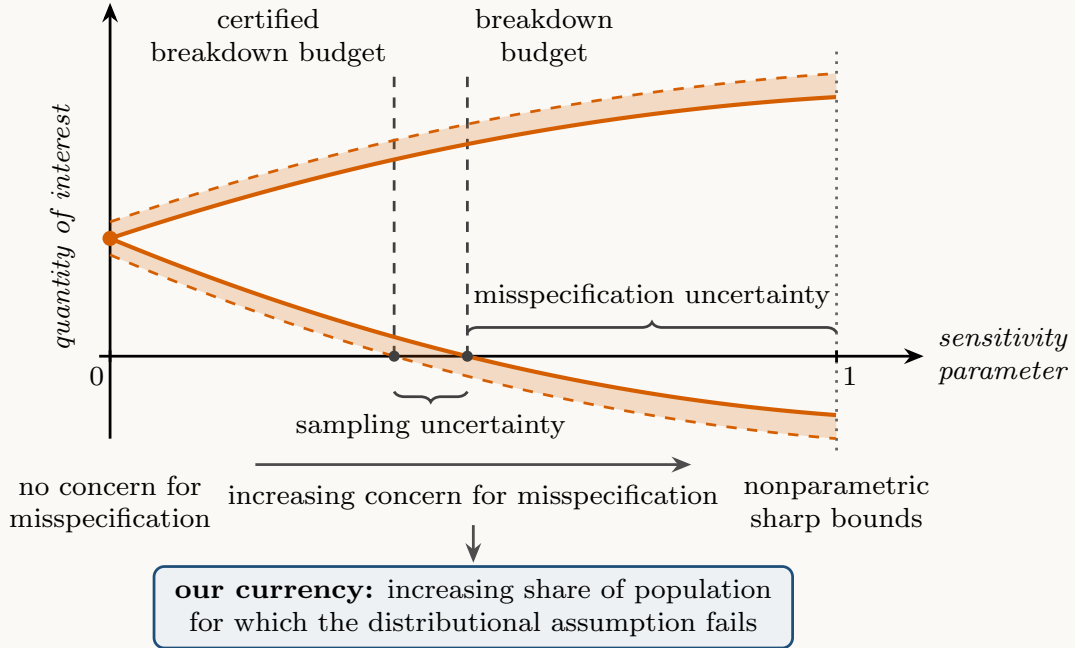
Our framework is also related to a large body of work that develops sensitivity exercises within specific identification problems: contaminated and corrupted sampling (Horowitz and Manski, 1995); missing data and attrition (Kline and Santos, 2013; Ober-Reynolds, 2026); selection on unobservables and unmeasured confounding (Rosenbaum and Rubin, 1983; Imbens, 2003; Altonji, Elder and Taber, 2005; Masten and Poirier, 2018; Bonvini and Kennedy, 2022; Masten, Poirier and Zhang, 2024; Masten, Poirier and Ren, 2025); the exclusion and exogeneity of instruments (Masten and Poirier, 2021; Diegert, Masten and Poirier, 2026); the distribution of the exogenous shocks that generate formula instruments (Sheng and Torgovitsky, 2026); random assignment and rank invariance (Masten and Poirier, 2020); and parallel trends in difference-in-differences designs (Rambachan and Roth, 2023). Relative to these papers, we provide a single framework that nests

many of the exercises described above, can include moment restrictions and structural parameters, provides a general estimation and inference theory, and allows the sensitivity parameter to maintain an analogous population share interpretation for a wide family of settings. In this sense, our sensitivity parameter is a *common currency* for misspecification.

Another strand of literature focused on local sensitivity analysis, which measures how quantities of interest respond to infinitesimal perturbations of the model or of its moment conditions (Kitamura, Otsu and Evdokimov, 2013; Andrews, Gentzkow and Shapiro, 2017; Mukhin, 2018; Andrews, Gentzkow and Shapiro, 2020; Armstrong and Kolesár, 2021; Bonhomme and Weidner, 2022), whereas we allow for global misspecification, in that δ can be large and does not shrink as the sample size grows. We show how the two approaches can be linked via the shadow price of robustness, the derivative of the sensitivity bounds with respect to δ , which is a local sensitivity measure at small budgets.

Finally, on the methodological side, we contribute to a line of work in economics that uses convex duality to turn optimization problems over infinite-dimensional sets of distributions into tractable ones (Hansen and Sargent, 2001; Ekeland et al., 2010; Schennach, 2014). Closest to our work, Christensen and Connault (2023) use duality to recast the optimization of a counterfactual over a ϕ -divergence neighborhood, subject to moment restrictions, as a finite-dimensional convex program. Their duality results are stated for divergences generated by strictly convex and differentiable functions, a class that excludes the distance on which our neighborhoods are built.

Figure 1: *Sensitivity analysis in population shares: a visual roadmap.*



Paper organization. Figure 1 collects the main output of a prototypical sensitivity exercise conducted with our framework and organizes the rest of the paper. Section 2 presents three leading applications that motivate our sensitivity exercise. Section 3 formally defines the framework nesting these applications and the optimization problems behind the thick lines, and explains why moving

rightward on the horizontal axis raises the share of the population for which the distributional assumption imposed by P_* is allowed to fail. Section 4 shows how to compute, estimate, and conduct inference on the sensitivity bounds, and thereby delivers the thick lines and the dashed bands around them. Section 5 proves that at $\delta = 1$ the bounds coincide with the exact nonparametric sharp bounds and discusses how to use them in practice by means of summary measures, such as the breakdown and certified breakdown budgets of Figure 1. The breakdown budget is the smallest share of the population for which the sensitivity bounds contain zero, whereas the certified breakdown budget, which also accounts for sampling uncertainty, is the smallest share for which the dashed bands do. Section 6 deploys the framework in the three empirical applications introduced in Section 2, and Section 7 concludes. The supplementary materials collect all proofs, establish the existence of solutions to the dual and outer problems, relax the simplifying conditions of Section 4, allow for moment inequalities and nuisance parameters, and work out further applications.

Notation. For a set $\mathcal{A}$, $\Delta(\mathcal{A})$ denotes the set of probability distributions on $\mathcal{A}$ and, for $P, Q \in \Delta(\mathcal{A})$, $\text{TV}(P \parallel Q) := \sup_{A \subseteq \mathcal{A}} |P(A) - Q(A)| \in [0, 1]$ denotes the total variation distance between them. We write $P \ll Q$ when P is absolutely continuous with respect to Q , dP/dQ for the corresponding density, $\text{supp}(P)$ for the support of P , $Pf := \mathbb{E}_P[f(Z)]$ for a random vector Z distributed according to P and a P -integrable function f , and $P\text{-esssup}_z f(z)$ for the smallest $c \in [-\infty, +\infty]$ such that $f \leq c$ holds P -almost surely. We let $[n] := \{1, \dots, n\}$ for $n \in \mathbb{N}$, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, $\mathbb{R}_+ := [0, \infty)$, 0_d be the d -dimensional zero vector, $\mathbf{1}\{\cdot\}$ the indicator function, $x^\top$ the transpose of x , and $\|\cdot\|$ the Euclidean norm. For a convex set $C \subseteq \mathbb{R}^d$, $\text{int}(C)$ and $\text{ri}(C)$ denote its interior and relative interior. Finally, $\rightsquigarrow$ denotes weak convergence, $o_{\mathbb{P}}(1)$ a term converging to zero in probability, and $\mathcal{N}(0, \Sigma)$ the Gaussian distribution with covariance matrix Σ .

2 Motivating Examples

In this section, we illustrate how three common economic applications rest on untestable distributional assumptions and are thus amenable to sensitivity analysis: a randomized controlled trial (RCT, henceforth) with attrition, counterfactuals in structural models following Christensen and Connault (2023), and the recentered instrumental variables estimator of Borusyak and Hull (2023). Many economically relevant scenarios beyond these three fit the framework (see Section SA6 of the supplementary materials).

Example 2.1 (RCTs with attrition). Consider a randomized experiment that grants half of the applicants access to a free job training course and then, two years later, runs a follow-up survey to record each applicant’s weekly wage.¹ Wages, however, are observed only for the applicants who follow up. Worse, trainees stay in touch with the agency while the control-group applicants are the hardest to track down, so respondents may end up being selected on particular dimensions, confounding the effects of the program. Practitioners typically assume attrition is ignorable or,

¹A famous program of this kind is Job Corps, which offers vocational training to disadvantaged youth aged 16–24 and remains active today with more than 120 campuses across the United States and Puerto Rico, roughly 25,000 students enrolled at a time, and a budget of about \$1.8 billion for fiscal year 2026 (U.S. Congress, 2026).

equivalently, that nonresponse is as good as random, an assumption known as “missing completely at random” (MCAR, henceforth) that we define precisely below.²

Formally, consider a population described by $(Y(0), Y(1), S(0), S(1), D) \sim P_0$, where $D \in \{0, 1\}$ is a binary treatment, $Y(d) \in \mathcal{Y}$, $d \in \{0, 1\}$, are potential outcomes taking values in a closed set $\mathcal{Y} \subseteq \mathbb{R}$, $S(d) \in \{0, 1\}$, $d \in \{0, 1\}$, are potential response indicators, and P_0 satisfies random assignment of D , that is $D \perp\!\!\!\perp (S(0), S(1), Y(0), Y(1))$. Observed and potential quantities are linked by $Y = DY(1) + (1 - D)Y(0)$ and $S = DS(1) + (1 - D)S(0)$. The econometrician observes (YS, S, D) for every applicant, as the outcome Y is observed only for respondents, that is, when $S = 1$. The pair of potential response indicators partitions the population into four groups:

$$G := \begin{cases} \text{AO} & \text{if } (S(0), S(1)) = (1, 1), \\ \text{C} & \text{if } (S(0), S(1)) = (0, 1), \\ \text{NO} & \text{if } (S(0), S(1)) = (0, 0), \\ \text{DF} & \text{if } (S(0), S(1)) = (1, 0), \end{cases}$$

namely the *always-observed*, who respond regardless of the arm they are assigned to; the *compliers*, who respond only if treated; the *never-observed*, who drop out of the sample regardless of the arm; and the *defiers*, who respond only if untreated.³ As a parameter of interest, we consider the average treatment effect (ATE, henceforth) for the always-observed,⁴ which is defined as

$$\tau_0 := \mathbb{E}_{P_0}[Y(1) - Y(0) \mid G = \text{AO}].$$

For simplicity, we assume throughout that treatment weakly encourages response, i.e., $S(1) \geq S(0)$ almost surely (**Mono**), an assumption known as “monotonicity,” which in this notation reads $P_0[G = \text{DF}] = 0$. In Remark SA-15 of the supplementary materials we explain how to conduct the same sensitivity exercise without **Mono**. Absent defiers, the respondents in the control arm are exactly the always-observed, hence

$$\mu^{\text{AO}}(0) := \mathbb{E}_{P_0}[Y(0) \mid G = \text{AO}] = \mathbb{E}_{P_0}[Y \mid D = 0, S = 1] \quad (2.1)$$

is nonparametrically identified, and we can focus on the treated arm.

²Perhaps surprisingly, this practice is the norm rather than the exception. In a systematic review of survey-based articles published in the top-five economics journals, Dutz, Huitfeldt, Lacouture, Mogstad, Torgovitsky and van Dijk (2026) find that nonresponse rates average 50% when researchers collect their own data and 19% when using pre-existing household surveys. Despite that, nearly half of the articles do not discuss nonresponse bias at all, while the remaining articles either assert that nonresponse is ignorable or correct for it by reweighting on observables. Linking their own survey to administrative records on the full population, they document substantial nonresponse bias that survives adjustment for observables, even in a survey run by a national statistical agency. There is further empirical evidence in economics that agents, when given the choice of whether to report, make that decision on the basis of observable or unobservable characteristics (Mu, Rubin and Zou, 2026; Bollinger, Hirsch, Hokayem and Ziliak, 2019; Crippa, Berg and Rigobon, 2025).

³This taxonomy deliberately mirrors the imperfect compliance taxonomy of Angrist, Imbens and Rubin (1996), with response to the follow-up survey playing the role of treatment take-up.

⁴Many more causal parameters are identified under MCAR. We focus on this particular one because it is often the object of interest of the attrition literature and because it can be nonparametrically partially identified without bounded $\mathcal{Y}$ when MCAR does not hold. We show how to conduct sensitivity analysis when the ATE over the whole population is the target estimand in Section SA6.4.3 of the supplementary materials.

In the treated arm, the data identify the conditional distribution of Y given $D = 1$ and $S = 1$ under P_0 , which we denote by P_T and whose mean we denote by $\mu_T := \mathbb{E}_{P_T}[Y]$, but not who among the treated respondents is a complier and who an always-observed. Let P_C and P_{AO} denote the conditional distributions of $Y(1)$ given $G = C$ and given $G = AO$ under P_0 , respectively, and let

$$\pi := P_0[G = C \mid D = 1, S = 1] = \frac{P_0[S = 1 \mid D = 1] - P_0[S = 1 \mid D = 0]}{P_0[S = 1 \mid D = 1]} \quad (2.2)$$

be the complier share among treated respondents, where the second equality follows from random assignment and **Mono**, so that π is identified by comparing the response rates of the two arms. We assume that $P_0[S = 1 \mid D = 0] > 0$, that treatment strictly raises the response rate, and that $\mathbb{E}_{P_0}[|Y(d)|] < \infty$, $d \in \{0, 1\}$, so that $\pi \in (0, 1)$ and the means above are finite. Since $S(1) = 1$ if and only if $G \in \{AO, C\}$, random assignment yields the mixture decomposition

$$P_T = \pi P_C + (1 - \pi) P_{AO}, \quad (2.3)$$

in which neither component is identified. The decomposition (2.3) is key for two reasons. First, **MCAR**, that is, the assumption that response is independent of potential outcomes, $G \perp\!\!\!\perp (Y(0), Y(1))$, implies that the distribution of $Y(1)$ is the same for compliers and always-observed, so that in the treated arm **MCAR** amounts to

$$P_C = P_{AO} = P_T, \quad (2.4)$$

which identifies the unobserved complier distribution with the observed distribution of treated respondents. Second, (2.3) ties the always-observed to the compliers: for any candidate complier distribution $Q \in \Delta(\mathcal{Y})$, the mixture identity forces the always-observed distribution to be $(P_T - \pi Q)/(1 - \pi)$, whose mean is $(\mu_T - \pi \mathbb{E}_Q[Y])/(1 - \pi)$. Subtracting (2.1), we obtain the map

$$\tau(Q) := \frac{\mu_T - \pi \mathbb{E}_Q[Y]}{1 - \pi} - \mu^{AO}(0), \quad Q \in \Delta(\mathcal{Y}), \quad (2.5)$$

in which $\mathbb{E}_Q[Y]$ is the only quantity not identified from the data. At the true complier distribution the map recovers the parameter of interest, $\tau(P_C) = \tau_0$. Under **MCAR**, (2.4) gives $\tau_0 = \tau(P_T)$, so that τ_0 is point identified by the difference in mean outcomes among respondents (Rubin, 1976),

$$\tau_{\text{MCAR}} := \tau(P_T) = \mathbb{E}_{P_0}[Y \mid D = 1, S = 1] - \mathbb{E}_{P_0}[Y \mid D = 0, S = 1].$$

However, **MCAR** is untestable and hard to justify by economic theory. Hence, a stakeholder might be interested in quantifying how far τ_0 can lie from τ_{MCAR} when **MCAR** fails.

The way we model relaxations of **MCAR** is by requiring that the outcome distributions of compliers and always-observed be close rather than equal. Put differently, we consider candidate pairs (Q_C, Q_{AO}) of distributions on $\mathcal{Y}$ that satisfy $P_T = \pi Q_C + (1 - \pi) Q_{AO}$ and $\text{TV}(Q_C \parallel Q_{AO}) \leq \delta$. As we show in Section 3, with this formulation, δ is the share of treated respondents or, equivalently, the share of compliers who may not satisfy **MCAR**.

The pair (Q_C, Q_{AO}) is fully described by its first component. Indeed, the mixture constraint (3.3) gives that $Q_{AO} = (P_T - \pi Q_C)/(1 - \pi)$ and that $\text{TV}(Q_C \parallel P_T) \leq (1 - \pi)\delta$.⁵ This reformulation

⁵The equivalence follows from the identity $\text{TV}(Q_C \parallel P_T) = (1 - \pi)\text{TV}(Q_C \parallel Q_{AO})$, which total variation inherits from being a norm-induced metric and which ϕ -divergences do not enjoy in general.

replaces the unidentified reference distribution Q_{AO} by the identified P_{T} , and lets us define the set of candidate complier distributions as

$$\begin{aligned} \mathcal{P}_{\text{RCT}}(\delta) &:= \{Q \in \Delta(\mathcal{Y}) : \text{TV}(Q \parallel P_{\text{T}}) \leq (1 - \pi)\delta, \\ &\quad P_{\text{T}} = \pi Q + (1 - \pi)Q_{\text{AO}} \text{ for some } Q_{\text{AO}}\}. \end{aligned} \quad (2.6)$$

The sensitivity bounds are then the extreme values that the causal estimand in (2.5) can take over the robustness set defined in (2.6). Formally, for $\delta \in [0, 1]$

$$\underline{\tau}(\delta) := \inf_{Q \in \mathcal{P}_{\text{RCT}}(\delta)} \tau(Q) \quad \text{and} \quad \bar{\tau}(\delta) := \sup_{Q \in \mathcal{P}_{\text{RCT}}(\delta)} \tau(Q).$$

◆

Example 2.2 (Counterfactuals in structural models). Counterfactual predictions from structural models depend on the distribution specified for latent variables such as taste shocks, unobserved heterogeneity, or productivity shocks, which is typically not nonparametrically identifiable and is therefore chosen for tractability. A stakeholder may want to know how sensitive the predictions are to this untestable choice. As a running example, we revisit one of the leading empirical applications of Christensen and Connault (2023) (CC23, henceforth): the marital college premium in the marriage-matching model that Chiappori, Salanié and Weiss (2017) estimate on American Community Survey data.

Consider a frictionless one-to-one matching model with transferable utility. Men and women each belong to one of $J \in \mathbb{N}$ education groups, indexed by $[J] := \{1, \dots, J\}$, which we term “types.” We let $j = 0$ denote the outside option of remaining unmatched. A type- $b \in [J]$ woman receives utility ν_{0b} if unmatched and $t_{jb} + \nu_{jb}$ if matched with a type- $j \in [J]$ man, where t_{jb} is the systematic component of match utility and ν_{jb} captures idiosyncratic marital preferences. For simplicity, we assume that the distribution of females’ preference shocks does not depend on their type, so we drop the b subscript and consider a single random vector $U = (\nu_0, \dots, \nu_J)^\top$.⁶ We let $\tilde{P}$ denote the distribution of latent preference shocks $(\nu_0, \dots, \nu_J)$ postulated by the econometrician, which we assume to be a Gumbel (type I extreme value) distribution.⁷

We fix two female types $b, b' \in [J]$, whose marital college premium is the object of interest, and let $\varrho \in \mathbb{R}^{2J}$ collect their observed marriage probabilities with each male type. The model pins down the structural parameter $\vartheta := (t_{jb}, t_{jb'}, j \in [J])$, ranging over a parameter space $\Theta \subseteq \mathbb{R}^{2J}$, via a set of moment equalities

$$\mathbb{E}_{\tilde{P}}[s(U; \vartheta)] = (\varrho, 0_{2(J+1)}), \quad s(U; \vartheta) = (s_1(U; \vartheta)^\top, s_2(U; \vartheta)^\top)^\top,$$

where $s(U; \vartheta)$ is the moment function, which we now describe in detail. First,

$$s_1(U; \vartheta) = \begin{bmatrix} \left(\mathbb{1} \left\{ j = \arg \max_{j' \in \{0, \dots, J\}} (t_{j'b} + \nu_{j'}) \right\} \right)_{j=1}^J \\ \left(\mathbb{1} \left\{ j = \arg \max_{j' \in \{0, \dots, J\}} (t_{j'b'} + \nu_{j'}) \right\} \right)_{j=1}^J \end{bmatrix}.$$

⁶See Section SA6.2, and Remark SA-16 in particular, in the supplementary materials for a discussion of this common-shock representation and its alternatives.

⁷Another common choice is to assume that $\tilde{P}$ is a standard Gaussian distribution, which gives a probit instead of a logit discrete choice model.

Ties are broken toward the lowest indexed alternative. Therefore, the moment conditions $\mathbb{E}_{\tilde{P}}[s_1(U; \vartheta)] = \varrho$ equate the observed marriage probabilities ϱ for female types b and b' to the model implied choice probabilities under $\tilde{P}$. Note that because the moment function must be a known function of (U, ϑ) alone, the observed marriage probabilities ϱ enter as the target of the moment conditions rather than as an argument of s_1 .

Second, one may impose some normalizations on the female shocks through

$$s_2(U; \vartheta) = (\nu_j, \nu_j^2 - \pi^2/6)_{j=0}^J.$$

Therefore, the moment conditions $\mathbb{E}_{\tilde{P}}[s_2(U; \vartheta)] = 0_{2(J+1)}$ impose that each shock must have mean zero and variance $\pi^2/6$ under $\tilde{P}$, the same variance as a standard Gumbel distribution.

Consider the marital college premium (i.e., the difference in indirect utilities) for female types b and b' ,

$$k(\vartheta; \tilde{P}) := \mathbb{E}_{\tilde{P}}[f(U; \vartheta)], \quad f(U; \vartheta) = \max_{j \in \{0, \dots, J\}} \{t_{jb} + \nu_j\} - \max_{j \in \{0, \dots, J\}} \{t_{jb'} + \nu_j\},$$

where $t_{0b} = t_{0b'} = 0$. A stakeholder interested in the output of the model above may ask what happens to the premium if not all women draw their shocks from a Gumbel distribution. To answer this question, we let P range over the robustness set

$$\begin{aligned} \mathcal{P}_{\text{STR}}(\vartheta; \delta) &:= \{P \in \Delta(\mathbb{R}^{J+1}) : \mathbb{E}_P[s(U; \vartheta)] = (\varrho, 0_{2(J+1)}), \\ &P = (1 - \delta)\tilde{P} + \delta R, \quad R \in \Delta(\mathbb{R}^{J+1})\}, \end{aligned} \quad (2.7)$$

which collects all the populations of women that (i) reproduce the observed marriage probabilities and satisfy the imposed normalizations and (ii) consist of a share $\delta \in [0, 1]$ of women with arbitrary shocks and a share $1 - \delta$ of women with Gumbel shocks. The sensitivity bounds on the premium are then

$$\underline{k}(\delta) := \inf_{\vartheta \in \Theta} \inf_{P \in \mathcal{P}_{\text{STR}}(\vartheta; \delta)} \mathbb{E}_P[f(U; \vartheta)], \quad \bar{k}(\delta) := \sup_{\vartheta \in \Theta} \sup_{P \in \mathcal{P}_{\text{STR}}(\vartheta; \delta)} \mathbb{E}_P[f(U; \vartheta)],$$

where the inner infimum and supremum are set to $+\infty$ and $-\infty$, respectively, whenever $\mathcal{P}_{\text{STR}}(\vartheta; \delta)$ is empty. Therefore, the values of ϑ at which no distribution within δ of the Gumbel baseline matches the observed marriage probabilities are infeasible. $\blacklozenge$

Example 2.3 (Recentered IV). Many causal questions in economics involve treatments or instruments that combine multiple sources of variation according to a known formula. Examples include shift-share (or Bartik) instruments, treatments capturing spillovers in social or transportation networks, or simulated instruments for policy eligibility (see Borusyak, Hull and Jaravel, 2025, for a review). Borusyak and Hull (2023) (BH23, henceforth) point out that exogeneity of the shocks entering the formula does not, by itself, deliver a consistent instrumental variables estimator, because units are nonrandomly exposed to the shocks. The solution BH23 propose is to recenter the instrument, subtracting from it the *expected instrument*, that is, the average instrument across counterfactual shocks. Computing the expected instrument requires the researcher to specify the

distribution of the shocks, or “shock assignment process.” In a randomized experiment the assignment process is given by the randomization protocol, but in natural experiments the assignment process is rarely known or justified by economic theory. A stakeholder may accordingly ask how much the recentered IV estimate would change under a different assignment process.

Formally, let $i \in [n], n \in \mathbb{N}$, index the units of the sample and consider the causal or structural model

$$y_i = \beta x_i + \varepsilon_i,$$

where $y_i \in \mathbb{R}$ is a demeaned outcome, $x_i \in \mathbb{R}$ is a demeaned endogenous regressor, $\varepsilon_i \in \mathbb{R}$ is an unobserved residual, and $\beta \in \mathbb{R}$ is the parameter of interest. Let $v = (v_1, \dots, v_K) \in \mathcal{V} \subseteq \mathbb{R}^K$, $K \in \mathbb{N}$, be a vector of exogenous shocks, let w collect predetermined covariates of unrestricted dimension, and let

$$z_i := f_i(v; w)$$

be the candidate instrument of unit $i \in [n]$, where $f_i(\cdot; w) : \mathcal{V} \rightarrow \mathbb{R}$ are known, bounded, non-stochastic functions.⁸ BH23 assume that the shocks are exogenous, that is,

$$v \perp\!\!\!\perp (\varepsilon_1, \dots, \varepsilon_n) \mid w, \quad (2.8)$$

and show that this alone does not make z_i orthogonal to ε_i , because units are nonrandomly exposed to the shocks through the formula.

Letting $\tilde{P} \in \Delta(\mathcal{V})$ denote the assignment process postulated by the econometrician, the expected instruments are $\mu_i(\tilde{P}) := \mathbb{E}_{\tilde{P}}[f_i(v; w)]$ and the recentered instruments are $\tilde{z}_i := z_i - \mu_i(\tilde{P})$, $i \in [n]$.⁹ The recentered IV estimator of BH23 is then

$$\hat{\beta}_* := \frac{\sum_{i=1}^n \tilde{z}_i y_i}{\sum_{i=1}^n \tilde{z}_i x_i}. \quad (2.9)$$

The setup also covers the case in which the treatment itself is generated by the formula, so that $z_i = x_i$ as in the empirical application of BH23 that we revisit in Section 6.3.

Our sensitivity exercise maintains exogeneity of the shocks in (2.8) and relaxes the assumption that their assignment process is known and equal to $\tilde{P}$. For this exercise, we hold the realized sample (outcomes, treatments, covariates, and the realized shocks $\check{v}$) fixed, so that the recentered estimate is a function of the assignment process alone. Because (2.9) depends on $\tilde{P}$ only through the expected instruments, replacing the postulated assignment process by a candidate one $P \in \Delta(\mathcal{V})$ produces the estimate

$$\hat{\beta}(P) := \frac{G_y(P)}{G_x(P)}, \quad (2.10)$$

⁸Boundedness holds in the empirical application of Section 6.3, where $\mathcal{V}$ is a finite set.

⁹We assume for simplicity that the distribution of $v \mid w$ does not depend on w , as is the case for the application we consider in Section 6.3.

where

$$G_y(P) := g_y(\check{v}) - \mathbb{E}_P[g_y(v)], \quad g_y(\cdot) := \sum_{i=1}^n y_i f_i(\cdot; w),$$

$$G_x(P) := g_x(\check{v}) - \mathbb{E}_P[g_x(v)], \quad g_x(\cdot) := \sum_{i=1}^n x_i f_i(\cdot; w).$$

By construction $\hat{\beta}(\tilde{P}) = \hat{\beta}_*$, so the exercise is centered at the reported estimate. For any $\delta \in [0, 1]$, we let the candidate assignment process P range over

$$\mathcal{P}_{\text{FI}}(\delta) := \{P \in \Delta(\mathcal{V}) : P = (1 - \delta)\tilde{P} + \delta R \text{ for some } R \in \Delta(\mathcal{V})\}. \quad (2.11)$$

Thus, up to a share δ of the shock draws may be assigned by a mechanism other than the one the researcher postulated. Throughout the exercise we maintain that the recentered first stage is uniformly bounded away from zero over the largest robustness set, that is,

$$\inf_{P \in \mathcal{P}_{\text{FI}}(1)} |G_x(P)| > 0. \quad (2.12)$$

As P ranges over $\mathcal{P}_{\text{FI}}(1) = \Delta(\mathcal{V})$, the expectation $\mathbb{E}_P[g_x(v)]$ ranges over an interval with endpoints $\inf_{v' \in \mathcal{V}} g_x(v')$ and $\sup_{v' \in \mathcal{V}} g_x(v')$, so that condition (2.12) holds if and only if $g_x(\check{v})$ lies strictly above this supremum or strictly below this infimum. In particular, it requires $\check{v} \notin \mathcal{V}$. Because the robustness sets are nested, condition (2.12) guarantees that the estimator is well defined under every candidate assignment process and at every budget.¹⁰ Since the analysis is conditional on the realized sample, the left-hand side of (2.12) is directly computable, and in the empirical application of Section 6.3 we verify that it holds. Finally, we define the sensitivity bounds as

$$\underline{\beta}(\delta) := \inf_{P \in \mathcal{P}_{\text{FI}}(\delta)} \hat{\beta}(P), \quad \overline{\beta}(\delta) := \sup_{P \in \mathcal{P}_{\text{FI}}(\delta)} \hat{\beta}(P).$$

At $\delta = 0$ the robustness set reduces to $\{\tilde{P}\}$ and both bounds collapse to $\hat{\beta}_*$. $\blacklozenge$

3 Setup

This section formally defines our framework for sensitivity analysis in terms of population shares not satisfying the distributional assumption, and shows how the three leading applications of Section 2 fit into it despite their apparent differences: the first has no structural parameter but an estimated baseline, the second has structural parameters but a known baseline distribution, and the third has a known baseline and no structural parameter. In the supplementary materials we present our fully fledged framework, which also allows for moment inequalities and external nuisance parameters.

Estimand. Let $\Theta \subseteq \mathbb{R}^{d_\theta}$, $d_\theta \in \mathbb{N}_0$ be the parameter space, $\mathcal{Z} \subseteq \mathbb{R}^{d_z}$, $d_z \in \mathbb{N}$ a closed set collecting the possible realizations of a random vector Z . The quantity of interest is a scalar expectation over a distribution $P \in \Delta(\mathcal{Z})$ of a known function $g : \mathcal{Z} \times \Theta \rightarrow \mathbb{R}$

$$\mathbb{E}_P[g(Z; \theta)] = \int_{\mathcal{Z}} g(z; \theta) dP(z). \quad (3.1)$$

¹⁰In Section SA6.3 of the supplementary materials we relax this assumption.

Leading examples are a counterfactual in a structural model, where θ plays the role of a structural parameter, a causal quantity such as an average treatment effect, or an estimated elasticity.

Moment restrictions. Let $m : \mathcal{Z} \times \Theta \rightarrow \mathbb{R}^{d_m}$, $d_m \in \mathbb{N}_0$ be a known moment function and $\vec{\rho}_0 \in \mathbb{R}^{d_m}$ a vector of target moments. We require that

$$\mathbb{E}_P[m(Z; \theta)] = \vec{\rho}_0. \quad (3.2)$$

In a structural model these are the conditions that identify the structural parameter θ , for instance by matching model implied choice probabilities with observed ones, while in a randomized experiment they can carry moments of pre-treatment covariates added to tighten the bounds. We allow for the components of $\vec{\rho}_0$ to be either known or estimated. Section 4.2 makes the distinction explicit. When no moment function is present, we set $d_m = 0$ and $\mathbb{R}^0 = \{0\}$, interpret (3.2) as vacuous, and suppress m and $\vec{\rho}_0$ from the notation.

Baseline distribution. The econometrician postulates that Z is distributed according to a *baseline distribution* P_* , the distribution she would use absent any concern for misspecification. We let $P_* \in \mathcal{P}_\theta(\mathcal{Z})$, where $\mathcal{P}_\theta(\mathcal{Z})$ denotes the set of distributions in $\Delta(\mathcal{Z})$ under which $|g(\cdot; \theta)|$ and $\|m(\cdot; \theta)\|$ are integrable at $\theta \in \Theta$. For brevity, we omit the dependence on θ and write $\mathcal{P}(\mathcal{Z}) \equiv \mathcal{P}_\theta(\mathcal{Z})$. We reserve P_0 for the true distribution of Z . The baseline may be a parametric specification for a distribution that is not nonparametrically identifiable (e.g., a Gumbel distribution for idiosyncratic taste shocks) or an identified feature of P_0 (e.g., the outcome distribution of observed respondents), in which case it must be estimated. In either case the identifying assumption is the equality between the law of Z and P_* , and the sensitivity exercise relaxes it. Both cases are covered by our framework.

Distributional assumptions and robustness sets. Different assumptions are represented by different distributions in $\mathcal{P}(\mathcal{Z})$. To control the degree of relaxation of the distributional assumption embedded in P_* , or equivalently the degree of misspecification, we introduce a *sensitivity parameter* $\delta \in [0, 1]$ to bound the distance between the baseline distribution P_* and any candidate alternative distribution $P \in \mathcal{P}(\mathcal{Z})$. In this paper, the distance is measured in terms of the share of the population in the alternative distribution P for which the distributional assumption embedded in P_* may fail. We call the set of distributions that satisfy the moment restrictions in (3.2) and are within a distance δ of the baseline distribution a *robustness set*. We consider two types of robustness sets, each suited to a different setting among those presented in Section 2.

First, consider settings where the statistical model naturally yields a mixture constraint

$$P_* = \kappa P + (1 - \kappa)R, \quad \text{for some } R \in \mathcal{P}(\mathcal{Z}), \quad (3.3)$$

for some known $\kappa \in (0, 1)$. In words, the population described by P_* consists of two sub-populations, or groups: a κ -share of its units belongs to the group whose distribution is P and the remaining $(1 - \kappa)$ -share belongs to a second group whose distribution is R . This is the case of Example 2.1, where (2.3) shows that the observed treated units are a mixture of two groups, the compliers and the always-observed units, with P playing the role of the complier distribution, R that of the

always-observed distribution, and $\kappa = \pi$ the complier share in (2.2). In this case, we define the *total variation robustness set* as

$$\begin{aligned} \mathcal{P}_{\text{TV}}(\theta; \delta) &:= \{P \in \mathcal{P}(\mathcal{Z}) : \text{TV}(P \parallel R) \leq \delta, \mathbb{E}_P[m(Z; \theta)] = \vec{\rho}_0, \\ &\quad P_* = \kappa P + (1 - \kappa)R \text{ for some } R \in \mathcal{P}(\mathcal{Z})\}. \end{aligned} \quad (3.4)$$

The reason (3.4) is appealing is that δ is the share of the population for which the distributional assumption in P_* may fail. The interpretation of δ just given deserves some elaboration. The constraint $\text{TV}(P \parallel R) = \delta$ only tells us that the distributions P and R of the two groups share a fraction $1 - \delta$ of their probability mass, in the sense that they admit the mixture representations $P = (1 - \delta)C + \delta Q_P$ and $R = (1 - \delta)C + \delta Q_R$ for a common distribution $C \in \mathcal{P}(\mathcal{Z})$ and some group-specific distributions $Q_P, Q_R \in \mathcal{P}(\mathcal{Z})$. Then, substituting the two representations above into the mixture constraint (3.3), we get

$$P_* = (1 - \delta)C + \delta(\kappa Q_P + (1 - \kappa)Q_R).$$

Put differently, for a $(1 - \delta)$ -share of the population, drawn from the common distribution C , the distribution of Z does not depend on the group a unit belongs to, and only the remaining δ -share, drawn from the group-specific $\kappa Q_P + (1 - \kappa)Q_R$, is allowed to differ across groups. The distributional assumption embedded in the baseline, namely that the two groups are drawn from the same distribution, is thus recovered at $\delta = 0$ and is gradually relaxed as δ grows.

Second, when the statistical model does not directly yield a mixture constraint, we consider a *contamination robustness set* in the spirit of Huber (1964), which is defined as

$$\begin{aligned} \mathcal{P}_{\text{C}}(\theta; \delta) &:= \{P \in \mathcal{P}(\mathcal{Z}) : P = (1 - \delta)P_* + \delta R \text{ for some } R \in \mathcal{P}(\mathcal{Z}), \\ &\quad \mathbb{E}_P[m(Z; \theta)] = \vec{\rho}_0\}. \end{aligned} \quad (3.5)$$

In words, a distribution P is a valid candidate distribution if it is a mixture of the baseline distribution P_* and an arbitrary distribution R , with δ being the share of the population that is drawn from R . In this case, the interpretation of δ is identical to the one given above for the total variation robustness set.

Sensitivity bounds. The *sensitivity bounds* are the extreme values the estimand (3.1) can take over the robustness sets (3.4) or (3.5), depending on the application:

$$\underline{\mathbf{b}}_{\diamond}(\delta) := \inf_{\theta \in \Theta} \inf_{P \in \mathcal{P}_{\diamond}(\theta; \delta)} \mathbb{E}_P[g(Z; \theta)] \quad \text{and} \quad \bar{\mathbf{b}}_{\diamond}(\delta) := \sup_{\theta \in \Theta} \sup_{P \in \mathcal{P}_{\diamond}(\theta; \delta)} \mathbb{E}_P[g(Z; \theta)], \quad (3.6)$$

for $\diamond \in \{\text{TV}, \text{C}\}$, with the convention that the inner infimum and supremum are $+\infty$ and $-\infty$, respectively, when the robustness set is empty at θ . For each fixed θ , we refer to the optimization over P as the *inner problem* and to the optimization over θ as the *outer problem*. Section 4 shows how to solve the inner problems in (3.6) exactly despite their infinite-dimensional nature.

The sensitivity exercise we propose in this work reports $[\underline{\mathbf{b}}_{\diamond}(\delta), \bar{\mathbf{b}}_{\diamond}(\delta)]$ for $\delta \in [0, 1]$. At $\delta = 0$ the

interval collapses to the baseline value of the estimand $\mathbb{E}_{P_*}[g(Z; \theta)]$.¹¹ As δ grows, the sensitivity bounds widen monotonically because the robustness sets are nested. The smallest δ at which the interval contains a given value, zero say, thus answers the question “how large does the share of the population for which the baseline assumption fails need to be to overturn the conclusion?” Section 5 formalizes this answer through summary measures.

We now map the three leading applications into this framework.

Example 2.1, continued. We can map the randomized experiment with attrition into the framework as follows. First, there is no structural parameter, so $\mathbf{d}_m = \mathbf{d}_\theta = 0$ and $\Theta = \{0\}$. The random vector of the framework is the treated potential outcome, that is $Z = Y(1)$ and $\mathcal{Z} = \mathcal{Y}$. The target estimand is $g(y; \theta) = y$. The baseline is the outcome distribution of the treated respondents, $P_* = P_T$, which is identified by the data and must therefore be estimated, whereas the distribution being perturbed is the complier distribution P_C . In this notation, MCAR in the form (2.4) is exactly the identifying equality between the distribution of Z , here P_C , and the baseline P_T . The mixture decomposition (2.3) is the mixture constraint (3.3) with $\kappa = \pi$, the complier share in (2.2), with P playing the role of the complier distribution and R that of the always-observed distribution. The relevant robustness set is thus the total variation one in (3.4), and in fact $\mathcal{P}_{TV}(\theta; \delta) = \mathcal{P}_{RCT}(\delta)$ for every $\delta \in [0, 1]$. In line with the interpretation given above, δ is the share of the treated respondents for whom MCAR may fail.¹² With these choices, the bounds in (3.6) are bounds on the complier mean,

$$\underline{\mathbf{b}}_{TV}(\delta) = \inf_{P \in \mathcal{P}_{RCT}(\delta)} \mathbb{E}_P[Y] \quad \text{and} \quad \bar{\mathbf{b}}_{TV}(\delta) = \sup_{P \in \mathcal{P}_{RCT}(\delta)} \mathbb{E}_P[Y].$$

The complier mean is the only quantity in (2.5) that the data do not identify, and $\tau(\cdot)$ is affine and decreasing in it, so the sensitivity bounds of Example 2.1 are

$$\underline{\tau}(\delta) = \frac{\mu_T - \pi \bar{\mathbf{b}}_{TV}(\delta)}{1 - \pi} - \mu^{AO}(0) \quad \text{and} \quad \bar{\tau}(\delta) = \frac{\mu_T - \pi \underline{\mathbf{b}}_{TV}(\delta)}{1 - \pi} - \mu^{AO}(0),$$

and both reduce to the MCAR estimand τ_{MCAR} at $\delta = 0$. Two features of this application are worth anticipating. First, the mixing weight π is estimated alongside the baseline. Second, when outcomes are continuously distributed, $\underline{\tau}(\delta)$ and $\bar{\tau}(\delta)$ admit a closed form, given in Theorem SA-34 of the supplementary materials, and $[\underline{\tau}(1), \bar{\tau}(1)]$ coincides with the sharp nonparametric bounds for τ_0 of Proposition 1a in Lee (2009). When pre-treatment covariates are added to tighten the bounds, there is no closed form for $\delta \in (0, 1)$, whereas at $\delta = 1$ the sharp nonparametric bounds of Proposition 1b in Lee (2009) and Lemma 3 in Semenova (2025) are recovered (see Section SA6.1.1 of the supplementary materials). ♦

Example 2.2, continued. In this case, the random vector Z is the vector of preference shocks, that is $Z = U$ and so $\mathcal{Z} = \mathbb{R}^{J+1}$. The structural parameter collects the systematic components of

¹¹This statement presumes that there exists a $\theta_* \in \Theta$ for which the baseline satisfies the moment conditions in (3.2), so that at $\delta = 0$ the baseline is the only distribution in the robustness set at the only feasible parameter value, a common assumption in the majority of the studies involving (generalized) method of moments estimators (see, e.g., Section 2.2.3 in Newey and McFadden, 1994).

¹²Equivalently, δ can be interpreted as the share of compliers or always-observed units for whom MCAR may fail.

match utility, $\theta = \vartheta$ and $\Theta \subseteq \mathbb{R}^{2J}$, the estimand is the marital college premium, $g = f$, and the moment function is $m = s$, so that $\mathbf{d}_m = 4J + 2$ and $\vec{\rho}_0 = (\varrho, 0_{2(J+1)})$. The first $2J$ components of $\vec{\rho}_0$ are the observed marriage probabilities, which are estimated from the data, whereas the remaining $2(J+1)$ components are the normalizations. The baseline is the Gumbel specification of the shocks, $P_* = \tilde{P}$, which is known rather than estimated. Unlike Example 2.1, the baseline is not itself a mixture of two groups, so the relevant robustness set is the contamination one: (2.7) is exactly (3.5) with the choices above, that is, $\mathcal{P}_{\text{STR}}(\vartheta; \delta) = \mathcal{P}_{\text{C}}(\vartheta; \delta)$ for every $\vartheta \in \Theta$ and $\delta \in [0, 1]$. Its two groups are the women whose shocks are Gumbel, a share $1 - \delta$ of the population, and the women whose shocks are drawn from an arbitrary distribution R , a share δ , so that δ is the share of women for whom the Gumbel assumption may fail. Finally, the sensitivity bounds are $\underline{b}_{\text{C}}(\delta) = \underline{k}(\delta)$ and $\overline{b}_{\text{C}}(\delta) = \overline{k}(\delta)$ for every $\delta \in [0, 1]$. In contrast with Example 2.1, these bounds admit no closed form, and the dual problems of Section 4 are the route through which we compute them in Section 6.2. ♦

Example 2.3, continued. Contrary to Examples 2.1 and 2.2, the target of this sensitivity exercise is the ratio $\hat{\beta}(P)$ of two expectations, which is not of the form (3.1). Its numerator and denominator, however, are both affine in the candidate assignment process P , and this is what we leverage to cast the problem in the desired form. Indeed, for $b \in \mathbb{R}$ we have $\hat{\beta}(P) = b$ if and only if $G_y(P) - bG_x(P) = 0$, and

$$G_y(P) - bG_x(P) = g_b(\tilde{v}) - \mathbb{E}_P[g_b(v)], \quad g_b(\cdot) := g_y(\cdot) - b g_x(\cdot),$$

in view of (2.10). In words, the recentered estimator attains the value b under the candidate assignment process P exactly when $\mathbb{E}_P[g_b(v)]$ equals the realized value $g_b(\tilde{v})$. Once b is fixed, g_b is a known function on $\mathcal{V}$, and the exercise fits the framework upon setting $Z = v$, $\mathcal{Z} = \mathcal{V}$, $g(v; \theta) = g_b(v)$, $\Theta = \{0\}$, $\mathbf{d}_m = \mathbf{d}_\theta = 0$, and $P_* = \tilde{P}$, the assignment process of the exogenous shocks postulated by the econometrician. With these choices, the robustness set (2.11) is the contamination set (3.5), that is $\mathcal{P}_{\text{FI}}(\delta) = \mathcal{P}_{\text{C}}(\theta; \delta)$ for every $\delta \in [0, 1]$, where we use that g_b is bounded, so that every $P \in \Delta(\mathcal{V})$ belongs to $\mathcal{P}(\mathcal{Z})$. The interpretation of δ is that the exogenous shocks are drawn a fraction $(1 - \delta)$ of the time from the postulated assignment process $\tilde{P}$ and a fraction δ of the time from an arbitrary assignment process, so that δ is the share of the population for which the assumption that the shocks are assigned by $\tilde{P}$ may fail. Write $\underline{b}_{\text{C}}(\delta; g_b)$ and $\overline{b}_{\text{C}}(\delta; g_b)$ for the bounds in (3.6) obtained with $g = g_b$, making the dependence on b explicit. Because $\mathcal{P}_{\text{FI}}(\delta)$ is convex and $P \mapsto \mathbb{E}_P[g_b(v)]$ is linear, the values of $\mathbb{E}_P[g_b(v)]$ generated by the robustness set form an interval whose endpoints are these two bounds. It follows that every value b taken by $\hat{\beta}(P)$ over $\mathcal{P}_{\text{FI}}(\delta)$ belongs to the set

$$\mathbf{B}(\delta) := \left\{ b \in \mathbb{R} : \underline{b}_{\text{C}}(\delta; g_b) \leq g_b(\tilde{v}) \leq \overline{b}_{\text{C}}(\delta; g_b) \right\}.$$

The sensitivity bounds of Example 2.3 are therefore the endpoints of $\mathbf{B}(\delta)$,

$$\underline{\beta}(\delta) = \inf \mathbf{B}(\delta) \quad \text{and} \quad \overline{\beta}(\delta) = \sup \mathbf{B}(\delta).$$

Computing them thus amounts to solving the two inner problems in (3.6) for each b . ♦

4 Implementation

The inner problems in (3.6) optimize over an infinite-dimensional space of distributions and, outside special cases such as the RCT with attrition of Example 2.1, admit no closed form. In this section we show that they can nonetheless be solved exactly through a convex program in $d_m + 2$ variables which, solved over a grid of values for δ , traces the thick orange lines of Figure 1. Section 4.1 derives the finite-dimensional dual of the inner problem for each robustness set of Section 3 and states conditions under which its value coincides with that of the primal in (3.6). Sections 4.2–4.3 then turn to finite samples: the former introduces plug-in estimators of the bounds, and the latter derives their limit distribution and proposes a resampling procedure to form valid confidence intervals.

4.1 Duality

Fix $\theta \in \Theta$ and $\diamond \in \{\text{TV}, \text{C}\}$, and write

$$\bar{V}_P^\diamond(\theta; \delta) := \sup_{P \in \mathcal{P}_\diamond(\theta; \delta)} \mathbb{E}_P[g(Z; \theta)] \quad (4.1)$$

for the value of the inner problem in the upper bound of (3.6). The lower bound is recovered by applying every result of this section to $-g$ in place of g , so we focus on the upper bound throughout.

The estimand and the moment restrictions are linear in P , and both robustness sets are convex, because the mixture representations in (3.4) and (3.5) are preserved under convex combinations and the total variation distance is jointly convex in its two arguments. The inner problem is therefore a convex program over probability distributions, and we construct its dual with elementary tools from convex analysis, following the logic of the duality results for distributionally robust optimization in Kuhn, Shafiee and Wiesemann (2025).¹³ The appeal of the dual is that it looks at the problem from the point of view of its constraints rather than that of the candidate distribution, and thereby reduces its dimensionality from that of the space of distributions to $d_m + 2$.

For $\theta \in \Theta$ and $\delta \in (0, 1]$, the optimization problems dual to (4.1) with robustness sets (3.4) and (3.5) are respectively

$$\begin{aligned} \bar{V}_D^{\text{TV}}(\theta; \delta) := \inf_{\substack{(\zeta, \eta, \lambda) \in \\ \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}}} & \left\{ \zeta + \eta(1 - \kappa)\delta + \lambda^\top \vec{\rho}_0 \right. \\ & \left. + \mathbb{E}_{P_*} [r_{\text{TV}}(g(Z; \theta) - \lambda^\top m(Z; \theta) - \zeta, \eta)] \right\} \end{aligned} \quad (4.2)$$

and

$$\begin{aligned} \bar{V}_D^{\text{C}}(\theta; \delta) := \inf_{\substack{(\zeta, \eta, \lambda) \in \\ \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}}} & \left\{ \zeta + \eta\delta + \lambda^\top \vec{\rho}_0 + \mathbb{E}_{P_*} [r_{\text{C}}(g(Z; \theta) - \lambda^\top m(Z; \theta) - \zeta, \eta)] \right\} \\ \text{s.to } & \zeta + \frac{\eta}{2} \geq \sup_{z \in \mathcal{Z}} \{g(z; \theta) - \lambda^\top m(z; \theta)\}, \end{aligned} \quad (4.3)$$

¹³For further details, we refer the interested reader to Section SA2 of the supplementary materials. For a textbook treatment of the relevant convex analysis tools, see Rockafellar (1970, Sections 6, 12 and 31) and Rockafellar (1974).

where, for $x \in \mathbb{R}$ and $\eta \geq 0$,

$$\begin{aligned} r_{\text{TV}}(x, \eta) &:= \max \left\{ -\frac{\eta}{2}, x, \frac{x}{\kappa} - \frac{(1-\kappa)\eta}{2\kappa} \right\}, \\ r_{\text{C}}(x, \eta) &:= \begin{cases} \max \left\{ (1-\delta)x - \frac{\delta\eta}{2}, x \right\}, & x \leq \eta/2, \\ +\infty, & x > \eta/2. \end{cases} \end{aligned} \quad (4.4)$$

Some comments on the two duals are in order. First, both programs have $d_m + 2$ variables, one per restriction of the primal problem other than the mixture restriction: $\zeta \in \mathbb{R}$ is the multiplier associated with the constraint that P is a probability distribution; $\eta \in \mathbb{R}_+$ to the budget constraint, which limits the share of the population whose distribution may depart from the baseline; $\lambda \in \mathbb{R}^{d_m}$ to the moment equalities. Second, the three expressions inside the maximum defining r_{TV} are the values of the associated Lagrangian when a unit of the baseline population is removed from the group described by P , kept at its baseline weight, or assigned the largest weight compatible with the mixture restriction; the two expressions in r_{C} are its values when the unit is scaled down to the share $1-\delta$ or kept at its baseline weight. Third, (4.3) carries one additional constraint for each point of $\mathcal{Z}$, because candidate distributions in $\mathcal{P}_{\text{C}}(\theta; \delta)$ may place mass outside the support of P_* , and the constraints make such relocations unprofitable. No such constraint appears in (4.2), because every $P \in \mathcal{P}_{\text{TV}}(\theta; \delta)$ is absolutely continuous with respect to P_* , the mixture restriction forcing $\kappa P \leq P_*$. Fourth, r_{C} embeds an implicit constraint of its own. Since $r_{\text{C}}(x, \eta) = +\infty$ whenever $x > \eta/2$, the objective of (4.3) is finite only if $\zeta + \eta/2 \geq g(z; \theta) - \lambda^\top m(z; \theta)$ for P_* -almost every $z \in \mathcal{Z}$. The pointwise constraint in (4.3) imposes the same inequality at every $z \in \mathcal{Z}$ and therefore implies the implicit one. Proposition 1 below gives a formal condition under which the converse also holds, so that the pointwise constraint is redundant and can be deleted.

The next theorem shows that solving (4.2) and (4.3) yields the same value as (4.1). Toward this goal, define the sets of attainable moments

$$\begin{aligned} \Gamma_{\text{TV}}(\theta; \delta) &:= \left\{ \mathbb{E}_P[m(Z; \theta)] : P \in \mathcal{P}(\mathcal{Z}), \text{TV}(P \| R) \leq \delta, \right. \\ &\quad \left. P_* = \kappa P + (1-\kappa)R \text{ for some } R \in \mathcal{P}(\mathcal{Z}) \right\}, \\ \Gamma_{\text{C}}(\theta; \delta) &:= \left\{ \mathbb{E}_P[m(Z; \theta)] : P \in \mathcal{P}(\mathcal{Z}), P = (1-\delta)P_* + \delta R \text{ for some } R \in \mathcal{P}(\mathcal{Z}) \right\}, \end{aligned}$$

which collect the moment vectors generated by the distributions satisfying every restriction defining the corresponding robustness set other than the moment equalities in (3.2).

Theorem 1 (Strong duality). *Let $\theta \in \Theta$ and $\delta \in (0, 1]$.*

- (i) *If $\vec{\rho}_0 \in \text{ri}(\Gamma_{\text{TV}}(\theta; \delta))$, then $\bar{V}_{\text{P}}^{\text{TV}}(\theta; \delta) = \bar{V}_{\text{D}}^{\text{TV}}(\theta; \delta)$.*
- (ii) *If $\sup_{z \in \mathcal{Z}} |g(z; \theta)| < \infty$ and $\vec{\rho}_0 \in \text{ri}(\Gamma_{\text{C}}(\theta; \delta))$, then $\bar{V}_{\text{P}}^{\text{C}}(\theta; \delta) = \bar{V}_{\text{D}}^{\text{C}}(\theta; \delta)$.*

The condition $\vec{\rho}_0 \in \text{ri}(\Gamma_{\diamond}(\theta; \delta))$ is a Slater condition. In words, it asks that the moment restrictions remain attainable, within the same budget δ , after a small perturbation of the target $\vec{\rho}_0$ in every direction along which the attainable moments vary. The relative interior formulation accommodates models whose moment components satisfy exact affine relations. When no moment restrictions are

present, the condition holds under the conventions of Section 3. The boundedness of g in part (ii) is the price of letting the candidate distributions place mass everywhere in $\mathcal{Z}$, even where P_* does not.

The pointwise constraint, potentially an infinite collection of inequalities, is what makes (4.3) more burdensome to compute than (4.2), and we now give a condition under which it can be dropped without affecting the value of the program. For $\theta \in \Theta$ and $\delta \in (0, 1]$ define

$$\bar{V}_P^{C,\ll}(\theta; \delta) := \sup_{P \in \mathcal{P}_C^{\ll}(\theta; \delta)} \mathbb{E}_P[g(Z; \theta)], \quad \mathcal{P}_C^{\ll}(\theta; \delta) := \mathcal{P}_C(\theta; \delta) \cap \{P : P \ll P_*\},$$

for the value of the inner problem over the contamination robustness set restricted to distributions dominated by the baseline, write $\bar{V}_D^{C,\ll}(\theta; \delta)$ for the infimum of the objective of (4.3) over all of $\mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}$, that is, for the value of (4.3) with the pointwise constraint deleted, and let

$$\Gamma_C^{\ll}(\theta; \delta) := \{ \mathbb{E}_P[m(Z; \theta)] : P \in \mathcal{P}(\mathcal{Z}), P \ll P_*, P = (1 - \delta)P_* + \delta R, R \in \mathcal{P}(\mathcal{Z}) \}$$

be the attainable moment set of the restricted problem, a subset of $\Gamma_C(\theta; \delta)$. The inequalities $\bar{V}_P^{C,\ll}(\theta; \delta) \leq \bar{V}_P^C(\theta; \delta)$ and $\bar{V}_D^{C,\ll}(\theta; \delta) \leq \bar{V}_D^C(\theta; \delta)$ hold in general, the first because $\mathcal{P}_C^{\ll}(\theta; \delta) \subseteq \mathcal{P}_C(\theta; \delta)$ and the second because deleting a constraint enlarges the feasible set.

Proposition 1 (Deleting the pointwise constraint). *Let $\theta \in \Theta, \delta \in (0, 1], \bar{\rho}_0 \in \text{ri}(\Gamma_C^{\ll}(\theta; \delta)), \sup_{z \in \mathcal{Z}} |g(z; \theta)| < \infty$, and, for every $\lambda \in \mathbb{R}^{d_m}$,*

$$\sup_{z \in \mathcal{Z}} \{g(z; \theta) - \lambda^\top m(z; \theta)\} = P_*\text{-ess sup}_{z \in \mathcal{Z}} \{g(z; \theta) - \lambda^\top m(z; \theta)\}, \quad (4.5)$$

then $\bar{V}_P^C(\theta; \delta) = \bar{V}_P^{C,\ll}(\theta; \delta) = \bar{V}_D^{C,\ll}(\theta; \delta)$.

Proposition 1 contains two results. The first equality states that, under its regularity conditions, removing the distributions $P \not\ll P_*$ from $\mathcal{P}_C(\theta; \delta)$ does not change the value of (4.1). The second states that this value is delivered by the dual (4.3) without the pointwise constraint. Condition (4.5) holds, for example, whenever $\text{supp}(P_*) = \mathcal{Z}$ and $z \mapsto g(z; \theta) - \lambda^\top m(z; \theta)$ is lower semicontinuous for every λ , as in the formula instrument setting of Example 2.3, and also for the matching indicators of Example 2.2, which are not continuous.

4.2 Estimation

Having recast the inner problems as tractable convex programs, we now turn to estimating the bounds, that is, the thick orange lines in Figure 1. Toward this goal, we introduce some notation. Partition the moment function as $m = (m_e^\top, m_k^\top)^\top$, where the target of the moment function $m_e : \mathcal{Z} \times \Theta \rightarrow \mathbb{R}^{d_e}, d_e \in \mathbb{N}_0$ is estimated and that of $m_k : \mathcal{Z} \times \Theta \rightarrow \mathbb{R}^{d_k}, d_k \in \mathbb{N}_0$ is known and normalized to zero. Throughout this section, $\bar{\rho}_0 := (\rho_0^\top, 0_{d_k}^\top)^\top$ where $\rho_0 \in \mathbb{R}^{d_e}$ denotes the moment target to be estimated. Accordingly, partition the multipliers as $\lambda = (\lambda_e^\top, \lambda_k^\top)^\top$ with $\lambda_e \in \mathbb{R}^{d_e}$ and $\lambda_k \in \mathbb{R}^{d_k}$. Throughout, we let $\rho \in \mathbb{R}^{d_e}$, $P \in \Delta(\mathcal{Z})$, and $\xi := (\zeta, \eta, \lambda)$ collect the dual

variables. To treat the two robustness sets at once, let $c_{\text{TV}} := 1 - \kappa$ and $c_{\text{C}} := 1$ be the factors that multiply the term $\eta\delta$ in (4.2) and (4.3), respectively. For $\diamond \in \{\text{TV}, \text{C}\}$, we write the objective of the corresponding dual problem as

$$\bar{\mathcal{J}}^\diamond(\xi, \theta; \rho, P) := \zeta + c_\diamond \eta\delta + \lambda_{\mathbf{e}}^\top \rho + P \bar{\ell}_{\xi, \theta}^\diamond, \quad \bar{\ell}_{\xi, \theta}^\diamond(z) := r_\diamond(g(z; \theta) - \lambda^\top m(z; \theta) - \zeta, \eta),$$

with $r_\diamond$ as in (4.4), and its feasible set as $\bar{\mathcal{F}}^{\text{TV}}(\theta) := \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}$ and

$$\bar{\mathcal{F}}^{\text{C}}(\theta) := \left\{ \xi \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m} : \zeta + \frac{\eta}{2} \geq \sup_{z \in \mathcal{Z}} \{g(z; \theta) - \lambda^\top m(z; \theta)\} \right\},$$

so that $\bar{\mathcal{F}}^{\text{C}}(\theta)$ is the set cut out by the pointwise constraint of (4.3), on which the integrand r_{C} is finite, whereas we recall that (4.2) has no such constraint. Symmetrically, let $\underline{\mathcal{J}}^\diamond$, $\underline{\ell}_{\xi, \theta}^\diamond$, and $\underline{\mathcal{F}}^\diamond(\theta)$ be defined as $\bar{\mathcal{J}}^\diamond$, $\bar{\ell}_{\xi, \theta}^\diamond$, and $\bar{\mathcal{F}}^\diamond(\theta)$ with $-g$ in place of g , so that, in particular, $\underline{\mathcal{F}}^{\text{TV}}(\theta) = \bar{\mathcal{F}}^{\text{TV}}(\theta)$. Further, let $\mathcal{P}_\diamond(\theta; \delta, \rho, P)$ denote the robustness set (3.4) or (3.5) with $(\vec{\rho}_0, P_*)$ replaced by $((\rho^\top, 0_{\mathbf{d}_k}^\top)^\top, P)$, so that $\mathcal{P}_\diamond(\theta; \delta)$ equals $\mathcal{P}_\diamond(\theta; \delta, \rho_0, P_*)$. In line with the convention following (3.6), we define the criterion functions

$$\bar{K}_\delta^\diamond(\theta; \rho, P) := \begin{cases} \inf_{\xi \in \bar{\mathcal{F}}^\diamond(\theta)} \bar{\mathcal{J}}^\diamond(\xi, \theta; \rho, P) & \text{if } \mathcal{P}_\diamond(\theta; \delta, \rho, P) \neq \emptyset, \\ -\infty & \text{otherwise,} \end{cases} \quad (4.6)$$

and

$$\underline{K}_\delta^\diamond(\theta; \rho, P) := \begin{cases} -\inf_{\xi \in \underline{\mathcal{F}}^\diamond(\theta)} \underline{\mathcal{J}}^\diamond(\xi, \theta; \rho, P) & \text{if } \mathcal{P}_\diamond(\theta; \delta, \rho, P) \neq \emptyset, \\ +\infty & \text{otherwise.} \end{cases} \quad (4.7)$$

Finally, let

$$\text{D}_\diamond(\theta; \rho, P) := \inf_{Q \in \mathcal{P}_\diamond(\theta; \delta, \rho, P)} \text{TV}(Q \| P) \text{ and } \Theta_\delta^\diamond(\rho, P) := \{\theta \in \Theta : \text{D}_\diamond(\theta; \rho, P) < c_\diamond \delta\}$$

be the smallest distance from P to the robustness set and the set of strictly feasible parameter values. The sensitivity bounds can thus be written as

$$\underline{\mathbf{b}}_\diamond(\delta, \rho, P) := \inf_{\theta \in \Theta_\delta^\diamond(\rho, P)} \underline{K}_\delta^\diamond(\theta; \rho, P) \quad \text{and} \quad \bar{\mathbf{b}}_\diamond(\delta, \rho, P) := \sup_{\theta \in \Theta_\delta^\diamond(\rho, P)} \bar{K}_\delta^\diamond(\theta; \rho, P),$$

for $\diamond \in \{\text{TV}, \text{C}\}$. At the true inputs (ρ_0, P_*) , these coincide with the sensitivity bounds $\underline{\mathbf{b}}_\diamond(\delta)$ and $\bar{\mathbf{b}}_\diamond(\delta)$ in (3.6). The plug-in estimators replace the two inputs by $\hat{\rho}_n$, an estimator of ρ_0 , and by the empirical measure $\hat{P}_n$ obtained from an i.i.d. sample from P_* , which we recall is an identified feature of P_0 , the true distribution of Z . Formally, the plug-in estimators are $\hat{\underline{\mathbf{b}}}_{\diamond, n}(\delta) := \underline{\mathbf{b}}_\diamond(\delta, \hat{\rho}_n, \hat{P}_n)$ and $\hat{\bar{\mathbf{b}}}_{\diamond, n}(\delta) := \bar{\mathbf{b}}_\diamond(\delta, \hat{\rho}_n, \hat{P}_n)$. When the baseline is instead known but the moment target is estimated, as in Example 2.2, the estimators are simply $\underline{\mathbf{b}}_\diamond(\delta, \hat{\rho}_n, P_*)$ and $\bar{\mathbf{b}}_\diamond(\delta, \hat{\rho}_n, P_*)$.

Remark 1 (Over-identified models). When the model is over-identified, i.e., when there are more moments ($\mathbf{d}_m$) than parameters ($\mathbf{d}_\theta$), there might not exist a $\theta \in \Theta$ for which $\mathbb{E}_{P_*}[m(Z; \theta)]$ equals the estimated target $(\hat{\rho}_n^\top, 0_{\mathbf{d}_k}^\top)^\top$. We refer to Section 3.3 in Christensen and Connault (2023) for two alternative solutions to handle this case. ♣

4.3 Large Sample Properties

Our large sample theory rests on two key properties of (4.6), which hold for both robustness sets: (i) the objective $\bar{J}^\diamond$ is affine in (ρ, P) ; (ii) the feasible set $\bar{\mathcal{F}}^\diamond(\theta)$ does not depend on (ρ, P) . As a consequence, the map $(\rho, P) \mapsto \bar{K}_\delta^\diamond(\theta; \rho, P)$ is concave with supergradient $(\lambda_e, \bar{\ell}_{\xi, \theta}^\diamond)$ at any dual minimizer, and derivatives of the bounds can be read off the dual solutions by an envelope argument. The same two properties hold for (4.7), so that $(\rho, P) \mapsto \underline{K}_\delta^\diamond(\theta; \rho, P)$ is convex with subgradient $-(\lambda_e, \underline{\ell}_{\xi, \theta}^\diamond)$.

The theory of this section provides large sample guarantees when both the moment target ρ_0 and the baseline P_* are estimated from the data. Remark 2 below explains how it simplifies when P_* is known, as in Example 2.2. In terms of Figure 1, we now turn to the dashed orange lines denoting the confidence intervals of the bounds.

Before stating the conditions needed by the large sample theory, we introduce some useful quantities. For $\diamond \in \{\text{TV}, \text{C}\}$, let

$$\bar{\Theta}_\delta^\diamond(\rho, P) := \arg \sup_{\theta \in \Theta} \bar{K}_\delta^\diamond(\theta; \rho, P) \quad \text{and} \quad \bar{\Xi}_\delta^\diamond(\theta; \rho, P) := \arg \inf_{\xi \in \bar{\mathcal{F}}^\diamond(\theta)} \bar{J}^\diamond(\xi, \theta; \rho, P)$$

be the set of outer optimizers of the dual problem and the set of optimal dual multipliers at (θ, ρ, P) , respectively, and

$$\bar{\Lambda}_\delta^\diamond(\theta; \rho, P) := \{\lambda_e \in \mathbb{R}^{\text{de}} : (\zeta, \eta, \lambda) \in \bar{\Xi}_\delta^\diamond(\theta; \rho, P) \text{ with } \lambda = (\lambda_e, \lambda_k) \text{ for some } (\zeta, \eta, \lambda_k)\}$$

be the projection of $\bar{\Xi}_\delta^\diamond(\theta; \rho, P)$ onto the space of multipliers λ_e . The quantities $\underline{\Theta}_\delta^\diamond$, $\underline{\Xi}_\delta^\diamond$, and $\underline{\Lambda}_\delta^\diamond$ are defined in the same way from the program (4.7). The following assumption lists all the regularity conditions needed for the large sample theory.

Assumption LS. *We observe an i.i.d. sample $\{Z_i\}_{i=1}^n$ from P_* . Furthermore, the following conditions hold:*

- (i) (Compactness.) $\mathcal{Z}$ and Θ are nonempty compact sets.
- (ii) (Continuity.) The functions $g(z; \theta)$ and $m(z; \theta)$ are jointly continuous on $\mathcal{Z} \times \Theta$ and there exists a constant $L < \infty$ such that for all $\theta, \theta' \in \Theta$

$$\sup_{z \in \mathcal{Z}} |g(z; \theta) - g(z; \theta')| \leq L \|\theta - \theta'\| \quad \text{and} \quad \sup_{z \in \mathcal{Z}} \|m(z; \theta) - m(z; \theta')\| \leq L \|\theta - \theta'\|.$$
- (iii) (Feasibility.) $\bar{\Theta}_\delta^\diamond(\rho_0, P_*) \neq \emptyset$.
- (iv) (Slater condition.) For each $\theta \in \underline{\Theta}_\delta^\diamond(\rho_0, P_*) \cup \bar{\Theta}_\delta^\diamond(\rho_0, P_*)$, $\vec{\rho}_0 \in \text{int}(\Gamma_\diamond(\theta; \delta))$.
- (v) (Asymptotic linearity.) We have $\hat{\rho}_n = \rho_0 + \hat{P}_n \psi + o_{\mathbb{P}}(n^{-1/2})$, with influence function $\psi : \mathcal{Z} \rightarrow \mathbb{R}^{\text{de}}$ satisfying $P_* \psi = 0$ and $P_* \|\psi\|^2 < \infty$.
- (vi) (Lower hemicontinuity.) The correspondences $\theta \mapsto \underline{\Xi}_\delta^\diamond(\theta; \rho_0, P_*)$ and $\theta \mapsto \bar{\Xi}_\delta^\diamond(\theta; \rho_0, P_*)$ are lower hemicontinuous at each $\theta \in \underline{\Theta}_\delta^\diamond(\rho_0, P_*)$ and each $\theta \in \bar{\Theta}_\delta^\diamond(\rho_0, P_*)$, respectively.

Assumption **LS(i)** greatly simplifies the integrability conditions and, together with Assumptions **LS(ii)** and **LS(iii)**, guarantees the existence of solutions to the outer problem.¹⁴ In Assumption **LS(ii)**, joint continuity is used, together with compactness, to control the supremum in the pointwise constraint of $\bar{\mathcal{F}}^C(\theta)$ and for the existence of solutions to the outer problem, whereas Lipschitz continuity is needed to invoke an appropriate functional central limit theorem and is vacuous when there is no structural parameter, as in Examples 2.1 and 2.3. Assumption **LS(iii)** ensures that the criterion functions are not identically $\pm\infty$. Assumption **LS(iv)** is the Slater-type condition discussed after Theorem 1, stated with the interior rather than the relative interior, which rules out exact affine relations among the moment components. It also guarantees that the dual minimizer sets $\Xi_\delta^\diamond(\theta; \rho_0, P_*)$ and $\Xi_\delta^\diamond(\theta; \rho_0, P_*)$ are nonempty, compact, and convex (Lemma SA-24 of the supplementary materials), so that the multiplier sets $\underline{\Lambda}_\delta^\diamond, \bar{\Lambda}_\delta^\diamond$ are well defined. Assumption **LS(v)** requires the estimator of ρ_0 to be asymptotically linear, so that the joint law of $\sqrt{n}(\hat{\rho}_n - \rho_0)$ and of the empirical process $\mathbb{G}_n := \sqrt{n}(\hat{P}_n - P_*)$ is amenable to the delta method used below. Finally, Assumption **LS(vi)** is used to establish the directional differentiability of the value function. Under Assumption **LS** the plug-in estimators for the bounds are consistent.

Theorem 2 (Consistency). *Let Assumption **LS** hold for $\diamond \in \{\text{TV}, \text{C}\}$. Then*

$$\underline{b}_\diamond(\delta, \hat{\rho}_n, \hat{P}_n) \xrightarrow{\mathbb{P}} \underline{b}_\diamond(\delta, \rho_0, P_*) \quad \text{and} \quad \bar{b}_\diamond(\delta, \hat{\rho}_n, \hat{P}_n) \xrightarrow{\mathbb{P}} \bar{b}_\diamond(\delta, \rho_0, P_*).$$

To establish a distributional approximation for the bounds, we rely on a delta method for maps that are only Hadamard directionally differentiable, that is, whose derivatives are continuous but not necessarily linear.¹⁵ The reason is that the bounds are values of optimization problems. The maps $(\rho, P) \mapsto (\underline{b}, \bar{b})(\delta, \rho, P)$ are Hadamard differentiable when these programs have a single optimizer, but generally only directionally so when they have several, as Christensen and Connault (2023, Section 6.1) already noted for the case of known P_* . We therefore use the delta method for Hadamard directionally differentiable functionals (see, e.g., Shapiro, 1991; Dümbgen, 1993). The next result first derives the directional derivative of the bounds at (ρ_0, P_*) and then establishes their distributional approximation.

Theorem 3 (Directional differentiability and distributional approximation). *Let Assumption **LS** hold for $\diamond \in \{\text{TV}, \text{C}\}$. Then the maps $(\rho, P) \mapsto (\underline{b}_\diamond, \bar{b}_\diamond)(\delta, \rho, P)$ are Hadamard directionally differ-*

¹⁴We stress that some of the motivating examples do not satisfy compactness of $\mathcal{Z}$ in Assumption **LS(i)**. In Example 2.1 the outcome space $\mathcal{Y}$ is a closed, possibly unbounded, subset of $\mathbb{R}$, whereas in Example 2.2, $\mathcal{Z} = \mathbb{R}^{J+1}$. For Example 2.1, Proposition SA-35 of the supplementary materials exploits the closed form of the bounds to replace compactness by a finite second moment of the outcome and a density positive at the trimming quantiles, and it also accounts for the estimation of the complier share. For Example 2.2, relaxing compactness of $\mathcal{Z}$ would require assuming a coercivity condition on the moment function, which we leave for future work. In the same example, the matching indicators in s_1 also violate the continuity requirement of Assumption **LS(ii)**, but only on the set of ties, which is null under the Gumbel baseline, so that this requirement can be relaxed as well.

¹⁵For a formal definition of Hadamard differentiability and Hadamard directional differentiability, see Section 2.2 in Fang and Santos (2019).

entiable at (ρ_0, P_*) , with

$$\begin{aligned} d\mathbf{b}_\diamond(\delta, \rho_0, P_*)[h, \gamma] &= \min_{\theta^* \in \Theta_\delta^\diamond(\rho_0, P_*)} \max_{\xi^* \in \Xi_\delta^\diamond(\theta^*; \rho_0, P_*)} \left\{ -(\lambda_{\mathbf{e}}^*)^\top h - \gamma \bar{\ell}_{\xi^*, \theta^*}^\diamond \right\}, \\ d\bar{\mathbf{b}}_\diamond(\delta, \rho_0, P_*)[h, \gamma] &= \max_{\theta^* \in \Theta_\delta^\diamond(\rho_0, P_*)} \min_{\xi^* \in \Xi_\delta^\diamond(\theta^*; \rho_0, P_*)} \left\{ (\lambda_{\mathbf{e}}^*)^\top h + \gamma \bar{\ell}_{\xi^*, \theta^*}^\diamond \right\}, \end{aligned}$$

and so

$$\sqrt{n} \begin{pmatrix} \mathbf{b}_\diamond(\delta, \hat{\rho}_n, \hat{P}_n) - \mathbf{b}_\diamond(\delta, \rho_0, P_*) \\ \bar{\mathbf{b}}_\diamond(\delta, \hat{\rho}_n, \hat{P}_n) - \bar{\mathbf{b}}_\diamond(\delta, \rho_0, P_*) \end{pmatrix} \rightsquigarrow \begin{pmatrix} d\mathbf{b}_\diamond(\delta, \rho_0, P_*)[Z_\psi, \mathbb{G}] \\ d\bar{\mathbf{b}}_\diamond(\delta, \rho_0, P_*)[Z_\psi, \mathbb{G}] \end{pmatrix},$$

where $\mathbb{G}$ is a P_* -Brownian bridge and $Z_\psi = \mathbb{G}\psi$, the pair $(Z_\psi, \mathbb{G})$ being jointly Gaussian with $\mathbb{E}[\mathbb{G}f Z_\psi] = P_*(f\psi)$.

The novelty of Theorem 3 lies in the fact that the limiting distribution carries the term $\mathbb{G}\bar{\ell}_{\xi^*, \theta^*}^\diamond$, which is also correlated with Z_ψ as both are functionals of the same sample. When the sets of outer maximizers and dual minimizers are singletons, the directional derivatives become linear and the limit is Gaussian with variance

$$\sigma_*^2 = \mathbb{V}_{P_*} \left((\lambda_{\mathbf{e}}^*)^\top \psi + \bar{\ell}_{\xi^*, \theta^*}^\diamond(Z) \right),$$

where the two terms in the variance capture the uncertainty due to the estimation of the moment target ρ_0 and of the baseline P_* , respectively, and their covariance reflects that both are estimated from the same sample.

Outside this singleton case the derivatives in Theorem 3 are, in general, nonlinear, and the standard bootstrap is invalid (Fang and Santos, 2019, Theorem 3.1). Indeed, the standard bootstrap may be interpreted as relying on a (numerical) derivative that consistently estimates the directional derivative only when the latter is linear (see Remark 3.5 in Fang and Santos, 2019). Accordingly, in the same spirit as Christensen and Connault (2023), we specialize the resampling procedure proposed in Fang and Santos (2019) to our case. Concretely, we estimate the directional derivative as

$$\begin{aligned} \widehat{d\mathbf{b}}_{\diamond, n}(\delta)[h, \gamma] &:= \sup_{\theta \in \widehat{\Theta}_{\delta, n}^\diamond} \inf_{\xi \in \Xi_\delta^\diamond(\theta; \hat{\rho}_n, \hat{P}_n)} \left\{ \lambda_{\mathbf{e}}^\top h + \gamma \bar{\ell}_{\xi, \theta}^\diamond \right\}, \\ \widehat{\Theta}_{\delta, n}^\diamond &:= \left\{ \theta \in \Theta_\delta^\diamond(\hat{\rho}_n, \hat{P}_n) : \bar{K}_\delta^\diamond(\theta; \hat{\rho}_n, \hat{P}_n) \geq \widehat{\mathbf{b}}_{\diamond, n}(\delta) - \widehat{\nu}_n \sqrt{\frac{\log n}{n}} \right\} \end{aligned}$$

where $\widehat{\nu}_n \geq 0$ is a tuning sequence that is bounded and bounded away from zero in probability. The lower-bound counterparts $\widehat{\Theta}_{\delta, n}^\diamond$ and $\widehat{d\mathbf{b}}_{\diamond, n}(\delta)$ are defined analogously, with $-g$ in place of g . Following Christensen and Connault (2023, Section 6.2), a convenient data-driven choice of $\widehat{\nu}_n$ is the smallest diagonal entry of the covariance matrix of $m(Z; \hat{\theta}_n)$ under $\hat{P}_n$, with $\hat{\theta}_n$ an approximate maximizer of the estimated outer problem, which requires no simulation. Algorithm 1 details all the steps of the resampling procedure, whose validity is proven in the supplementary materials under a strengthening of Assumption LS. In particular, we show that the bootstrap statistics $\widehat{T}_n^{*, b, \text{low}}$ and $\widehat{T}_n^{*, b, \text{up}}$ of Algorithm 1 are consistent for their population counterparts in Theorem 3, and the confidence sets they deliver have asymptotic coverage at least $1 - \alpha$ whenever the distribution functions of those limits are continuous and strictly increasing at the quantiles used.

Algorithm 1 Confidence sets for the sensitivity bounds

- 1: Compute $\hat{\rho}_n$, $\hat{P}_n$, the bounds $\hat{\underline{b}}_{\diamond,n}(\delta)$, $\hat{\bar{b}}_{\diamond,n}(\delta)$, the near-argmax sets $\hat{\underline{\Theta}}_{\delta,n}^{\diamond}$, $\hat{\bar{\Theta}}_{\delta,n}^{\diamond}$, and the dual minimizer sets $\Xi_{\delta}^{\diamond}(\theta; \hat{\rho}_n, \hat{P}_n)$ and $\Xi_{\delta}^{\diamond}(\theta; \hat{\rho}_n, \hat{P}_n)$.
- 2: **for** $b = 1, \dots, B$ **do**
- 3: Draw an i.i.d. sample $\{Z_i^{*,b}\}_{i=1}^n$ from $\hat{P}_n$ and form $\hat{P}_n^{*,b}$ and $\hat{\rho}_n^{*,b}$.
- 4: Set $\hat{T}_n^{*,b,\text{up}} := \hat{d}\hat{\bar{b}}_{\diamond,n}(\delta)[\sqrt{n}(\hat{\rho}_n^{*,b} - \hat{\rho}_n), \sqrt{n}(\hat{P}_n^{*,b} - \hat{P}_n)]$ and the analog $\hat{T}_n^{*,b,\text{low}}$.
- 5: **end for**
- 6: For $u \in (0, 1)$, let $\hat{\mathbf{c}}_u^{\text{low}}$ and $\hat{\mathbf{c}}_u^{\text{up}}$ be the empirical u -quantiles of $\{\hat{T}_n^{*,b,\text{low}}\}_{b=1}^B$ and $\{\hat{T}_n^{*,b,\text{up}}\}_{b=1}^B$, respectively. Valid lower and upper $100(1 - \alpha)\%$ confidence sets are

$$[\hat{\underline{b}}_{\diamond,n}(\delta) - \hat{\mathbf{c}}_{1-\alpha}^{\text{low}}/\sqrt{n}, +\infty) \quad \text{and} \quad (-\infty, \hat{\bar{b}}_{\diamond,n}(\delta) - \hat{\mathbf{c}}_{\alpha}^{\text{up}}/\sqrt{n}].$$

Remark 2 (Known baseline). When the baseline is known, the same argument with P held fixed at P_* yields consistency, while the directional derivative becomes

$$d\bar{\mathbf{b}}_{\diamond}(\delta, \rho_0, P_*)[h] = \max_{\theta^* \in \bar{\Theta}_{\delta}^{\diamond}(\rho_0, P_*)} \min_{\lambda^* \in \bar{\Lambda}_{\delta}^{\diamond}(\theta^*; \rho_0, P_*)} (\lambda^*)^{\top} h,$$

which is identical to that of Christensen and Connault (2023, Theorem 6.2) even though total variation sits outside the class of divergences considered there. In this case, the Lipschitz requirement of Assumption **LS(ii)** is not needed, and the lower hemicontinuity of Assumption **LS(vi)** may be imposed on the multiplier projections $\underline{\Lambda}_{\delta}^{\diamond}, \bar{\Lambda}_{\delta}^{\diamond}$ alone, since the derivative then sees the dual optimizer only through them. Moreover, since expectations against P_* are computed exactly, the data enter only through $\hat{\rho}_n$. Accordingly, Assumption **LS(v)** may be weakened to the requirement that $\sqrt{n}(\hat{\rho}_n - \rho_0) \rightsquigarrow \mathbf{N}(0, \Sigma)$ for a finite $\mathbf{d}_e \times \mathbf{d}_e$ matrix Σ , with $\hat{\rho}_n$ computed from whatever sample identifies ρ_0 , which need not be a sample of Z . In Example 2.2, Z collects latent preference shocks that are never observed while ρ_0 is estimated from data on realized matches. Then, Step 3 of Algorithm 1 draws only a bootstrap replicate of $\hat{\rho}_n$ and the resulting procedure coincides with the algorithm of Christensen and Connault (2023, Section 6.2) (see Remark SA-11 of the supplementary materials for formal statements). ♣

5 How to Use the Sensitivity Bounds?

This section explains how the output of our sensitivity framework is meant to be used. In terms of Figure 1, Section 5.1 studies the right endpoint of the horizontal axis, where at $\delta = 1$ the bounds coincide with the exact nonparametric sharp bounds, and Section 5.2 introduces some reportable summary measures.

5.1 Exact Nonparametric Sharp Bounds

Besides admitting the population share interpretation, measuring the relaxation of the distributional assumption through shares has an important byproduct: the sensitivity parameter never

needs to exceed one. At $\delta = 1$, the total variation constraint in (3.4) holds for every pair of groups, because the total variation distance never exceeds one, and the mixture representation in (3.5) reads $P = R$, so that it places no restriction on P . What remains binding at $\delta = 1$ are the restrictions that the model itself imposes, namely that Z takes values in $\mathcal{Z}$, that the moment restrictions (3.2) hold at θ , and, for the total variation robustness set, the mixture constraint (3.3). Therefore, the sensitivity bounds at $\delta = 1$ coincide with the nonparametric sharp bounds on the estimand (3.1). The next theorem formalizes this intuition. Toward this goal, we define

$$\mathcal{E}(\theta) := \{P \in \mathcal{P}(\mathcal{Z}) : \mathbb{E}_P[m(Z; \theta)] = \vec{\rho}_0\},$$

the set of distributions compatible with the moment restrictions at θ ,

$$\mathcal{D}_{\text{TV}}(P_*) := \{P \in \mathcal{P}(\mathcal{Z}) : \kappa P \leq P_*\} \quad \text{and} \quad \mathcal{D}_{\text{C}}(P_*) := \mathcal{P}(\mathcal{Z}).$$

Theorem 4 (Nonparametric sharp bounds). *Let $\diamond \in \{\text{TV}, \text{C}\}$. Then, for $\theta \in \Theta$,*

$$\mathcal{P}_{\diamond}(\theta; 1) = \mathcal{D}_{\diamond}(P_*) \cap \mathcal{E}(\theta).$$

Consequently, the sensitivity bounds (3.6) at $\delta = 1$ satisfy

$$\underline{\mathbf{b}}_{\diamond}(1) = \inf_{\theta \in \Theta} \inf_{P \in \mathcal{D}_{\diamond}(P_*) \cap \mathcal{E}(\theta)} \mathbb{E}_P[g(Z; \theta)], \quad \bar{\mathbf{b}}_{\diamond}(1) = \sup_{\theta \in \Theta} \sup_{P \in \mathcal{D}_{\diamond}(P_*) \cap \mathcal{E}(\theta)} \mathbb{E}_P[g(Z; \theta)].$$

Theorem 4 follows directly from the definitions of the two robustness sets. In words, $\underline{\mathbf{b}}_{\text{C}}(1)$ and $\bar{\mathbf{b}}_{\text{C}}(1)$ are the infimum and the supremum of $\mathbb{E}_P[g(Z; \theta)]$ over all $\theta \in \Theta$ and all $P \in \mathcal{E}(\theta)$, that is, over every distribution on $\mathcal{Z}$ that satisfies the moment restrictions at some parameter value, whereas $\underline{\mathbf{b}}_{\text{TV}}(1)$ and $\bar{\mathbf{b}}_{\text{TV}}(1)$ further require the baseline to admit the mixture representation (3.3). In both cases the bounds are sharp, meaning that no distribution compatible with the maintained restrictions generates a value of the estimand outside $[\underline{\mathbf{b}}_{\diamond}(1), \bar{\mathbf{b}}_{\diamond}(1)]$, so that neither bound can be tightened without imposing restrictions beyond those maintained.

Theorem 4 goes beyond sensitivity analysis as it makes the framework a device for computing nonparametric sharp bounds. In some problems those bounds have known closed-form expressions. For instance, in Example 2.1 the interval $[\underline{\mathbf{b}}_{\text{TV}}(1), \bar{\mathbf{b}}_{\text{TV}}(1)]$ delivers the trimming bounds of Lee (2009) when outcomes are continuously distributed (see Theorem SA-34 of the supplementary materials). Similarly, in the missing data problem of Section SA6.4.3 of the supplementary materials, where a share p of the population is observed and its outcome law is the baseline, the contamination bounds $[\underline{\mathbf{b}}_{\text{C}}(1 - p), \bar{\mathbf{b}}_{\text{C}}(1 - p)]$ at a budget equal to the missing share coincide with the sharp nonparametric bounds of Manski (1994, 2003). In other problems, such as the marital college premium of Example 2.2, no closed form is available, and the dual problems of Section 4 return the sharp nonparametric bounds numerically.

5.2 Summary Measures

Throughout this subsection we let $\diamond \in \{\text{TV}, \text{C}\}$ and suppress it from the notation, so that $\underline{\mathbf{b}}(\delta)$ and $\bar{\mathbf{b}}(\delta)$ denote the bounds in (3.6). We measure the robustness of a conclusion against a *reference*

value that the analyst wants to rule out, such as zero for the sign of the estimand or a policy relevant threshold. For simplicity, we set it to zero and take the effect to be positive at the baseline, $\underline{b}(0) > 0$, the opposite case being symmetric. The estimated counterparts we report in Section 6 are plug-in estimators of the four summary measures defined below. Section SA5 of the supplementary materials studies the four measures and establishes the large sample properties of the estimated certified breakdown budget. Its formal results are stated for the total variation robustness set without structural parameters, which is the case of the attrition application of Section 6.1, but the properties below hold for the contamination robustness set by the same arguments.

Breakdown and certified breakdown. The *breakdown budget* is the smallest budget at which the sensitivity bounds include the reference value, that is,

$$\delta_b \equiv \delta_b(0) := \inf \{ \delta \in [0, 1] : \underline{b}(\delta) \leq 0 \}, \quad (5.1)$$

with the convention that the infimum over an empty set equals 1. The breakdown point goes back to the robust statistics literature (Hampel, 1968; Donoho and Huber, 1983) and has been adopted in econometrics to quantify how much a maintained assumption can be relaxed before a conclusion is overturned (see, among many, Horowitz and Manski, 1995; Kline and Santos, 2013; Masten and Poirier, 2020; Ober-Reynolds, 2026). By the population share interpretation of Section 3, no departure from the identifying assumption involving less than a δ_b share of the population can overturn the conclusion.

The breakdown budget captures solely misspecification uncertainty. To account for sampling uncertainty as well, for $\alpha \in (0, 1)$ and a sample size n held fixed, we define the *certified breakdown budget*

$$\delta_{cb}(\alpha; n) := \inf \left\{ \delta \in [0, 1] : \underline{b}(\delta) - \frac{c_{1-\alpha/2} \sigma(\delta)}{\sqrt{n}} \leq 0 \right\}, \quad (5.2)$$

where $\sigma(\delta)$ is the asymptotic standard deviation of the plug-in estimator $\widehat{\underline{b}}_n(\delta)$ and $c_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -th quantile of the standard normal distribution.¹⁶ The critical value is two-sided because the exercise does not commit in advance to the sign of $\underline{b}(\delta)$. We say that a study *certifies* robustness at a budget δ and level α when its conclusion survives both misspecification of size δ and sampling uncertainty at level $1 - \alpha$. Similar measures appear in Cinelli and Hazlett (2020) for unobserved confounding and in Rambachan and Roth (2023) for violations of parallel trends.

By construction, $\delta_{cb}(\alpha; n) \leq \delta_b$. Moreover, $\delta_{cb}(\alpha; n)$ is nondecreasing in n and converges to δ_b as $n \rightarrow \infty$, and, under the conditions of Proposition SA-32 of the supplementary materials, its sample counterpart is consistent for δ_b . Sampling uncertainty thus vanishes but robustness does not, because δ_b is a feature of the design alone. Reporting the pair $(\widehat{\delta}_b, \widehat{\delta}_{cb})$ therefore separates fragility by design (a small breakdown budget) from fragility due to imprecision (a large gap between the two). The two budgets can also rank studies differently: a decision maker weighing whether to fund a second round of data collection should prefer the study with the larger breakdown budget,

¹⁶This definition presumes that the problems defining $\underline{b}(\delta)$ have a unique optimizer, so that the limit of $\sqrt{n}(\widehat{\underline{b}}_n(\delta) - \underline{b}(\delta))$ is Gaussian by Theorem 3; see Remark SA-13 of the supplementary materials for the attrition application and the general case.

whereas one who has to act on the evidence in hand should prefer the study with the larger certified budget.

The shadow price of robustness. We hold the structural parameter fixed and suppress it from the notation. Under appropriate regularity conditions, the bound paths are finite, monotone, and concave ($\bar{\mathbf{b}}$) and convex ($\underline{\mathbf{b}}$), respectively, on the range of budgets at which the robustness set is nonempty, so that each path is differentiable at all but at most countably many budgets. At every such budget we define the *shadow price of robustness* as

$$\bar{\eta}(\delta) := \bar{\mathbf{b}}'(\delta) \quad \text{and} \quad \underline{\eta}(\delta) := -\underline{\mathbf{b}}'(\delta).$$

Both shadow prices are nonnegative and nonincreasing in the budget, so that the first unit of misspecification is the most damaging. For the total variation robustness set, the budget enters (4.2) only through the additive term $\eta(1 - \kappa)\delta$, so that η prices the total variation constraint as a Lagrange multiplier prices a resource constraint, and $\bar{\eta}(\delta) = (1 - \kappa)\eta^*$ wherever the optimal multiplier η^* is unique (Proposition SA-31 of the supplementary materials).

Unlike the breakdown budget, which concerns the sign of an effect, the shadow price also speaks to its magnitude. Suppose that a decision maker is unwilling to commit to a single budget and instead holds beliefs μ on $[0, 1]$, with distribution function F_μ . Then, provided that $\underline{\mathbf{b}}$ is finite and continuous on $[0, 1]$, she could evaluate the design by integrating the worst case sensitivity bound $\underline{\mathbf{b}}(\delta)$ over her beliefs, that is

$$\mathbb{E}_\mu [\underline{\mathbf{b}}(\delta)] = \underline{\mathbf{b}}(0) - \int_0^1 \underline{\eta}(s) (1 - F_\mu(s)) \, ds,$$

so that the price of the s -th unit of misspecification is weighted by the probability she assigns to a violation at least that large.

The certification frontier. For a budget $\delta < \delta_{\mathbf{b}}$, the *certification frontier*

$$n^*(\delta; \alpha) := \frac{\mathfrak{c}_{1-\alpha/2}^2 \sigma^2(\delta)}{\underline{\mathbf{b}}(\delta)^2} \tag{5.3}$$

is the sample size a study would need in order to certify robustness at budget δ and level α . Under the conditions of Proposition SA-33 of the supplementary materials, its semi-elasticity with respect to δ behaves as $2/(\delta_{\mathbf{b}} - \delta)$ as $\delta \rightarrow \delta_{\mathbf{b}}$, so that each additional percentage point of certified robustness becomes increasingly expensive in observations. More data, however, cannot make a study more robust: as $n \rightarrow \infty$ the certified budget converges to $\delta_{\mathbf{b}}$ and no further, because misspecification is of fixed size while sampling uncertainty shrinks at rate $n^{-1/2}$. Put differently, the frontier prices certification, not robustness itself.

6 Empirical Applications

We present three empirical applications. The first revisits the online tutoring program of Carlana and La Ferrara (2025), a randomized experiment in which attrition is moderate but strongly differential across arms. The second revisits the marital college premium application of Christensen and

Connault (2023) and focuses on interpretation and computational cost. The third turns to formula instruments: it assesses the sensitivity of the estimated elasticity of Borusyak and Hull (2023) to the postulated assignment process of the shocks. The supplementary materials contain two further exercises on attrition. Section SA7.1 revisits the transport subsidy experiment of Christensen, Nino and Osman (2026), in which attrition is severe but nearly balanced across arms, and Section SA7.2 runs the same sensitivity exercise on every eligible trial in RCT-Bench, the public collection of randomized clinical trials assembled by Shao, Lyu, Yu and Wang (2026), which currently comprises 125 trials, asking how often attrition threatens the conclusions of medical experiments in practice.

6.1 Attrition in Randomized Experiments

We revisit the Tutoring Online Program experiment of Carlana and La Ferrara (2025), maintaining the setup of Example 2.1 throughout.

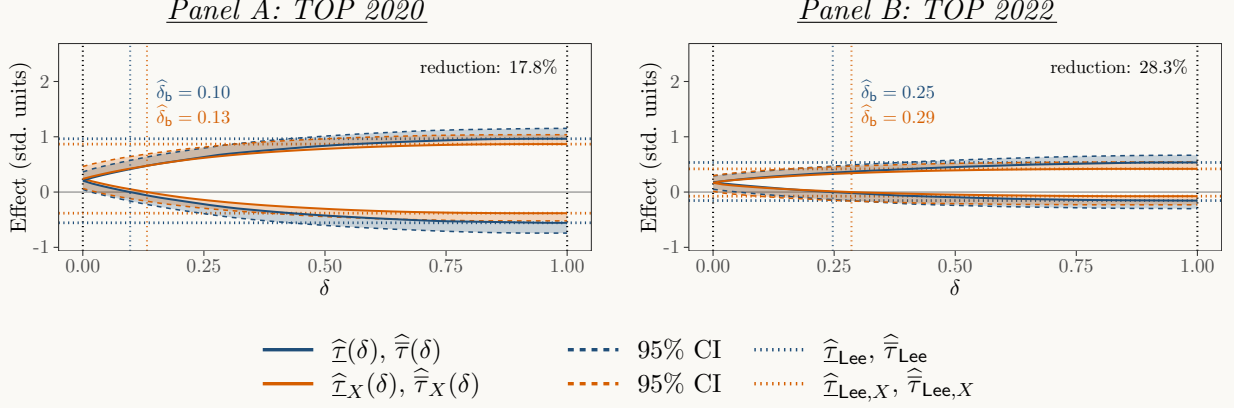
Setup. Carlana and La Ferrara (2025) evaluate the Tutoring Online Program (TOP), which offered free individual online tutoring by university student volunteers to disadvantaged Italian middle schoolers. The experiment ran in two waves, with individual-level randomization in each: TOP 2020, during the COVID-19 school closures (530 treated, 529 control), and TOP 2022, during regular in-person schooling (607 treated, 530 control). The outcome is a standardized mathematics performance index that is observed only for test takers (i.e., respondents). Table 1 shows that attrition varies across waves and arms: the response rates of the treated and control arms are 87.9% and 46.3% in 2020 and 92.3% and 72.3% in 2022, implying estimated complier shares $\hat{\pi}$ of 0.47 and 0.22 in the 2020 and 2022 samples, respectively. Our sensitivity exercise asks how much of the experimental conclusion survives departures from ignorable attrition. In addition, our framework allows us to compare and rank the two waves in terms of how robust their conclusions are.

Table 1: *Attrition and sample size by wave in the Tutoring Online Program.*

Wave	<i>Treatment</i>		<i>Control</i>		$\hat{\pi}$
	<i>n</i>	Attrition	<i>n</i>	Attrition	
TOP 2020	530	12.1%	529	53.7%	0.47
TOP 2022	607	7.7%	530	27.7%	0.22

Results. Figure 2 reports the bounds for TOP 2020 (Panel A) and TOP 2022 (Panel B). In TOP 2020, the bounds start at $\delta = 0$ from the difference in means among respondents, which corresponds to an estimated 0.207 standard deviations, and at $\delta = 1$ reach the Lee bounds, $[-0.556, 0.964]$. The sign of the effect survives at most 10% of the compliers departing from the MCAR assumption. Once sampling uncertainty is taken into account, the sign is certified at the 5% level only up to $\hat{\delta}_{cb} = 0.017$. We do not report the certified breakdown budgets in Figure 2 to avoid cluttering the panels. Conditioning on a handful of pre-treatment covariates tightens the bounds at every budget, shrinking their width at $\delta = 1$ by 18%, and increases the breakdown budget to 0.13 and

Figure 2: TV sensitivity bounds in the Tutoring Online Program.



Notes: Computed from the replication data of Carlana and La Ferrara (2025). Each panel plots the plug-in estimators of $\tau(\delta)$ and $\bar{\tau}(\delta)$ described in Section 4.2 over $\delta \in \{0, 0.02, \dots, 1\}$, without and with covariates, together with the bounds of Lee (2009) and the outer 95% confidence bounds from 1,000 iterations of the bootstrap of Algorithm 1. The vertical dotted lines mark the estimated breakdown budgets $\hat{\delta}_b$ without and with covariates, and the annotation in each panel reports the reduction in the width of the bounds at $\delta = 1$ achieved by the covariates. The covariate cells are the full interaction of indicators for grades six and seven, a median split of the baseline INVALSI math score, mother’s high school completion, mother’s white collar occupation, and single parent household.

the certified breakdown budget to 0.028.¹⁷ In TOP 2022, the bounds start from a difference in means of 0.176 standard deviations and reach the Lee bounds, $[-0.154, 0.535]$. The sign of the effect survives at most 25% of the compliers departing from the MCAR assumption, and it is certified at the 5% level up to $\hat{\delta}_{cb} = 0.051$. Covariates tighten the bounds at every budget above $\delta = 0.1$, shrinking their width at $\delta = 1$ by 28%, and increase the breakdown budget to 0.29. The sensitivity exercise thus ranks the two waves: TOP 2020 is the more fragile one, even though its difference in means among respondents is larger. In hindsight, this is not surprising, as TOP 2020 ran during the school closures and is also the wave with the larger response differential across arms. Neither the point estimates nor the Lee bounds, which do not sign the effect in either wave, can make this distinction.

Table 2 reports, for each specification, some of the summary measures of Section 5.2: the breakdown budget $\hat{\delta}_b$, the certified breakdown budget $\hat{\delta}_{cb}$ at the 5% level, the shadow price of robustness at the breakdown budget $\hat{\eta}(\hat{\delta}_b)$, and the certification frontier evaluated five percentage points above the certified budget, together with its implied cost (see the table notes). Three comments are in order. First, in both waves the certified budget is far below the breakdown budget, so that most of the apparent fragility is due to sampling uncertainty, not to misspecification uncertainty. Second, certifying five additional percentage points of robustness requires roughly 8,000 additional students (\$80k per point) in TOP 2020, but only about 1,200 (\$12k per point) in TOP 2022, whose smaller complier share makes the bounds widen more slowly. The frontier thus answers the second round budgeting question raised in Section 5.2 and tells a decision maker funding a hypothetical second round where an additional dollar buys the most certified robustness. Third, the shadow price $\hat{\eta}(\hat{\delta}_b)$

¹⁷We explain in detail how to add covariates in Section SA6.1.1 of the supplementary materials.

is minus the slope of the estimated lower bound in Figure 2 at the budget where it crosses zero. At the breakdown budget, each additional percentage point of compliers departing from MCAR lowers the worst case by 0.017 standard deviations in TOP 2020, against 0.005 in TOP 2022. As shadow prices are nonincreasing in the budget, the lower bound of TOP 2020 falls at least this fast at every smaller budget, while that of TOP 2022 never falls faster than its initial shadow price of 1.00. This is why TOP 2020 breaks down first despite its larger difference in means.

Table 2: *Summary measures in the Tutoring Online Program.*

	$\hat{\delta}_b$	$\hat{\delta}_{cb}$	$\hat{\eta}(\hat{\delta}_b)$	n	n^*	Δn	\$/pp
TOP 2020	0.098	0.017	1.74	1,059	9,048	+7,989	80k
TOP 2022	0.247	0.051	0.49	1,137	2,326	+1,189	12k

Notes: $\hat{\delta}_b$ is the plug-in breakdown budget (5.1) and $\hat{\delta}_{cb}$ the certified breakdown budget (5.2) at $\alpha = 0.05$, both without covariates; $\hat{\eta}(\hat{\delta}_b)$ is the estimated shadow price of robustness at $\hat{\delta}_b$; n is the realized sample size and n^* the certification frontier (5.3) evaluated at $\hat{\delta}_{cb} + 0.05$, rounded up, with $\Delta n = n^* - n$; \$/pp is the implied cost per percentage point of certified robustness at a marginal cost of \$50 per participant enrolled and surveyed (Puri and Rathinam, 2019).

Finally, we use the shadow price of robustness to offer a different comparison of the two waves. Consider a decision maker who has no information on the extent of the departure from MCAR and holds uniform beliefs μ on $[0, 1]$, so that $F_\mu(s) = s$ and the evaluation of Section 5.2 reads $\mathbb{E}_\mu[\tau(\delta)] = \tau(0) - \int_0^1 \eta(s)(1 - s) ds$. We compute the shadow prices as the slopes of the plug-in lower bound, without covariates, between consecutive budgets of the grid of Figure 2. In TOP 2020 the shadow price starts at 2.74, so that each of the first two percentage points of compliers departing from MCAR lowers the worst case by about 0.027 standard deviations, and the weighted integral equals 0.556, which gives $\mathbb{E}_\mu[\hat{\tau}(\delta)] = 0.207 - 0.556 = -0.349$. In TOP 2022 the shadow price starts at 1.00 and the integral equals 0.237, which gives $\mathbb{E}_\mu[\hat{\tau}(\delta)] = 0.176 - 0.237 = -0.061$. The decision maker thus ranks the waves in the opposite order to their differences in means because the larger point estimate of TOP 2020 is more than offset by a shadow price that starts almost three times as large.

6.2 Counterfactuals in Structural Models

We revisit one of the leading applications of CC23: the sensitivity of the marital college premium in the matching model of Chiappori et al. (2017) to the Gumbel specification of the taste shocks.

Setup. For illustrative purposes, we use the data for a single cohort instead of the 28 separate cohorts of CC23, namely cohort 14, composed of white men born in 1953 and white wives one year younger. There are $J = 5$ education groups (high school dropouts, high school graduates, some college, college graduates, and postgraduates), and we measure the premium between college graduate (b) and some college (b') women. We set $\Theta = [-12, 12]^{2J}$ and plug in the estimated matching frequencies for ϱ . At the Gumbel baseline the premium is -0.221 . We obtain confidence bounds from the version of Algorithm 1 where P_* is known. Expectations over P_* are approximated

by 5,000 scrambled Halton draws, while the outer optimization problem is solved with ten random restarts per bound and warm starts along the δ -grid. We compare two sensitivity exercises using the same data, moments, and simulation draws. The first reports the bounds of CC23, which let the candidate distributions range over the robustness set

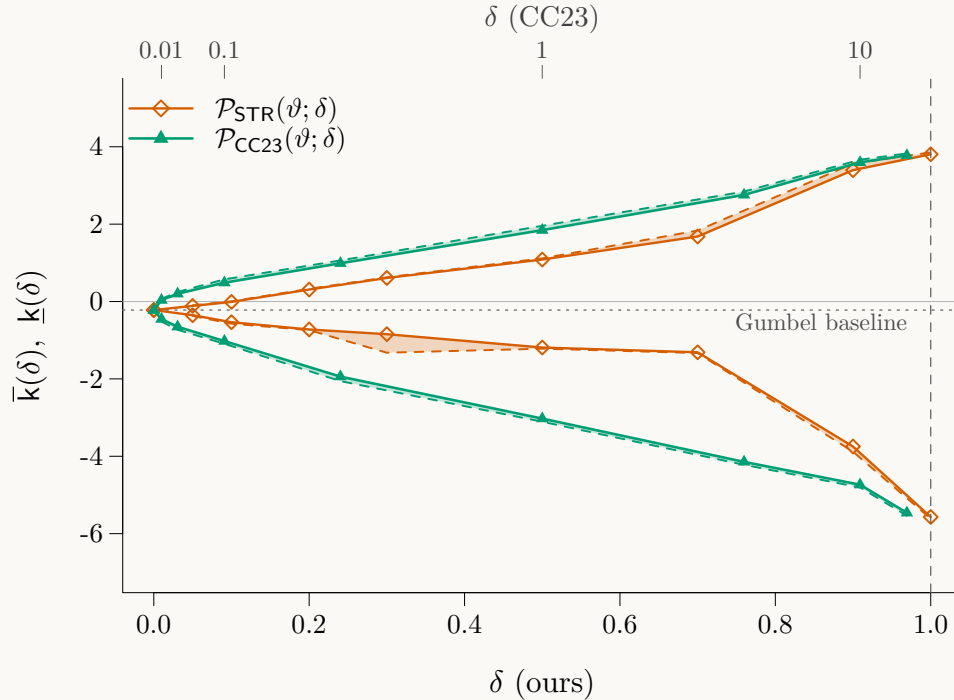
$$\mathcal{P}_{\text{CC23}}(\vartheta; \delta) := \left\{ P \in \mathcal{P}(\mathbb{R}^{J+1}) : P \ll P_*, \mathbb{E}_P[s(U; \vartheta)] = (\varrho, 0_{2(J+1)}), \right. \\ \left. \mathbb{E}_{P_*} \left[\phi_{\text{CC23}} \left(\frac{dP}{dP_*} \right) \right] \leq \delta \right\}, \quad (6.1)$$

where

$$\phi_{\text{CC23}}(x) := \begin{cases} x \log x - x + 1, & 0 \leq x \leq e, \\ \frac{1}{2e}(x - e)^2 + (x - e) + 1, & x > e. \end{cases}$$

In words, the divergence used to measure distances between a candidate P and the baseline P_* coincides with the Kullback–Leibler divergence for $x \leq e$ and continues with a χ^2 -type quadratic tail for $x > e$. CC23 adopt this hybrid because one of their regularity conditions fails for pure Kullback–Leibler at the Gumbel baseline. The second exercise delivers the bounds $\underline{k}(\delta)$ and $\bar{k}(\delta)$ of Example 2.2 over the robustness set $\mathcal{P}_{\text{STR}}(\vartheta; \delta)$ in (2.7), in which $P = (1 - \delta)P_* + \delta R$, so that δ is the share of women drawing non-Gumbel shocks.

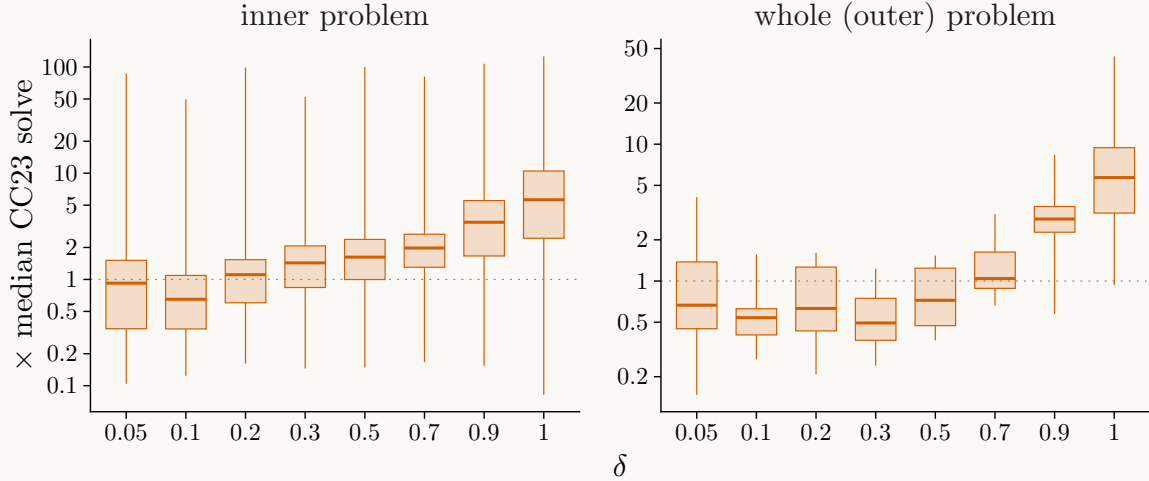
Figure 3: *Sensitivity of the marital college premium on a common axis.*



Notes: Computed from the cohort-14 marriage probabilities of white couples in the data of Chiappori et al. (2017), as prepared by the replication package of CC23. The figure plots the bounds $\underline{k}(\delta)$ and $\bar{k}(\delta)$ of Example 2.2 over the robustness set $\mathcal{P}_{\text{STR}}(\vartheta; \delta)$ of (2.7) (solid orange lines) and the bounds of CC23 over their robustness set $\mathcal{P}_{\text{CC23}}(\vartheta; \delta)$ of (6.1) (solid green lines). Our bounds are plotted against the bottom x -axis, the budget δ of (2.7). The CC23 bounds are drawn at the horizontal position $\delta/(1 + \delta)$, which maps their budget $\delta \in [0, \infty)$ onto $[0, 1]$, and the top x -axis reports the original values of δ . Dashed lines are 95% confidence bounds from 1,000 draws of Algorithm 1.

Results. Figure 3 plots the two sets of bounds on a common axis. Three comments are in order. First, the sign of the premium is fragile. The upper bound $\bar{k}(\delta)$ is still negative at $\delta = 0.1$, where it equals -0.010 , but it is already positive at $\delta = 0.2$, where it equals 0.310 . Hence, a share of women between a tenth and a fifth drawing non-Gumbel shocks suffices to overturn the sign of the baseline premium. Second, by Theorem 4, $\underline{k}(1)$ and $\bar{k}(1)$ are the sharp nonparametric bounds on the premium. As expected, the CC23 bounds approach the same envelope only in the large budget limit, without attaining it at any finite budget, coming very close at a budget of roughly 100, the largest value CC23 use in their application. Since our bounds attain the envelope, they can also be used as a diagnostic for how well a divergence-based approximation does at any given budget. Third, along the whole path sampling uncertainty is an order of magnitude smaller than misspecification uncertainty, suggesting that ambiguity, not noise, drives the width of the bounds. Figure 4 compares solve times. A CC23 inner solve takes on the order of a hundredth of a second at 5,000 draws, and the outer problem for one bound a few seconds (median between 0.4 and 5.3 seconds). Our programs solve in comparable time: up to $\delta = 0.7$, the median inner solve takes less than twice as long as that of CC23, while the median whole problem takes roughly the same time. Both ratios rise to between five and six at $\delta = 1$. Computing one bound over the eight-point budget grid takes about 3 minutes against 2.5 for CC23.

Figure 4: *Computational cost of the sensitivity bounds.*



Notes: Computed from the replication data of CC23, sharing the simulation draws of Figure 3. The panels report solve times for the bounds over the robustness set $\mathcal{P}_{\text{STR}}(\vartheta; \delta)$ of (2.7), for the inner problem at a given ϑ (left panel) and for the whole problem (right panel), in multiples of the median solve time of the corresponding CC23 problem (dotted horizontal line) using $\mathcal{P}_{\text{CC23}}(\vartheta; \delta)$ of (6.1), over the budget grid $\delta \in \{0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 0.9, 1\}$, displayed at equally spaced positions. All programs are solved with Knitro 15.1 in Julia 1.12 on an Apple M5 laptop.

6.3 Formula Instruments

We revisit one of the leading applications of BH23: the effect of market access growth on the growth of urban employment across Chinese prefectures between 2007 and 2016.

Setup. BH23 exploit the construction of the country’s high-speed rail network. The market access of prefecture $i \in [n]$ in year $t \in \{2007, 2016\}$ is

$$\text{MA}_{it} \equiv \text{MA}_i(v_t) = \sum_{j \in [n]} \frac{\text{Pop}_j}{\text{Time}(v_t, \text{Loc}_i, \text{Loc}_j)},$$

where Pop_j is the population of prefecture j in the 2000 census and $\text{Time}(\cdot)$ is the predicted travel time between the locations Loc_i and Loc_j given the status v_t of the railway lines in year t . The endogenous regressor is $x_i = \log \text{MA}_{i,2016} - \log \text{MA}_{i,2007}$, the outcome y_i is the growth of urban employment over the same period, and the sample consists of the $n = 275$ prefectures for which the employment outcome is recorded. Because market access grew most in the economic center of the country, a comparison with the periphery may be confounded by unobserved determinants of employment growth.

BH23 therefore use the timing of construction, rather than the location of the network, as the source of exogenous variation. The shocks $v \in \{0, 1\}^K$ are the 2016 completion indicators of the $K = 149$ lines that were in service by 2016 (83) or completed or under construction by 2019 (66), and $\mathcal{V} = \{v^{(1)}, \dots, v^{(S)}\}$ collects the $S = 1,999$ counterfactual networks obtained by permuting completion status within groups of lines that connect the same number of prefectures. The formula of Example 2.3 is $f_i(v'; w) = \log \text{MA}_i(v') - \log \text{MA}_{i,2007}$, where $w = ((\text{Loc}_j, \text{Pop}_j)_{j \in [n]}, v_{2007})$ is predetermined, so that $z_i = f_i(\check{v}; w) = x_i$ at the realized shocks $\check{v}$. Following BH23, $\tilde{P}$ is the uniform distribution on $\mathcal{V}$, so that $\mu_i(\tilde{P}) = S^{-1} \sum_{s=1}^S f_i(v^{(s)}; w)$. We consider both specifications of BH23, the first controlling for a constant only and the second adding distance to Beijing, latitude, and longitude. Recentering lowers the estimated elasticity from 0.232 to $\hat{\beta}_* = 0.084$ without controls and from 0.133 to $\hat{\beta}_* = 0.056$ with them. Condition (2.12) holds in both specifications. The recentered first stage at the postulated assignment process is $G_x(\tilde{P}) = 9.92$ without controls and 9.24 with them, whereas $\sup_{v' \in \mathcal{V}} g_x(v') - \mathbb{E}_{\tilde{P}}[g_x(v)]$ equals 8.34 and 5.66, respectively.

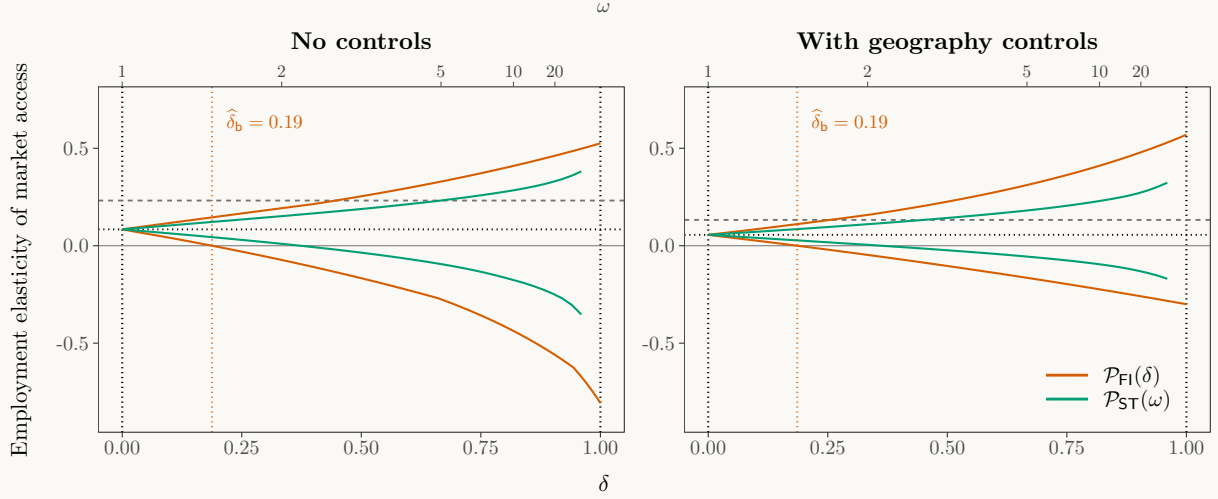
Figure 5 reports two types of bounds. The first are our sensitivity bounds $\underline{\beta}(\delta)$ and $\overline{\beta}(\delta)$ over the robustness set $\mathcal{P}_{\text{FI}}(\delta)$ of (2.11) (Proposition SA-37 in the supplementary materials characterizes them and their computation). The second are the bounds of Sheng and Torgovitsky (2026) (ST26, henceforth), who ask the same question for this very application and vary the assignment process over what they call “the joint sensitivity set”

$$\mathcal{P}_{\text{ST}}(\omega) := \left\{ P \in \Delta(\mathcal{V}) : P \ll \tilde{P}, \omega^{-1} \leq dP/d\tilde{P} \leq \omega, \tilde{P}\text{-a.s.} \right\}, \quad (6.2)$$

for $\omega \geq 1$, which rescales the postulated probability of each configuration by a factor of at most ω in either direction. Like ours, and unlike their marginal sensitivity set, this set constrains the law of the configurations.

Results. Three features of this exercise are worth commenting on. First, the sign of the estimate is fragile: the sensitivity bounds contain zero from a budget of $\hat{\delta}_b = 0.19$ onwards, both without and with controls. In other words, a mechanism other than the postulated one governing roughly a fifth of the shock draws suffices for the estimated employment effect to lose its sign. Second, the sensitivity bounds include the unadjusted least squares estimate at $\delta = 0.45$ without controls, and

Figure 5: *Sensitivity bounds for the employment effect of market access growth.*



Notes: Computed from the replication data of BH23. Each panel plots the bounds $\underline{\beta}(\delta)$ and $\overline{\beta}(\delta)$ of Example 2.3 over the robustness set $\mathcal{P}_{FI}(\delta)$ of (2.11) (solid orange lines), and the ST26 bounds over the set (6.2) (solid green lines), the latter drawn against the rescaling $(\omega - 1)/(\omega + 1)$ of their budget, whose original values are reported on the top horizontal axis. The dotted horizontal line is the recentered estimate $\hat{\beta}_*$ of (2.9) and the dashed horizontal line the unadjusted OLS estimate. The orange vertical dotted line marks the estimated breakdown budget $\hat{\delta}_b$ for the sign of the effect, defined for $\hat{\beta}_*$ as in (5.1). The bounds carry no sampling uncertainty (see Remark SA-18 in the supplementary materials).

already at $\delta = 0.25$ with them. Put differently, the gap between the OLS and the recentered IV estimates, which is what the recentering is designed to remove, can be reproduced by assignment processes that are not far from the postulated one. Third, the ST26 bounds are nested in ours even though they need not be in general. In this application, the nesting reflects the difference in the least (best) favorable assignment processes attaining each type of bound: the distributions attaining ours concentrate mass on a few configurations, which the multiplicative constraint in (6.2) rules out by capping the reweighting of every configuration at the factor ω .

7 Conclusion

This paper developed a framework for global sensitivity analysis in which the sensitivity parameter δ is the share of the population allowed to violate a distributional assumption. The sensitivity bounds are the values of finite-dimensional programs, and at $\delta = 1$ they coincide with the exact nonparametric sharp bounds. We proposed plug-in estimators of the bounds and established their consistency, limit distribution, and the validity of a bootstrap procedure. We also proposed summary measures that separate misspecification from sampling uncertainty and make studies comparable. Applications to attrition in randomized experiments, counterfactuals in structural models, and formula instruments illustrate the questions these tools answer: which of two experiments to trust, how much of a study's fragility is due to sampling uncertainty, and what a second round would have to spend to certify more robustness.

References

- Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools,” *Journal of Political Economy*, 2005, 113 (1), 151–184.
- Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro, “Measuring the Sensitivity of Parameter Estimates to Estimation Moments,” *Quarterly Journal of Economics*, 2017, 132 (4), 1553–1592.
- , —, and —, “On the Informativeness of Descriptive Statistics for Structural Estimates,” *Econometrica*, 2020, 88 (6), 2231–2258.
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin, “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association*, 1996, 91 (434), 444–455.
- Armstrong, Timothy B. and Michal Kolesár, “Sensitivity Analysis Using Approximate Moment Condition Models,” *Quantitative Economics*, 2021, 12 (1), 77–108.
- Bollinger, Christopher R., Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak, “Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch,” *Journal of Political Economy*, 2019, 127 (5), 2143–2185.
- Bonhomme, Stéphane and Martin Weidner, “Minimizing Sensitivity to Model Misspecification,” *Quantitative Economics*, 2022, 13 (3), 907–954.
- Bonvini, Matteo and Edward H. Kennedy, “Sensitivity Analysis via the Proportion of Unmeasured Confounding,” *Journal of the American Statistical Association*, 2022, 117 (539), 1540–1550.
- Borusyak, Kirill and Peter Hull, “Nonrandom Exposure to Exogenous Shocks,” *Econometrica*, 2023, 91 (6), 2155–2185.
- , —, and Xavier Jaravel, “Design-Based Identification with Formula Instruments: A Review,” *Econometrics Journal*, 2025, 28 (1), 83–108.
- Carlana, Michela and Eliana La Ferrara, “Apart but Connected: Online Tutoring, Cognitive Outcomes, and Soft Skills,” *American Economic Review*, 2025, 115 (10), 3487–3513.
- Chiappori, Pierre-André, Bernard Salanié, and Yoram Weiss, “Partner Choice, Investment in Children, and the Marital College Premium,” *American Economic Review*, 2017, 107 (8), 2109–2167.
- Christensen, Peter, Gustavo Nino, and Adam Osman, “The Demand for Mobility: Evidence from an Experiment with Uber Riders,” *Journal of Political Economy*, 2026. Forthcoming.
- Christensen, Timothy and Benjamin Connault, “Counterfactual Sensitivity and Robustness,” *Econometrica*, 2023, 91 (1), 263–298.
- Cinelli, Carlos and Chad Hazlett, “Making Sense of Sensitivity: Extending Omitted Variable Bias,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2020, 82 (1), 39–67.
- Crippa, Giulia, Florian Berg, and Roberto Rigobon, “Corporate Omissions: Correcting the Bias in Carbon Reporting,” 2025. SSRN Working Paper 5503418.
- Diegert, Paul, Matthew A. Masten, and Alexandre Poirier, “Assessing Sensitivity to IV Exclusion and Exogeneity without First Stage Monotonicity,” 2026. arXiv:2604.07604.
- Donoho, David L. and Peter J. Huber, “The Notion of Breakdown Point,” in Peter J. Bickel, Kjell A. Doksum, and Joseph L. Hodges, Jr., eds., *A Festschrift for Erich L. Lehmann*, Belmont, CA: Wadsworth, 1983, pp. 157–184.
- Dümbgen, Lutz, “On Nondifferentiable Functions and the Bootstrap,” *Probability Theory and Related Fields*, 1993, 95 (1), 125–140.
- Dutz, Deniz, Ingrid Huitfeldt, Santiago Lacouture, Magne Mogstad, Alexander Torgovitsky, and Winnie van Dijk, “Selection in Surveys: Using Randomized Incentives to Detect and Account for Nonresponse Bias,” *Review of Economic Studies*, 2026. Forthcoming.
- Ekeland, Ivar, Alfred Galichon, and Marc Henry, “Optimal transportation and the falsifiability of incompletely specified economic models,” *Economic Theory*, 2010, 42 (2), 355–374.
- Fang, Zheng and Andres Santos, “Inference on Directionally Differentiable Functions,” *Review of Economic Studies*, 2019, 86 (1), 377–412.
- Gu, Jiaying and Thomas Russell, “Wasserstein-Robust Counterfactuals,” 2024. SSRN Working Paper 4517842.
- Hampel, Frank R., “Contributions to the Theory of Robust Estimation.” PhD dissertation, University of California, Berkeley 1968.
- Hansen, Lars Peter and Thomas J. Sargent, “Robust Control and Model Uncertainty,” *American Economic Review*, 2001, 91 (2), 60–66.
- Horowitz, Joel L. and Charles F. Manski, “Identification and Robustness with Contaminated and Corrupted Data,” *Econometrica*, 1995, 63 (2), 281–302.
- Huber, Peter J., “Robust Estimation of a Location Parameter,” *Annals of Mathematical Statistics*, 1964, 35 (1), 73–101.
- Imbens, Guido W., “Sensitivity to Exogeneity Assumptions in Program Evaluation,” *American Economic Review*, 2003, 93 (2), 126–132.
- Kitamura, Yuichi, Taisuke Otsu, and Kirill Evdokimov, “Robustness, Infinitesimal Neighborhoods, and Moment Restrictions,” *Econometrica*, 2013, 81 (3), 1185–1201.
- Kline, Patrick and Andres Santos, “Sensitivity to Missing Data Assumptions: Theory and an Evaluation of the U.S. Wage Structure,” *Quantitative Economics*, 2013, 4 (2), 231–267.

- Kuhn, Daniel, Soroosh Shafiee, and Wolfram Wiesemann**, “Distributionally Robust Optimization,” *Acta Numerica*, 2025, *34*, 579–804.
- Leamer, Edward E.**, “Let’s Take the Con Out of Econometrics,” *American Economic Review*, 1983, *73* (1), 31–43.
- , “Sensitivity Analyses Would Help,” *American Economic Review*, 1985, *75* (3), 308–313.
- Lee, David S.**, “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects,” *Review of Economic Studies*, 2009, *76* (3), 1071–1102.
- Lieber, Jonas, Pietro Tebaldi, and Alexander Torgovitsky**, “Counterfactuals in Discrete Choice Models under Flexible Assumptions on Random Utility,” 2026. Working paper, Imperial College London, Columbia University, and University of Chicago.
- Manski, Charles F.**, “The Selection Problem,” in Christopher A. Sims, ed., *Advances in Econometrics: Sixth World Congress*, Econometric Society Monographs, Cambridge: Cambridge University Press, 1994, pp. 143–170.
- , *Partial Identification of Probability Distributions*, New York: Springer, 2003.
- Masten, Matthew A., Alexandre Poirier, and Linqi Zhang**, “Assessing Sensitivity to Unconfoundedness: Estimation and Inference,” *Journal of Business & Economic Statistics*, 2024, *42* (1), 1–13.
- , —, and **Muyang Ren**, “A General Approach to Relaxing Unconfoundedness,” 2025. arXiv:2501.15400.
- and —, “Identification of Treatment Effects under Conditional Partial Independence,” *Econometrica*, 2018, *86* (1), 317–351.
- and —, “Inference on Breakdown Frontiers,” *Quantitative Economics*, 2020, *11* (1), 41–111.
- and —, “Salvaging Falsified Instrumental Variable Models,” *Econometrica*, 2021, *89* (3), 1449–1469.
- Mu, Yingfei, Edward Rubin, and Eric Yongchen Zou**, “What’s Missing in Environmental Self-Monitoring: Evidence from Strategic Shutdowns of Pollution Monitors,” *Review of Economics and Statistics*, 2026, *108* (3), 597–612.
- Mukhin, Yaroslav**, “Sensitivity of Regular Estimators,” 2018. arXiv:1805.08883.
- Newey, Whitney K. and Daniel McFadden**, “Large Sample Estimation and Hypothesis Testing,” in Robert F. Engle and Daniel McFadden, eds., *Handbook of Econometrics*, Vol. 4, Amsterdam: Elsevier, 1994, pp. 2111–2245.
- Ober-Reynolds, Daniel**, “Robustness to Missing Data: Breakdown Point Analysis,” *Journal of Econometrics*, 2026, *253*, 106151.
- Puri, Jyotsna and Francis Rathinam**, “Challenges in Real-World Impact Evaluations: Some Learning on Costs and Timeliness,” 2019. IEU Working Paper 3, Independent Evaluation Unit, Green Climate Fund.
- Rambachan, Ashesh and Jonathan Roth**, “A More Credible Approach to Parallel Trends,” *Review of Economic Studies*, 2023, *90* (5), 2555–2591.
- Rockafellar, R. Tyrrell**, *Convex Analysis*, Princeton: Princeton University Press, 1970.
- , *Conjugate Duality and Optimization*, Philadelphia: Society for Industrial and Applied Mathematics, 1974.
- and **Roger J.-B. Wets**, *Variational Analysis*, Berlin: Springer, 2009.
- Rosenbaum, Paul R. and Donald B. Rubin**, “Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 1983, *45* (2), 212–218.
- Rubin, Donald B.**, “Inference and Missing Data,” *Biometrika*, 1976, *63* (3), 581–592.
- Schennach, Susanne M.**, “Entropic Latent Variable Integration via Simulation,” *Econometrica*, 2014, *82* (1), 345–385.
- Semenova, Vira**, “Generalized Lee Bounds,” *Journal of Econometrics*, 2025, *251*, 106055.
- Shao, Yulin, Liangbo Lyu, Menggang Yu, and Bingkai Wang**, “How Should Covariates Be Handled in Randomized Trials? Empirical Evidence from 50 Trials and Recommendations for Practice,” *Journal of Clinical Epidemiology*, 2026, *197*, 112374.
- Shapiro, Alexander**, “Asymptotic Analysis of Stochastic Programs,” *Annals of Operations Research*, 1991, *30* (1), 169–186.
- Sheng, Peizan and Alexander Torgovitsky**, “What’s the Magic Formula Instrument?,” 2026. arXiv:2606.21569.
- Tamer, Elie**, “Sensitivity Analysis in Some Econometric Models,” 2015. Cowles Lecture, 11th World Congress of the Econometric Society, Montréal.
- U.S. Congress**, “Consolidated Appropriations Act, 2026,” 2026. Public Law 119–75, 140 Stat. 173 (H.R. 7148), enacted February 3, 2026.

Supplemental Appendix for “Sensitivity Analysis in Population Shares”

Filippo Palomba

Department of Economics, Princeton University

“Statistical inferences are based only in part upon the observations. An equally important base is formed by prior assumptions about the underlying situation.”

Peter J. Huber, *Robust Statistics*

“A fragile inference is not worth taking seriously. All scientific disciplines routinely subject their inferences to studies of fragility. Why should economics be different?”

Edward E. Leamer, *Sensitivity Analyses Would Help*

Abstract

This supplement collects all proofs, additional results, and technical details on convex duality, feasibility, estimation, inference, summary measures, and applications of sensitivity analysis under total variation and contamination neighborhoods. Section SA1 fixes notation and sets up the statistical framework used throughout. Section SA2 develops the core duality theory and introduces the total variation and contamination neighborhoods. Section SA3 establishes existence of solutions. Section SA4 develops the large sample theory. Section SA5 introduces the summary measures for sensitivity analysis and derives their large sample properties. Section SA6 works through a sequence of relevant applications. Section SA7 collects two additional empirical applications on attrition. Section SA8 contains all the proofs.

Contents

SA1 Introduction	SA-1
SA1.1 Notation	SA-1
SA1.2 Setup	SA-2
SA1.3 Correspondence with the Paper	SA-4
SA2 Convex Duality	SA-6
SA2.1 A Primer in Convex Analysis	SA-6
SA2.2 Weak Duality	SA-9
SA2.3 Strong Duality	SA-9
SA2.4 Duality Theory Under Mixture Constraints	SA-11
SA2.5 Ignoring the Pointwise Dual Constraint	SA-13
SA2.6 Total Variation and Contamination Neighborhoods	SA-15
SA3 Existence of Solutions to the Inner and Outer Problems	SA-17
SA3.1 Existence of Dual Solutions	SA-17
SA3.2 Existence of Solutions to the Outer Problem	SA-19
SA4 Large Sample Properties	SA-21
SA4.1 A Family of Robustness Sets	SA-21
SA4.2 Regularity Conditions	SA-24
SA4.3 Auxiliary Results	SA-25
SA4.4 Consistency, Directional Differentiability, and Distributional Approximation	SA-27
SA4.5 Bootstrap	SA-28
SA5 Summary Measures for Sensitivity Analysis	SA-31
SA5.1 The Shadow Price of Robustness	SA-31
SA5.2 Breakdown and Certified Breakdown	SA-32
SA5.3 The Certification Frontier	SA-33
SA6 Applications of the Framework	SA-35
SA6.1 RCTs with Attrition	SA-35
SA6.2 Counterfactuals in Structural Models	SA-41

SA6.3 Formula and Shift-Share Instruments	SA-42
SA6.4 Other Applications	SA-45
SA7 Additional Empirical Applications	SA-53
SA7.1 RCT with Attrition I: Christensen, Nino and Osman (2026)	SA-53
SA7.2 RCT with Attrition II: Shao, Lyu, Yu and Wang (2026)	SA-55
SA8 Proofs	SA-58
SA8.1 Proof of Lemma SA-1	SA-58
SA8.2 Proof of Lemma SA-2	SA-58
SA8.3 Proof of Lemma SA-3	SA-58
SA8.4 Proof of Lemma SA-4	SA-58
SA8.5 Proof of Corollary SA-7	SA-59
SA8.6 Proof of Theorem SA-9	SA-59
SA8.7 Proof of Lemma SA-10	SA-62
SA8.8 Proof of Theorem SA-11	SA-63
SA8.9 Proof of Lemma SA-12	SA-64
SA8.10 Proof of Proposition SA-13	SA-65
SA8.11 Proof of Theorem SA-14	SA-65
SA8.12 Proofs of Theorem SA-15 and Proposition SA-16	SA-65
SA8.13 Proof of Lemma SA-17	SA-67
SA8.14 Proof of Lemma SA-18	SA-67
SA8.15 Proof of Theorem SA-19	SA-73
SA8.16 Proof of Lemma SA-20	SA-73
SA8.17 Proof of Lemma SA-21	SA-74
SA8.18 Proof of Theorem SA-22	SA-76
SA8.19 Proof of Lemma SA-23	SA-77
SA8.20 Proof of Lemma SA-24	SA-77
SA8.21 Proof of Lemma SA-25	SA-78
SA8.22 Proof of Lemma SA-26	SA-81
SA8.23 Details on the Construction of Ξ and $\mathcal{H}$	SA-82
SA8.24 Proof of Lemma SA-27	SA-82
SA8.25 Proof of Theorem SA-28	SA-85
SA8.26 Proof of Theorem SA-29	SA-85
SA8.27 Proof of Theorem SA-30	SA-86
SA8.28 Proof of Proposition SA-31	SA-89
SA8.29 Proof of Proposition SA-32	SA-90
SA8.30 Proof of Proposition SA-33	SA-91
SA8.31 Proof of Theorem SA-34	SA-91
SA8.32 Proof of Proposition SA-35	SA-92
SA8.33 Proof of Lemma SA-36	SA-94
SA8.34 Proof of Proposition SA-37	SA-94
SA8.35 Proof of Lemma SA-38	SA-95
SA8.36 Proof of Theorem SA-39	SA-97

SA1 Introduction

This section introduces the notation and sets up the statistical framework we rely on throughout this supplement.

SA1.1 Notation

In this supplement, we let n denote the sample size.

Sets and spaces. We write $\mathbb{R}$ for the real line, $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ for the extended reals, $\mathbb{R}_+$ and $\mathbb{R}_{++}$ for the nonnegative and positive reals, and $\mathbb{N}, \mathbb{Z}$ for the naturals and integers. For $J \in \mathbb{N}$, we write $[J] := \{1, \dots, J\}$. For a closed set $\mathcal{Z}$ of a Euclidean space, we write $\mathcal{F}$ for its Borel σ -algebra, $\Delta(\mathcal{Z})$ for the probability measures on $(\mathcal{Z}, \mathcal{F})$, and $\mathcal{M}_+(\mathcal{Z})$ for the cone of finite nonnegative measures on $(\mathcal{Z}, \mathcal{F})$. We write $C(\mathcal{Z}; \mathbb{R}^k)$ for the space of continuous functions from $\mathcal{Z}$ to $\mathbb{R}^k$, and $C(\mathcal{Z}) := C(\mathcal{Z}; \mathbb{R})$. For $1 \leq p < \infty$, we denote by $L_p(\mu)$ the space of measurable f with $\int_{\mathcal{Z}} |f|^p d\mu < \infty$. A subset K of a real vector space is a cone if $\lambda x \in K$ whenever $x \in K$ and $\lambda > 0$, so that K is a union of half lines emanating from the origin.

Linear algebra. 0_k and 1_k are the k -dimensional vectors of zeros and ones, e_j is the j -th standard basis vector of a Euclidean space whose dimension is clear from the context, and $^\top$ denotes transposition. We write $|\cdot|$ or $\|\cdot\|_2$ for the Euclidean norm and $\|\cdot\|_\infty$ for the sup-norm, and set $[x]_+ = \max\{x, 0\}$ and $[x]_- = \max\{-x, 0\}$. For a point $x \in \mathbb{R}^k$ and a radius $r > 0$, we write

$$\mathcal{B}_r(x) := \{u \in \mathbb{R}^k : \|u - x\|_2 < r\} \quad \text{and} \quad \overline{\mathcal{B}}_r(x) := \{u \in \mathbb{R}^k : \|u - x\|_2 \leq r\}$$

for the open and the closed ball of radius r centered at x , respectively. A neighborhood of x is any set that contains $\mathcal{B}_r(x)$ for some $r > 0$, so that a property holds on a neighborhood of x if and only if it holds on $\mathcal{B}_r(x)$ for some $r > 0$. We use the same notation on any space endowed with a norm or a seminorm other than the Euclidean one, in which case we name the norm or seminorm that replaces $\|\cdot\|_2$ in the two displayed sets. For a point a and a nonempty set B of a Euclidean space, $d(a, B) := \inf_{b \in B} \|a - b\|_2$ is the distance from a to B , and for two nonempty compact sets A, B of a Euclidean space

$$d_H(A, B) := \max \left\{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \right\}$$

is their Hausdorff distance.

Probability and measure. Expectation and variance under a distribution P are $\mathbb{E}_P[\cdot]$ and $\mathbb{V}_P[\cdot]$, respectively, and $\perp\!\!\!\perp$ denotes independence. Without a subscript, $\mathbb{P}$ and $\mathbb{E}$ denote probability and expectation with respect to the joint distribution of the random variables they involve. For two measures, $P \ll Q$ means P is absolutely continuous with respect to Q , with Radon–Nikodym derivative dP/dQ , while $P \perp Q$ means they are mutually singular, i.e. there is a measurable $\mathcal{D} \in \mathcal{F}$ with $P(\mathcal{D}) = 0$ and $Q(\mathcal{Z} \setminus \mathcal{D}) = 0$. The total variation distance between $P, Q \in \Delta(\mathcal{Z})$ is $\text{TV}(P \| Q) := \sup_{\mathcal{A} \in \mathcal{F}} |P(\mathcal{A}) - Q(\mathcal{A})|$, which lies in $[0, 1]$. A point $z \in \mathcal{Z}$ is an atom of $Q \in \Delta(\mathcal{Z})$ if $Q(\{z\}) > 0$, and the support $\text{supp}(\mu)$ of a Borel measure μ on $\mathcal{Z}$ is the smallest closed set of full μ -measure. For an event or condition A , $\mathbb{1}\{A\}$ equals 1 if A holds and 0 otherwise. We write $X \stackrel{d}{=} Y$ when X and Y have the same distribution, and $X | A \stackrel{d}{=} Y | B$ when the conditional distribution of X given A equals that of Y given B . For a distribution P on a closed set $\mathcal{Y} \subseteq \mathbb{R}$, $F_P(y) := P(\mathcal{Y} \cap (-\infty, y])$, $y \in \mathbb{R}$, is its distribution function and $F_P^{-1}(u) := \inf\{y \in \mathcal{Y} : F_P(y) \geq u\}$, $u \in (0, 1)$, is its quantile function. We write $\text{Be}(p)$ for the Bernoulli distribution with success probability $p \in [0, 1]$ and $\text{Unif}([a, b])$ for the uniform distribution on an interval $[a, b] \subset \mathbb{R}$ with $a < b$. Finally, δ_x^{Dirac} denotes the Dirac measure at a point x .

Convex analysis. For a set $C \subseteq \mathbb{R}^k$ we write $\text{co}(C)$ for its convex hull, the set of finite convex combinations of points of C , and $\text{aff}(C)$ for its affine hull. The interior, closure, and relative interior of C are

$$\begin{aligned} \text{int}(C) &:= \{x : \exists \varepsilon > 0, \mathcal{B}_\varepsilon(x) \subseteq C\}, & \text{cl}(C) &:= \bigcap \{F \supseteq C : F \text{ closed}\}, \\ \text{ri}(C) &:= \{x \in \text{aff}(C) : \exists \varepsilon > 0, \mathcal{B}_\varepsilon(x) \cap \text{aff}(C) \subseteq C\}. \end{aligned}$$

For $f : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$, the effective domain and the subdifferential are

$$\text{dom}(f) := \{z \in \mathbb{R}^k : f(z) < \infty\}, \quad \partial f(z) := \{v \in \mathbb{R}^k : f(z') \geq f(z) + v^\top(z' - z) \ \forall z' \in \mathbb{R}^k\},$$

and f is proper if $\text{dom}(f) \neq \emptyset$ and $f(z) > -\infty$ for all $z \in \mathbb{R}^k$. For a function f defined on an interval of $\mathbb{R}$, we write $\partial_+ f(x)$ and $\partial_- f(x)$ for its right and left derivatives at x , that is, the limits of $\{f(x+h) - f(x)\}/h$ as $h \downarrow 0$ and as $h \uparrow 0$, respectively, whenever they exist. Finally, clf is the closure (lower semicontinuous hull) of f , and f is closed if $\text{clf} = f$, which for a proper convex function is equivalent to lower semicontinuity, and P -esssup denotes the essential supremum with respect to the distribution P .

Asymptotic statements. We write $x_n \rightarrow x$ for limits, including one-sided ones, and use $x_n \downarrow x$ or $x_n \uparrow x$ only when the direction of approach matters for the argument. For positive sequences $\{a_n\}, \{b_n\}$ we write $a_n = O(b_n)$ if $a_n \leq M b_n$ for some M and all large n , and $a_n = o(b_n)$ if $a_n/b_n \rightarrow 0$. For sequences of random variables we write $A_n = o_{\mathbb{P}}(B_n)$ and $A_n = O_{\mathbb{P}}(B_n)$ for the usual stochastic-order analogs. We denote convergence in probability by $\xrightarrow{\mathbb{P}}$ and convergence in distribution by $\rightsquigarrow$, and write $\mathbf{N}(a, B)$ for a (possibly multivariate) Gaussian with mean a and variance-covariance B .

Empirical process theory. For a probability measure P on $\mathcal{Z}$ and a P -integrable $f : \mathcal{Z} \rightarrow \mathbb{R}$ we write $Pf := \int_{\mathcal{Z}} f dP$, so that $\mathbb{E}_P[f] = Pf$. For a class $\mathcal{H}$ of functions and two measures P, P' we write $\|P - P'\|_{\mathcal{H}} := \sup_{f \in \mathcal{H}} |Pf - P'f|$, and $\ell^\infty(\mathcal{H})$ denotes the space of bounded real functions on $\mathcal{H}$ equipped with the supremum norm. Given observations $Z_1, \dots, Z_n$ drawn i.i.d. from P , we write $\hat{P}_n := n^{-1} \sum_{i=1}^n \delta_{Z_i}^{\text{Dirac}}$ for their empirical measure and $\mathbb{G}_n := \sqrt{n}(\hat{P}_n - P)$ for the empirical process, viewed as the random map $f \mapsto \mathbb{G}_n f = \sqrt{n}(\hat{P}_n f - Pf)$. Following Section 2.1 in van der Vaart and Wellner (1996), a class $\mathcal{H} \subseteq L_2(P)$ of measurable functions with $\sup_{f \in \mathcal{H}} |f(z) - Pf| < \infty$ for every $z \in \mathcal{Z}$ is called P -Donsker if $\mathbb{G}_n \rightsquigarrow \mathbb{G}$ in $\ell^\infty(\mathcal{H})$, where $\rightsquigarrow$ is weak convergence in the sense of Hoffmann-Jørgensen and the limit $\mathbb{G}$ is a mean zero Gaussian process with covariance $\mathbb{E}[\mathbb{G}f \mathbb{G}f'] = P(ff') - PfPf'$, whose sample paths lie almost surely in $UC(\mathcal{H}, e_P)$, the space of bounded functions on $\mathcal{H}$ uniformly continuous for the intrinsic semimetric $e_P(f, f')^2 := \mathbb{V}_P(f - f')$. The class $\mathcal{H}$ is P -Glivenko–Cantelli if $\|\hat{P}_n - P\|_{\mathcal{H}} \rightarrow 0$ in outer probability.

SA1.2 Setup

Let $\Theta \subseteq \mathbb{R}^{d_\theta}$, $d_\theta \in \mathbb{N}$, be the parameter space and $\mathcal{Z} \subseteq \mathbb{R}^{d_z}$, $d_z \in \mathbb{N}$, a closed set collecting the possible realizations of a random vector $Z \in \mathcal{Z}$. We equip $\mathcal{Z}$ with its Borel σ -algebra $\mathcal{F}$. The econometrician postulates that Z is distributed according to a baseline distribution $P_* \in \mathcal{P}(\mathcal{Z})$. The class $\mathcal{P}(\mathcal{Z}) \subseteq \Delta(\mathcal{Z})$, formally defined below in (1) once the estimand and moment function have been introduced, collects the Borel probability distributions under which the absolute value of the estimand's integrand and the norm of the moment function are integrable. We impose this restriction so that the expectations, moment restrictions, optimization problems, and associated quantities introduced below are well defined.

The baseline P_* is the model one would use absent any concern for misspecification, and the analysis asks how a conclusion drawn from P_* moves as the distribution is allowed to depart from it while still satisfying the model restriction.

Remark SA-1 (Interpretation of the baseline). We denote the baseline distribution by P_* rather than by P_0 , which we reserve for the distribution that generates the data. The starred notation emphasizes that the baseline plays a role analogous to that of a pseudo-true value in maximum likelihood theory: it is the model the econometrician posits, and it need not coincide with the truth. In the RCTs with attrition example of Section SA6.1, for instance, the joint distribution of potential outcomes, response indicators, and treatment assignment is P_0 , and the data are drawn from the distribution of observables that P_0 induces. The baseline is P_T , the identified distribution of the outcome among treated respondents, and the point identifying assumption that observations are missing completely at random is the equality $P_C = P_T$ between the unidentified outcome distribution for the compliers and the baseline. The sensitivity exercise is accordingly conducted around P_T , which incorporates the econometrician's point identifying assumption, and not around P_0 . In the counterfactuals in structural models example of Section SA6.2 the baseline is instead the distribution that the econometrician imposes on the latent idiosyncratic utility shocks. ♣

Estimand. The quantity of interest is a scalar *estimand*, obtained by integrating an $\mathcal{F}$ -measurable *integrand* $g : \mathcal{Z} \times \Theta \rightarrow \mathbb{R}$ against the distribution of the data,

$$\mathbb{E}_P[g(Z; \theta)] = \int_{\mathcal{Z}} g(z; \theta) dP(z).$$

A leading case is a counterfactual or policy relevant quantity reported by the analyst, with θ indexing the structural parameter on which g depends.

Model restrictions. The model restricts the parameter through moment conditions. Let $m : \mathcal{Z} \times \Theta \rightarrow \mathbb{R}^{d_m}$ be an $\mathcal{F}$ -measurable *moment function*. We partition the moment function into four blocks,

$$m(\cdot; \theta) = (m_1(\cdot; \theta)^\top, m_2(\cdot; \theta)^\top, m_3(\cdot; \theta)^\top, m_4(\cdot; \theta)^\top)^\top,$$

where $m_i : \mathcal{Z} \times \Theta \rightarrow \mathbb{R}^{d_i}$, $d_i \in \{0, 1, \dots\}$, and $d_m = \sum_{i=1}^4 d_i$. The first and third blocks contain moment inequalities, whereas the second and fourth contain moment equalities. We collect the right-hand sides of the first two blocks in $\rho = (\rho_1^\top, \rho_2^\top)^\top \in \mathbb{R}^{d_\rho}$, where $\rho_1 \in \mathbb{R}^{d_1}$ and $\rho_2 \in \mathbb{R}^{d_2}$ are unknown vectors of target moments, and $d_\rho = d_1 + d_2$. We also normalize the known right-hand sides of the last two blocks to zero. The moment constraint set is thus

$$\mathcal{M}(\rho) = \left(\bigtimes_{j=1}^{d_1} (-\infty, \rho_{1,j}] \right) \times \{\rho_2\} \times (-\infty, 0]^{d_3} \times \{0\}^{d_4} \subseteq \mathbb{R}^{d_m}.$$

Thus,

$$\mathbb{E}_P[m(Z; \theta)] \in \mathcal{M}(\rho) \iff \begin{pmatrix} \mathbb{E}_P[m_1(Z; \theta)] \leq \rho_1 \\ \mathbb{E}_P[m_2(Z; \theta)] = \rho_2 \\ \mathbb{E}_P[m_3(Z; \theta)] \leq 0 \\ \mathbb{E}_P[m_4(Z; \theta)] = 0 \end{pmatrix},$$

where vector inequalities are interpreted componentwise. We also define $\vec{\rho} := (\rho_1^\top, \rho_2^\top, 0_{d_3}^\top, 0_{d_4}^\top)^\top \in \mathbb{R}^{d_m}$. We write $\rho_0 \in \mathbb{R}^{d_\rho}$ for the population value of the target vector and $\vec{\rho}_0 := (\rho_0^\top, 0_{d_3}^\top, 0_{d_4}^\top)^\top$. Let $\lambda \in \mathbb{R}^{d_m}$ denote the

multiplier attached to the moment restriction in the dual programs of Section SA2. Partitioning it conformably as $\lambda = (\lambda_1^\top, \lambda_2^\top, \lambda_3^\top, \lambda_4^\top)^\top$, define $\lambda_{12} := (\lambda_1^\top, \lambda_2^\top)^\top$ and

$$\mathcal{R}_\lambda := \mathbb{R}_+^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}_+^{d_3} \times \mathbb{R}^{d_4}.$$

For a set $C \subseteq \mathbb{R}^k$, we write ι_C for its convex indicator, which equals 0 on C and $+\infty$ off C , and $\iota_C^*(y) := \sup_{x \in C} y^\top x$ for its support function, which is the convex conjugate of ι_C in the sense of Definition SA-1. The support function of the moment constraint set is

$$\iota_{\mathcal{M}(\rho)}^*(\lambda) = \begin{cases} \lambda_{12}^\top \rho, & \lambda \in \mathcal{R}_\lambda, \\ +\infty, & \lambda \notin \mathcal{R}_\lambda. \end{cases}$$

For the θ under consideration, we write $\mathcal{M}_{+,m}(\mathcal{Z}) \subseteq \mathcal{M}_+(\mathcal{Z})$ for the subcone of the measures ν with finite moment $\int_{\mathcal{Z}} \|m(z; \theta)\| d\nu(z) < \infty$ and we omit the dependence on θ for brevity. When no moment function is present, all four block dimensions are zero. We then use the conventions $m \equiv 0$, $\mathcal{M}(\rho) = \mathbb{R}^0$, and $\mathcal{R}_\lambda = \mathbb{R}^0$, so that the moment restriction is vacuous and its support function contribution is zero. In this case we also suppress m, ρ, λ , and λ_{12} from the notation, interpret their inner products as zero, and set $\mathcal{M}_{+,m}(\mathcal{Z}) = \mathcal{M}_+(\mathcal{Z})$.

For each $\theta \in \Theta$, and suppressing the dependence on θ in the notation, we formally define the class of admissible distributions as

$$\mathcal{P}(\mathcal{Z}) \equiv \mathcal{P}_\theta(\mathcal{Z}) := \{P \in \Delta(\mathcal{Z}) : \mathbb{E}_P[|g(Z; \theta)|] < \infty \text{ and } \mathbb{E}_P[\|m(Z; \theta)\|] < \infty\}. \quad (1)$$

The first integrability requirement makes the estimand finite, while the second makes the moment vector finite. Thus, every expectation entering the moment restrictions and sensitivity bounds is well defined. Throughout this supplement, unless explicitly stated otherwise, we maintain without further mention the measurability and integrability conditions needed for all displayed objects, optimization problems, and invoked results, and omit their routine verification from the proofs. We say that a distribution $P \in \mathcal{P}(\mathcal{Z})$ and a parameter θ are compatible with the model when $\mathbb{E}_P[m(Z; \theta)] \in \mathcal{M}(\rho)$.

Robustness set. Sensitivity to deviations from P_* is gauged by means of a ϕ -divergence. Let $\phi : \mathbb{R} \rightarrow \bar{\mathbb{R}}$ be an entropy function (see Definition SA-4) and $D_\phi(\cdot \| P_*)$ the associated divergence from the baseline (see Definition SA-5). For a sensitivity parameter $\delta \in \mathbb{R}_+$, also referred to as the divergence budget, the *robustness set* collects the distributions that are both compatible with the model and within δ of the baseline in the divergence induced by ϕ :

$$\mathcal{P}_\phi(\theta; \rho, P_*, \delta) := \{P \in \mathcal{P}(\mathcal{Z}) : \mathbb{E}_P[m(Z; \theta)] \in \mathcal{M}(\rho), D_\phi(P \| P_*) \leq \delta\}. \quad (2)$$

For an entropy function ψ and a distribution $P \in \Delta(\mathcal{Z})$, the moment vectors attainable within a budget δ of P form the set

$$\Gamma_\psi(\theta; P, \delta) := \{\mathbb{E}_Q[m(Z; \theta)] : Q \in \mathcal{P}(\mathcal{Z}), D_\psi(Q \| P) \leq \delta\} \subseteq \mathbb{R}^{d_m}, \quad (3)$$

and $\Gamma_\psi^{\leq}(\theta; P, \delta)$ denotes its subset obtained by restricting to the $Q \ll P$. We drop ψ when $\psi = \phi$, P when $P = P_*$, and θ when it is held fixed, so that, for instance, $\Gamma(\delta)$ collects the moment vectors attainable within divergence δ of the baseline. When no moment function is present, we set all these sets equal to $\{0\}$. We do not require the baseline distribution itself to satisfy the moment restrictions at any parameter value. Under total variation, that is, for $\phi = \phi_{\text{TV}}$, where $\phi_{\text{TV}}(s) = \frac{1}{2}|s - 1|$ for $s \geq 0$ is the total variation entropy of Example SA-1, the smallest divergence needed to reconcile the baseline with the model is $\inf_{\theta \in \Theta} D(\theta; \rho, P_*)$, where D is the minimum-divergence function of (30) with $\mathbb{U} = [0, +\infty)$ (see Remark SA-7). When the baseline satisfies the moment restrictions at some parameter value, this quantity is zero, and the usual sensitivity analysis centered at a model-compatible baseline is obtained as a special case.

Sensitivity bounds. Letting both P and the compatible θ vary, the range of the estimand over the robustness set is delimited by its largest and smallest attainable values, namely the lower and upper sensitivity bounds, formally defined as

$$\inf_{\theta \in \Theta} \inf_{P \in \mathcal{P}_\phi(\theta; \rho, P_*, \delta)} \mathbb{E}_P[g(Z; \theta)] \quad \text{and} \quad \sup_{\theta \in \Theta} \sup_{P \in \mathcal{P}_\phi(\theta; \rho, P_*, \delta)} \mathbb{E}_P[g(Z; \theta)], \quad (4)$$

with the understanding that when the robustness set is empty for a given $\theta \in \Theta$, the corresponding infimum or supremum is taken to be $+\infty$ or $-\infty$, respectively, so that, should the set be empty for every $\theta \in \Theta$, the upper bound is $-\infty$ and the lower $+\infty$. For each fixed $\theta \in \Theta$, the inner optimization over $P \in \mathcal{P}_\phi(\theta; \rho, P_*, \delta)$ in (4) is the *inner problem*, whose optimal value is the most extreme value the estimand attains over the robustness set at that θ , while the outer optimization over $\theta \in \Theta$ is the *outer problem*. We use these names for the two programs in (4) throughout. Furthermore, the lower bound can be obtained from the upper bound by replacing the integrand g with $-g$, since $\inf_{\theta, P} \mathbb{E}_P[g(Z; \theta)] = -\sup_{\theta, P} \mathbb{E}_P[-g(Z; \theta)]$. We accordingly develop the theory for the upper bound and recover the lower bound by this substitution.

Addition of “external” nuisance parameters. In some applications, the integrand and the moment function carry an additional nuisance parameter $\varsigma \in \mathcal{W}$, with $\mathcal{W}$ a metric space, not necessarily finite-dimensional, so that the primitives of the problem become $g(z; \theta, \varsigma)$ and $m(z; \theta, \varsigma)$. Leading examples of such a parameter are a law of motion for state variables, a transition or a conditional choice probability in dynamic discrete choice models, or a mixing proportion such as the share of compliers π in the RCTs-with-attrition application of Section SA6.1. Unlike θ , the nuisance parameter is not optimized over. Rather, it has a true value $\varsigma_0 \in \mathcal{W}$, which is consistently estimable from

the data in a preliminary step, separately from the rest of the model, hence the terminology “external” nuisance parameter. The model restrictions are accordingly required to hold at the true value, $\mathbb{E}_P[m(Z; \theta, \varsigma_0)] \in \mathcal{M}(\rho)$, the estimand becomes $\mathbb{E}_P[g(Z; \theta, \varsigma_0)]$, and the optimization in (4) continues to run over (θ, P) only, with the nuisance held fixed at ς_0 . The definition of $\mathcal{P}(\mathcal{Z})$ is then read with $g(\cdot; \theta, \varsigma_0)$ and $m(\cdot; \theta, \varsigma_0)$ in place of $g(\cdot; \theta)$ and $m(\cdot; \theta)$, with ς_0 held fixed. In estimation, ς_0 is replaced with a consistent first-step estimator $\hat{\varsigma}$. With the nuisance parameter held fixed, the results of Sections SA2 and SA3 are unaffected and it would suffice to read (θ, ς_0) in place of θ in all of the assumptions and results. The large sample theory of Section SA4 is a different matter: there, the first-step estimator $\hat{\varsigma}$ carries sampling uncertainty of its own, which propagates to the estimated bounds, as we discuss in Remark SA-10.

SA1.3 Correspondence with the Paper

The paper considers two robustness sets, the total variation robustness set $\mathcal{P}_{\text{TV}}(\theta; \delta)$ and the contamination robustness set $\mathcal{P}_{\text{C}}(\theta; \delta)$, defined in (3.4) and (3.5) of the main text. This supplement instead works, in some of its parts, with a generic entropy function ϕ . We now outline how the general results developed here cover both robustness sets of the main text, and Table SA-1 then shows where to locate in this supplement the results reported in the paper.

Both sets are constrained total variation neighborhoods. Each of the two sets is defined through a mixture restriction. In the total variation robustness set, the baseline is a mixture of the candidate distribution and of the distribution of a second group, $P_* = \kappa P + (1 - \kappa)R$ with $\kappa \in (0, 1)$, and the two groups are within total variation distance δ of each other. In the contamination robustness set, the candidate distribution is a mixture of the baseline and of an unrestricted contaminating distribution, $P = (1 - \delta)P_* + \delta R$. Lemma SA-17 shows that each of these sets is a total variation neighborhood of the baseline intersected with a one-sided dominance constraint. The first one collects the distributions within total variation distance $(1 - \kappa)\delta$ of the baseline that satisfy $\kappa P \leq P_*$, whereas the second one collects the distributions within total variation distance δ of the baseline that satisfy $P \geq (1 - \delta)P_*$, in which case the total variation constraint is redundant. The formal mappings are in (19) and (20).

The mixture restrictions are absorbed into the entropy. By Lemma SA-12, each dominance constraint is a pointwise bound on the density of the candidate distribution with respect to the baseline. Such a bound is the convex indicator of an interval evaluated at the density, and it can therefore be added to the total variation entropy. Proposition SA-13 shows that the resulting functions,

$$\bar{\phi}_{\text{TV}} := \phi_{\text{TV}} + \iota_{[0, 1/\kappa]} \quad \text{and} \quad \underline{\phi}_{\text{TV}} := \phi_{\text{TV}} + \iota_{[1 - \delta, \infty)},$$

are again entropy functions in the sense of Definition SA-4, and that the two robustness sets of the main text are the divergence neighborhoods they induce, that is,

$$\mathcal{P}_{\text{TV}}(\theta; \delta) = \mathcal{P}_{\bar{\phi}_{\text{TV}}}(\theta; \rho_0, P_*, (1 - \kappa)\delta) \quad \text{and} \quad \mathcal{P}_{\text{C}}(\theta; \delta) = \mathcal{P}_{\underline{\phi}_{\text{TV}}}(\theta; \rho_0, P_*, \delta).$$

Duality. Because both sets are robustness sets of the form (2), the supremum and the infimum of the estimand over them are dualized by the general theory of Section SA2, with no argument specific to either set. Weak duality is Theorem SA-9, strong duality is Theorem SA-11, and Theorem SA-14 specializes them to the two induced entropies, its first part covering the total variation robustness set and its second part the contamination one. The two duals differ in one respect. With the total variation robustness set, the inner dual is an unconstrained convex program. In the contamination robustness set the contaminating distribution may place mass outside the support of the baseline at the finite cost $\phi_{\text{TV}}^\infty(1) = 1/2$, and the dual problem has an additional constraint. Section SA2.5 provides conditions under which this constraint can be deleted without changing the value of the dual (Theorem SA-15 and Proposition SA-16).

Existence. Section SA3 establishes existence of dual solutions and of solutions to the outer problem for generic entropies, under Assumption SA-1 when the recession is finite and under Assumption SA-2 when it is not. The total variation robustness set falls under case (b) of Assumption SA-2(i) with $\bar{x} = 1/\kappa$, because $\bar{\phi}_{\text{TV}}$ is finite at both endpoints of its domain. Therefore Theorem SA-19, Lemma SA-21, and Theorem SA-22 apply to it, as recorded in Remark SA-6. The contamination robustness set is not covered by these statements, because $\underline{\phi}_{\text{TV}}(0) = +\infty$ violates Assumption SA-1(ii). Existence of dual solutions for this set is instead established in Lemma SA-24(iii), which shows that the set of dual minimizers is nonempty, compact, and convex for every member of a family of robustness sets containing both sets of the main text. Its proof is direct and relies only on a weak duality inequality with explicit slack and on a constraint qualification, under the compact support and continuity conditions maintained in the large sample theory.

Inference. Section SA4 does not treat the two robustness sets separately. It works with the family of Section SA4.1, indexed by an interval $\mathbb{U}$ of admissible values for the density of the candidate distribution with respect to the baseline, of which the total variation robustness set is the member $\mathbb{U} = [0, 1/\kappa]$ and the contamination robustness set is the member $\mathbb{U} = [1 - \delta, +\infty)$. By Lemma SA-23, every member of the family is a divergence neighborhood whose dual integrand is a maximum of finitely many affine functions, and by Lemma SA-24 strong duality and existence of dual solutions hold for every member under a single constraint qualification. The large sample arguments use only these common features, together with the fact that the pointwise dual constraint, present for the contamination

Table SA-1: Correspondence between the formal results in the paper and this supplement.

Paper	Supplement	Supporting results
Theorem 1	Theorem SA-14	Theorems SA-9 and SA-11, Lemma SA-12, Proposition SA-13, Example SA-3, Lemmas SA-17 and SA-23
Proposition 1	Theorem SA-15 and Proposition SA-16	Proposition SA-13, Corollary SA-7, Example SA-3, Lemmas SA-17 and SA-23
Theorem 2	Theorem SA-28	Lemmas SA-26 and SA-27
Theorem 3	Theorem SA-29	Lemmas SA-26, SA-27 and SA-24
Theorem 4	Theorem SA-39	Lemmas SA-1 and SA-12, Proposition SA-13

robustness set and absent for the total variation one, depends neither on the estimated moments nor on the baseline. Hence consistency (Theorem SA-28), the asymptotic distribution (Theorem SA-29), and the validity of the bootstrap (Theorem SA-30) hold for both robustness sets at once. The first two results are obtained under Assumption SA-4, and the third one additionally requires Assumption SA-5. The case of a known baseline distribution is discussed in Remark SA-11.

SA2 Convex Duality

This section introduces some key ideas from convex analysis on which the rest of this supplement relies. Section SA2.1 first defines the concepts of convex conjugate, recession function, perspective function, and entropy function, and then proceeds to discuss different definitions of ϕ -divergences that have been used in the literature. Section SA2.2 leverages this framework to recast the infinite-dimensional inner problem in (4) as a tractable finite-dimensional program. Section SA2.3 identifies the conditions under which strong duality holds. Section SA2.4 introduces an additional mixture constraint to the problem in (4). Both robustness sets of the main text are of this form, and the formulation recurs in all applications of Section SA6. Section SA2.5 provides primitive conditions under which the pointwise constraint appearing in the duals derived here can be ignored, that is, deleted without affecting the value of the dual problem. Finally, Section SA2.6 closes the section by specializing to total variation and recasting the contamination neighborhoods as constrained total variation neighborhoods.

SA2.1 A Primer in Convex Analysis

We start by defining some fundamental concepts from convex analysis.

Definition SA-1 (Convex conjugate). *The convex conjugate of a function $f : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$, $k \in \mathbb{N}$, is the function $f^* : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$*

$$f^*(y) = \sup_{z \in \mathbb{R}^k} \{y^\top z - f(z)\}.$$

The function f^{**} , termed the convex biconjugate of f , is the convex conjugate of f^* , that is

$$f^{**}(z) = \sup_{y \in \mathbb{R}^k} \{y^\top z - f^*(y)\}.$$

The convex conjugate is a way of representing the epigraph of a function via supporting hyperplanes. For each slope y , it records the maximal gap between the linear function $z \mapsto y^\top z$ and f .

Definition SA-2 (Recession function, Rockafellar (1970, Theorem 8.5)). *Let $f : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$, $k \in \mathbb{N}$, be proper, convex, and lower semicontinuous. The recession function of f is the function $f^\infty : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$ defined as*

$$f^\infty(z) = \lim_{\alpha \rightarrow \infty} \frac{1}{\alpha} \{f(z_0 + \alpha z) - f(z_0)\},$$

where $z_0 \in \mathbb{R}^k$ is any point with $f(z_0) < \infty$ (under the stated conditions on f , the limit exists in $\overline{\mathbb{R}}$ and does not depend on the choice of z_0).

The recession function f^∞ captures the asymptotic behavior (as $\alpha \rightarrow \infty$) of the function f along the direction z . It is a positively homogeneous function, i.e., $f^\infty(\alpha z) = \alpha f^\infty(z)$ for all $\alpha > 0$ and $z \in \mathbb{R}^k$. The recession function is also convex and lower semicontinuous.

Definition SA-3 (Perspective function). *Let $f : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$, $k \in \mathbb{N}$, be proper, convex, and lower semicontinuous. The perspective of f is the function $f^\pi : \mathbb{R}^k \times \mathbb{R} \rightarrow \overline{\mathbb{R}}$ defined as*

$$f^\pi(z, t) = \begin{cases} tf(z/t), & t > 0, \\ f^\infty(z), & t = 0, \\ +\infty, & t < 0. \end{cases}$$

The perspective function f^π is useful to define the concept of generalized ϕ -divergence we rely on throughout this supplement (see Definition SA-5). It is itself convex and lower semicontinuous (see Lemma SA-3 for the perspectives of an entropy function and of its conjugate).

Before introducing the generalized ϕ -divergence, we need to define the concept of entropy function.

Definition SA-4 (Entropy function). *An entropy function $\phi : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ is a nonnegative lower semicontinuous convex function with $\phi(1) = 0$ and $\phi(s) = +\infty$ for all $s < 0$.*

Table SA-2 below collects the entropy functions ϕ , their convex conjugates ϕ^* , and their recession constants $\phi^\infty(1)$ for popular ϕ -divergences.

Now we have all the ingredients to define the family of generalized ϕ -divergences (Csiszár, 1963; Morimoto, 1963; Ali and Silvey, 1966; Kuhn, Shafiee and Wiesemann, 2025, Definition 2.5).

Definition SA-5 (Generalized ϕ -divergence). *The generalized ϕ -divergence of $P \in \mathcal{M}_+(\mathcal{Z})$ with respect to $P_* \in \Delta(\mathcal{Z})$ is given by*

$$D_\phi(P \| P_*) = \int_{\mathcal{Z}} \phi^\pi \left(\frac{dP}{d\varrho}(z), \frac{dP_*}{d\varrho}(z) \right) d\varrho(z),$$

where ϕ is an entropy function and $\varrho \in \mathcal{M}_+(\mathcal{Z})$ is any measure dominating both P and P_* , that is, $P, P_* \ll \varrho$.

Table SA-2: Entropy functions ϕ , their convex conjugates ϕ^* , and recession constants $\phi^\infty(1)$.

Divergence	$\phi(t)$	$\phi^*(t)$	$\phi^\infty(1)$
Total variation	$\frac{1}{2} t-1 $	$\begin{cases} -\frac{1}{2}, & t < -\frac{1}{2} \\ t, & -\frac{1}{2} \leq t \leq \frac{1}{2} \\ +\infty, & t > \frac{1}{2} \end{cases}$	$\frac{1}{2}$
Kullback–Leibler	$t \log t - t + 1$	$e^t - 1$	∞
Likelihood	$-\log t + t - 1$	$\begin{cases} -\log(1-t), & t < 1 \\ +\infty, & t \geq 1 \end{cases}$	1
Pearson χ^2	$(t-1)^2$	$\begin{cases} -1, & t < -2 \\ t + \frac{t^2}{4}, & t \geq -2 \end{cases}$	∞
Neyman χ^2	$\frac{(t-1)^2}{t}$	$\begin{cases} 2 - 2\sqrt{1-t}, & t \leq 1 \\ +\infty, & t > 1 \end{cases}$	1
Squared Hellinger	$(1 - \sqrt{t})^2$	$\begin{cases} \frac{t}{1-t}, & t < 1 \\ +\infty, & t \geq 1 \end{cases}$	1
Jensen–Shannon	$t \log\left(\frac{2t}{t+1}\right) + \log\left(\frac{2}{t+1}\right)$	$\begin{cases} -\log(2 - e^t), & t < \log 2 \\ +\infty, & t \geq \log 2 \end{cases}$	$\log 2$

Remark SA-2. The divergence does not depend on the choice of the dominating measure ϱ , see Proposition 2.6 in Kuhn, Shafiee and Wiesemann (2025) or Theorem 7.26 in Polyanskiy and Wu (2025). $\clubsuit$

The next lemma expresses the divergence of a measure through its Lebesgue decomposition with respect to the baseline. We use it repeatedly below.

Lemma SA-1 (Lebesgue decomposition of the divergence). *Let ϕ be an entropy function in the sense of Definition SA-4, $P_* \in \Delta(\mathcal{Z})$, and $\nu \in \mathcal{M}_+(\mathcal{Z})$ with Lebesgue decomposition $\nu = wP_* + \nu^\perp$, where $\nu^\perp \perp P_*$ and $w := d\nu_{\text{ac}}/dP_*$ is the Radon–Nikodym derivative of the absolutely continuous part of ν with respect to P_* . Then*

$$D_\phi(\nu \| P_*) = \int_{\mathcal{Z}} \phi(w(z)) dP_*(z) + \phi^\infty(1) \nu^\perp(\mathcal{Z}),$$

where the product $\phi^\infty(1) \nu^\perp(\mathcal{Z})$ evaluates to 0 if $\nu^\perp(\mathcal{Z}) = 0$.

[Proof]

Definition SA-5 above is very general and allows for entropy functions that do not grow superlinearly (e.g., the total variation entropy). We say that ϕ grows superlinearly if $\lim_{s \rightarrow \infty} \phi(s)/s = +\infty$, or equivalently $\phi^\infty(1) = +\infty$. When the entropy function grows superlinearly, the divergence $D_\phi(P \| P_*)$ is finite only if $P \ll P_*$. Indeed, let $P = wP_* + P^\perp$ be the Lebesgue decomposition of $P \in \mathcal{M}_+(\mathcal{Z})$ with respect to P_* . By Lemma SA-1,

$$D_\phi(P \| P_*) = \int_{\mathcal{Z}} \phi(w(z)) dP_*(z) + \phi^\infty(1) P^\perp(\mathcal{Z}),$$

where the integral is nonnegative because $\phi \geq 0$, and the second term equals $+\infty$ as soon as $P^\perp(\mathcal{Z}) > 0$, because $\phi^\infty(1) = +\infty$. Hence a finite divergence forces $P^\perp = 0$, that is $P \ll P_*$, in which case $w = dP/dP_*$ and the second term vanishes. This also yields the following more common definition for ϕ -divergences (Hansen and Sargent, 2001; Maccheroni, Marinacci and Rustichini, 2006):

$$D_\phi(P \| P_*) = \begin{cases} \int \phi\left(\frac{dP}{dP_*}(z)\right) dP_*(z), & P \ll P_*, \\ +\infty, & \text{otherwise.} \end{cases} \quad (5)$$

Most importantly for our purposes, the definition in (5) assigns a finite value only when $P \ll P_*$. Therefore, any neighborhood of finite radius under this definition does not include distributions that are not absolutely continuous with respect to the baseline, which would be possible under divergences with finite recession, such as total variation or squared Hellinger distance.

Example SA-1 (Absolute continuity and total variation). The total variation entropy is $\phi_{\text{TV}}(s) = \frac{1}{2}|s-1|$ for $s \geq 0$ and $\phi_{\text{TV}}(s) = +\infty$ for $s < 0$. Its recession function is $\phi_{\text{TV}}^\infty(s) = s/2$ for $s \geq 0$ and $\phi_{\text{TV}}^\infty(s) = +\infty$ for $s < 0$. Let $Q, R \in \Delta(\mathcal{Z})$ be mutually singular, let $a \in [0, 1]$, and define $P_a = (1-a)Q + aR$. Taking $\varrho = Q + R$ as the dominating measure, mutual singularity gives

$$\frac{dQ}{d\varrho} = 1, \quad \frac{dP_a}{d\varrho} = 1 - a \quad Q\text{-almost surely,}$$

whereas

$$\frac{dQ}{d\varrho} = 0, \quad \frac{dP_a}{d\varrho} = a \quad R\text{-almost surely.}$$

Consequently,

$$D_{\phi_{\text{TV}}}(P_a \parallel Q) = \phi_{\text{TV}}(1 - a) + \phi_{\text{TV}}^\infty(a) = \frac{a}{2} + \frac{a}{2} = a.$$

Thus, for every $a > 0$, the divergence is finite although $P_a \not\ll Q$. Taking $Q = \delta_0^{\text{Dirac}}$, $R = \delta_1^{\text{Dirac}}$, and $a = 1$ gives an example with discrete variables, whereas taking $Q = \text{Unif}([0, 1])$, $R = \delta_2^{\text{Dirac}}$, and $a \in (0, 1)$ gives an example with continuous variables. $\blacklozenge$

We also record a property of total variation that is used repeatedly below.

Lemma SA-2 (Total variation as a ϕ -divergence). *Let $P, Q \in \Delta(\mathcal{Z})$. If $\varrho \in \mathcal{M}_+(\mathcal{Z})$ dominates P and Q , with $p = dP/d\varrho$ and $q = dQ/d\varrho$, then*

$$\text{TV}(P \parallel Q) = \frac{1}{2} \int |p - q| d\varrho = D_{\phi_{\text{TV}}}(P \parallel Q),$$

and the supremum defining $\text{TV}(P \parallel Q)$ is attained at $\{p \geq q\}$.

[Proof]

The next series of lemmas establishes some useful properties of the convex conjugate, the recession function, and perspective function of an entropy function ϕ .

Lemma SA-3 (Properties of entropy functions). *Let ϕ be an entropy function in the sense of Definition SA-4. Then ϕ^* , ϕ^π , and $(\phi^*)^\pi$ are proper, convex, and lower semicontinuous.*

[Proof]

Lemma SA-4 (A lower bound for the perspective of the conjugate function). *Let ϕ be an entropy function in the sense of Definition SA-4. Then, for all $z \in \mathbb{R}$, $t \in \mathbb{R}$, and $\tilde{x} \in \text{dom}(\phi)$,*

$$(\phi^*)^\pi(z, t) \geq z\tilde{x} - t\phi(\tilde{x}).$$

Thus, for $\tilde{x} = 1$, it follows that

$$(\phi^*)^\pi(z, t) \geq z.$$

[Proof]

Lemma SA-5 (Kuhn et al. (2025, Lemma 4.11)). *If ϕ is an entropy function in the sense of Definition SA-4, $t \in \mathbb{R}$, $\beta \in \mathbb{R}_+$, and $\eta \in \mathbb{R}_{++}$, then*

$$\sup_{\alpha \in \mathbb{R}} \{t\alpha - \eta\phi^\pi(\alpha, \beta)\} = \begin{cases} \beta\eta\phi^*(t/\eta) & \text{if } \beta > 0, \\ \eta t_{\text{cl}(\text{dom}(\phi^*))}(t/\eta) & \text{if } \beta = 0. \end{cases}$$

Lemma SA-6 (Kuhn et al. (2025, Lemma 4.12)). *If ϕ is an entropy function in the sense of Definition SA-4, then*

$$\text{cl}(\text{dom}(\phi^*)) = (-\infty, \phi^\infty(1)],$$

where $(-\infty, +\infty]$ is understood as $\mathbb{R}$.

The next corollary shows that if the domain of the conjugate of an entropy function is bounded from above, then the recession function of ϕ must be such that $\phi^\infty(1) < \infty$. Furthermore, the threshold beyond which the conjugate ϕ^* is infinite is precisely $\phi^\infty(1)$. The equivalence plays a central role in Section SA2.5, where it identifies the implicit constraint that the conjugate embeds in the objective function of the dual of the problem in (4).

Corollary SA-7 (Implicit regularization of the entropy). *Let ϕ be an entropy function in the sense of Definition SA-4. Then, the following are equivalent:*

- (1) $\phi^\infty(1) < \infty$;
- (2) $\sup \text{dom}(\phi^*) < +\infty$;
- (3) there exists $M \in \mathbb{R}$ such that $\phi^*(t) = +\infty$ for all $t > M$.

Moreover, if any of the above holds, then $\sup \text{dom}(\phi^*) = \phi^\infty(1)$, so that in (3) we can pick $M = \phi^\infty(1)$.

[Proof]

For any function $f : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$, the definition of the convex conjugate implies the biconjugate inequality $f^{**} \leq f$. This inequality will suffice for weak duality. Strong duality instead relies on the following optimization result. A convex set is polyhedral if it is the intersection of finitely many closed half spaces. A proper convex function is polyhedral if its epigraph $\text{epi } f := \{(x, t) \in \mathbb{R}^{k+1} : t \geq f(x)\}$ is a polyhedral convex set. In particular, the indicator ι_C of a polyhedral convex set C is a polyhedral convex function.

Lemma SA-8 (Fenchel's duality theorem, Rockafellar (1970, Theorem 31.1)). *Let $f, q : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$ be proper convex functions. If*

$$\text{ri}(\text{dom}(f)) \cap \text{ri}(\text{dom}(q)) \neq \emptyset,$$

then

$$\inf_{x \in \mathbb{R}^k} \{f(x) + q(x)\} = \sup_{x^* \in \mathbb{R}^k} \{-f^*(x^*) - q^*(-x^*)\}.$$

If q is polyhedral, the same conclusion holds under the weaker condition

$$\text{ri}(\text{dom}(f)) \cap \text{dom}(q) \neq \emptyset.$$

Whenever the common value is finite, the supremum is attained.

SA2.2 Weak Duality

Throughout this and the next two subsections we suppress the dependence on θ and on the fixed baseline P_* to reduce notational clutter, abbreviating $\mathcal{P}_\phi(\theta; \rho, P_*, \delta)$ as $\mathcal{P}_\phi(\rho, \delta)$. Accordingly, every result below should be interpreted as holding for a fixed $\theta \in \Theta$ and the given baseline P_* . The inner problem over $\mathcal{P}_\phi(\rho, \delta)$ in (4) is an infinite-dimensional optimization problem that we show can be recast as a finite-dimensional one. The estimand

$$\mathbb{E}_P[g(Z)] = \int_{\mathcal{Z}} g(z) dP(z)$$

is linear in the underlying measure $P \in \mathcal{M}_+(\mathcal{Z})$, and the form of the resulting dual depends only on the entropy function ϕ . Heuristically, the next (weak) duality result presented in Theorem SA-9 is obtained by first recasting the inner problem as a linear program over the space of signed measures, then obtaining the conjugate and biconjugate of the primal value function, and concluding with the biconjugate inequality stated above.

Theorem SA-9 (Weak duality). *Let ϕ be an entropy function in the sense of Definition SA-4, and let g be integrable with respect to every $\nu \in \mathcal{M}_{+,m}(\mathcal{Z})$ satisfying $D_\phi(\nu \| P_*) < \infty$. Then*

$$\sup_{P \in \mathcal{P}_\phi(\rho, \delta)} \mathbb{E}_P[g(Z)] \leq \begin{cases} \inf_{(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda} \lambda_{12}^\top \rho + \zeta + \eta \delta + \mathbb{E}_{P_*} \left[(\phi^*)^\pi (g(Z) - \lambda^\top m(Z) - \zeta, \eta) \right], \\ \text{s.t. } \zeta + \eta \phi^\infty(1) \geq \sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\}. \end{cases}$$

The product $\eta \phi^\infty(1)$ is understood to equal $+\infty$ when $\eta = 0$ and $\phi^\infty(1) = +\infty$.

[Proof]

Two special cases are worth noting: (i) if $\phi^\infty(1) = +\infty$, the pointwise constraint is vacuous and the dual is an unconstrained problem; (ii) if instead attention is restricted to distributions that are absolutely continuous with respect to the baseline P_* , the constraint is already implied by the perspective and can be dropped from the dual problem. For a formal explanation, see Remark SA-3 below. A detailed explanation of when the pointwise constraint can be deleted without affecting the value of the dual is given in Theorem SA-15 of Section SA2.5.

Remark SA-3 (Only absolutely continuous distributions). Consider only distributions that are absolutely continuous with respect to the baseline P_* . In this case, the robustness set becomes

$$\mathcal{P}_\phi^{\ll}(\rho, \delta) := \{P \in \mathcal{P}(\mathcal{Z}) : P \ll P_*, \mathbb{E}_P[m(Z)] \in \mathcal{M}(\rho), D_\phi(P \| P_*) \leq \delta\}.$$

For every such P the singular part of P with respect to P_* vanishes, so that Lemma SA-1 gives $D_\phi(P \| P_*) = \int \phi(dP/dP_*) dP_*$. It is the singular part that generates the pointwise constraint in the dual problem. In the proof of Theorem SA-9, the constraint stems from the second integral in (65), which runs over the set where the density of P_* with respect to the dominating measure ϱ vanishes. For measures that are absolutely continuous with respect to P_* we may take $\varrho = P_*$, so that this set is empty and the integral disappears. ♣

SA2.3 Strong Duality

For a divergence budget $b \in \mathbb{R}_+$, recall from (3) the set $\Gamma(b) \equiv \Gamma_\psi(\theta; P_*, b)$ of moments attainable when $m(z)$ is integrated against distributions within divergence b of P_* . It is convex, because the set of distributions with $D_\phi(P \| P_*) \leq b$ is convex, in view of the convexity of $D_\phi(\cdot \| P_*)$, and $\Gamma(b)$ is its image under the linear map $P \mapsto \mathbb{E}_P[m(Z)]$. It is nonempty because $\mathbb{E}_{P_*}[m(Z)] \in \Gamma(b)$ as $D_\phi(P_* \| P_*) = 0 \leq b$. The value function underlying the inner problem is

$$h(a, b, x) = \inf_{\nu \in \mathcal{M}_{+,m}(\mathcal{Z})} \left\{ - \int g(z) d\nu(z) : \int m(z) d\nu(z) = x, \int d\nu(z) = a, D_\phi(\nu \| P_*) \leq b \right\}, \quad (6)$$

which is jointly convex in (a, b, x) by the argument used in the proof of Theorem SA-9 to establish the convexity of the value function. Let

$$K_\rho := \{1\} \times \{\delta\} \times \mathcal{M}(\rho).$$

With this notation, the primal value function equals $-\inf_{y \in K_\rho} h(y)$. Because K_ρ is polyhedral, so is its indicator ι_{K_ρ} . Therefore, once we show that $K_\rho \cap \text{ri}(\text{dom}(h)) \neq \emptyset$ and h is proper, we can apply Lemma SA-8 to $h + \iota_{K_\rho}$. The next lemma shows that both requirements hold.

Lemma SA-10 (Requirements for Fenchel duality). *Let $\delta > 0$, let ϕ be an entropy function in the sense of Definition SA-4 that is continuous at 1, and suppose $\int_{\mathcal{Z}} |g(z)| d\nu(z) < \infty$ for every $\nu \in \mathcal{M}_+(\mathcal{Z})$ with $D_\phi(\nu \| P_*) < \infty$. Then, the value function h in (6) is proper and*

$$(1, \delta, \bar{m}) \in \text{ri}(\text{dom}(h)) \quad \text{for every } \bar{m} \in \text{ri}(\Gamma(\delta)).$$

[Proof]

We can now state the main result of this section.

Theorem SA-11 (Strong duality). *Let $\delta > 0$, let ϕ be an entropy function in the sense of Definition SA-4 that is continuous at 1, and let $\int_{\mathcal{Z}} |g(z)| d\nu(z) < \infty$ for every $\nu \in \mathcal{M}_+(\mathcal{Z})$ satisfying $D_\phi(\nu \| P_*) < \infty$, and suppose*

$$\mathcal{M}(\rho) \cap \text{ri}(\Gamma(\delta)) \neq \emptyset.$$

Then

$$\sup_{P \in \mathcal{P}_\phi(\rho, \delta)} \mathbb{E}_P[g(Z)] = \begin{cases} \inf_{(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda} \lambda_{12}^\top \rho + \zeta + \eta \delta + \mathbb{E}_{P_*} \left[(\phi^*)^\pi (g(Z) - \lambda^\top m(Z) - \zeta, \eta) \right], \\ \text{s.t. } \zeta + \eta \phi^\infty(1) \geq \sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\}. \end{cases}$$

The product $\eta \phi^\infty(1)$ is understood to equal $+\infty$ when $\eta = 0$ and $\phi^\infty(1) = +\infty$.

[Proof]

Remark SA-4 (Meaning of the compatibility condition). The set $\Gamma(\delta)$ collects all moment vectors that distributions within the divergence budget δ from the baseline P_* can attain. Thus, the condition

$$\mathcal{M}(\rho) \cap \text{ri}(\Gamma(\delta)) \neq \emptyset$$

requires at least one of the moment vectors compatible with the model constraints to lie in the relative interior of the attainable moment set. Put differently, small perturbations, in every direction along which the attainable moments vary, of some vector in $\mathcal{M}(\rho)$ must remain attainable for the same divergence budget. Note that the interior is relative to the affine hull of $\Gamma(\delta)$, so the condition accommodates equality restrictions and lower-dimensional moment sets. This condition is useful because it is precisely the qualification under which Lemma SA-8 applies, so that strong duality holds without requiring $\mathcal{M}(\rho)$ to be compact or full dimensional. When no moment restrictions are present, the condition is automatically satisfied under the convention introduced in Section SA1.2. ♣

We close this part by making the dual in Theorem SA-11 explicit for the total variation distance. Throughout this illustrative example, we implicitly maintain the regularity and compatibility conditions of that theorem.

Example SA-2 (Total variation I). Recall that in Example SA-1 we recorded that the recession function of the total variation entropy ϕ_{TV} is $\phi_{\text{TV}}^\infty(s) = s/2$ if $s \geq 0$ and $+\infty$ otherwise, whereas the convex conjugate of ϕ_{TV} is

$$\phi_{\text{TV}}^*(y) = \sup_{s \geq 0} \left\{ ys - \frac{1}{2} |s - 1| \right\} = \begin{cases} -\frac{1}{2}, & y < -\frac{1}{2}, \\ y, & -\frac{1}{2} \leq y \leq \frac{1}{2}, \\ +\infty, & y > \frac{1}{2}, \end{cases}$$

so that $\text{dom}(\phi_{\text{TV}}^*) = (-\infty, \frac{1}{2}]$, in agreement with Lemma SA-6. Taking the perspective of ϕ_{TV}^* in the sense of Definition SA-3 gives

$$(\phi_{\text{TV}}^*)^\pi(z, t) = \begin{cases} -\frac{t}{2}, & t > 0, z < -\frac{t}{2}, \\ z, & t > 0, -\frac{t}{2} \leq z \leq \frac{t}{2}, \\ +\infty, & t > 0, z > \frac{t}{2}, \\ 0, & t = 0, z \leq 0, \\ +\infty, & t = 0, z > 0, \\ +\infty, & t < 0. \end{cases}$$

Because $\phi_{\text{TV}}^\infty(1) = \frac{1}{2} < \infty$, the pointwise constraint of Theorem SA-11 is not vacuous and reads $\zeta + \frac{\eta}{2} \geq \sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\}$. Theorem SA-15 below provides conditions under which it can be deleted without affecting

the value of the dual. Under this constraint the perspective integrand collapses to $\max\{g(Z) - \lambda^\top m(Z) - \zeta, -\eta/2\}$. Recalling from Section SA1.2 that $\iota_{\mathcal{M}(\rho)}^*(\lambda) = \lambda_{12}^\top \rho$ for $\lambda \in \mathcal{R}_\lambda$, the dual of Theorem SA-11 becomes

$$\sup_{P \in \mathcal{P}_{\phi_{\text{TV}}(\rho, \delta)}} \mathbb{E}_P[g(Z)] = \begin{cases} \inf_{(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda} \lambda_{12}^\top \rho + \zeta + \eta\delta + \mathbb{E}_{P_*} \left[\max\{g(Z) - \lambda^\top m(Z) - \zeta, -\frac{\eta}{2}\} \right], \\ \text{s.t. } \zeta + \frac{\eta}{2} \geq \sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\}. \end{cases}$$

◆

SA2.4 Duality Theory Under Mixture Constraints

We now modify the geometry of $\mathcal{P}_\phi(\rho, \delta)$ by adding a new constraint in the form of a mixture restriction. We do so in two complementary ways, each dictated by a distinct and economically relevant class of applications. In particular, for some fixed $\kappa \in (0, 1)$, we define

$$\mathcal{P}_{\text{mix}}^{\leq}(\rho, \delta) := \left\{ P \in \mathcal{P}(\mathcal{Z}) : D_\phi(P \| P_*) \leq \delta, \mathbb{E}_P[m(Z)] \in \mathcal{M}(\rho), \exists R \in \Delta(\mathcal{Z}) : P_* = \kappa P + (1 - \kappa)R \right\} \quad (7)$$

and

$$\mathcal{P}_{\text{mix}}(\rho, \delta) := \left\{ P \in \mathcal{P}(\mathcal{Z}) : D_\phi(P \| P_*) \leq \delta, \mathbb{E}_P[m(Z)] \in \mathcal{M}(\rho), \exists R \in \Delta(\mathcal{Z}) : P = \kappa P_* + (1 - \kappa)R \right\}. \quad (8)$$

Our notation suppresses the dependence of the robustness sets on κ . This is a deliberate choice because we interpret κ as a fixed nuisance parameter which will be treated as known in the large sample theory of Section SA4. In some special cases we also quantify the uncertainty in estimating κ (see Section SA6.1 and Proposition SA-35).

In the first robustness set, the baseline distribution must be a mixture of the candidate distribution P and some other distribution R . This is the geometry of randomized experiments with attrition (Section SA6.1), contaminated data (Section SA6.4.2), and departures from randomization (Section SA6.4.4). In the second robustness set, the roles are reversed and the baseline is instead a component of the mixture. This is the geometry of missing data (Section SA6.4.3). As Proposition SA-13 below establishes, each mixture restriction amounts to replacing ϕ by a modified entropy, after which both duals are obtained by the dual construction underlying Theorem SA-9 together with the strong duality of Theorem SA-11, applied to that modified entropy.

We stress that the two restrictions above, besides altering the geometry of the robustness set, also change the interpretation of the sensitivity parameter δ , a point we return to in Section SA2.6 and, in particular, in Remark SA-5.

Before deriving the duals under the two mixture restrictions, we present two auxiliary results that give a different interpretation to the mixture constraints and help clarify their implications. The first, Lemma SA-12, shows that although each mixture restriction is phrased through the existence of a mixing distribution R , it is in fact equivalent to a pointwise bound on the Radon–Nikodym derivative of the candidate distribution with respect to the baseline distribution. The two cases are mirror images of one another: when the baseline is itself the mixture, the Radon–Nikodym derivative is bounded above by $1/\kappa$, whereas when the baseline is one of the mixing components, the density of the absolutely continuous part is bounded below by κ . We record both characterizations together, as parts (i) and (ii) below.

Lemma SA-12 (Mixture restrictions as Radon–Nikodym bounds). *Fix $\kappa \in (0, 1)$ and let $P \in \Delta(\mathcal{Z})$.*

(i) **Baseline as the mixture.** *The following are equivalent:*

(1) *there exists $R \in \Delta(\mathcal{Z})$ such that*

$$P_* = \kappa P + (1 - \kappa)R;$$

(2) *$\kappa P \leq P_*$ as measures, that is,*

$$\kappa P(A) \leq P_*(A) \quad \forall A \in \mathcal{F};$$

(3) *$P \ll P_*$ and*

$$\frac{dP}{dP_*} \leq \frac{1}{\kappa} \quad P_*\text{-a.s.}$$

(ii) **Baseline as a mixing component.** *The following are equivalent:*

(1) *there exists $R \in \Delta(\mathcal{Z})$ such that*

$$P = \kappa P_* + (1 - \kappa)R;$$

(2) *$P \geq \kappa P_*$ as measures, that is,*

$$P(A) \geq \kappa P_*(A) \quad \forall A \in \mathcal{F};$$

(3) *the density u of the absolutely continuous part of P in its Lebesgue decomposition $P = uP_* + P^\perp$ with respect to P_* , $P^\perp \perp P_*$, satisfies*

$$u \geq \kappa \quad P_*\text{-a.s.}$$

In particular, if $P \ll P_*$, then this restriction is equivalent to

$$\frac{dP}{dP_*} \geq \kappa \quad P_*\text{-a.s.}$$

[Proof]

The second auxiliary result, Proposition SA-13, builds on the characterization presented in Lemma SA-12 to show that the mixture restrictions can be absorbed into the entropy function, inducing a new modified entropy function. This is because each mixture restriction is the convex indicator of an interval evaluated at the density, which can be added to the entropy function. The result allows us to treat the robustness sets $\mathcal{P}_{\text{mix}}^{\leq}(\rho, \delta)$ and $\mathcal{P}_{\text{mix}}(\rho, \delta)$ as standard robustness sets of the form defined in (2) with the original entropy function replaced by a modified one.

Proposition SA-13 (Induced entropy functions). *Fix $\kappa \in (0, 1)$ and $\delta \in \mathbb{R}_+$, and let ϕ be an entropy function in the sense of Definition SA-4.*

- (i) **Baseline as the mixture.** *The function $\bar{\phi} := \phi + \iota_{[0, 1/\kappa]}$ is an entropy function in the sense of Definition SA-4, with recession $\bar{\phi}^\infty(1) = +\infty$. Moreover*

$$\mathcal{P}_{\text{mix}}^{\leq}(\rho, \delta) = \mathcal{P}_{\bar{\phi}}(\rho, \delta),$$

where $\mathcal{P}_{\bar{\phi}}(\rho, \delta)$ is the robustness set (2) for the entropy $\bar{\phi}$.

- (ii) **Baseline as a mixing component.** *The function $\underline{\phi} := \phi + \iota_{[\kappa, \infty)}$ is an entropy function in the sense of Definition SA-4, with recession $\underline{\phi}^\infty(1) = \phi^\infty(1)$. Moreover*

$$\mathcal{P}_{\text{mix}}(\rho, \delta) = \mathcal{P}_{\underline{\phi}}(\rho, \delta),$$

where $\mathcal{P}_{\underline{\phi}}(\rho, \delta)$ is the robustness set (2) for the entropy $\underline{\phi}$.

[Proof]

We now present the main result of this subsection, which provides the duals of the inner problem in (4) when the robustness set is defined via either (7) or (8). The proof of the result follows the steps of that of Theorem SA-11, with the entropy function replaced by the one identified in Proposition SA-13, which arises from the mixture restriction. The two geometries differ, however, in the assumptions needed to guarantee strong duality. When the baseline distribution is itself the mixture, the restriction is equivalent to the bound on the Radon–Nikodym derivative $dP/dP_* \leq 1/\kappa$ (see Lemma SA-12(i)). As such, the candidate distributions are forced to be absolutely continuous with respect to P_* , the modified entropy $\bar{\phi}$ satisfies $\bar{\phi}^\infty(1) = +\infty$, and the pointwise dual constraint in Theorem SA-11 is vacuous, as its left-hand side equals $+\infty$ for every $\eta \geq 0$ under the convention on the product $\eta \bar{\phi}^\infty(1)$. When instead the baseline is a component of the mixture, the restriction only bounds from below the density u of the absolutely continuous part of P , $u \geq \kappa$ P_* -almost surely (see Lemma SA-12(ii)), and leaves the singular part of P unrestricted. The resulting modified entropy $\underline{\phi}$ inherits the recession of ϕ , $\underline{\phi}^\infty(1) = \phi^\infty(1)$ (equal to $1/2$ for the total variation entropy). When this recession is finite, the constraint in Theorem SA-11 is not vacuous because singular mass has finite cost $\phi^\infty(1)$.

Theorem SA-14 (Strong duality under mixture constraints). *Fix $\kappa \in (0, 1)$ and $\delta > 0$, let ϕ be an entropy function in the sense of Definition SA-4 that is continuous at 1, and, for any entropy function ψ , let $\Gamma_\psi(\delta)$ be the attainable moment set of (3).*

- (i) **Baseline as the mixture.** *Let $\bar{\phi} := \phi + \iota_{[0, 1/\kappa]}$ be the induced entropy of Proposition SA-13(i), and suppose*
- $$\mathcal{M}(\rho) \cap \text{ri}(\Gamma_{\bar{\phi}}(\delta)) \neq \emptyset.$$

Then

$$\sup_{P \in \mathcal{P}_{\text{mix}}^{\leq}(\rho, \delta)} \mathbb{E}_P[g(Z)] = \inf_{(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda} \lambda_{12}^\top \rho + \zeta + \eta \delta + \mathbb{E}_{P_*} \left[(\bar{\phi}^*)^\pi (g(Z) - \lambda^\top m(Z) - \zeta, \eta) \right],$$

where, by definition of $\bar{\phi}$,

$$\bar{\phi}^*(y) = \sup_{0 \leq s \leq 1/\kappa} \{ys - \phi(s)\}.$$

- (ii) **Baseline as a mixing component.** *Let $\underline{\phi} := \phi + \iota_{[\kappa, \infty)}$ be the induced entropy of Proposition SA-13(ii), and suppose $\int_{\mathcal{Z}} |g(z)| d\nu(z) < \infty$ for every $\nu \in \mathcal{M}_+(\mathcal{Z})$ with $D_\phi(\nu \| P_*) < \infty$ and*

$$\mathcal{M}(\rho) \cap \text{ri}(\Gamma_{\underline{\phi}}(\delta)) \neq \emptyset.$$

Then

$$\sup_{P \in \mathcal{P}_{\text{mix}}(\rho, \delta)} \mathbb{E}_P[g(Z)] = \begin{cases} \inf_{(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda} \lambda_{12}^\top \rho + \zeta + \eta \delta + \mathbb{E}_{P_*}[(\underline{\phi}^*)^\pi(g(Z) - \lambda^\top m(Z) - \zeta, \eta)], \\ \text{s.t. } \zeta + \eta \phi^\infty(1) \geq \sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\}, \end{cases}$$

where, by definition of $\underline{\phi}$,

$$\underline{\phi}^*(y) = \sup_{s \geq \kappa} \{ys - \phi(s)\}.$$

The product $\eta \phi^\infty(1)$ is understood to equal $+\infty$ when $\eta = 0$ and $\phi^\infty(1) = +\infty$.

[Proof]

Part (i) requires no integrability condition beyond the maintained $P_* \in \mathcal{P}(\mathcal{Z})$, whereas part (ii) does. The asymmetry between the integrability conditions of the two parts reflects the geometry of the two neighborhoods. In part (i), the domain of $\bar{\phi}$ is the bounded interval $[0, 1/\kappa]$, so that $\bar{\phi}^\infty(1) = +\infty$. Lemma SA-1 then implies that every $\nu \in \mathcal{M}_+(\mathcal{Z})$ with $D_{\bar{\phi}}(\nu \| P_*) < \infty$ has no singular part with respect to P_* and a density $d\nu/dP_* \leq 1/\kappa$ P_* -almost surely. Hence $\int_{\mathcal{Z}} |g(z)| d\nu(z) \leq \kappa^{-1} \mathbb{E}_{P_*}[|g(Z)|] < \infty$ for every such ν . In part (ii), the domain of $\underline{\phi}$ is $[\kappa, \infty)$, so the same lemma bounds the density only from below, and ν may carry a singular part when $\phi^\infty(1) < \infty$. Thus no such implication holds.

We conclude this subsection by making the induced entropies $\underline{\phi}$ and $\bar{\phi}$ explicit for the total variation entropy.

Example SA-3 (Total variation II). For the total variation entropy ϕ_{TV} of Example SA-1, both modified entropies admit closed form conjugates. For the upper truncation $\bar{\phi}_{\text{TV}} = \phi_{\text{TV}} + \iota_{[0, 1/\kappa]}$ of Theorem SA-14(i),

$$\begin{aligned} \bar{\phi}_{\text{TV}}^*(y) &= \sup_{0 \leq u \leq 1/\kappa} \left\{ yu - \frac{1}{2}|u - 1| \right\} = \begin{cases} -\frac{1}{2}, & y < -\frac{1}{2}, \\ y, & -\frac{1}{2} \leq y \leq \frac{1}{2}, \\ \frac{y}{\kappa} + \frac{\kappa-1}{2\kappa}, & y > \frac{1}{2}, \end{cases} \\ (\bar{\phi}_{\text{TV}}^*)^\pi(z, t) &= \begin{cases} \max \left\{ -\frac{t}{2}, z, \frac{z}{\kappa} - \frac{(1-\kappa)t}{2\kappa} \right\}, & t \geq 0, \\ +\infty, & t < 0. \end{cases} \end{aligned}$$

At $t = 0$ the first branch equals $\max\{z, 0\}/\kappa$, the recession function of $\bar{\phi}_{\text{TV}}^*$, and for $t \geq 0$ it coincides with the total variation integrand $r_{\text{TV}}(z, t)$ in (4.4) of the main text. For the lower truncation $\underline{\phi}_{\text{TV}} = \phi_{\text{TV}} + \iota_{[\kappa, \infty)}$ of Theorem SA-14(ii), since $\phi_{\text{TV}}^\infty(1) = \frac{1}{2}$ the pointwise constraint reads $\zeta + \eta/2 \geq \sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\}$, and

$$\begin{aligned} \underline{\phi}_{\text{TV}}^*(y) &= \sup_{u \geq \kappa} \left\{ yu - \frac{1}{2}|u - 1| \right\} = \begin{cases} \kappa y + \frac{\kappa-1}{2}, & y < -\frac{1}{2}, \\ y, & -\frac{1}{2} \leq y \leq \frac{1}{2}, \\ +\infty, & y > \frac{1}{2}, \end{cases} \\ (\underline{\phi}_{\text{TV}}^*)^\pi(z, t) &= \begin{cases} \kappa z + \frac{\kappa-1}{2}t, & t > 0, z < -t/2, \\ z, & t > 0, -t/2 \leq z \leq t/2, \\ +\infty, & t > 0, z > t/2, \\ \kappa z, & t = 0, z \leq 0, \\ +\infty, & t = 0, z > 0, \\ +\infty, & t < 0. \end{cases} \end{aligned}$$

◆

SA2.5 Ignoring the Pointwise Dual Constraint

The general dual of Theorems SA-9 and SA-11 carries the pointwise constraint

$$\zeta + \eta \phi^\infty(1) \geq \sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\}, \quad (9)$$

which prices the singular mass that a candidate distribution may place at any point of $\mathcal{Z}$. Relocating mass toward a point z gains $g(z) - \lambda^\top m(z)$ per unit in the Lagrangian, and the multipliers on the normalization constraint, ζ , and divergence constraint, η , must render such relocations unprofitable everywhere on $\mathcal{Z}$. As in Theorem SA-9, here and in the rest of this subsection the product $\eta \phi^\infty(1)$ evaluates to $+\infty$ when $\eta = 0$ and $\phi^\infty(1) = +\infty$. When the entropy function grows superlinearly, as in Christensen and Connault (2023), singular mass is infinitely expensive because $\phi^\infty(1) = +\infty$ and the constraint becomes vacuous. For entropies with linear growth, such as ϕ_{TV} , the recession function is such that $\phi^\infty(1) < \infty$, and thus the constraint is not vacuous.

Handling (9) in practice can be demanding from a computational perspective, since it adds one feasibility constraint per support point. It is therefore natural to ask when (9) can be ignored, that is, deleted from the dual without

changing its value. Beyond the superlinear case, the two results of this subsection provide sufficient conditions for doing so. Under regularity conditions on the primitives of the problem, Theorem SA-15 shows that restricting the robustness set to distributions that are absolutely continuous with respect to P_* leaves the value of the inner primal problem unchanged. In turn, Proposition SA-16 shows that this new restricted inner primal problem admits a dual problem that is identical to the dual we obtained in Theorem SA-9 but without (9).

Before stating the results of this subsection, we introduce two quantities. The first quantity is the restriction of the robustness set to absolutely continuous distributions, $\mathcal{P}_\phi^{\ll}(\rho, \delta)$, defined in Remark SA-3 as the set of $P \in \mathcal{P}_\phi(\rho, \delta)$ with $P \ll P_*$. The second is the moment set generated by absolutely continuous distributions, $\Gamma^{\ll}(b) \subseteq \Gamma(b)$ of (3), the analog of $\Gamma(b)$ that considers only absolutely continuous distributions with respect to P_* . The set $\Gamma^{\ll}(b)$ is convex by the same argument used for $\Gamma(b)$, as mixtures of distributions that are absolutely continuous with respect to P_* are absolutely continuous themselves.

We now define two pairs of objects. On the primal side, for a divergence budget $\delta \in \mathbb{R}_+$, we let

$$V_P(\delta) := \sup_{P \in \mathcal{P}_\phi(\rho, \delta)} \mathbb{E}_P[g(Z)], \quad V_P^{\ll}(\delta) := \sup_{P \in \mathcal{P}_\phi^{\ll}(\rho, \delta)} \mathbb{E}_P[g(Z)] \quad (10)$$

denote the value of the inner problem over the full robustness set and over its absolutely continuous restriction, respectively. On the dual side, we let

$$f(\zeta, \eta, \lambda; \delta) := \lambda_{12}^\top \rho + \zeta + \eta \delta + \mathbb{E}_{P_*} [(\phi^*)^\pi(g(Z) - \lambda^\top m(Z) - \zeta, \eta)] \quad (11)$$

be the dual objective of Theorem SA-9 and we let

$$\mathcal{F} := \left\{ (\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda : \zeta + \eta \phi^\infty(1) \geq \sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\} \right\} \quad (12)$$

collect the dual variables that satisfy (9). The two dual values of interest in this section are then

$$V_D(\delta) := \inf_{(\zeta, \eta, \lambda) \in \mathcal{F}} f(\zeta, \eta, \lambda; \delta), \quad V_D^\circ(\delta) := \inf_{(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda} f(\zeta, \eta, \lambda; \delta), \quad (13)$$

that is, the value of the dual of Theorem SA-9 and the value of the same program with (9) deleted. In Section SA3 we reinstate the dependence on the parameter and write $V_P(\theta; \delta)$ and $V_D(\theta; \delta)$ for these same two values.

Theorem SA-15 (Restriction to absolutely continuous distributions). *Let ϕ be an entropy function in the sense of Definition SA-4 and let $\delta \in \mathbb{R}_+$. Suppose that at least one of the following two conditions holds:*

(a) $\phi^\infty(1) = +\infty$;

(b) $\delta > 0$, ϕ is continuous at 1, $\int_{\mathcal{Z}} |g(z)| d\nu(z) < \infty$ for every $\nu \in \mathcal{M}_+(\mathcal{Z})$ with $D_\phi(\nu \| P_*) < \infty$,
 $\mathcal{M}(\rho) \cap \text{ri}(\Gamma^{\ll}(\delta)) \neq \emptyset$,

and, for every $\lambda \in \mathcal{R}_\lambda$,

$$\sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\} = P_*\text{-ess sup}_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\}.$$

Then

$$V_P(\delta) = V_P^{\ll}(\delta).$$

[Proof]

Theorem SA-15 provides two sets of sufficient conditions that allow us to discard from the robustness set all the distributions that are not absolutely continuous with respect to the baseline while leaving the value of the inner primal problem unchanged. This result alone is not constructive from a practical standpoint and does not answer the central question of this subsection: when can we ignore (9)? The next proposition addresses this question and shows that the restricted primal problem admits a strong dual in which the constraint (9) is absent, so that the restricted bound can be computed by an unconstrained convex program.

Proposition SA-16. *Let ϕ be an entropy function in the sense of Definition SA-4, let $\int_{\mathcal{Z}} |g(z)| d\nu(z) < \infty$ for every $\nu \in \mathcal{M}_+(\mathcal{Z})$ satisfying $D_\phi(\nu \| P_*) < \infty$, and suppose that $\delta > 0$, ϕ is continuous at 1, and*

$$\mathcal{M}(\rho) \cap \text{ri}(\Gamma^{\ll}(\delta)) \neq \emptyset.$$

Then

$$V_P^{\ll}(\delta) = V_D^\circ(\delta).$$

[Proof]

Two comments are in order. First, the comparison between the constrained and unconstrained dual objectives is pointwise in λ . Therefore, the two dual programs agree on the multiplier set $\mathcal{R}_\lambda$ generated by the maintained moment equalities and inequalities. Second, we stress that full support and lower semicontinuity together suffice to verify the essential supremum identity of condition (b). Indeed, whenever $\text{supp}(P_*) = \mathcal{Z}$ and the map $z \mapsto g(z) - \lambda^\top m(z)$

is lower semicontinuous for every $\lambda \in \mathcal{R}_\lambda$, as happens when g is lower semicontinuous and m is continuous, the displayed essential-supremum identity of condition (b) is automatically satisfied. We illustrate a scenario in which condition (b) holds in Section SA6.2.1.

SA2.6 Total Variation and Contamination Neighborhoods

We close this section by specializing the entropy function to the total variation one, that is $\phi = \phi_{\text{TV}}$, and by introducing the robustness sets on which the remainder of this supplement relies, namely the total variation neighborhood and the two neighborhoods generated by the mixture restrictions of Section SA2.4. We then show that each of the latter amounts to intersecting the total variation neighborhood with a one-sided contamination neighborhood.

Fix a sensitivity parameter $\delta \in [0, 1]$ and a mixing weight $\kappa \in [0, 1]$. We denote the set of distributions satisfying the moment restrictions by

$$\mathcal{E}(\rho) := \{P \in \mathcal{P}(\mathcal{Z}) : \mathbb{E}_P[m(Z)] \in \mathcal{M}(\rho)\},$$

and we define the following three sets:

$$\begin{aligned} \mathcal{T}_\delta(P_*) &:= \{P \in \Delta(\mathcal{Z}) : \text{TV}(P \| P_*) \leq \delta\}, \\ \mathcal{C}_\kappa(P_*) &:= \{P \in \Delta(\mathcal{Z}) : P = \kappa P_* + (1 - \kappa)R, R \in \Delta(\mathcal{Z})\}, \\ \mathcal{C}_\kappa^{\leq}(P_*) &:= \{P \in \Delta(\mathcal{Z}) : P_* = \kappa P + (1 - \kappa)R, R \in \Delta(\mathcal{Z})\}. \end{aligned}$$

The first one is a total variation neighborhood. The second one is a contamination neighborhood in the spirit of Huber (1964) in which the baseline distribution is a mixing component of the candidate distribution with weight κ , so that $1 - \kappa$ is the contamination share. In the third one, the roles are inverted and the candidate distribution P is a mixing component of the baseline with the same weight κ . We refer to it as the decomposition neighborhood, since each of its elements is obtained by decomposing the baseline population into a retained subpopulation, distributed according to P , and a discarded one, distributed according to R . The two-parameter families

$$\mathcal{C}_{\kappa,\delta}(P_*) := \mathcal{T}_\delta(P_*) \cap \mathcal{C}_\kappa(P_*) \quad \text{and} \quad \mathcal{C}_{\kappa,\delta}^{\leq}(P_*) := \mathcal{T}_\delta(P_*) \cap \mathcal{C}_\kappa^{\leq}(P_*) \quad (14)$$

are of particular interest to us. These families nest the three sets above: at $\kappa = 0$ both mixture constraints are vacuous, so that $\mathcal{C}_{0,\delta}(P_*) = \mathcal{C}_{0,\delta}^{\leq}(P_*) = \mathcal{T}_\delta(P_*)$, whereas at $\delta = 1$ the total variation constraint is vacuous, so that $\mathcal{C}_{\kappa,1}(P_*) = \mathcal{C}_\kappa(P_*)$ and $\mathcal{C}_{\kappa,1}^{\leq}(P_*) = \mathcal{C}_\kappa^{\leq}(P_*)$. They also coincide, for $\phi = \phi_{\text{TV}}$, with the robustness sets of Section SA2.4. Indeed, $\mathcal{P}_{\text{mix}}(\rho, \delta)$ of (8) and $\mathcal{P}_{\text{mix}}^{\leq}(\rho, \delta)$ of (7) can be rewritten as

$$\mathcal{P}_{\text{mix}}(\rho, \delta) = \mathcal{C}_{\kappa,\delta}(P_*) \cap \mathcal{E}(\rho) \quad \text{and} \quad \mathcal{P}_{\text{mix}}^{\leq}(\rho, \delta) = \mathcal{C}_{\kappa,\delta}^{\leq}(P_*) \cap \mathcal{E}(\rho), \quad (15)$$

with δ free and κ fixed.

The next lemma leverages the equivalence between mixture constraints and dominance constraints established in Lemma SA-12 to provide a characterization of the sets in (14).

Lemma SA-17 (Contamination as constrained TV). *Let $\kappa \in [0, 1]$ and $\delta \in [0, 1]$. If $\kappa < 1$, then for every $P \in \mathcal{C}_\kappa(P_*)$ the mixing distribution is uniquely determined as $R = (P - \kappa P_*)/(1 - \kappa)$, for every $P \in \mathcal{C}_\kappa^{\leq}(P_*)$ as $R = (P_* - \kappa P)/(1 - \kappa)$, and the following characterizations hold:*

(i) Baseline as the mixture.

$$\begin{aligned} \mathcal{C}_{\kappa,\delta}^{\leq}(P_*) &= \left\{ P \in \Delta(\mathcal{Z}) : P_* = \kappa P + (1 - \kappa)R, R \in \Delta(\mathcal{Z}), \text{TV}(P \| R) \leq \frac{\delta}{1 - \kappa} \right\} \\ &= \mathcal{T}_\delta(P_*) \cap \{P : \kappa P \leq P_*\}, \end{aligned} \quad (16)$$

and the mixing distribution of every $P \in \mathcal{C}_\kappa^{\leq}(P_*)$ satisfies $\text{TV}(P \| P_*) = (1 - \kappa)\text{TV}(P \| R)$ and $\text{TV}(R \| P_*) = \kappa \text{TV}(P \| R)$;

(ii) Baseline as a mixing component.

$$\begin{aligned} \mathcal{C}_{\kappa,\delta}(P_*) &= \left\{ \kappa P_* + (1 - \kappa)R : R \in \Delta(\mathcal{Z}), \text{TV}(R \| P_*) \leq \frac{\delta}{1 - \kappa} \right\} \\ &= \mathcal{T}_\delta(P_*) \cap \{P : P \geq \kappa P_*\}, \end{aligned} \quad (17)$$

and the mixing distribution of every $P \in \mathcal{C}_\kappa(P_*)$ satisfies $\text{TV}(P \| P_*) = (1 - \kappa)\text{TV}(R \| P_*)$.

Both dominance constraints are interpreted as measure inequalities, that is, $\kappa P \leq P_*$ means $\kappa P(A) \leq P_*(A)$ for every $A \in \mathcal{F}$, and analogously for $P \geq \kappa P_*$. Moreover, the rescaling identities imply $\mathcal{C}_{\kappa,\delta}(P_*) \cup \mathcal{C}_{\kappa,\delta}^{\leq}(P_*) \subseteq \mathcal{T}_{\min\{\delta, 1 - \kappa\}}(P_*)$, and whenever $\delta \geq 1 - \kappa$ the total variation constraint is redundant, in the sense that $\mathcal{C}_{\kappa,\delta}(P_*) = \mathcal{C}_\kappa(P_*)$ and $\mathcal{C}_{\kappa,\delta}^{\leq}(P_*) = \mathcal{C}_\kappa^{\leq}(P_*)$. Finally, if $\kappa = 1$, then $\mathcal{C}_{1,\delta}(P_*) = \mathcal{C}_{1,\delta}^{\leq}(P_*) = \{P_*\}$ for every $\delta \in [0, 1]$.

[Proof]

Lemma SA-17 shows that intersecting a total variation neighborhood with a contamination neighborhood produces

again a contamination class in which the mixing distribution is itself disciplined in total variation. The two parameters thus play distinct roles: $1 - \kappa$ caps the share of the population that may deviate from the baseline, while δ caps the total probability mass that may be misplaced in doing so. Since, by Lemma SA-17, the total variation constraint is redundant whenever $\delta \geq 1 - \kappa$, (15) shows that, for $\phi = \phi_{\text{TV}}$ and $\delta = 1 - \kappa$, the robustness sets of Section SA2.4 are pure contamination neighborhoods, that is,

$$\mathcal{P}_{\text{mix}}(\rho, 1 - \kappa) = \mathcal{C}_{\kappa}(P_*) \cap \mathcal{E}(\rho) \quad \text{and} \quad \mathcal{P}_{\text{mix}}^{\leq}(\rho, 1 - \kappa) = \mathcal{C}_{\kappa}^{\leq}(P_*) \cap \mathcal{E}(\rho). \quad (18)$$

We now map the sets defined above into the two robustness sets of the main text. There, all moment restrictions are equalities, that is, $\mathbf{d}_1 = \mathbf{d}_3 = 0$ and ρ_0 reduces to its equality block, so that $\mathcal{M}(\rho_0) = \{\bar{\rho}_0\}$ and $\mathcal{E}(\rho_0)$ collects the distributions satisfying $\mathbb{E}_P[m(Z; \theta)] = \bar{\rho}_0$. The total variation robustness set in equation (3.4) of the main text requires the baseline to be a mixture, $P_* = \kappa P + (1 - \kappa)R$ with $\kappa \in (0, 1)$, and the two mixing components to satisfy $\text{TV}(P \| R) \leq \delta$. Therefore,

$$\mathcal{P}_{\text{TV}}(\theta; \delta) = \mathcal{C}_{\kappa, (1-\kappa)\delta}^{\leq}(P_*) \cap \mathcal{E}(\rho_0) = \mathcal{P}_{\text{mix}}^{\leq}(\rho_0, (1 - \kappa)\delta), \quad (19)$$

where the first equality follows from the first equality in (16) applied with budget $(1 - \kappa)\delta$, and the second one from (15). In words, a budget δ on the distance between the two groups is a budget $(1 - \kappa)\delta$ on the distance between the candidate and the baseline distributions, in view of the rescaling identity of Lemma SA-17(i). The contamination robustness set in equation (3.5) of the main text instead ties the mixing weight to the sensitivity parameter, $\kappa = 1 - \delta$, and leaves the mixing distribution unrestricted. Hence, for every $\delta \in (0, 1)$,

$$\mathcal{P}_{\text{C}}(\theta; \delta) = \mathcal{C}_{1-\delta}(P_*) \cap \mathcal{E}(\rho_0) = \mathcal{P}_{\text{mix}}(\rho_0, \delta) \quad \text{with} \quad \kappa = 1 - \delta, \quad (20)$$

where the first equality holds by definition of $\mathcal{C}_{1-\delta}(P_*)$, and the second one follows from (18). In both sets, the paper draws the mixing distribution R from the integrability class $\mathcal{P}(\mathcal{Z})$, which is without loss here, because $(1 - \kappa)R \leq P_*$ in the first and $\delta R \leq P$ in the second make R inherit the integrability of $P_* \in \mathcal{P}(\mathcal{Z})$ and of $P \in \mathcal{E}(\rho_0) \subseteq \mathcal{P}(\mathcal{Z})$, respectively. Both robustness sets of the main text are thus members of the two-parameter families in (14), intersected with $\mathcal{E}(\rho_0)$.

Remark SA-5 (Different interpretations). One key implication of the difference between the total variation neighborhood $\mathcal{T}_{\delta}(P_*)$ and the two mixture-based sets lies in their interpretation. If $P \in \mathcal{C}_{1-\delta}(P_*)$, not only can we say that P and P_* share at least a fraction $1 - \delta$ of their mass, but also, in view of the mixture structure, that a $(1 - \delta)$ -fraction of the population behaves exactly as P_* . Similarly, if $P \in \mathcal{C}_{1-\delta}^{\leq}(P_*)$, then P and P_* share at least a fraction $1 - \delta$ of their mass, and a $(1 - \delta)$ -fraction of the population behaves exactly as P . If instead all we know is that $\text{TV}(P \| P_*) = \delta$, with $\delta \in (0, 1)$, then both distributions can be written as mixtures over a common component,

$$P = (1 - \delta)Q + \delta R_P \quad \text{and} \quad P_* = (1 - \delta)Q + \delta R_*,$$

where $Q \in \Delta(\mathcal{Z})$ has density $\min\{dP/d\varrho, dP_*/d\varrho\}/(1 - \delta)$ with respect to the dominating measure $\varrho := (P + P_*)/2$, and $R_P, R_* \in \Delta(\mathcal{Z})$ carry the mass on which the two distributions disagree. Hence P and P_* share exactly a fraction $1 - \delta$ of their mass, but the shared component Q need not coincide with either P or P_* . In general, neither distribution need be a mixture component of the other. Indeed, not every distribution within total variation distance δ of the baseline admits such a mixture representation. For example, for $\mathcal{Z} = \{0, 1\}$, $P_* = \text{Be}(1/2)$, and $P(\{1\}) = 0.7$, so that $\text{TV}(P \| P_*) = 0.2$, no $R \in \Delta(\mathcal{Z})$ satisfies $P = 0.8P_* + 0.2R$ or $P_* = 0.8P + 0.2R$, as either would give R a negative mass. Consequently, P belongs to $\mathcal{T}_{0.2}(P_*)$ but to neither $\mathcal{C}_{0.8}(P_*)$ nor $\mathcal{C}_{0.8}^{\leq}(P_*)$. ♣

SA3 Existence of Solutions to the Inner and Outer Problems

This section provides sufficient conditions for the dual inner problem and the outer problem of (4) to admit solutions. Accordingly, we make dependence of quantities on θ explicit again. Section SA3.1 treats the dual inner problem holding the parameter θ fixed. Section SA3.2 then turns to the outer problem, which optimizes the inner value over θ . In this section, we keep suppressing the dependence of the robustness set on the baseline distribution P_* for notational convenience.

SA3.1 Existence of Dual Solutions

We first establish sufficient conditions for the inner problem of (4) to have solutions in its dual form, holding $\theta \in \Theta$ fixed. We write $V_P(\theta; \delta)$ for the value $\sup_{P \in \mathcal{P}_\phi(\theta; \rho, P_*, \delta)} \mathbb{E}_P[g(Z; \theta)]$ of the inner problem and $V_D(\theta; \delta)$ for the value of the associated dual inner problem, displayed in (21) below, where we recall the convention of Theorem SA-9 that the product between η and $\phi^\infty(1)$ evaluates to $+\infty$ if $\eta = 0$ and $\phi^\infty(1) = +\infty$. Throughout this section, we write $\mathcal{N}(\theta) := \{m(z; \theta) : z \in \mathcal{Z}\} \subseteq \mathbb{R}^{d_m}$ for the image of the moment function at θ and, whenever $\text{dom}(\phi)$ is bounded above, $\bar{x} := \sup \text{dom}(\phi)$, and we recall the padded target $\vec{\rho}$ of Section SA1.2.

We start by introducing the conditions on the primitives of the problem that are maintained throughout this subsection. We rely on two sets of conditions, which handle different scenarios that might occur in practice and differ according to the properties of the recession of the entropy function. Formally, Assumption SA-1 covers entropies with finite recession ($\phi^\infty(1) < \infty$), for which the pointwise constraint in (21) is not vacuous, whereas Assumption SA-2 covers superlinear entropies (with $\phi^\infty(1) = +\infty$), for which the constraint is vacuous.

Assumption SA-1 (Primitives I: finite recession entropies). *Let $\theta \in \Theta$. Suppose:*

- (i) (Boundedness from above.) $\bar{g} := \sup_{z \in \mathcal{Z}} g(z; \theta) < \infty$.
- (ii) (Entropy.) ϕ is an entropy function in the sense of Definition SA-4 with $\phi^\infty(1) \in (0, +\infty)$ and $\phi(0) < \infty$.
- (iii) (Feasibility.) There exists a distribution $P^\circ \in \mathcal{P}(\mathcal{Z})$ with $\mathbb{E}_{P^\circ}[m(Z; \theta)] \in \mathcal{M}(\rho)$ and $D_\phi(P^\circ \| P_*) < \delta$.
- (iv) (Slater condition.) If $d_m \geq 1$, then

$$0 \in \text{int}(\text{co}(\mathcal{N}(\theta)) - \vec{\rho}).$$

Assumption SA-2 (Primitives II: superlinear entropies). *Let $\theta \in \Theta$. Suppose:*

- (i) (Entropy.) ϕ is an entropy function in the sense of Definition SA-4 with $\phi^\infty(1) = +\infty$ and $0 \in \text{dom}(\phi)$, and either (a) $[0, \infty) \subseteq \text{dom}(\phi)$, or (b) $\bar{x} \in (1, \infty)$ with $\phi(\bar{x}) < \infty$.
- (ii) (Dual anchor.) There exists $\xi^* = (\zeta^*, \eta^*, \lambda^*) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$ with
$$\mathbb{E}_{P_*}[(\phi^*)^\pi(g(Z; \theta) - \lambda^{*\top} m(Z; \theta) - \zeta^*, \eta^*)] < \infty.$$
- (iii) (Feasibility.) There exists $P^\circ \in \mathcal{P}(\mathcal{Z})$ with $\mathbb{E}_{P^\circ}[m(Z; \theta)] \in \mathcal{M}(\rho)$ and $D_\phi(P^\circ \| P_*) < \delta$.
- (iv) (Integrability.) $\int_{\mathcal{Z}} |g(z; \theta)| d\nu(z) < \infty$ for every $\nu \in \mathcal{M}_+(\mathcal{Z})$ with $D_\phi(\nu \| P_*) < \infty$.
- (v) (Support and continuity.) In case (a) of Assumption SA-2(i), $g(\cdot; \theta)$ and $m(\cdot; \theta)$ are continuous on $\mathcal{Z}$ and $\text{supp}(P_*) = \mathcal{Z}$.
- (vi) (Slater condition.) If $d_m \geq 1$, the following holds: In case (a) of Assumption SA-2(i), $0 \in \text{int}(\text{co}(\mathcal{N}(\theta)) - \vec{\rho})$. In case (b) of Assumption SA-2(i), letting

$$\mathcal{N}_{\bar{x}}(\theta) := \left\{ \mathbb{E}_Q[m(Z; \theta)] : Q \in \Delta(\mathcal{Z}), Q \ll P_*, \frac{dQ}{dP_*} \leq \bar{x} \text{ } P_*\text{-a.s.} \right\} \subseteq \mathbb{R}^{d_m},$$

$$0 \in \text{int}(\mathcal{N}_{\bar{x}}(\theta) - \vec{\rho}).$$

Cases (a) and (b) in Assumption SA-2(i) cover the two ways a superlinear entropy can handle large density ratios. Case (a) allows arbitrarily large likelihood ratios at finite cost, as with the Kullback–Leibler or χ^2 divergences. Case (b) accommodates entropies whose domain bounds the likelihood ratio at some $\bar{x} \in (1, \infty)$, as with the modified entropy $\bar{\phi}$ induced by the baseline-as-mixture restriction of Section SA2.4, whose cap is $\bar{x} = 1/\kappa$.

As already highlighted, the two sets of assumptions differ mainly in the recession of the entropy function: under Assumption SA-1 the pointwise constraint in (21) is not vacuous, whereas under Assumption SA-2 it is vacuous. They also differ in how they control the objective. Assumption SA-1 requires $g(\cdot; \theta)$ to be bounded above, while Assumption SA-2 removes every boundedness requirement at the price of the stronger integrability condition of Assumption SA-2(iv) and of the support and continuity condition of Assumption SA-2(v). The feasibility margins of Assumptions SA-1(iii) and SA-2(iii) coincide as they both ask for a distribution that satisfies the moment restrictions strictly inside the divergence budget, which holds at $P^\circ = P_*$ whenever $\delta > 0$ and $\mathbb{E}_{P_*}[m(Z; \theta)] \in \mathcal{M}(\rho)$, as $D_\phi(P_* \| P_*) = \phi(1) = 0$, but also accommodates a misspecified baseline. Finally, both sets close with a Slater

condition, which requires room to perturb the moments in every direction simultaneously and is generically satisfied when the moment functions are sufficiently rich on $\mathcal{Z}$.

We now turn to the main results of this section. Define the dual support function

$$\sigma(\lambda, \theta) := \sup_{z \in \mathcal{Z}} \{g(z; \theta) - \lambda^\top m(z; \theta)\}, \quad \lambda \in \mathcal{R}_\lambda,$$

and, within the ambient cone $\mathcal{L} := \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$ of dual multipliers, the dual feasibility set

$$\mathcal{F}(\theta) := \left\{ (\zeta, \eta, \lambda) \in \mathcal{L} : \zeta + \eta \phi^\infty(1) \geq \sigma(\lambda, \theta) \right\},$$

and, gathering the multipliers into $\xi := (\zeta, \eta, \lambda)$, also define the dual objective as

$$f(\xi, \theta) := \lambda_{12}^\top \rho + \zeta + \eta \delta + \mathbb{E}_{P_*} [(\phi^*)^\pi (g(Z; \theta) - \lambda^\top m(Z; \theta) - \zeta, \eta)].$$

As we show in the proof of Lemma SA-18 below, the function $f(\cdot, \theta)$ is proper, convex, and lower semicontinuous, so that its recession function $f^\infty(\cdot, \theta)$ is defined as in Definition SA-2, and the set $\mathcal{F}(\theta)$ is closed and convex. The recession cone of a nonempty closed convex set $C \subseteq \mathbb{R}^k$ is $\text{rec}(C) := \{d \in \mathbb{R}^k : C + \tau d \subseteq C \text{ for all } \tau \geq 0\}$, and we call its elements the recession directions of C .

The next lemma presents some auxiliary results that will be useful in the arguments that follow. In particular, it establishes (i) the existence of a pair of distributions (Q_-, Q_+) at which the moment inequalities are strictly slack and the moment equalities are perturbed to either side of their targets, (ii) the existence of a strictly feasible point of $\mathcal{F}(\theta)$, and (iii) the strict positivity of the recession function of the dual objective along every nonzero recession direction of $\mathcal{F}(\theta)$.

Lemma SA-18 (Slater in dual form). *Suppose either Assumption SA-1 or Assumption SA-2 holds at $\theta \in \Theta$. Then:*

- (i) *(Perturbed-moment distributions.) When $d_m \geq 1$, there exist $Q_-, Q_+ \in \Delta(\mathcal{Z})$ and $\varepsilon > 0$ such that the inequality blocks are strictly slack at both distributions, that is*

$$\mathbb{E}_{Q_\pm} [m_1(Z; \theta)] < \rho_1, \quad \mathbb{E}_{Q_\pm} [m_3(Z; \theta)] < 0,$$

while the equality blocks are perturbed to either side of their targets, that is

$$\mathbb{E}_{Q_\pm} [m_2(Z; \theta)] = \rho_2 \pm 1_{d_2} \varepsilon \quad \text{and} \quad \mathbb{E}_{Q_\pm} [m_4(Z; \theta)] = \pm 1_{d_4} \varepsilon.$$

- (ii) *(Strictly feasible point.) Under Assumption SA-1, $\mathcal{F}(\theta)$ admits a strictly feasible point, that is, there exists $\xi^\circ = (\zeta^\circ, \eta^\circ, \lambda^\circ)$ in $\mathbb{R} \times \mathbb{R}_{++} \times \mathcal{R}_\lambda$ such that*

$$\zeta^\circ + \eta^\circ \phi^\infty(1) > \sigma(\lambda^\circ, \theta).$$

Under Assumption SA-2, the pointwise constraint is vacuous and $\mathcal{F}(\theta) = \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$.

- (iii) *(Positive recession function.) The recession function $f^\infty(\cdot, \theta)$ is strictly positive on every nonzero recession direction of $\mathcal{F}(\theta)$.*

[Proof]

The three parts of Lemma SA-18 play distinct roles in what follows. Part (i) is the device behind part (iii): the two distributions perturb the moment equalities to either side of their targets, which is what makes the recession function of the dual objective strictly positive along every nonzero recession direction of the feasible set. Part (ii) provides a strictly feasible dual point, which Lemma SA-21 uses to show that the dual value is finite and upper semicontinuous in θ . Part (iii) is what Theorem SA-19 invokes: by Proposition 3.2.2 in Bertsekas (2009), it is equivalent to the set of minimizers being nonempty and compact.

Theorem SA-19 (Existence of dual solutions). *Suppose either Assumption SA-1 or Assumption SA-2 holds at $\theta \in \Theta$. Then the set of minimizers of*

$$\begin{aligned} & \inf_{\substack{(\zeta, \eta, \lambda) \in \\ \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda}} \lambda_{12}^\top \rho + \zeta + \eta \delta + \mathbb{E}_{P_*} [(\phi^*)^\pi (g(Z; \theta) - \lambda^\top m(Z; \theta) - \zeta, \eta)] \\ & \text{s.t.} \quad \zeta + \eta \phi^\infty(1) \geq \sup_{z \in \mathcal{Z}} \{g(z; \theta) - \lambda^\top m(z; \theta)\}, \end{aligned} \tag{21}$$

is nonempty, compact, and convex.

[Proof]

We conclude by showing that the feasibility and Slater conditions of Assumptions SA-1 and SA-2 can be packaged, when all moment restrictions are equalities, into a single interior condition on the target, which is the form the main text adopts. The next lemma records the implication, together with a primitive sufficient condition.

Lemma SA-20 (Constraint qualification with moment equalities). *Let $\theta \in \Theta$ and $\delta > 0$, let ϕ be an entropy function in the sense of Definition SA-4, and suppose that all moment restrictions are equalities, so that $\mathcal{M}(\rho) = \{\bar{\rho}\}$ and $\mathcal{R}_\lambda = \mathbb{R}^{d_m}$. Let $\Gamma(\delta)$ be the attainable moment set of (3) at θ , and, whenever $\bar{x} < \infty$, let $\mathcal{N}_{\bar{x}}(\theta)$ be the set of Assumption SA-2(vi).*

- (i) (Consequences.) If $\vec{\rho} \in \text{int}(\Gamma(\delta))$, then (a) there exists $P^\circ \in \mathcal{P}(\mathcal{Z})$ with $\mathbb{E}_{P^\circ}[m(Z; \theta)] = \vec{\rho}$ and $D_\phi(P^\circ \| P_*) < \delta$; (b) $0 \in \text{int}(\text{co}(\mathcal{N}(\theta)) - \vec{\rho})$; and (c) if $\bar{x} < \infty$, then $0 \in \text{int}(\mathcal{N}_{\bar{x}}(\theta) - \vec{\rho})$.
- (ii) (A sufficient condition.) Suppose that $\phi(0) < \infty$ and $\phi^\infty(1) < \infty$. If there exists $P^\circ \in \mathcal{P}(\mathcal{Z})$ with $\mathbb{E}_{P^\circ}[\|m(Z; \theta)\|^2] < \infty$, $\mathbb{E}_{P^\circ}[m(Z; \theta)] = \vec{\rho}$, $D_\phi(P^\circ \| P_*) < \delta$, and a nonsingular covariance matrix $\mathbb{V}_{P^\circ}[m(Z; \theta)]$, then $\vec{\rho} \in \text{int}(\Gamma(\delta))$. If, moreover, $P^\circ \ll P_*$, the same conclusion holds for the set obtained from $\Gamma(\delta)$ by restricting to distributions dominated by P_* .

[Proof]

SA3.2 Existence of Solutions to the Outer Problem

The previous subsection established existence of dual solutions to the inner problem of (4) for each fixed value of the parameter $\theta \in \Theta$. We now turn to the outer problem. We first introduce the additional regularity conditions used for upper semicontinuity.

Assumption SA-3 (Outer regularity). Suppose that either Assumption SA-1 holds at every $\theta \in \Theta$ or Assumption SA-2 holds at every $\theta \in \Theta$, and that:

- (i) (Continuity in the parameter.) For each $\theta \in \Theta$, the map $\theta' \mapsto (g(z; \theta'), m(z; \theta'))$ is continuous at θ for P_* -almost every $z \in \mathcal{Z}$.
- (ii) (Upper semicontinuity of the dual support function.) If $\phi^\infty(1) < \infty$, then, for every $\lambda \in \mathcal{R}_\lambda$, the map $\theta \mapsto \sigma(\lambda, \theta)$ is upper semicontinuous at every $\theta \in \Theta$ at which it is finite.
- (iii) (Local domination.) If $\phi^\infty(1) = +\infty$, then, in case (a) of Assumption SA-2(i), for every $(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathcal{R}_\lambda$ and every $\theta \in \Theta$ there exist a radius $r > 0$ and a function $G \in L_1(P_*)$ such that

$$\eta \phi^* \left(\frac{g(z; \theta') - \lambda^\top m(z; \theta') - \zeta}{\eta} \right) \leq G(z)$$

for all $\theta' \in \mathcal{B}_r(\theta) \cap \Theta$ and P_* -almost every $z \in \mathcal{Z}$. In case (b) of Assumption SA-2(i), for each $\theta \in \Theta$ there exist a radius $r > 0$ and functions $G_g, G_m \in L_1(P_*)$ with

$$\sup_{\theta' \in \mathcal{B}_r(\theta) \cap \Theta} |g(z; \theta')| \leq G_g(z), \quad \sup_{\theta' \in \mathcal{B}_r(\theta) \cap \Theta} \|m(z; \theta')\| \leq G_m(z),$$

for P_* -almost every $z \in \mathcal{Z}$.

The three conditions play distinct roles, and we comment on each in turn. Assumption SA-3(i) asks for continuity in θ only at P_* -almost every point of the support, which accommodates, for example, indicator-type moment functions whose discontinuity locus is P_* -null, as in the Gumbel marital-premium example of Section SA6.2. Assumption SA-3(ii) is stated on the derived support function σ rather than on the primitives (g, m) . A sufficient condition on the primitives is that $\mathcal{Z}$ be compact and $(z, \theta) \mapsto (g(z; \theta), m(z; \theta))$ be jointly continuous, in which case $\theta \mapsto \sigma(\lambda, \theta)$ is continuous by Berge's maximum theorem. Finally, Assumption SA-3(iii) is the price of superlinearity: the constraint is then vacuous and the dual integrand must be controlled by a local domination condition. It has no counterpart under finite recession, where the dual constraint itself supplies the dominating bound.

Lemma SA-21 (Regularity of the dual value function). (i) (Finiteness.) If either Assumption SA-1 or Assumption SA-2 holds at $\theta \in \Theta$, then $V_P(\theta; \delta) \in \mathbb{R}$ and $V_D(\theta; \delta) \in \mathbb{R}$;

- (ii) (Upper semicontinuity.) Under Assumption SA-3, the dual value function $V_D(\cdot; \delta)$ defined in (21) is upper semicontinuous on Θ .

[Proof]

Combining the preceding auxiliary results with Theorem SA-19 gives the following theorem.

Theorem SA-22 (Existence of solutions and strong duality). Let $\Theta \subseteq \mathbb{R}^{d_\theta}$ be nonempty and compact and let Assumption SA-3 hold. Then the following hold:

- (i) (Existence of solutions.) The set of maximizers

$$\arg \sup_{\theta \in \Theta} V_D(\theta; \delta)$$

is nonempty and compact and, at every maximizer, the set of minimizers of the dual inner problem (21) is nonempty, compact, and convex.

- (ii) (Strong duality.) If, in addition, $\int_{\mathcal{Z}} |g(z; \theta)| d\nu(z) < \infty$ for every $\theta \in \Theta$ and every $\nu \in \mathcal{M}_+(\mathcal{Z})$ satisfying $D_\phi(\nu \| P_*) < \infty$, then $V_D(\theta; \delta) = V_P(\theta; \delta)$ for every $\theta \in \Theta$ and, consequently,

$$\arg \sup_{\theta \in \Theta} V_D(\theta; \delta) = \arg \sup_{\theta \in \Theta} V_P(\theta; \delta).$$

Remark SA-6 (Dual feasibility under mixture constraints). The results for the dual problem of this section cover the baseline-as-mixture sets whenever $0, 1/\kappa \in \text{dom}(\phi)$. Indeed, $\bar{\phi}$ has infinite recession, $\bar{x} = \sup \text{dom}(\bar{\phi}) = 1/\kappa \in (1, \infty)$, and both $\bar{\phi}(0) = \phi(0)$ and $\bar{\phi}(1/\kappa) = \phi(1/\kappa)$ are finite, so it falls under case (b) of Assumption SA-2(i). Lemma SA-21 and Theorem SA-22 then apply whenever the remaining conditions of Assumption SA-2 and of Assumption SA-3 hold, with the local domination condition of Assumption SA-3(iii) in its case-(b) form. The additional integrability condition in Theorem SA-22(ii) is automatic in this case as any $\nu \in \mathcal{M}_+(\mathcal{Z})$ with $D_{\bar{\phi}}(\nu \| P_*) < \infty$ admits a density with respect to P_* bounded by $1/\kappa$, so the condition follows from $\mathbb{E}_{P_*}[|g(Z; \theta)|] < \infty$, which holds as $P_* \in \mathcal{P}(\mathcal{Z})$. The contamination neighborhoods are outside the scope of these statements because the induced entropy ϕ has $\phi(0) = +\infty$. Since $\phi^\infty(1) = \phi^\infty(1)$ by Proposition SA-13(ii), this violates Assumption SA-1(ii) when $\phi^\infty(1) < \infty$, and Assumption SA-2(i), which requires 0 to lie in the domain, when $\phi^\infty(1) = +\infty$. ♣

SA4 Large Sample Properties

In this section, we develop the asymptotic theory for the plug-in estimators of the sensitivity bounds. In particular, we (i) establish consistency, (ii) characterize the asymptotic distribution, and (iii) propose a bootstrap procedure to construct valid confidence intervals. We work with a family of robustness sets which contains the two robustness sets (3.4) and (3.5) of the main text and is slightly less general than those defined via ϕ -divergences in the previous sections. Throughout, the baseline distribution P_* is an identified feature of the law P_0 of the observed data introduced in Remark SA-1, and it is estimated nonparametrically by the empirical measure $\hat{P}_n = n^{-1} \sum_{i=1}^n \delta_{Z_i}^{\text{Dirac}}$ of an i.i.d. sample $Z_1, \dots, Z_n$ from P_* . The case in which P_* is known, as in the structural model of Section SA6.2, is treated as a special case (see Remark SA-11). Sampling uncertainty comes from the estimator of the moment targets $\hat{\rho}_n$ and the empirical baseline $\hat{P}_n$.

Roadmap. Section SA4.1 introduces the family of robustness sets, the common form of their duals, and the two auxiliary objects on which the theory rests: a constraint qualification, which packages the Slater and feasibility-margin conditions of Section SA3.1, and a minimum-divergence function, which governs which parameter values are feasible within the budget. Section SA4.2 records the regularity conditions. Section SA4.3 proves that the integrands through which the baseline enters the dual form a Donsker class, and that the bound is continuous in (ρ, P) at the population value. From these we deduce, in turn, consistency, Hadamard directional differentiability, and the asymptotic distribution, the latter obtained from the joint empirical process limit through an appropriate directional delta method (Section SA4.4), and the validity of our bootstrap procedure (Section SA4.5). Throughout, the lower bound is obtained from the upper bound by substituting g with $-g$.

SA4.1 A Family of Robustness Sets

Fix $\underline{u}$ and $\bar{u}$ with $\underline{u} \in [0, 1)$ and $\bar{u} \in (1, +\infty]$ and write $\mathbb{U} := [\underline{u}, \bar{u}]$, understood as $[\underline{u}, +\infty)$ when $\bar{u} = +\infty$. The endpoint $\bar{u} = 1$ is excluded, as it would force $Q = P$ in the set defined next. Hence 1 is an interior point of $\mathbb{U}$, which makes the entropy function of Lemma SA-23 continuous at 1 and the set of dual minimizers in Lemma SA-24(iii) compact. For a distribution $P \in \Delta(\mathcal{Z})$, let

$$\mathcal{P}_{\mathbb{U}}(P) := \{Q \in \Delta(\mathcal{Z}) : \underline{u}P \leq Q \leq \bar{u}P\},$$

where the upper inequality is omitted when $\bar{u} = +\infty$. By the equivalence of (2) and (3) in Lemma SA-12, which holds for every $\kappa > 0$, applied with (P, Q) in place of (P_*, P) , with $\kappa = \underline{u}$ in part (ii) when $\underline{u} > 0$, and with $\kappa = 1/\bar{u}$ in part (i) when $\bar{u} < \infty$, a distribution Q belongs to $\mathcal{P}_{\mathbb{U}}(P)$ if and only if the density w of the absolutely continuous part of Q with respect to P satisfies $w \geq \underline{u}$ P -almost surely and, when $\bar{u} < \infty$, $Q \ll P$ with $w \leq \bar{u}$ P -almost surely. When $\bar{u} = +\infty$ the singular part of Q is unrestricted. For $\theta \in \Theta$, $\rho \in \mathbb{R}^{d_\theta}$, and a budget $\delta \in (0, 1]$ (larger budgets add nothing, as $\text{TV} \leq 1$), the robustness set is

$$\mathcal{P}_{\mathbb{U}}(\theta; \rho, P, \delta) := \{Q \in \mathcal{P}_{\mathbb{U}}(P) : \mathbb{E}_Q[m(Z; \theta)] \in \mathcal{M}(\rho), \text{TV}(Q \| P) \leq \delta\}. \quad (22)$$

Three members of the family are of interest:

1. when $\mathbb{U} = [0, +\infty)$ no weight is restricted and (22) is the plain total variation robustness set (2) with $\phi = \phi_{\text{TV}}$;
2. when $\mathbb{U} = [0, 1/\kappa]$ with $\kappa \in (0, 1)$, it is the set $\mathcal{P}_{\text{mix}}^{\leq}(\rho, \delta)$ of (7) with $\phi = \phi_{\text{TV}}$, which coincides with the decomposition neighborhood of Section SA2.6. At budget $(1 - \kappa)\delta$, it is the total variation robustness set (3.4) of the main text at budget δ , by (19);
3. when $\mathbb{U} = [1 - \delta, +\infty)$, it is the contamination set $\mathcal{C}_{1-\delta}(P) \cap \mathcal{E}(\rho)$ of Section SA2.6 and of the main text.

The next lemma identifies the family with the ϕ -divergence neighborhoods of Section SA2, which is what lets us import the duality theory developed there. Its proof applies Proposition SA-13 once for each endpoint of $\mathbb{U}$ and computes the conjugate directly.

Lemma SA-23. *Let $\phi_{\mathbb{U}} := \phi_{\text{TV}} + \iota_{\mathbb{U}}$. Then $\phi_{\mathbb{U}}$ is an entropy function in the sense of Definition SA-4, continuous at 1, with recession $\phi_{\mathbb{U}}^\infty(1) = 1/2$ if $\bar{u} = +\infty$ and $\phi_{\mathbb{U}}^\infty(1) = +\infty$ if $\bar{u} < \infty$. For every $P \in \Delta(\mathcal{Z})$ and $Q \in \Delta(\mathcal{Z})$,*

$$D_{\phi_{\mathbb{U}}}(Q \| P) = \begin{cases} \text{TV}(Q \| P), & Q \in \mathcal{P}_{\mathbb{U}}(P), \\ +\infty, & \text{otherwise,} \end{cases}$$

so that, whenever either $g(\cdot; \theta)$ and $m(\cdot; \theta)$ are bounded or $\bar{u} < \infty$ and $P \in \mathcal{P}(\mathcal{Z})$, $\mathcal{P}_{\mathbb{U}}(\theta; \rho, P, \delta) = \{Q \in \mathcal{P}(\mathcal{Z}) : \mathbb{E}_Q[m(Z; \theta)] \in \mathcal{M}(\rho), D_{\phi_{\mathbb{U}}}(Q \| P) \leq \delta\}$. Finally, the perspective of the conjugate of $\phi_{\mathbb{U}}$ is the function $r : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ given by

$$r(x, \eta) := \begin{cases} \max\{\underline{u}x - \frac{1-\underline{u}}{2}\eta, x, \bar{u}x - \frac{\bar{u}-1}{2}\eta\}, & \bar{u} < \infty, \\ \max\{\underline{u}x - \frac{1-\underline{u}}{2}\eta, x\}, & \bar{u} = +\infty \text{ and } x \leq \eta/2, \\ +\infty, & \bar{u} = +\infty \text{ and } x > \eta/2, \end{cases} \quad (23)$$

that is, $(\phi_{\mathbb{U}}^*)^\pi(x, \eta) = r(x, \eta)$ for every $x \in \mathbb{R}$ and $\eta \geq 0$.

For the three cases above, the integrand r reads $\max\{x, -\eta/2\}$ for the plain total variation robustness set, $\max\{-\eta/2, x, x/\kappa - (1 - \kappa)\eta/(2\kappa)\}$ for the decomposition neighborhood, and $\max\{(1 - \delta)x - \delta\eta/2, x\}$ for the contamination robustness set. In the first and third cases $\bar{u} = +\infty$, so that these expressions apply for $x \leq \eta/2$, whereas $r(x, \eta) = +\infty$ for $x > \eta/2$. The last two are the integrands r_{TV} and r_{C} in (4.4) of the main text.

Write $u^\dagger := \bar{u}$ if $\bar{u} < \infty$ and $u^\dagger := 1$ if $\bar{u} = +\infty$ for the largest finite slope. Four properties of r are used repeatedly and follow directly from (23):

- (r1) $r(x, \eta) = \sup_{u \in \mathbb{U}} \{ux - \eta|u - 1|/2\}$ for every $x \in \mathbb{R}$ and $\eta \geq 0$;
- (r2) r is jointly convex and positively homogeneous of degree one in (x, η) , and $r(x, \cdot)$ is nonincreasing;
- (r3) on its effective domain, r is Lipschitz: $|r(x, \eta) - r(x', \eta')| \leq u^\dagger|x - x'| + \frac{u^\dagger}{2}|\eta - \eta'|$, and $|r(x, \eta)| \leq u^\dagger|x| + \frac{u^\dagger}{2}\eta$;
- (r4) $r(x, \eta) \geq x$ and $r(x, \eta) \geq \underline{u}x - \frac{1-\underline{u}}{2}\eta$. If $\bar{u} < \infty$, also $r(x, \eta) \geq \bar{u}x - \frac{\bar{u}-1}{2}\eta$.

The dual. For $\xi = (\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}$ and $\theta \in \Theta$, define the dual integrand

$$\ell_{\xi, \theta}(z) := r(g(z; \theta) - \lambda^\top m(z; \theta) - \zeta, \eta), \quad z \in \mathcal{Z}, \quad (24)$$

the dual objective

$$J(\xi, \theta; \rho, P) := \zeta + \eta\delta + \lambda_{12}^\top \rho + P\ell_{\xi, \theta}, \quad (25)$$

and, with $\sigma(\lambda, \theta)$ the dual support function of Section SA3.1, the dual feasible set

$$\mathcal{F}(\theta) := \begin{cases} \{\xi \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda : \zeta + \frac{\eta}{2} \geq \sigma(\lambda, \theta)\}, & \bar{u} = +\infty, \\ \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda, & \bar{u} < \infty. \end{cases} \quad (26)$$

In view of Lemma SA-23, the program $\inf_{\xi \in \mathcal{F}(\theta)} J(\xi, \theta; \rho, P)$ is the dual of Theorem SA-11 for the entropy $\phi_{\mathbb{U}}$ and the baseline P : the pointwise constraint $\zeta + \eta\phi_{\mathbb{U}}^\infty(1) \geq \sigma(\lambda, \theta)$ is present when $\bar{u} = +\infty$, because singular mass is then admissible and priced at the finite recession $1/2$, and vacuous when $\bar{u} < \infty$, because its left-hand side is then $+\infty$ for every $\eta \geq 0$ under the convention of that theorem. On $\mathcal{F}(\theta)$ the integrand (24) is finite, since the constraint forces $g - \lambda^\top m - \zeta \leq \eta/2$ on $\mathcal{Z}$ when $\bar{u} = +\infty$. For the total variation robustness set of the main text, the budget in (25) is $(1 - \kappa)\delta$ by the second item of the list above, so that the term $(1 - \kappa)\eta\delta$ is the term $c_{\text{TV}}\eta\delta$ of (4.2) of the main text, with $c_{\text{TV}} = 1 - \kappa$. We call

$$\bar{K}_\delta(\theta; \rho, P) := \inf_{\xi \in \mathcal{F}(\theta)} J(\xi, \theta; \rho, P) \in [-\infty, +\infty) \quad (27)$$

the criterion function. It is well defined at every (θ, ρ, P) at which $g(\cdot; \theta)$ is bounded above, because $\mathcal{F}(\theta)$ then contains the point $(\sup_{z \in \mathcal{Z}} g(z; \theta), 0, 0)$, at which J is finite by (r3). The representation (27) has two structural properties on which every large sample argument below relies:

(L) Joint linearity. For each ξ and θ , $J(\xi, \theta; \rho, P)$ is affine in ρ and linear in P . Consequently, by (I), $(\rho, P) \mapsto \bar{K}_\delta(\theta; \rho, P)$ is concave, as an infimum over the fixed set $\mathcal{F}(\theta)$ of jointly affine maps, with supergradient $(\lambda_{12}, \ell_{\xi, \theta})$ at any minimizer ξ .

(I) Independence. The feasible set $\mathcal{F}(\theta)$ depends only on $g, m, \theta, \mathcal{Z}$, and neither on ρ nor on P .

Attainable moments and the constraint qualification. Let

$$\Gamma_{\mathbb{U}}(\theta; P, \delta) := \{\mathbb{E}_Q[m(Z; \theta)] : Q \in \mathcal{P}_{\mathbb{U}}(P), \text{TV}(Q \| P) \leq \delta\} \subseteq \mathbb{R}^{d_m} \quad (28)$$

be the set of moment vectors attainable within the budget, a convex set because $\mathcal{P}_{\mathbb{U}}(P)$ is convex and $\text{TV}(\cdot \| P)$ is convex. By Lemma SA-23, it is the counterpart of the attainable moment set (3) for the entropy $\phi_{\mathbb{U}}$ and the baseline P . We say that the constraint qualification holds at (θ, ρ, P) with radius $r > 0$ if there is $w_0 \in \mathcal{M}(\rho)$ with

$$\mathcal{B}_r(w_0) \subseteq \Gamma_{\mathbb{U}}(\theta; P, \delta), \quad (29)$$

and that it holds at (θ, ρ, P) if it does so for some radius. Equivalently, $\mathcal{M}(\rho) \cap \text{int}(\Gamma_{\mathbb{U}}(\theta; P, \delta)) \neq \emptyset$. When no moment restrictions are present, the condition is vacuous, as we then set $\Gamma_{\mathbb{U}}(\theta; P, \delta) := \{0\}$ in line with the conventions of Section SA1.2. In words, the moment restrictions must be attainable within the budget with room to perturb the attained moment vector in every direction. Condition (29) and the feasibility and Slater conditions of Section SA3.1 differ in one respect. With inequality restrictions, the Slater conditions require the specific point $\vec{\rho}$ to be interior to the budget-free attainable set, whereas (29) only requires some point of $\mathcal{M}(\rho)$ to be interior to the attainable set within the budget, which is weaker. With equality restrictions only, $\mathcal{M}(\rho) = \{\vec{\rho}\}$ and the two coincide.

Feasible parameters and the objects of interest. Define the minimum-divergence function, that is, the smallest distance from the baseline P to a candidate distribution whose moments satisfy the restrictions at θ ,

$$D(\theta; \rho, P) := \inf \{\text{TV}(Q \| P) : Q \in \mathcal{P}_{\mathbb{U}}(P), \mathbb{E}_Q[m(Z; \theta)] \in \mathcal{M}(\rho)\}, \quad (30)$$

with $D(\theta; \rho, P) = +\infty$ when no such Q exists. Its dual representation, given in Lemma SA-25 below, carries a

pointwise restriction on its test functions when the weights are unbounded above, which we record through the set

$$\mathcal{F}_D(\theta) := \begin{cases} \{(\zeta, \lambda) \in \mathbb{R} \times \mathcal{R}_\lambda : \zeta + \frac{1}{2} \geq \sigma^\infty(\lambda, \theta)\}, & \bar{u} = +\infty, \\ \mathbb{R} \times \mathcal{R}_\lambda, & \bar{u} < \infty, \end{cases} \quad \sigma^\infty(\lambda, \theta) := \sup_{z \in \mathcal{Z}} \{-\lambda^\top m(z; \theta)\},$$

the analog for the projection problem of the constraint defining $\mathcal{F}(\theta)$, obtained by setting $g \equiv 0$ and $\eta = 1$ there. We then define the strictly feasible parameter set, the bound, the outer argmax, the dual optimizer set, and the set of optimal estimated-block multipliers as

$$\begin{aligned} \Theta_\delta(\rho, P) &:= \{\theta \in \Theta : D(\theta; \rho, P) < \delta\}, & \bar{b}(\delta, \rho, P) &:= \sup_{\theta \in \Theta_\delta(\rho, P)} \bar{K}_\delta(\theta; \rho, P), \\ \bar{\Theta}_\delta(\rho, P) &:= \arg \sup_{\theta \in \Theta} \bar{K}_\delta(\theta; \rho, P), & \bar{\Xi}_\delta(\theta; \rho, P) &:= \arg \inf_{\xi \in \mathcal{F}(\theta)} J(\xi, \theta; \rho, P), \end{aligned}$$

and $\bar{\Lambda}_\delta(\theta; \rho, P) := \Pi_{12}(\bar{\Xi}_\delta(\theta; \rho, P))$, where $\Pi_{12}(\zeta, \eta, \lambda) = \lambda_{12}$ is the projection onto the estimated-block multipliers, that is, the multipliers λ_{12} of Section SA1.2, which are attached to the moments whose targets $\rho = (\rho_1^\top, \rho_2^\top)^\top$ are estimated. The lower bound objects $\underline{K}_\delta, \underline{b}, \underline{\Theta}_\delta, \underline{\Xi}_\delta, \underline{\Lambda}_\delta$ are obtained by using $-g$ in place of g and, writing $\tilde{K}, \tilde{b}, \tilde{\Theta}^*, \tilde{\Xi}^*, \tilde{\Lambda}^*$ for $\bar{K}_\delta, \bar{b}, \bar{\Theta}_\delta, \bar{\Xi}_\delta, \bar{\Lambda}_\delta$, setting $\underline{K}_\delta := -\tilde{K}$ and $\underline{b} := -\tilde{b}$, so that $\underline{b}(\delta, \rho, P) = \inf_{\theta \in \Theta_\delta(\rho, P)} \underline{K}_\delta(\theta; \rho, P)$ is the lower sensitivity bound. Finally, the plug-in estimators of the bounds are

$$\hat{\bar{b}}_n(\delta) := \bar{b}(\delta, \hat{\rho}_n, \hat{P}_n), \quad \hat{\underline{b}}_n(\delta) := \underline{b}(\delta, \hat{\rho}_n, \hat{P}_n).$$

Two auxiliary results. The first lemma specializes the duality theory of Section SA2 to the family of robustness sets considered in this section.

Lemma SA-24 (Duality for the family). *Fix $\theta \in \Theta$, let $\mathcal{Z}$ be compact, let $g(\cdot; \theta)$ and $m(\cdot; \theta)$ be continuous on $\mathcal{Z}$, and let $\delta \in (0, 1]$, $\rho \in \mathbb{R}^{d^*}$, and $P \in \Delta(\mathcal{Z})$.*

- (i) (Weak duality with slack.) *For every $Q \in \mathcal{P}_U(P)$ and every $\xi = (\zeta, \eta, \lambda) \in \mathcal{F}(\theta)$,*

$$J(\xi, \theta; \rho, P) - \mathbb{E}_Q[g(Z; \theta)] \geq \eta(\delta - \text{TV}(Q \| P)) + (\lambda_{12}^\top \rho - \lambda^\top \mathbb{E}_Q[m(Z; \theta)]),$$
and both terms on the right are nonnegative when $Q \in \mathcal{P}_U(\theta; \rho, P, \delta)$. In particular, $\bar{K}_\delta(\theta; \rho, P)$ is at least the upper bound over (22), and it is finite whenever (22) is nonempty.
- (ii) (Strong duality.) *If the constraint qualification holds at (θ, ρ, P) , then*

$$\bar{K}_\delta(\theta; \rho, P) = \sup_{Q \in \mathcal{P}_U(\theta; \rho, P, \delta)} \mathbb{E}_Q[g(Z; \theta)].$$
- (iii) (Existence of dual solutions.) *The recession cone of $\mathcal{F}(\theta)$ is $0^+ \mathcal{F}(\theta) = \{d = (\zeta_d, \eta_d, \lambda_d) : \eta_d \geq 0, \lambda_d \in \mathcal{R}_\lambda, \zeta_d + \eta_d/2 \geq \sigma^\infty(\lambda_d, \theta)\}$ if $\bar{u} = +\infty$ and $0^+ \mathcal{F}(\theta) = \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$ if $\bar{u} < \infty$, and the recession function of $J(\cdot, \theta; \rho, P)$ on $\mathcal{F}(\theta)$ is*

$$J^\infty(d, \theta; \rho, P) = \zeta_d + \eta_d \delta + \lambda_{d,12}^\top \rho + P r(-\lambda_d^\top m(\cdot; \theta) - \zeta_d, \eta_d), \quad d \in 0^+ \mathcal{F}(\theta).$$
If the constraint qualification holds at (θ, ρ, P) , then $J^\infty(d, \theta; \rho, P) > 0$ for every nonzero $d \in 0^+ \mathcal{F}(\theta)$, and the set $\bar{\Xi}_\delta(\theta; \rho, P)$ of minimizers of (27) is nonempty, compact, and convex.
- (iv) (Equivalent form.) *The constraint qualification holds at (θ, ρ, P) if and only if there exists $Q^\circ \in \mathcal{P}_U(P)$ with $\mathbb{E}_{Q^\circ}[m(Z; \theta)] \in \mathcal{M}(\rho)$ and $\text{TV}(Q^\circ \| P) < \delta$, and $\mathcal{M}(\rho) \cap \text{int}(\Gamma_U(\theta; P, 1)) \neq \emptyset$. In particular, it implies $D(\theta; \rho, P) < \delta$, that is, $\theta \in \Theta_\delta(\rho, P)$.*

[Proof]

Two comments are in order. First, part SA-24(iii) is the counterpart of Theorem SA-19. The latter cannot be invoked directly for every member because the entropy ϕ_U is infinite at zero when $\underline{u} > 0$, which violates Assumption SA-1(ii), whereas the direct argument given in the proof only uses the slack inequality of part SA-24(i) together with the room provided by the constraint qualification. Second, part SA-24(iv) connects the constraint qualification to the primitives.

The second lemma collects the properties of the minimum-divergence function on which the theory relies: its dual representation, how it relates to the feasibility of the criterion function, and its stability under perturbations of (θ, ρ, P) . The last property is what keeps the population optimizers feasible when the inputs are replaced by their sample analogs. To measure the baseline perturbation we use the class of recession integrands

$$\mathcal{H}_\mathcal{A}^\infty := \{z \mapsto r(-\lambda^\top m(z; \theta) - \zeta, \eta) : \theta \in \Theta, \xi = (\zeta, \eta, \lambda) \in \mathcal{A}, \text{ and } \zeta + \frac{\eta}{2} \geq \sigma^\infty(\lambda, \theta) \text{ when } \bar{u} = +\infty\}, \quad (31)$$

indexed by a compact set $\mathcal{A} \subseteq \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}$, in which the multiplier λ is not restricted to $\mathcal{R}_\lambda$. When $\bar{u} = +\infty$, the restriction on the index places the argument of r in its domain at every z . When $\bar{u} < \infty$, the function r is finite everywhere by (23), and no restriction is needed. Lemma SA-26 below shows that the class is P_* -Donsker.

Lemma SA-25 (The minimum-divergence function). *Fix $\theta \in \Theta$ and $\delta \in (0, 1]$, let $\mathcal{Z}$ be compact and $m(\cdot; \theta) \in C(\mathcal{Z}; \mathbb{R}^{d_m})$, and let $\rho \in \mathbb{R}^{d^*}$ and $P \in \Delta(\mathcal{Z})$. Whenever the moment-constrained set in (30) is nonempty, the infimum defining $D(\theta; \rho, P)$ is attained.*

- (i) (Dual representation.) If there is $w_0 \in \mathcal{M}(\rho)$ with $\mathcal{B}_r(w_0) \subseteq \Gamma_{\mathbb{U}}(\theta; P, 1)$ for some $r > 0$, in particular if the constraint qualification holds at (θ, ρ, P) with radius r , then

$$D(\theta; \rho, P) = \max_{(\zeta, \lambda) \in \mathcal{F}_D(\theta)} v_P(\zeta, \lambda), \quad v_P(\zeta, \lambda) := -\mathbb{E}_P[r(-\zeta - \lambda^\top m(Z; \theta), 1)] - \zeta - \lambda_{12}^\top \rho. \quad (32)$$

Moreover, every maximizer (ζ^*, λ^*) satisfies $\|\lambda^*\| \leq 1/r$ and $|\zeta^*| \leq \frac{1}{2} + \sup_{z \in \mathcal{Z}} \|m(z; \theta)\|/r$.

- (ii) (Criterion dichotomy.) Let $g(\cdot; \theta) \in C(\mathcal{Z}; \mathbb{R})$. If $D(\theta; \rho, P) \leq \delta$, then $\bar{K}_\delta(\theta; \rho, P) \in \mathbb{R}$. If the hypothesis of (i) holds and $D(\theta; \rho, P) > \delta$, then $\bar{K}_\delta(\theta; \rho, P) = -\infty$.
- (iii) (Persistence and local Lipschitz continuity.) Suppose m is jointly continuous on $\mathcal{Z} \times \Theta$ with Θ compact, and that the constraint qualification holds at (θ, ρ, P) with radius r . Then there exist $\varepsilon > 0$, a constant $C < \infty$, and a compact set $\mathcal{A}_D \subseteq \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}$ such that the constraint qualification holds, with radius $r/3$, at every $(\theta', \rho', P') \in \Theta \times \mathbb{R}^{d_0} \times \Delta(\mathcal{Z})$ with

$$\|\theta' - \theta\| + \|\rho' - \rho\| + \|P' - P\|_{\mathcal{H}_{\mathcal{A}_D}^\infty} < \varepsilon,$$

and, for all such triples (θ_1, ρ_1, P_1) and (θ_2, ρ_2, P_2) ,

$$|D(\theta_1; \rho_1, P_1) - D(\theta_2; \rho_2, P_2)| \leq C \left(\|\rho_1 - \rho_2\| + \sup_{z \in \mathcal{Z}} \|m(z; \theta_1) - m(z; \theta_2)\| + \|P_1 - P_2\|_{\mathcal{H}_{\mathcal{A}_D}^\infty} \right).$$

The constant C and the set $\mathcal{A}_D$ depend only on r , δ , $\mathbb{U}$, and $\sup_{(z, \theta') \in \mathcal{Z} \times \Theta} \|m(z; \theta')\|$, while ε depends in addition on the modulus of continuity of m on $\mathcal{Z} \times \Theta$.

[Proof]

Remark SA-7 (A specification statistic in probability units). The minimum-divergence function of (30) might be a quantity of interest in itself. Indeed, it measures the smallest probability mass that must be reallocated from the baseline, within the family of distributions selected by $\mathbb{U}$, for the moment restrictions to hold at θ . Its infimum over θ is, under total variation, the smallest divergence needed to reconcile the baseline with the model (Section SA1.2). This is the total variation analogue of the optimally-transported generalized method of moments criterion of Schennach and Starck (2026), which reconciles model and data through the least mean squared transport cost. ♣

SA4.2 Regularity Conditions

The following assumption collects the regularity conditions of the large sample theory. It is the analog of Assumption M in Christensen and Connault (2023) (CC23, henceforth), modified to accommodate the non-smoothness of the conjugate, the estimation of the baseline, and the family of robustness sets.

Assumption SA-4 (Estimation regularity). *The following conditions hold.*

- (i) (Primitive regularity.) $\mathcal{Z}$ is compact; Θ is a compact subset of $\mathbb{R}^{d_0}$; the maps $(z, \theta) \mapsto g(z; \theta)$ and $(z, \theta) \mapsto m(z; \theta)$ are jointly continuous on $\mathcal{Z} \times \Theta$ and Lipschitz in θ uniformly in z : there is $L < \infty$ with $\sup_{z \in \mathcal{Z}} |g(z; \theta) - g(z; \theta')| \leq L\|\theta - \theta'\|$ and $\sup_{z \in \mathcal{Z}} \|m(z; \theta) - m(z; \theta')\| \leq L\|\theta - \theta'\|$ for all $\theta, \theta' \in \Theta$.
- (ii) (Constraint qualification.) The constraint qualification (29) holds at (θ, ρ_0, P_*) for each $\theta \in \underline{\Theta}_\delta(\rho_0, P_*) \cup \bar{\Theta}_\delta(\rho_0, P_*)$.
- (iii) (Feasibility and strictly feasible optimizers.) $\underline{\Theta}_\delta(\rho_0, P_*) \cup \bar{\Theta}_\delta(\rho_0, P_*) \subseteq \Theta_\delta(\rho_0, P_*) \neq \emptyset$.
- (iv) (Asymptotic linearity.) The estimator $\hat{\rho}_n$ of ρ_0 is asymptotically linear, $\hat{\rho}_n = \rho_0 + \hat{P}_n \psi + o_{\mathbb{P}}(n^{-1/2})$, with influence function $\psi : \mathcal{Z} \rightarrow \mathbb{R}^{d_0}$ measurable and satisfying $P_* \psi = 0$ and $P_* \|\psi\|^2 < \infty$. The known block carries no sampling uncertainty.
- (v) (Dual optimizer lower hemicontinuity.) The maps $\theta \mapsto \underline{\Xi}_\delta(\theta; \rho_0, P_*)$ and $\theta \mapsto \bar{\Xi}_\delta(\theta; \rho_0, P_*)$ are lower hemicontinuous at each $\theta \in \underline{\Theta}_\delta(\rho_0, P_*)$ and $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$, respectively.

We now comment on each piece in turn.

On SA-4(i). Compactness of $\mathcal{Z}$ is the substantive requirement, with no counterpart in CC23, who impose compactness only on Θ (their Assumption M(v)). It lets us control the support supremum $\sigma(\lambda, \theta)$ in the pointwise constraint through Berge's maximum theorem, and it bounds the dual integrand (24) uniformly over compact sets of multipliers. Proposition SA-35 relaxes it in the RCT with attrition application. Continuity of g and m is the analog of CC23's Assumption M(i)–(ii), and the Lipschitz requirement makes the integrand class Donsker (Lemma SA-26).

On SA-4(ii). This condition is the analog of CC23's Condition S' (their Assumption M(iii)), except that we impose it only at the population optimizers rather than on all of Θ_δ . Its validity at nearby triples (θ, ρ, P) is supplied by Lemma SA-25(iii). By Lemma SA-24(iv) it is a classical strict-feasibility condition, and by Lemma SA-24(iii) it makes the dual objective coercive, so that $\underline{\Xi}_\delta(\theta; \rho_0, P_*)$ and $\bar{\Xi}_\delta(\theta; \rho_0, P_*)$ are nonempty, compact, and convex.

On SA-4(iii). The inclusion is implied by SA-4(ii) through Lemma SA-24(iv) and is recorded for ease of reference. Nonemptiness of $\Theta_\delta(\rho_0, P_*)$ ensures that the criterion is not identically $-\infty$. Strictness of the budget inequality at the optimizers keeps them feasible under small perturbations of (ρ, P) (Lemma SA-27(i)), on which the consistency and directional-differentiability arguments rely. It is the analog of CC23's Assumption M(vi).

On SA-4(iv). Under this condition, $\sqrt{n}(\hat{\rho}_n - \rho_0) = \mathbb{G}_n\psi + o_{\mathbb{P}}(1)$, so that $\mathbb{Z} := \mathbb{G}\psi \sim \mathcal{N}(0, P_*\psi\psi^\top)$ and the process limit $\mathbb{G}$ are jointly Gaussian with cross-covariance $\mathbb{E}[\mathbb{G}f\mathbb{Z}] = P_*(f\psi)$. We impose the linear expansion rather than marginal asymptotic normality because the delta method of Theorem SA-29 uses the joint law of $(\sqrt{n}(\hat{\rho}_n - \rho_0), \sqrt{n}(\hat{P}_n - P_*))$, which determines the asymptotic variance of the bounds (Remark SA-9) and the consistency of the bootstrap (Theorem SA-30). It is the counterpart of the $\sqrt{n}$ -Gaussian hypothesis that CC23's Theorem 6.2 places on the estimated vector of targeted moments, the analog of $\hat{\rho}_n$.

On SA-4(v). This condition is the counterpart of CC23's Assumption M(vii), and its strengthening for the bootstrap (Assumption SA-5) is the counterpart of their Assumption M(vii'). It is imposed on $\bar{\Xi}_\delta$ rather than on $\bar{\Lambda}_\delta$ because the derivative depends on the entire dual optimizer ξ through the integrand $\ell_{\xi, \theta}$. We use it only to establish the directional differentiability of the value function. It holds automatically when the dual optimizer is unique on $\Theta_\delta \cup \bar{\Theta}_\delta$, since the optimizer sets are nonempty and locally bounded (Lemma SA-27(ii)) and upper hemicontinuous in θ (Lemma SA-27(iii)).

SA4.3 Auxiliary Results

We first record the Donsker property of the integrand classes through which the baseline enters the program. The objective J , the minimum-divergence function D in its dual form (32), and the recession function of Lemma SA-24(iii) each depend on P only through an expectation, respectively of the dual integrands $\ell_{\xi, \theta}$ of (24), of the divergence integrands $r(-\zeta - \lambda^\top m(z; \theta), 1)$, and of the recession integrands (31). All are of the form $z \mapsto r(a(z; \theta, \zeta, \lambda), \eta)$, where

$$a(z; \theta, \zeta, \lambda) := g(z; \theta) - \lambda^\top m(z; \theta) - \zeta,$$

and $g \equiv 0$ for the last two classes, so a single lemma covers all of them.

Lemma SA-26 (Donsker property). *Let Assumption SA-4(i) hold and let $\mathcal{A} \subseteq \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}$ be compact. Then each of the classes*

$$\mathcal{H}_\mathcal{A}^J := \{\ell_{\xi, \theta} : \theta \in \Theta, \xi \in \mathcal{A} \cap \mathcal{F}(\theta)\},$$

$$\mathcal{H}_\mathcal{A}^D := \{z \mapsto r(-\zeta - \lambda^\top m(z; \theta), 1) : \theta \in \Theta, (\zeta, \lambda) \in \Pi_{\zeta\lambda}(\mathcal{A}) \cap \mathcal{F}_D(\theta)\},$$

with $\Pi_{\zeta\lambda}(\mathcal{A})$ the projection of $\mathcal{A}$ onto its (ζ, λ) coordinates, as well as the class of recession integrands $\mathcal{H}_\mathcal{A}^\infty$ of (31), is P_* -Donsker with bounded envelope, hence P_* -Glivenko–Cantelli.

[Proof]

Lemma SA-24(iii) gives, at each population optimizer and at the population value (ρ_0, P_*) , a nonempty compact set of dual minimizers, and Lemma SA-25(iii) shows how the constraint qualification and of the strict feasibility of the optimizers are maintained locally. With an estimated baseline, these conclusions must hold uniformly over joint perturbations of (ρ, P) , which index the Donsker class used below. Properties (L) and (I) reduce this requirement to measuring changes in P in a seminorm under which $P \mapsto P\ell$ is continuous on the relevant integrands ℓ .

The estimated baseline enters the large sample analysis through the empirical process $\mathbb{G}_n = \sqrt{n}(\hat{P}_n - P_*)$, which we view as a random element of $\ell^\infty(\mathcal{H})$ for a class $\mathcal{H}$ of functions. We take

$$\mathcal{H} := \mathcal{H}_\mathcal{A}^J \cup \mathcal{H}_\mathcal{A}^{J,-} \cup \mathcal{H}_\mathcal{A}^D \cup \mathcal{H}_\mathcal{A}^\infty,$$

where $\mathcal{A} \subseteq \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}$ is a compact index set chosen below, $\mathcal{H}_\mathcal{A}^J$, $\mathcal{H}_\mathcal{A}^D$, and $\mathcal{H}_\mathcal{A}^\infty$ are the classes of Lemma SA-26, and $\mathcal{H}_\mathcal{A}^{J,-}$ is the class $\mathcal{H}_\mathcal{A}^J$ formed with $-g$ in place of g , the substitution that defines the lower bound objects. Whatever the compact set $\mathcal{A}$, the class $\mathcal{H}$ is a finite union of the P_* -Donsker classes of Lemma SA-26 and so itself P_* -Donsker. For any class of functions $\mathcal{G}$, with $\|P - P'\|_\mathcal{G} = \sup_{f \in \mathcal{G}} |Pf - P'f|$, define

$$|||(\rho, P) - (\rho', P')|||_\mathcal{G} := \|\rho - \rho'\| + \|P - P'\|_\mathcal{G}$$

on the pair (ρ, P) . We equip the pair (ρ, P) with the seminorm $|||\cdot|||_\mathcal{H}$, and throughout the rest of the section the convergence $(\rho, P) \rightarrow (\rho_0, P_*)$ is understood as $|||(\rho, P) - (\rho_0, P_*)|||_\mathcal{H} \rightarrow 0$.

It remains to choose the index set $\mathcal{A}$, which has to serve two purposes. First, the integrand $\ell_{\xi, \theta}$ belongs to $\mathcal{H}$ only if $\xi \in \mathcal{A}$, so $\mathcal{A}$ must contain the dual optimizers of the upper- and lower bound programs realized near the population value. Second, the baseline enters the minimum-divergence function only through expectations of recession integrands, and Lemma SA-25(iii) states the persistence of the constraint qualification and of the strict feasibility of the optimizers in a norm over compact families of such integrands, so $\mathcal{A}$ must contain the indices of these families. Accordingly, we fix any compact set

$$\mathcal{A} \subseteq \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m} \quad \text{with} \quad \mathcal{A} \supseteq \Xi \cup \Xi^- \cup \mathcal{A}_0,$$

where Ξ and Ξ^- are bounded sets collecting the dual optimizers of the two programs and $\mathcal{A}_0$ is a compact set of indices of recession integrands. We define these sets in turn.

The sets Ξ and Ξ^- . Let $\mathcal{H}_0 := \mathcal{H}_{\mathcal{A}_0}^\infty$, with $\mathcal{A}_0$ as defined below. For each $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$ there is a radius $\varepsilon_\theta > 0$ such that, over all triples (θ', ρ', P') with $\|\theta' - \theta\| + \|\rho' - \rho_0\| + \|P' - P_*\|_{\mathcal{H}_0} < \varepsilon_\theta$, the sets $\bar{\Xi}_\delta(\theta'; \rho', P')$ are nonempty and compact and their union is bounded, as we verify in Section SA8.23. The balls $\mathcal{B}_{\varepsilon_\theta}(\theta)$ cover the compact set $\bar{\Theta}_\delta(\rho_0, P_*)$, so we can extract a finite subcover, with centers θ_k and radii $\varepsilon_k := \varepsilon_{\theta_k}$, collect the associated sets of triples into the set

$$\mathcal{N}_0 := \bigcup_k \{(\theta', \rho', P') \in \Theta \times \mathbb{R}^{\mathbf{d}_\theta} \times \Delta(\mathcal{Z}) : \|\theta' - \theta_k\| + \|\rho' - \rho_0\| + \|P' - P_*\|_{\mathcal{H}_0} < \varepsilon_k\},$$

and let

$$\Xi := \bigcup_{(\theta', \rho', P') \in \mathcal{N}_0} \bar{\Xi}_\delta(\theta'; \rho', P') \subseteq \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{\mathbf{d}_m},$$

which is bounded as a finite union of bounded sets. Repeating the construction with $-g$ in place of g yields a bounded set Ξ^- containing the optimizers of the lower program over an analogous cover of $\bar{\Theta}_\delta(\rho_0, P_*)$.

The set $\mathcal{A}_0$. This set fixes the reference points at which the minimum-divergence function is controlled. By Lemma SA-25(iii), each $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$ admits a radius $\varepsilon_\theta^D > 0$, a constant $C_\theta < \infty$, and a compact index set $\mathcal{A}_{D, \theta}$ such that the constraint qualification, with radius $r_\theta/3$, and the Lipschitz estimate of that lemma are available at every (θ', ρ', P') with $\|\theta' - \theta\| + \|\rho' - \rho_0\| + \|P' - P_*\|_{\mathcal{H}_{\mathcal{A}_{D, \theta}}^\infty} < \varepsilon_\theta^D$, where $r_\theta > 0$ is a radius with which the constraint qualification of Assumption SA-4(ii) holds at (θ, ρ_0, P_*) , fixed once and for all. Since $\bar{\Theta}_\delta(\rho_0, P_*)$ is compact, finitely many of the balls $\mathcal{B}_{\varepsilon_\theta^D/2}(\theta)$, with centers $\theta_1, \dots, \theta_J$, cover it, and we fix such a subcover once and for all. Let

$$\mathcal{A}_0 := \bigcup_{j=1}^J \mathcal{A}_{D, \theta_j},$$

which is compact as a finite union of compact subsets of $\mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{\mathbf{d}_m}$. Since $\mathcal{H}_{\mathcal{A}}^\infty$ grows with its index set $\mathcal{A}$, the single class $\mathcal{H}_0$ contains $\mathcal{H}_{\mathcal{A}_{D, \theta_j}}^\infty$ for every $j \in [J]$, and thus controls the perturbation of the baseline at every center of the subcover. For the same reason $\mathcal{A}_0 \subseteq \mathcal{A}$ gives $\mathcal{H}_0 \subseteq \mathcal{H}$, so that $\|\cdot\|_{\mathcal{H}_0} \leq \|\cdot\|_{\mathcal{H}}$ and every property established on a $\|\cdot\|_{\mathcal{H}_0}$ -neighborhood of (ρ_0, P_*) holds on the $\|\cdot\|_{\mathcal{H}}$ -neighborhood of the same radius.

The next lemma collects the regularity properties of the dual program and of the bound on which the consistency and directional-differentiability arguments rest. The large sample analysis perturbs the pair (ρ, P) around its population value and propagates the perturbation through three nested objects: the feasible set $\Theta_\delta(\rho, P)$ of the outer program, the dual optimizers $\bar{\Xi}_\delta(\theta; \rho, P)$, and the value $\bar{\mathbf{b}}(\delta, \rho, P)$ of the bound. For the directional functional delta method we need each of these objects to behave continuously in (ρ, P) near (ρ_0, P_*) . Part (i) guarantees that no optimizer is lost from the feasible set under small perturbations. Parts (ii)–(iv) show that the optimizer set is nonempty, compact, and locally uniformly bounded, varies upper hemicontinuously in (ρ, P) , and delivers an inner value continuous in θ . Part (v) proves continuity of the bound itself, which is what consistency requires. Part (vi) shows that the near-optimizers of every perturbed program must belong to any prescribed neighborhood of the population argmax, which is what lets the directional-differentiability and bootstrap arguments focus on compact neighborhoods.

Lemma SA-27 (Regularity of the empirical-baseline program). *Let Assumptions SA-4(i)–SA-4(iii) hold. Then:*

- (i) *(Feasibility persistence at the optimizers.) There is $r > 0$ such that $\bar{\Theta}_\delta(\rho_0, P_*) \subseteq \Theta_\delta(\rho, P)$ for every $(\rho, P) \in \mathcal{B}_r(\rho_0, P_*)$, where the ball is taken with respect to $\|\cdot\|_{\mathcal{H}}$. In particular, for every sequence $(\rho_n, P_n) \rightarrow (\rho_0, P_*)$, $\bar{\Theta}_\delta(\rho_0, P_*) \subseteq \Theta_\delta(\rho_n, P_n)$ for all n large.*
- (ii) *(Nonemptiness, compactness, and local boundedness.) For any $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$ there is $\varepsilon > 0$ such that $\bar{\Xi}_\delta(\theta'; \rho', P')$ is nonempty and compact, and their union is bounded, over all (θ', ρ', P') in the ball $\mathcal{B}_\varepsilon(\theta, \rho_0, P_*)$ of points with $\|\theta' - \theta\| + \|\rho' - \rho_0\| + \|P' - P_*\|_{\mathcal{H}} < \varepsilon$.*
- (iii) *(Joint upper hemicontinuity.) For any $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$, the correspondence $(\theta', \rho, P) \mapsto \bar{\Xi}_\delta(\theta'; \rho, P)$, and hence its λ_{12} -projection $\bar{\Lambda}_\delta(\theta'; \rho, P)$, is upper hemicontinuous at (θ, ρ_0, P_*) .*
- (iv) *(Continuity of the inner value.) If, in addition, Assumption SA-4(v) holds, then for each $h \in \mathbb{R}^{\mathbf{d}_\theta}$ and each $\gamma \in UC(\mathcal{H}, e_{P_*})$ the map $\theta \mapsto v(\theta; h, \gamma) := \min_{\xi \in \bar{\Xi}_\delta(\theta; \rho_0, P_*)} \{\lambda_{12}^\top h + \gamma \ell_{\xi, \theta}\}$, well defined for $\theta \in \mathcal{N} \cap \Theta$ with $\mathcal{N}$ as in part (v), where $\bar{\Xi}_\delta(\theta; \rho_0, P_*)$ is nonempty, compact, and contained in $\Xi \subseteq \mathcal{A}$, is continuous at every $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$.*
- (v) *(Continuity of the bound.) The map $(\rho, P) \mapsto \bar{\mathbf{b}}(\delta, \rho, P)$ is continuous at (ρ_0, P_*) along every sequence $(\rho_n, P_n) \rightarrow (\rho_0, P_*)$ such that $\|P_n - P_*\|_{\mathcal{H}_K} \rightarrow 0$ for every compact $K \subseteq \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{\mathbf{d}_m}$. Moreover, there are a compact neighborhood $\mathcal{N}$ of $\bar{\Theta}_\delta(\rho_0, P_*)$, $r > 0$, and $M < \infty$ such that, for all $\theta \in \mathcal{N} \cap \Theta$ and all (ρ, P) and (ρ', P') in the $\|\cdot\|_{\mathcal{H}}$ -ball $\mathcal{B}_r(\rho_0, P_*)$, the set $\bar{\Xi}_\delta(\theta; \rho, P)$ is nonempty, compact, and contained in $\Xi \subseteq \mathcal{A}$, every $\lambda_{12} \in \bar{\Lambda}_\delta(\theta; \rho, P)$ satisfies $\|\lambda_{12}\| \leq M$, and*

$$|\bar{K}_\delta(\theta; \rho, P) - \bar{K}_\delta(\theta; \rho', P')| \leq M \|\rho, P - \rho', P'\|_{\mathcal{H}}.$$

(vi) (*Localization of near-optimizers.*) Let $(\rho_n, P_n) \rightarrow (\rho_0, P_*)$ be such that $\|P_n - P_*\|_{\mathcal{H}_K^1} \rightarrow 0$ for every compact $K \subseteq \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}$, and let $\nu_n \geq 0$ with $\nu_n \rightarrow 0$. Then, for every open set $\mathcal{U} \supseteq \bar{\Theta}_\delta(\rho_0, P_*)$, there is n_0 such that, for all $n \geq n_0$, every $\theta \in \Theta_\delta(\rho_n, P_n)$ with $\bar{K}_\delta(\theta; \rho_n, P_n) \geq \bar{\mathbf{b}}(\delta, \rho_n, P_n) - \nu_n$ belongs to $\mathcal{U}$.

The same conclusions hold for the lower bound, with $\underline{K}_\delta, \underline{\mathbf{b}}, \underline{\Theta}_\delta, \underline{\Xi}_\delta, \underline{\Lambda}_\delta$ in place of $\bar{K}_\delta, \bar{\mathbf{b}}, \bar{\Theta}_\delta, \bar{\Xi}_\delta, \bar{\Lambda}_\delta$.

[Proof]

SA4.4 Consistency, Directional Differentiability, and Distributional Approximation

We now establish the consistency of $\hat{\mathbf{b}}_n(\delta)$ and $\hat{\bar{\mathbf{b}}}_n(\delta)$ via the joint continuity of $(\rho, P) \mapsto \bar{\mathbf{b}}(\delta, \rho, P)$ at (ρ_0, P_*) (Lemma SA-27(v)), together with the convergence of the empirical baseline uniformly over the integrand classes that the bound sees, which is what the Glivenko–Cantelli half of Lemma SA-26 provides.

Theorem SA-28 (Consistency). *Under Assumptions SA-4(i)–SA-4(iv),*

$$\hat{\mathbf{b}}_n(\delta) = \underline{\mathbf{b}}(\delta, \hat{\rho}_n, \hat{P}_n) \xrightarrow{\mathbb{P}} \underline{\mathbf{b}}(\delta, \rho_0, P_*) \quad \text{and} \quad \hat{\bar{\mathbf{b}}}_n(\delta) = \bar{\mathbf{b}}(\delta, \hat{\rho}_n, \hat{P}_n) \xrightarrow{\mathbb{P}} \bar{\mathbf{b}}(\delta, \rho_0, P_*).$$

[Proof]

We next provide an asymptotically valid distributional approximation for the plug-in estimators, which rests on the Hadamard directional differentiability of the bounds at the population value. Recall that a map $f : \mathbb{D} \rightarrow \mathbb{R}$ on a normed space $\mathbb{D}$ is Hadamard directionally differentiable at x_0 tangentially to a set $\mathbb{D}_0 \subseteq \mathbb{D}$ if there is a continuous map $df_{x_0} : \mathbb{D}_0 \rightarrow \mathbb{R}$ with $\{f(x_0 + t_n h_n) - f(x_0)\}/t_n \rightarrow df_{x_0}[h]$ for every $h \in \mathbb{D}_0$ and all sequences $t_n \downarrow 0$ and $h_n \rightarrow h$ in $\mathbb{D}$. For f to be Hadamard differentiable, df_{x_0} would in addition have to be linear. See also Shapiro (1990). The reason why the bounds are in general directionally, rather than fully, differentiable lies in their nested variational form. For fixed θ and ξ , the dual objective $\zeta + \eta\delta + \lambda_{12}^\top \rho + P\ell_{\xi, \theta}$ is affine in (ρ, P) , so that $\bar{K}_\delta(\theta; \rho, P)$ is a pointwise infimum of affine maps of (ρ, P) , and $\bar{\mathbf{b}}(\delta, \rho, P)$ is a supremum over θ of such infima. Under the regularity of Lemma SA-27, value functions of this kind are directionally differentiable, with derivative obtained by differentiating the criterion at the optimizers only, in the spirit of Danskin’s theorem. At a direction (h, γ) , the derivative of the inner infimum is the minimum over $\xi^* \in \bar{\Xi}_\delta(\theta; \rho_0, P_*)$ of the linear map $(h, \gamma) \mapsto (\lambda_{12}^*)^\top h + \gamma \ell_{\xi^*, \theta}$, and the derivative of the outer supremum is the maximum of these inner derivatives over $\theta^* \in \bar{\Theta}_\delta(\rho_0, P_*)$. A minimum (maximum) of linear maps over a set is positively homogeneous and continuous, and it is linear whenever the set is a singleton. It fails in general to be additive otherwise, because the optimizer selected by the extremum may vary with the direction. Hence, when the outer or the dual optimizer set at the population value contains more than one element, the derivative is in general nonlinear. Neither form of nonuniqueness is exceptional in our setting. The outer program is a nonconvex program in θ whose supremum may be attained at several points, and the dual program, although convex in ξ , need not be strictly convex, so that its minimizers may form a nontrivial compact set (Lemma SA-27(ii)).

To derive the asymptotic distribution of $\hat{\bar{\mathbf{b}}}_n(\delta)$, it is useful to consider the bound as a functional on the normed space $\mathbb{D} := \mathbb{R}^{d_e} \times \ell^\infty(\mathcal{H})$. Two localizations make this extension well behaved. First, because an element $\mathbf{b}$ evaluates only integrands with index in $\mathcal{A}$, the extended criterion restricts the dual program to $\mathcal{F}(\theta) \cap \mathcal{A}$, and the extended minimum-divergence function is defined through its dual form (32) restricted to the indices of $\mathcal{H}_\mathcal{A}^\mathbb{D}$. Second, the outer supremum is restricted to a fixed compact neighborhood of the population argmax within Θ . The union of the finitely many θ -balls extracted in the construction of Ξ in Section SA4.3 is an open set that contains the compact set $\bar{\Theta}_\delta(\rho_0, P_*)$, so we can fix $\varepsilon_\Theta > 0$ such that the compact set

$$\mathcal{N}_\Theta := \{\theta \in \Theta : d(\theta, \bar{\Theta}_\delta(\rho_0, P_*)) \leq \varepsilon_\Theta\}$$

is contained in that union. We take the neighborhood relative to Θ , rather than requiring $\bar{\Theta}_\delta(\rho_0, P_*)$ to lie in the interior of $\mathcal{N}_\Theta$, so as to allow population optimizers on the boundary of Θ . By compactness of $\mathcal{N}_\Theta$, the continuous map $\theta \mapsto \max_k \{\varepsilon_k - \|\theta - \theta_k\|\}$, with θ_k and ε_k the centers and radii of that finite subcover, is bounded away from zero on $\mathcal{N}_\Theta$, so there is $\varepsilon_* > 0$, which we shrink further so that the persistence neighborhoods of Lemma SA-25(iii) at the centers $\theta_1, \dots, \theta_J$ also apply. Therefore, whenever $\theta \in \mathcal{N}_\Theta$ and $\|(\rho, P) - (\rho_0, P_*)\|_{\mathcal{H}_0} < \varepsilon_*$, the optimizer set $\bar{\Xi}_\delta(\theta; \rho, P)$ is nonempty and contained in $\Xi \subseteq \mathcal{A}$, and every maximizer of the dual (32) at (θ, ρ, P) lies in $\Pi_{\zeta\lambda}(\mathcal{A}_0)$, by the bounds of Lemma SA-25(i) at the persisting radius $r_{\theta_j}/3$ and the construction of $\mathcal{A}_0$. Concretely, we set

$$\begin{aligned} \bar{K}_\delta(\theta; \rho, \mathbf{b}) &:= \inf_{\xi \in \mathcal{F}(\theta) \cap \mathcal{A}} \{\zeta + \eta\delta + \lambda_{12}^\top \rho + \mathbf{b}(\ell_{\xi, \theta})\}, \\ \mathbf{D}(\theta; \rho, \mathbf{b}) &:= \sup \{-\mathbf{b}(\mathbf{r}(-\zeta - \lambda^\top m(\cdot; \theta), 1)) - \zeta - \lambda_{12}^\top \rho : (\zeta, \lambda) \in \Pi_{\zeta\lambda}(\mathcal{A}) \cap \mathcal{F}_\mathbb{D}(\theta)\}, \\ \bar{\mathbf{b}}(\delta, \rho, \mathbf{b}) &:= \sup \{\bar{K}_\delta(\theta; \rho, \mathbf{b}) : \theta \in \mathcal{N}_\Theta, \mathbf{D}(\theta; \rho, \mathbf{b}) < \delta\}, \end{aligned}$$

with the analogous localization around $\underline{\Theta}_\delta(\rho_0, P_*)$ for the lower bound. Note that for $\theta \in \mathcal{N}_\Theta$ and (ρ, P) as above, restricting the infimum to $\mathcal{F}(\theta) \cap \mathcal{A}$ does not change its value and, similarly, restricting the dual of $\mathbf{D}$ to $\Pi_{\zeta\lambda}(\mathcal{A})$ does not change its value either, so, at $\mathbf{b} = (P\ell)_{\ell \in \mathcal{H}}$, $\bar{K}_\delta(\theta; \rho, \mathbf{b}) = \bar{K}_\delta(\theta; \rho, P)$ and $\mathbf{D}(\theta; \rho, \mathbf{b}) = \mathbf{D}(\theta; \rho, P)$ for every

Algorithm SA-1 Confidence sets for the sensitivity bounds

- 1: Compute $\hat{\rho}_n$, $\hat{P}_n$, the bounds $\hat{\underline{b}}_n(\delta)$, $\hat{\bar{b}}_n(\delta)$, the near-argmax sets $\hat{\Theta}_{\delta,n}$, $\hat{\bar{\Theta}}_{\delta,n}$, and the dual minimizer sets $\bar{\Xi}_\delta(\theta; \hat{\rho}_n, \hat{P}_n)$ and $\Xi_\delta(\theta; \hat{\rho}_n, \hat{P}_n)$.
- 2: **for** $b = 1, \dots, B$ **do**
- 3: Draw an i.i.d. sample $\{Z_i^{*,b}\}_{i=1}^n$ from $\hat{P}_n$ and form $\hat{P}_n^{*,b}$ and $\hat{\rho}_n^{*,b}$.
- 4: Set $\hat{T}_n^{*,b,\text{up}} := \hat{\bar{b}}_n[\sqrt{n}(\hat{\rho}_n^{*,b} - \hat{\rho}_n), \sqrt{n}(\hat{P}_n^{*,b} - \hat{P}_n)]$ and the analog $\hat{T}_n^{*,b,\text{low}}$.
- 5: **end for**
- 6: For $u \in (0, 1)$, let $\hat{\mathfrak{c}}_u^{\text{low}}$ and $\hat{\mathfrak{c}}_u^{\text{up}}$ be the empirical u -quantiles of $\{\hat{T}_n^{*,b,\text{low}}\}_{b=1}^B$ and $\{\hat{T}_n^{*,b,\text{up}}\}_{b=1}^B$, respectively. Valid lower and upper $100(1 - \alpha)\%$ confidence sets are

$$[\hat{\underline{b}}_n(\delta) - \hat{\mathfrak{c}}_{1-\alpha}^{\text{low}}/\sqrt{n}, +\infty) \quad \text{and} \quad (-\infty, \hat{\bar{b}}_n(\delta) - \hat{\mathfrak{c}}_\alpha^{\text{up}}/\sqrt{n}].$$

$\theta \in \mathcal{N}_\Theta$. In particular, writing $\mathbf{b}_0 := (P_*\ell)_{\ell \in \mathcal{H}}$ for the element of $\ell^\infty(\mathcal{H})$ induced by the population baseline P_* , we obtain these two identities at $(\rho_0, \mathbf{b}_0)$ for every $\theta \in \mathcal{N}_\Theta$, and hence, since $\bar{\Theta}_\delta(\rho_0, P_*) \subseteq \mathcal{N}_\Theta$, $\bar{\mathbf{b}}(\delta, \rho_0, \mathbf{b}_0) = \bar{\mathbf{b}}(\delta, \rho_0, P_*)$ and $\bar{\Xi}_\delta(\theta; \rho_0, \mathbf{b}_0) = \bar{\Xi}_\delta(\theta; \rho_0, P_*)$ for every $\theta \in \mathcal{N}_\Theta$. The main result of this subsection is the following.

Theorem SA-29 (Distributional approximation). *Under Assumption SA-4, the map $(\rho, P) \mapsto (\underline{\mathbf{b}}(\delta, \rho, P), \bar{\mathbf{b}}(\delta, \rho, P))$ is Hadamard directionally differentiable at (ρ_0, P_*) tangentially to $\mathbb{R}^{\mathbf{d}_\Theta} \times UC(\mathcal{H}, e_{P_*})$. At a direction $(h, \gamma) \in \mathbb{R}^{\mathbf{d}_\Theta} \times UC(\mathcal{H}, e_{P_*})$, the upper and lower bound derivatives are*

$$\begin{aligned} d\bar{\mathbf{b}}_{(\rho_0, P_*)}[h, \gamma] &= \max_{\theta^* \in \bar{\Theta}_\delta(\rho_0, P_*)} \min_{\xi^* \in \bar{\Xi}_\delta(\theta^*; \rho_0, P_*)} \{(\lambda_{12}^*)^\top h + \gamma \ell_{\xi^*, \theta^*}\}, \\ d\underline{\mathbf{b}}_{(\rho_0, P_*)}[h, \gamma] &= \min_{\theta^* \in \underline{\Theta}_\delta(\rho_0, P_*)} \max_{\xi^* \in \underline{\Xi}_\delta(\theta^*; \rho_0, P_*)} \{-(\lambda_{12}^*)^\top h - \gamma \ell_{\xi^*, \theta^*}^-\}, \end{aligned} \tag{33}$$

where $\ell_{\xi, \theta}^-(z) := \mathbf{r}(-g(z; \theta) - \lambda^\top m(z; \theta) - \zeta, \eta)$ is the dual integrand (24) of the program obtained by the substitution $g \mapsto -g$. With $(\mathbb{Z}, \mathbb{G})$ the joint Gaussian limit of $\sqrt{n}((\hat{\rho}_n - \rho_0), (\hat{P}_n - P_*))$ in $\mathbb{R}^{\mathbf{d}_\Theta} \times \ell^\infty(\mathcal{H})$, so that $\mathbb{G}$ is the P_* -Brownian bridge indexed by $\mathcal{H}$ and $\mathbb{Z} = \mathbb{G}\psi$, we have

$$\sqrt{n} \begin{pmatrix} \hat{\underline{b}}_n(\delta) - \underline{\mathbf{b}}(\delta, \rho_0, P_*) \\ \hat{\bar{b}}_n(\delta) - \bar{\mathbf{b}}(\delta, \rho_0, P_*) \end{pmatrix} \rightsquigarrow \begin{pmatrix} d\underline{\mathbf{b}}_{(\rho_0, P_*)}[\mathbb{Z}, \mathbb{G}] \\ d\bar{\mathbf{b}}_{(\rho_0, P_*)}[\mathbb{Z}, \mathbb{G}] \end{pmatrix}.$$

[Proof]

The limiting law carries the bridge term $\mathbb{G}\ell_{\xi^*, \theta^*}$ alongside the finite-dimensional term $(\lambda_{12}^*)^\top \mathbb{Z}$, and the two are correlated, as both are functionals of the same sample. When the sets of outer maximizers and dual minimizers are singletons the derivative is linear and the limit is Gaussian. We return to this case in Remark SA-9.

SA4.5 Bootstrap

Because the derivative (33) is generally nonlinear in (h, γ) , the naive bootstrap is not consistent (see Theorem 3.1 in Fang and Santos, 2019), and we thus specialize the general resampling procedure of Fang and Santos (2019) to our setting. In particular, we draw $Z_1^*, \dots, Z_n^*$ i.i.d. from $\hat{P}_n$, yielding $\hat{P}_n^* = n^{-1} \sum_i \delta_{Z_i^*}^{\text{Dirac}}$, and recompute the estimator on the resample, which gives $\hat{\rho}_n^*$. In the leading case in which $\hat{\rho}_n$ is the sample mean of a function of the data, the influence function ψ of Assumption SA-4(iv) is that function minus its mean ρ_0 under P_* , so that $\hat{\rho}_n = \hat{P}_n(\rho_0 + \psi)$ and $\hat{\rho}_n^* = \hat{P}_n^*(\rho_0 + \psi)$. Hence, $\sqrt{n}(\hat{\rho}_n^* - \hat{\rho}_n) = \mathbb{G}_n^* \psi$, where $\mathbb{G}_n^* := \sqrt{n}(\hat{P}_n^* - \hat{P}_n)$ is the bootstrapped empirical process. For a general asymptotically linear $\hat{\rho}_n$, we instead require the bootstrap linearization $\sqrt{n}(\hat{\rho}_n^* - \hat{\rho}_n) = \mathbb{G}_n^* \psi + o_{\mathbb{P}}(1)$, which is the hypothesis imposed by Theorem SA-30 below. The estimated derivative replaces $\bar{\Theta}_\delta(\rho_0, P_*)$ and $\bar{\Xi}_\delta(\theta; \rho_0, P_*)$ by their sample analogs

$$\hat{\bar{\mathbf{b}}}_n[h, \gamma] := \sup_{\theta \in \hat{\bar{\Theta}}_{\delta,n}} \inf_{\xi \in \bar{\Xi}_\delta(\theta; \hat{\rho}_n, \hat{P}_n)} \{\lambda_{12}^\top h + \gamma \ell_{\xi, \theta}\},$$

with the estimated argmax set

$$\hat{\bar{\Theta}}_{\delta,n} := \{\theta \in \Theta_\delta(\hat{\rho}_n, \hat{P}_n) : \bar{K}_\delta(\theta; \hat{\rho}_n, \hat{P}_n) \geq \hat{\bar{b}}_n(\delta) - \hat{\nu}_n \sqrt{\frac{\log n}{n}}\}, \quad \hat{\nu}_n \geq 0, \quad \hat{\nu}_n = O_{\mathbb{P}}(1), \quad \hat{\nu}_n^{-1} = O_{\mathbb{P}}(1),$$

where $\hat{\nu}_n$ is a tuning sequence whose choice is discussed in Remark SA-8 below, and $\hat{\bar{\Lambda}}_{\delta,n}(\theta) := \Pi_{12}(\bar{\Xi}_\delta(\theta; \hat{\rho}_n, \hat{P}_n))$. The resulting resampling procedure and the confidence sets it delivers are summarized in Algorithm SA-1, which reproduces Algorithm 1 of the main text in the notation of this section, and whose validity we establish in Theorem SA-30 below.

Remark SA-8 (Choice of the tuning sequence $\hat{\nu}_n$). The sequence $\hat{\nu}_n$ governs the slack used to form the estimated argmax set $\hat{\bar{\Theta}}_{\delta,n}$, and such slack, namely $\hat{\nu}_n \sqrt{\log n/n}$, must vanish slowly enough that $\hat{\bar{\Theta}}_{\delta,n}$ contains the population

set $\bar{\Theta}_\delta(\rho_0, P_*)$ with probability approaching one, yet fast enough that $\hat{\bar{\Theta}}_{\delta,n}$ excludes strictly suboptimal θ in the limit. Both requirements are made precise in the proof of Theorem SA-30. For concrete guidance on this choice we refer the reader to Christensen and Connault (2023, Section 6.2), who discuss the selection of the analogous tuning sequence in their setting. Following their implementation, we set $\hat{\nu}_n = \min_{1 \leq j \leq d_m} (\hat{\Sigma}_m)_{jj}$, the smallest diagonal entry of the covariance matrix of the moment function under the empirical baseline

$$\hat{\Sigma}_m := \hat{P}_n[(m(\cdot; \hat{\theta}_n) - \hat{P}_n m(\cdot; \hat{\theta}_n))(m(\cdot; \hat{\theta}_n) - \hat{P}_n m(\cdot; \hat{\theta}_n))^\top]$$

evaluated at an approximate maximizer $\hat{\theta}_n \in \Theta_\delta(\hat{\rho}_n, \hat{P}_n)$ of the estimated outer problem, that is, one with $\bar{K}_\delta(\hat{\theta}_n; \hat{\rho}_n, \hat{P}_n) \geq \hat{\mathbf{b}}_n(\delta) - 1/n$. This choice meets the requirement stated with $\hat{\bar{\Theta}}_{\delta,n}$. It is bounded, as m is bounded on the compact set $\mathcal{Z} \times \Theta$ under Assumption SA-4(i). Moreover, Lemma SA-27(vi), applied along almost surely convergent subsequences as in the proof of Theorem SA-30, gives $d(\hat{\theta}_n, \bar{\Theta}_\delta(\rho_0, P_*)) \xrightarrow{\mathbb{P}} 0$. By a uniform law of large numbers over the Lipschitz class $\{m(\cdot; \theta) : \theta \in \Theta\}$ and the compactness of $\bar{\Theta}_\delta(\rho_0, P_*)$, $\hat{\nu}_n$ is then bounded away from zero in probability, provided that no component of $m(Z; \theta)$ is P_* -almost surely constant at any $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$. A constant component, such as a normalization, must therefore be excluded from the minimum. ♣

Bootstrap consistency requires a joint strengthening of the pointwise lower hemicontinuity of Assumption SA-4(v), needed for the convergence of the estimated derivative at every direction of the tangent set. Like that assumption, it is imposed on $\bar{\Xi}_\delta$ rather than on $\bar{\Lambda}_\delta$, for the reason given in Section SA4.2.

Assumption SA-5 (Joint lower hemicontinuity). *The correspondences $(\theta, \rho, P) \mapsto \bar{\Xi}_\delta(\theta; \rho, P)$ and $(\theta, \rho, P) \mapsto \bar{\Xi}_\delta(\theta; \rho, P)$ are lower hemicontinuous jointly at every (θ, ρ_0, P_*) with $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$ (resp. $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$).*

Theorem SA-30 establishes the validity of Algorithm SA-1.

Theorem SA-30 (Bootstrap consistency). *Suppose Assumptions SA-4 and SA-5 hold, and that the bootstrap estimator satisfies the linearization $\sqrt{n}(\hat{\rho}_n^* - \hat{\rho}_n) = \mathbb{G}_n^* \psi + o_{\mathbb{P}}(1)$, with $\mathbb{G}_n^* := \sqrt{n}(\hat{P}_n^* - \hat{P}_n)$. Then, conditionally on the data,*

$$\widehat{d\mathbf{b}}_n[\sqrt{n}(\hat{\rho}_n^* - \hat{\rho}_n), \sqrt{n}(\hat{P}_n^* - \hat{P}_n)] \rightsquigarrow d\bar{\mathbf{b}}_{(\rho_0, P_*)}[\mathbb{Z}, \mathbb{G}].$$

If, in addition, the distribution function of $d\bar{\mathbf{b}}_{(\rho_0, P_*)}[\mathbb{Z}, \mathbb{G}]$ is continuous and strictly increasing at its α - and $(\alpha/2)$ -quantiles and the distribution function of $d\mathbf{b}_{(\rho_0, P_*)}[\mathbb{Z}, \mathbb{G}]$ is continuous and strictly increasing at its $(1 - \alpha)$ - and $(1 - \alpha/2)$ -quantiles, then the one-sided confidence sets of Algorithm SA-1, as well as the two-sided set $[\hat{\mathbf{b}}_n(\delta) - \hat{\mathbf{c}}_{1-\alpha/2}^{\text{low}}/\sqrt{n}, \hat{\mathbf{b}}_n(\delta) - \hat{\mathbf{c}}_{\alpha/2}^{\text{up}}/\sqrt{n}]$ formed from its critical values, have asymptotic coverage at least $1 - \alpha$. In particular

$$\liminf_{n \rightarrow \infty} \mathbb{P}(\bar{\mathbf{b}}(\delta, \rho_0, P_*) \in (-\infty, \hat{\mathbf{b}}_n(\delta) - \hat{\mathbf{c}}_{\alpha}^{\text{up}}/\sqrt{n})] \geq 1 - \alpha,$$

with analogous statements for the lower and two-sided sets.

[Proof]

Remark SA-9 (Singleton dual). Suppose that

$$\bar{\Theta}_\delta(\rho_0, P_*) = \{\theta^*\}, \quad \bar{\Xi}_\delta(\theta^*; \rho_0, P_*) = \{\xi^*\}.$$

Then the derivative $d\bar{\mathbf{b}}_{(\rho_0, P_*)}[h, \gamma] = (\lambda_{12}^*)^\top h + \gamma \ell_{\xi^*, \theta^*}$ is linear, so $\hat{\mathbf{b}}_n(\delta)$ is asymptotically Gaussian,

$$\sqrt{n}(\hat{\mathbf{b}}_n(\delta) - \bar{\mathbf{b}}(\delta, \rho_0, P_*)) \rightsquigarrow \mathbf{N}(0, \sigma_*^2), \quad \sigma_*^2 = \mathbb{V}((\lambda_{12}^*)^\top \mathbb{Z} + \mathbb{G} \ell_{\xi^*, \theta^*}),$$

where, by Assumption SA-4(iv), $\sigma_*^2 = \mathbb{V}_{P_*}((\lambda_{12}^*)^\top \psi + \ell_{\xi^*, \theta^*})$ combines the variance of the estimated moments, the variance of the empirical baseline, and their covariance. The bootstrap of Algorithm SA-1 reduces to the standard nonparametric bootstrap in this case. ♣

Remark SA-10 (Uncertainty in external nuisance parameters). Throughout this section the only objects carrying sampling uncertainty are the estimated moment block $\hat{\rho}_n$ and the empirical baseline $\hat{P}_n$, and the large sample theory of Theorems SA-28–SA-30 is stated for the bound as a functional of the pair (ρ, P) alone. Some applications, however, also feature an auxiliary parameter ς that must itself be estimated and that enters the program through g , m , or the geometry of the neighborhood, generally in a nonlinear fashion. We derive the asymptotic distribution assuming ς is known and suppressing the dependence of all quantities on it, so that the limit law and the bootstrap of Algorithm SA-1 reflect only the sampling uncertainty of $(\hat{\rho}_n, \hat{P}_n)$. We leave to future work accommodating sampling variation in $\hat{\varsigma}$ in a tractable and general manner.

In specific applications, however, it is possible to leverage additional structure to quantify the uncertainty from estimating these parameters. The RCT-with-attrition application of Section SA6.1 is a leading instance. There the complier share π , the share of compliers among the observed treated, is exactly such an auxiliary parameter: it enters the bound nonlinearly, through the cap $1/\pi$ and the effective budget $(1 - \pi)\delta$, as well as through the affine map relating the complier mean to the always observed mean, and it is estimated from the same sample as the

baseline distribution. Section SA6.1.2 takes up exactly this point, exploiting the explicit structure of the bound to account for the sampling uncertainty in $\hat{\pi}$ that the general theory holds fixed. Proposition SA-35 there shows that estimating π adds the term $\varkappa(\delta)\psi_\pi$ to the influence function of the plug-in bound, and that the nonparametric bootstrap accounts for it once $\hat{\pi}$ is recomputed in each resample. ♣

Remark SA-11 (Known baseline distribution). When P_* is known, as with the Gumbel shocks of Section SA6.2, expectations against it carry no sampling uncertainty, as they can be computed exactly or by Monte Carlo integration with an arbitrarily large number of draws, and the data enter only through $\hat{\rho}_n$. Writing $\bar{K}_\delta(\theta; \rho) := \bar{K}_\delta(\theta; \rho, P_*)$, and likewise for $\bar{D}$, $\bar{\Theta}_\delta$, $\bar{\mathbf{b}}$, $\bar{\Theta}_\delta$, $\bar{\Xi}_\delta$, and $\bar{\Lambda}_\delta$, the plug-in estimator is $\hat{\mathbf{b}}_n(\delta) = \bar{\mathbf{b}}(\delta, \hat{\rho}_n, P_*)$, and Theorems SA-28–SA-30 hold with the baseline slot frozen under a weakened version of Assumption SA-4: the Lipschitz requirement in θ of Assumption SA-4(i) is dropped; Assumption SA-4(iv) is replaced by $\sqrt{n}(\hat{\rho}_n - \rho_0) \rightsquigarrow \mathbf{N}(0, \Sigma)$ for some positive semidefinite $\Sigma \in \mathbb{R}^{d_* \times d_*}$, so that $\hat{\rho}_n$ may be computed from data on a random vector other than Z ; and Assumptions SA-4(v) and SA-5 may be imposed on the multiplier projections $\bar{\Lambda}_\delta$ rather than on $\bar{\Xi}_\delta$. The derivative becomes

$$d\bar{\mathbf{b}}_{\rho_0}[h] = \max_{\theta^* \in \bar{\Theta}_\delta(\rho_0)} \min_{\lambda^* \in \bar{\Lambda}_\delta(\theta^*; \rho_0)} (\lambda^*)^\top h,$$

which coincides with that of Christensen and Connault (2023, Theorem 6.2), so that $\sqrt{n}(\hat{\mathbf{b}}_n(\delta) - \bar{\mathbf{b}}(\delta, \rho_0)) \rightsquigarrow d\bar{\mathbf{b}}_{\rho_0}[\mathbb{Z}]$ with $\mathbb{Z} \sim \mathbf{N}(0, \Sigma)$, and Algorithm SA-1 reduces to resampling $\hat{\rho}_n$ alone, by any bootstrap satisfying Assumption 3 of Fang and Santos (2019), such as the Gaussian draw $\hat{\rho}_n^* \sim \mathbf{N}(\hat{\rho}_n, \hat{\Sigma}/n)$ with $\hat{\Sigma} \xrightarrow{\mathbb{P}} \Sigma$, and evaluating $\sup_{\theta \in \bar{\Theta}_{\delta, n}} \inf_{\lambda \in \bar{\Lambda}_\delta(\theta; \hat{\rho}_n)} \lambda^\top h$ at $h = \sqrt{n}(\hat{\rho}_n^* - \hat{\rho}_n)$, which is the procedure of Christensen and Connault (2023, Section 6.2). The tuning sequence of Remark SA-8 is computed from the Monte Carlo draws used to evaluate $\mathbb{E}_{P_*}[\cdot]$. ♣

SA5 Summary Measures for Sensitivity Analysis

Our framework outputs a map $\delta \mapsto [\underline{\tau}(\delta), \bar{\tau}(\delta)]$ that traces out the range of the estimand as the sensitivity parameter varies over $[0, 1]$. Reporting the bounds alone is not always the most effective way of communicating the results, and it may hide information that is relevant for decision making. This section develops the four summary measures of Section 5.2 of the main text, which are computed directly from the bounds. The first is the *breakdown budget*, the largest δ at which a conclusion still holds. The second is the *certified breakdown budget*, the largest δ at which it also holds once we account for sampling uncertainty at level $1 - \alpha$. The third is the *shadow price of robustness*, the marginal widening of either bound with respect to δ , which Proposition SA-31 identifies up to a scalar with the optimal multiplier on the divergence constraint. The fourth is the *certification frontier*, the sample size a study would need to certify robustness at a given budget. Proposition SA-33 computes its semi-elasticity with respect to δ , that is, the proportional increase in sample size needed to certify one additional percentage point of misspecified probability mass. We study the shadow price first, in Section SA5.1, because the asymptotic analysis of the certified breakdown budget in Section SA5.2 rests on it.

Throughout, we work in a simplified version of the framework of Section SA1.2 in which the integrand does not depend on a structural parameter, no moment restriction is imposed, and the robustness set is the total variation robustness set of the main text. Formally, we let $\mathcal{Z} \subseteq \mathbb{R}^{\mathcal{d}_z}$ be closed, let $P_* \in \Delta(\mathcal{Z})$ be the baseline distribution, let $g : \mathcal{Z} \rightarrow \mathbb{R}$ be such that $\mathbb{E}_{P_*}[|g(Z)|] < \infty$, and we let $\mathbf{d}_m = 0$ for simplicity. For a known mixing weight $\kappa \in (0, 1)$ and a budget $\delta \geq 0$, we consider the robustness set

$$\begin{aligned} \mathcal{P}_{\text{TV}}(\delta) &:= \{P \in \mathcal{P}(\mathcal{Z}) : \text{TV}(P \| R) \leq \delta, P_* = \kappa P + (1 - \kappa)R \text{ for some } R \in \mathcal{P}(\mathcal{Z})\} \\ &= \{P \in \mathcal{P}(\mathcal{Z}) : \kappa P \leq P_*, \text{TV}(P \| P_*) \leq (1 - \kappa)\delta\}, \end{aligned} \quad (34)$$

where the second equality follows by Lemma SA-17(i) for $\delta \leq 1$. For $\delta > 1$, both sets equal $\{P \in \mathcal{P}(\mathcal{Z}) : \kappa P \leq P_*\}$, as $\text{TV}(P \| R) \leq 1$ for all P and R , and as $\kappa P \leq P_*$ implies $\text{TV}(P \| P_*) \leq 1 - \kappa$ by the same lemma. The set thus does not change for $\delta \geq 1$, and we allow $\delta > 1$ only to let the one-sided derivatives at $\delta = 1$ be well defined. The above set coincides with the robustness set $\mathcal{P}_{\text{mix}}^{\leq((1 - \kappa)\delta)}$ of (7) with entropy ϕ_{TV} , which Proposition SA-13(i) identifies with the neighborhood generated by the upper-truncated entropy $\bar{\phi}_{\text{TV}}$ of Section SA1.3 at budget $(1 - \kappa)\delta$. The sensitivity bounds are thus

$$\bar{\tau}(\delta) := \sup_{P \in \mathcal{P}_{\text{TV}}(\delta)} \mathbb{E}_P[g(Z)] \quad \text{and} \quad \underline{\tau}(\delta) := \inf_{P \in \mathcal{P}_{\text{TV}}(\delta)} \mathbb{E}_P[g(Z)]. \quad (35)$$

This is the scenario of the RCT with attrition application of Section SA6.1. Absent a structural parameter, the dual problem is only two-dimensional and uniqueness of the optimal budget multiplier can be verified directly. A structural parameter raises difficulties of a different nature, which we leave to future work. For example, the bounds become upper envelopes over the parameter space of the values of the inner problems, and an envelope of concave functions of the budget need not be concave, so that the shadow price is defined only at each outer maximizer and may jump when the maximizer changes with the budget.

SA5.1 The Shadow Price of Robustness

In this simplified scenario, the dual of the inner problem at budget $\delta > 0$ is, by Theorem SA-14(i) and Lemma SA-23,

$$\bar{\tau}(\delta) = \inf_{(\zeta, \eta) \in \mathbb{R} \times \mathbb{R}_+} \left\{ \zeta + \eta(1 - \kappa)\delta + \mathbb{E}_{P_*}[\mathbf{r}_{\text{TV}}(g(Z) - \zeta, \eta)] \right\}, \quad \mathbf{r}_{\text{TV}}(x, \eta) := \max \left\{ -\frac{\eta}{2}, x, \frac{x}{\kappa} - \frac{(1 - \kappa)\eta}{2\kappa} \right\}. \quad (36)$$

The budget enters only through the term $\eta(1 - \kappa)\delta$, so that η prices the total variation constraint as a Lagrange multiplier prices a resource constraint. The next result shows that $\bar{\tau}$ is concave and $\underline{\tau}$ convex in δ , so that their one-sided derivatives exist, and that these derivatives equal, up to the factor $1 - \kappa$, the optimal multipliers, whence the name shadow price.

Proposition SA-31 (The shadow price of the robustness budget). *For $\delta > 0$, let $\bar{\Xi}_\delta$ and $\underline{\Xi}_\delta$ denote the solution sets of the dual problem (36) with budget δ , computed at g and at $-g$, respectively. Let*

$$\begin{aligned} \bar{\eta}_\delta^{\min} &:= \min\{\eta : (\zeta, \eta) \in \bar{\Xi}_\delta\}, & \bar{\eta}_\delta^{\max} &:= \max\{\eta : (\zeta, \eta) \in \bar{\Xi}_\delta\}, \\ \underline{\eta}_\delta^{\min} &:= \min\{\eta : (\zeta, \eta) \in \underline{\Xi}_\delta\}, & \underline{\eta}_\delta^{\max} &:= \max\{\eta : (\zeta, \eta) \in \underline{\Xi}_\delta\} \end{aligned}$$

denote the smallest and the largest optimal budget multiplier of each program, which are well defined because each solution set is nonempty and compact. Then the following hold:

- (i) (Regularity.) *On $[0, \infty)$ the map $\bar{\tau}$ is finite, nondecreasing, concave, and continuous, and the map $\underline{\tau}$ is finite, nonincreasing, convex, and continuous. Both admit finite one-sided derivatives at every $\delta \in (0, \infty)$, both are differentiable at all but at most countably many budgets, and both are constant on $[1, \infty)$.*
- (ii) (Shadow prices.) *For every $\delta \in (0, \infty)$,*

$$\partial_+ \bar{\tau}(\delta) = (1 - \kappa) \bar{\eta}_\delta^{\min}, \quad \partial_- \bar{\tau}(\delta) = (1 - \kappa) \bar{\eta}_\delta^{\max}, \quad -\partial_+ \underline{\tau}(\delta) = (1 - \kappa) \underline{\eta}_\delta^{\min}, \quad -\partial_- \underline{\tau}(\delta) = (1 - \kappa) \underline{\eta}_\delta^{\max}.$$
- (iii) (Declining prices.) *For every $0 < \delta_1 < \delta_2 < \infty$, both $\bar{\eta}_{\delta_1}^{\min} \geq \bar{\eta}_{\delta_2}^{\max}$ and $\underline{\eta}_{\delta_1}^{\min} \geq \underline{\eta}_{\delta_2}^{\max}$.*

(iv) (A single shadow price.) The bound $\bar{\tau}$ is differentiable at $\delta \in (0, \infty)$ if and only if $\bar{\eta}_\delta^{\min} = \bar{\eta}_\delta^{\max}$, and $\underline{\tau}$ is differentiable at δ if and only if $\underline{\eta}_\delta^{\min} = \underline{\eta}_\delta^{\max}$, that is, exactly when the optimal budget multiplier of the corresponding program is unique. At every such budget we write

$$\bar{\eta}(\delta) := \bar{\tau}'(\delta) = (1 - \kappa) \bar{\eta}_\delta^{\min}, \quad \underline{\eta}(\delta) := -\underline{\tau}'(\delta) = (1 - \kappa) \underline{\eta}_\delta^{\min}.$$

Both $\bar{\eta}$ and $\underline{\eta}$ are nonnegative, defined at all but at most countably many $\delta \in (0, \infty)$, and nonincreasing on the set of budgets at which they are defined.

[Proof]

Beyond $\delta = 1$ the bounds are constant, so that $\partial_+ \bar{\tau}(1) = 0$ and the smallest optimal multiplier at $\delta = 1$ vanishes, whereas the largest need not.

SA5.2 Breakdown and Certified Breakdown

Let $\tau_\star \in \mathbb{R}$ be a reference value for the estimand, typically $\tau_\star = 0$ when the sign of a treatment effect is of interest. The bounds in (35) are finite and continuous on $[0, 1]$ by Proposition SA-31(i), and the breakdown budget for $\tau_\star$ is

$$\delta_b \equiv \delta_b(\tau_\star) := \inf\{\delta \in [0, 1] : \underline{\tau}(\delta) \leq \tau_\star \leq \bar{\tau}(\delta)\}, \quad (37)$$

with the convention that the infimum over an empty set equals 1. Since the bounds are monotone in δ , at most one of the two inequalities fails at $\delta = 0$, and the same one fails for every budget below the breakdown. If $\underline{\tau}(0) > \tau_\star$, the case on which we focus, (37) reduces to $\inf\{\delta \in [0, 1] : \underline{\tau}(\delta) \leq \tau_\star\}$. The case $\bar{\tau}(0) < \tau_\star$ is symmetric, and if $\tau_\star \in [\underline{\tau}(0), \bar{\tau}(0)]$, then $\delta_b = 0$.

The breakdown budget ignores sampling uncertainty. For $\alpha \in (0, 1)$ and a sample size n held fixed, the population certified breakdown budget is

$$\delta_{cb}(\alpha; n) := \inf\left\{\delta \in [0, 1] : \underline{\tau}(\delta) - \frac{\mathfrak{c}_{1-\alpha/2} \sigma(\delta)}{\sqrt{n}} \leq \tau_\star\right\}, \quad (38)$$

where the infimum over an empty set equals 1 as in (37), $\sigma(\delta)$ is the standard deviation at δ of the Gaussian limit of $\sqrt{n}(\hat{\tau}_n - \underline{\tau})$, with $\hat{\tau}_n$ the estimated lower bound of (39), and $\mathfrak{c}_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -th quantile of the standard normal distribution. It is the smallest budget at which a study of size n would fail to certify the conclusion, were the bounds and their standard deviations known. When $\underline{\tau}(0) > \tau_\star$, $\underline{\tau}$ is nonincreasing and continuous with $\delta_b \in (0, 1)$, and σ is continuous and strictly positive on $[0, 1]$, as Assumption SA-6 below imposes, $\delta_{cb}(\alpha; n) \leq \delta_b$ for every n , because $\underline{\tau}(\delta_b) = \tau_\star$ places δ_b in the set in (38), the map $n \mapsto \delta_{cb}(\alpha; n)$ is nondecreasing, and $\delta_{cb}(\alpha; n) \rightarrow \delta_b$ as $n \rightarrow \infty$. The sample analogue replaces the population lower bound and its standard deviation by estimators,

$$\hat{\delta}_{cb}(\alpha) := \inf\left\{\delta \in [0, 1] : \hat{\tau}_n(\delta) - \frac{\mathfrak{c}_{1-\alpha/2} \hat{\sigma}_n(\delta)}{\sqrt{n}} \leq \tau_\star\right\}, \quad (39)$$

where $\hat{\tau}_n(\delta)$ and $\hat{\sigma}_n(\delta)$ estimate $\underline{\tau}(\delta)$ and $\sigma(\delta)$, for instance through the bootstrap of Section SA4 or the influence-function estimator of Proposition SA-35. It is the infimum of the budgets at which $\tau_\star$ is not rejected at level α . In the case on which we focus, the plug-in estimator of the breakdown budget is

$$\hat{\delta}_b := \inf\{\delta \in [0, 1] : \hat{\tau}_n(\delta) \leq \tau_\star\}.$$

Remark SA-12 (Coverage of the identified set, not of a point effect). In the literature, there are at least two approaches to constructing confidence intervals for a partially identified parameter. The first, and the one used throughout this work, covers the entire identified interval by placing a two-sided critical value at each estimated endpoint. The second, following Imbens and Manski (2004), covers a single point of that interval, such as the ATE τ_0 in the attrition example of Section SA6.1, through a critical value that lies between the one- and two-sided normal quantiles and is therefore narrower at the same nominal level. Lee (2009) favors the latter because his object of interest is the treatment effect, a point inside the identified set, and not the region of all rationalizable effects. Here the object of interest is precisely that region: the sensitivity analysis reports the identified set $[\underline{\tau}(\delta), \bar{\tau}(\delta)]$ as a function of the budget δ , and it is the bounds delimiting this set whose sampling uncertainty we quantify. ♣

Assumption SA-6 (Regularity at the breakdown). For the reference value $\tau_\star \in \mathbb{R}$ of (37) the following hold:

- (i) (Population bound.) $\underline{\tau} : [0, 1] \rightarrow \mathbb{R}$ is nonincreasing and continuous, $\underline{\tau}(0) > \tau_\star$, $\delta_b \in (0, 1)$, and $\underline{\tau}$ is differentiable at δ_b with $\underline{\eta}(\delta_b) = -\underline{\tau}'(\delta_b) \in (0, \infty)$.
- (ii) (Estimator.) $\sqrt{n}(\hat{\tau}_n - \underline{\tau}) \rightsquigarrow \mathbb{G}$ in $\ell^\infty([0, 1])$, where $\mathbb{G}$ is a tight zero-mean Gaussian process with continuous sample paths on $[0, 1]$ whose standard deviation function $\sigma(\delta) := \{\mathbb{E}[\mathbb{G}(\delta)^2]\}^{1/2}$ is continuous and strictly positive. Moreover, $\delta \mapsto \hat{\tau}_n(\delta)$ has continuous sample paths.
- (iii) (Standard errors.) $\sup_{\delta \in [0, 1]} |\hat{\sigma}_n(\delta) - \sigma(\delta)| = o_{\mathbb{P}}(1)$, and $\delta \mapsto \hat{\sigma}_n(\delta)$ has continuous sample paths.

In this scenario, the monotonicity and continuity of $\underline{\tau}$ and its differentiability off a countable set are automatic by Proposition SA-31(i), so that Assumption SA-6(i) only requires an interior breakdown budget at which $\underline{\tau}$ is

differentiable, that is, by Proposition SA-31(iv), at which the optimal multiplier is unique, and a strictly positive shadow price there, which rules out a bound path that is flat at the crossing. Assumption SA-6(ii) imposes a Gaussian limit uniformly over $[0, 1]$, which we discuss in the next remark.

Remark SA-13 (Uniqueness and the certified breakdown budget). The certified breakdown budgets (38) and (39) build the lower confidence bound on $\underline{\tau}(\delta)$ from a normal critical value, which presumes that $\sqrt{n}(\widehat{\tau}_n(\delta) - \underline{\tau}(\delta))$ has a Gaussian limit. As there is no outer maximization in this scenario, Remark SA-9 delivers such a limit, with $(1 - \alpha/2)$ -quantile $\mathbf{c}_{1-\alpha/2} \sigma(\delta)$, whenever the dual solution set Ξ_δ is a singleton. Non-uniqueness of the budget multiplier is one way in which this property can fail, and by Proposition SA-31(iv) it occurs exactly at the budgets at which $\underline{\tau}$ has a kink. The RCT application of Section SA6.1 is not affected. There, the baseline has a density that is positive at the trimming quantiles of the lower bound, under the version of Assumption SA-7 for the lower bound given in Section SA6.1.2, so that the optimizer set is a singleton and the plug-in bound is asymptotically normal (Proposition SA-35). When the dual solution is not unique, the limit is in general not Gaussian, and $\mathbf{c}_{1-\alpha/2} \sigma(\delta)$ should be replaced by the $(1 - \alpha/2)$ -quantile of the limit, estimated by the Fang-Santos bootstrap as in Remark SA-14. ♣

Remark SA-14 (Bootstrap critical values). The certified breakdown budget can also be built from the quantiles delivered by Algorithm SA-1, without estimating $\sigma(\delta)$. Letting $\widehat{\mathbf{c}}_{1-\alpha/2}^{\text{low}}(\delta)$ denote the critical value of Algorithm SA-1 for the lower bound at budget δ , one replaces $\mathbf{c}_{1-\alpha/2} \widehat{\sigma}_n(\delta)$ in (39) by $\widehat{\mathbf{c}}_{1-\alpha/2}^{\text{low}}(\delta)$, that is,

$$\widehat{\delta}_{\text{cb}}(\alpha) := \inf \left\{ \delta \in [0, 1] : \widehat{\tau}_n(\delta) - \frac{\widehat{\mathbf{c}}_{1-\alpha/2}^{\text{low}}(\delta)}{\sqrt{n}} \leq \tau_\star \right\}.$$

At each fixed budget this critical value is consistent even when the dual solution is not unique (Theorem SA-30). When the dual solution is unique, it estimates the same quantity as $\mathbf{c}_{1-\alpha/2} \widehat{\sigma}_n(\delta)$, and so does the distance between $\widehat{\tau}_n(\delta)$ and the lower endpoint of a percentile bootstrap band, the $\alpha/2$ -quantile of the bootstrap distribution of $\widehat{\tau}_n(\delta)$. The certified breakdown budgets reported in Section SA7 are computed from such a percentile band. ♣

We define the robustness standard error as

$$\varsigma_b := \frac{\sigma(\delta_b)}{\underline{\eta}(\delta_b)}, \quad (40)$$

which converts the standard error of the lower bound at the breakdown from estimand units into budget units through the shadow price. The next result shows that the certified breakdown budget is consistent for δ_b , that it has a Gaussian limit whose location and scale are governed by ς_b , and that $[\widehat{\delta}_{\text{cb}}(\alpha), 1]$ is an asymptotically valid lower confidence interval for δ_b with coverage $1 - \alpha/2$.

Proposition SA-32 (The certified breakdown budget). *Let Assumption SA-6 hold and $\alpha \in (0, 1)$. Then, as $n \rightarrow \infty$:*

- (i) $\widehat{\delta}_{\text{cb}}(\alpha) \xrightarrow{\mathbb{P}} \delta_b$;
- (ii) $\sqrt{n} (\widehat{\delta}_{\text{cb}}(\alpha) - \delta_b) \rightsquigarrow \mathbf{N}(-\mathbf{c}_{1-\alpha/2} \varsigma_b, \varsigma_b^2)$;
- (iii) $\lim_{n \rightarrow \infty} \mathbb{P}(\widehat{\delta}_{\text{cb}}(\alpha) \leq \delta_b) = 1 - \alpha/2$.

[Proof]

By Proposition SA-32(ii), $\delta_b - \widehat{\delta}_{\text{cb}}(\alpha) = \varsigma_b (\mathbf{c}_{1-\alpha/2} + \mathbb{Z}_n)/\sqrt{n} + o_{\mathbb{P}}(n^{-1/2})$ with $\mathbb{Z}_n \rightsquigarrow \mathbf{N}(0, 1)$, so that the first-order certification gap $\mathbf{c}_{1-\alpha/2} \varsigma_b/\sqrt{n}$ increases with the confidence level and with the robustness standard error and decreases with the sample size. Proposition SA-32(iii) says that the certified budget is a one-sided $1 - \alpha/2$ confidence bound on the population breakdown.

SA5.3 The Certification Frontier

For $\delta \in [0, \delta_b)$, the certification frontier

$$n^\star(\delta; \alpha) := \frac{\mathbf{c}_{1-\alpha/2}^2 \sigma^2(\delta)}{(\underline{\tau}(\delta) - \tau_\star)^2} \quad (41)$$

is the sample size at which the lower confidence bound in (38) equals the reference value at budget δ . We treat it as a real valued quantity. The smallest certifying integer sample size is $\lfloor n^\star(\delta; \alpha) \rfloor + 1$. The next result computes how the frontier grows with the robustness target, that is, the marginal rate of substitution between observations and misspecified probability mass along the frontier.

Proposition SA-33 (Semi-elasticity of the certification frontier). *Let Assumption SA-6 hold, and suppose in addition that $\underline{\tau}$ is convex on $[0, 1]$ and that σ is continuously differentiable on $[0, 1]$. Fix $\alpha \in (0, 1)$. Then, at every*

$\delta \in (0, \delta_{\mathbf{b}})$ at which $\underline{\tau}$ is differentiable, $n^*(\cdot; \alpha)$ is differentiable with

$$\frac{d \log n^*(\delta; \alpha)}{d\delta} = 2 \left[\frac{\underline{\eta}(\delta)}{\underline{\tau}(\delta) - \tau_{\star}} + \frac{\sigma'(\delta)}{\sigma(\delta)} \right], \quad (42)$$

and as $\delta \uparrow \delta_{\mathbf{b}}$

$$\frac{d \log n^*(\delta; \alpha)}{d\delta} = \frac{2}{\delta_{\mathbf{b}} - \delta} \{1 + o(1)\}.$$

[Proof]

The convexity of $\underline{\tau}$ is automatic in this scenario by Proposition SA-31(i). Formula (42) decomposes the rate into the ratio of the shadow price to the slack $\underline{\tau}(\delta) - \tau_{\star}$, which dominates near the breakdown, and the precision correction σ'/σ , which stays bounded while the slack vanishes. Near the breakdown, one additional percentage point of certified robustness thus costs approximately $2/(\delta_{\mathbf{b}} - \delta)$ percent additional sample, so that closing the certification gap from four to three percentage points requires multiplying the sample size by approximately 1.78.

SA6 Applications of the Framework

In this section we show how the framework covers several settings that are common in applied economics. In Section 6 of the main text we develop three applications in detail: randomized experiments with attrition (Section SA6.1), counterfactuals in structural models (Section SA6.2), and formula and shift-share instruments (Section SA6.3). Here we provide further details on these applications together with formal results. Section SA6.4 then collects a series of further applications in a deliberately succinct format. A closing subsection shows that the sensitivity bounds at the endpoint $\delta = 1$ coincide with the nonparametric sharp bounds.

SA6.1 RCTs with Attrition

This section maintains the setup of Example 2.1 of the main text, which we briefly recall, and expands on it in three directions. First, we show that the sensitivity bounds have a closed form and coincide with the trimming bounds of Lee (2009) at $\delta = 1$. Section SA6.1.1 shows how to add pre-treatment covariates. Section SA6.1.2 develops inference on the sensitivity bounds that accounts for the estimation of the nuisance parameter π .

Setup. Let $D \in \{0, 1\}$ be a binary treatment, $Y(d) \in \mathcal{Y} \subseteq \mathbb{R}, d \in \{0, 1\}$ be potential outcomes, and $S(d) \in \{0, 1\}, d \in \{0, 1\}$ be potential response indicators, with

$$Y = DY(1) + (1 - D)Y(0), \quad S = DS(1) + (1 - D)S(0).$$

The outcome is recorded only when $S = 1$, so the observed data are (YS, S, D) . The distribution P_0 of the population vector $(Y(0), Y(1), S(0), S(1), D)$ satisfies random treatment assignment, that is, $(S(0), S(1), Y(0), Y(1)) \perp\!\!\!\perp D$. The potential response indicators partition the population into four groups, which we index by the latent type

$$G := \begin{cases} \text{AO} & \text{if } (S(0), S(1)) = (1, 1), \\ \text{C} & \text{if } (S(0), S(1)) = (0, 1), \\ \text{NO} & \text{if } (S(0), S(1)) = (0, 0), \\ \text{DF} & \text{if } (S(0), S(1)) = (1, 0), \end{cases}$$

namely the always observed, the compliers, the never observed, and the defiers. The causal estimand of interest is the average treatment effect (ATE) on the always observed units (Lee, 2009; Semenova, 2025),¹

$$\tau_0 := \mathbb{E}_{P_0}[Y(1) - Y(0) \mid G = \text{AO}].$$

Two assumptions are routinely imposed in this setting. The first, which dates back to Rubin (1976), is that observations are missing completely at random (MCAR, henceforth), that is

$$(S(0), S(1)) \perp\!\!\!\perp (Y(0), Y(1)), \quad (\text{MCAR})$$

which point identifies τ_0 . The second, proposed in Lee (2009), is monotonicity of selection,

$$S(1) \geq S(0) \quad \text{a.s.}, \quad (\text{Mono})$$

which rules out defiers, $P_0(G = \text{DF}) = 0$, and delivers sharp nonparametric bounds for τ_0 . We maintain **Mono** throughout and relax **MCAR**.

Under **Mono** a control unit is observed only if it is always observed, so that

$$\mu^{\text{AO}}(0) := \mathbb{E}_{P_0}[Y(0) \mid G = \text{AO}] = \mathbb{E}_{P_0}[Y \mid D = 0, S = 1] \quad (43)$$

is identified, whereas the treated respondents comprise both compliers and always observed, so that $\mu^{\text{AO}}(1) := \mathbb{E}_{P_0}[Y(1) \mid G = \text{AO}]$ is not. Let P_{T} denote the conditional distribution of Y given $D = 1$ and $S = 1$ under P_0 , with mean $\mu_{\text{T}} := \mathbb{E}_{P_{\text{T}}}[Y]$, let P_{C} and P_{AO} denote the conditional distributions of $Y(1)$ given $G = \text{C}$ and given $G = \text{AO}$ under P_0 , respectively, and let

$$\pi := P_0(G = \text{C} \mid D = 1, S = 1) = \frac{P_0(S = 1 \mid D = 1) - P_0(S = 1 \mid D = 0)}{P_0(S = 1 \mid D = 1)}$$

be the complier share among treated respondents, where the second equality follows from random assignment and **Mono**. Since $S(1) = 1$ if and only if $G \in \{\text{AO}, \text{C}\}$, random assignment yields the mixture decomposition

$$P_{\text{T}} = \pi P_{\text{C}} + (1 - \pi) P_{\text{AO}}, \quad (44)$$

in which only P_{T} is identified. In the treated arm, **MCAR** implies, as G is a function of $(S(0), S(1))$, that

$$P_{\text{C}} = P_{\text{AO}} = P_{\text{T}},$$

that is, $\text{TV}(P_{\text{C}} \parallel P_{\text{AO}}) = 0$, whereas **Mono** leaves the distance between the two distributions unrestricted. We interpolate between the two by imposing $\text{TV}(P_{\text{C}} \parallel P_{\text{AO}}) \leq \delta \in [0, 1]$. In view of (44), we get

$$\text{TV}(P_{\text{C}} \parallel P_{\text{T}}) = (1 - \pi) \text{TV}(P_{\text{C}} \parallel P_{\text{AO}}),$$

so that the restriction can be centered at the identified law P_{T} and only one distribution is left to vary.²

¹A different estimand, the average treatment effect for the whole population $\mathbb{E}_{P_0}[Y(1) - Y(0)]$, can be accommodated as well. Because in that case the baseline is a mixing component of the candidate distribution rather than the mixture itself, we defer its analysis to the missing data case analyzed in Section SA6.4.3.

²Wasserstein-1 would also have this property, whereas for the Kullback–Leibler divergence convexity yields only an inequality. In general,

Sensitivity bounds. For any candidate complier distribution $Q \in \Delta(\mathcal{Y})$, the mixture identity (44) forces the always observed distribution to be $(P_T - \pi Q)/(1 - \pi)$. Subtracting the identified control mean (43), we obtain the map

$$\tau(Q) := \frac{\mu_T - \pi \mathbb{E}_Q[Y]}{1 - \pi} - \mu^{\text{AO}}(0), \quad Q \in \Delta(\mathcal{Y}) \text{ with } \mathbb{E}_Q[|Y|] < \infty, \quad (45)$$

in which $\mathbb{E}_Q[Y]$ is the only quantity not identified from the data, with $\tau_0 = \tau(P_C)$ and, under **MCAR**, $\tau_0 = \tau(P_T) = \mu_T - \mu^{\text{AO}}(0)$. For $\delta \in [0, 1]$, the set of candidate complier distributions is

$$\begin{aligned} \mathcal{P}_{\text{RCT}}(\delta) &:= \{Q \in \Delta(\mathcal{Y}) : \text{TV}(Q \parallel P_T) \leq (1 - \pi)\delta, \exists R \in \Delta(\mathcal{Y}) : P_T = \pi Q + (1 - \pi)R\} \\ &= \{Q \in \Delta(\mathcal{Y}) : \text{TV}(Q \parallel P_T) \leq (1 - \pi)\delta, \pi Q \leq P_T\}, \end{aligned}$$

where the second equality follows from Lemma SA-12(i). When $\mathbb{E}_{P_T}[|Y|] < \infty$, every $Q \in \mathcal{P}_{\text{RCT}}(\delta)$ has $\mathbb{E}_Q[|Y|] \leq \pi^{-1} \mathbb{E}_{P_T}[|Y|] < \infty$ because $\pi Q \leq P_T$, so that τ is well defined on $\mathcal{P}_{\text{RCT}}(\delta)$, and $\mathcal{P}_{\text{RCT}}(\delta)$ is the set $\mathcal{P}_{\text{TV}}(\delta)$ of (34) with $(\kappa, P_*) = (\pi, P_T)$. The sensitivity bounds on the complier mean are

$$\underline{\mu}^C(1; \delta) := \inf_{Q \in \mathcal{P}_{\text{RCT}}(\delta)} \mathbb{E}_Q[Y], \quad \bar{\mu}^C(1; \delta) := \sup_{Q \in \mathcal{P}_{\text{RCT}}(\delta)} \mathbb{E}_Q[Y],$$

and, through the mixture identity, they translate into bounds on the always observed mean $\mu^{\text{AO}}(1)$,

$$\underline{\mu}^{\text{AO}}(1; \delta) := \frac{\mu_T - \pi \bar{\mu}^C(1; \delta)}{1 - \pi}, \quad \bar{\mu}^{\text{AO}}(1; \delta) := \frac{\mu_T - \pi \underline{\mu}^C(1; \delta)}{1 - \pi}, \quad (46)$$

where the first argument records the treatment arm and the second the budget. At $\delta = 0$ we have $\mathcal{P}_{\text{RCT}}(0) = \{P_T\}$ and the two bounds coincide. In this application $g(y) = y$ and there is no moment function, so that $\mathbf{d}_m = 0$ and the moment restriction is vacuous. Since $\tau(Q)$ is a decreasing affine function of $\mathbb{E}_Q[Y]$, the sensitivity bounds on the estimand of interest are

$$\underline{\tau}(\delta) := \inf_{Q \in \mathcal{P}_{\text{RCT}}(\delta)} \tau(Q) = \underline{\mu}^{\text{AO}}(1; \delta) - \mu^{\text{AO}}(0), \quad \bar{\tau}(\delta) := \sup_{Q \in \mathcal{P}_{\text{RCT}}(\delta)} \tau(Q) = \bar{\mu}^{\text{AO}}(1; \delta) - \mu^{\text{AO}}(0). \quad (47)$$

The strong duality result of Theorem SA-14(i) applies, and the next theorem shows that in this particular case the dual problem admits a closed form solution.

Theorem SA-34 (TV sensitivity bounds for the average treatment effect on the always observed). *Assume **Mono** holds and $\pi \in (0, 1)$. Fix $\delta \in (0, 1]$, and suppose $\mathbb{E}_{P_T}[|Y|] < \infty$ and that $\mu^{\text{AO}}(0)$ is finite. Define the trimming quantiles*

$$t_L(\delta) := F_{P_T}^{-1}((1 - \pi)\delta), \quad t_U(\delta) := F_{P_T}^{-1}(1 - \pi\delta), \quad s_L(\delta) := F_{P_T}^{-1}(\pi\delta), \quad s_U(\delta) := F_{P_T}^{-1}(1 - (1 - \pi)\delta),$$

and assume P_T has no atoms at any of them. Then, the sensitivity bounds on τ_0 defined in (47) are

$$\begin{aligned} \underline{\tau}(\delta) &= \mathbb{E}_{P_T}[Y \mathbf{1}\{t_L(\delta) < Y < t_U(\delta)\}] + \frac{1}{1 - \pi} \mathbb{E}_{P_T}[Y \mathbf{1}\{Y \leq t_L(\delta)\}] - \mu^{\text{AO}}(0), \\ \bar{\tau}(\delta) &= \mathbb{E}_{P_T}[Y \mathbf{1}\{s_L(\delta) < Y < s_U(\delta)\}] + \frac{1}{1 - \pi} \mathbb{E}_{P_T}[Y \mathbf{1}\{Y \geq s_U(\delta)\}] - \mu^{\text{AO}}(0). \end{aligned}$$

Moreover, both bounds are attained, and the set of values of τ_0 consistent with the observed distribution P_T , the complier share π , the identified control mean $\mu^{\text{AO}}(0)$, and the sensitivity restriction $\text{TV}(P_C \parallel P_{\text{AO}}) \leq \delta$ is exactly the closed interval $[\underline{\tau}(\delta), \bar{\tau}(\delta)]$.

[Proof]

When $\delta = 1$, the trimming quantiles coincide, $t_L(1) = t_U(1) =: y_{1-\pi}$, the upper bound of the mean for compliers becomes $\bar{\mu}^C(1; 1) = \frac{1}{\pi} \mathbb{E}_{P_T}[Y \mathbf{1}\{Y \geq y_{1-\pi}\}]$, and so the lower bound for always observed is

$$\underline{\mu}^{\text{AO}}(1; 1) = \frac{\mu_T - \pi \bar{\mu}^C(1; 1)}{1 - \pi} = \mathbb{E}_{P_T}[Y \mid Y \leq y_{1-\pi}].$$

Symmetrically, at $\delta = 1$ we have $s_L(1) = s_U(1) =: y_\pi$ and the upper bound reads $\bar{\mu}^{\text{AO}}(1; 1) = \mathbb{E}_{P_T}[Y \mid Y \geq y_\pi]$. After subtracting the identified control mean as in (47), the sensitivity bounds on τ_0 at $\delta = 1$ read

$$\begin{aligned} \underline{\tau}(1) &= \mathbb{E}_{P_T}[Y \mid Y \leq y_{1-\pi}] - \mathbb{E}_{P_0}[Y \mid D = 0, S = 1] =: \underline{\tau}_{\text{Lee}}, \\ \bar{\tau}(1) &= \mathbb{E}_{P_T}[Y \mid Y \geq y_\pi] - \mathbb{E}_{P_0}[Y \mid D = 0, S = 1] =: \bar{\tau}_{\text{Lee}}, \end{aligned}$$

which are precisely the sharp bounds on the average treatment effect for the always observed of Proposition 1a in Lee (2009) under **Mono**. We note that the requirement that P_T place no atom at the trimming quantiles is weaker than the continuity of the potential outcomes maintained there.

Total variation thus interpolates between the point identified estimand under **MCAR** at $\delta = 0$ and the Lee bounds at $\delta = 1$. For an intermediate budget, say $\delta = 0.3$, the bounds allow at most 30% of the treated respondents to depart from **MCAR**. This population reading of δ rests on the mixture structure (44), since on its own a total variation neighborhood of radius δ bounds the probability mass that is moved, not a share of a population (see

any metric that is norm-induced would preserve the equality, whereas ϕ -divergences will not in general. Total variation enjoys this property because it is not only a ϕ -divergence but also an optimal transport metric (see, e.g., Proposition 2.24 in Kuhn et al., 2025).

Remark SA-5). Finally, the fact that $[\underline{\tau}(1), \bar{\tau}(1)]$ coincides exactly with the sharp nonparametric bounds, and does not merely approximate them, is an instance of the general result of Section SA6.4.7.

Remark SA-15 (Relaxing monotonicity). We have maintained **Mono** throughout, but the approach does not hinge on it. Once defiers are allowed, the control respondents are a mixture of always observed and defiers, and the control arm can be handled exactly as the treated arm, with a second sensitivity parameter bounding the total variation distance between the outcome distributions of defiers and always observed. The reduction to a single distribution and the closed form trimmed means of Theorem SA-34 then carry over to each arm, and bounds on τ_0 follow by combining the bounds on $\mu^{\text{AO}}(1)$ and $\mu^{\text{AO}}(0)$, with one budget per arm. The only new ingredient is that the response rates of the two arms no longer pin down the shares of the latent types. The share of defiers can be anything between zero and the minimum between the control response rate and the treated nonresponse rate. For bounds on τ_0 that do not require monotonicity, see also Zhang and Rubin (2003) and Imai (2008). ♣

SA6.1.1 The role of covariates

Suppose we also observe a vector X of pre-treatment covariates with finite support $\mathcal{X}$. With a slight abuse of notation we now use P_0 as the distribution of the vector $(Y(0), Y(1), S(0), S(1), D, X)$, and we maintain that random assignment holds jointly with the covariates, $D \perp\!\!\!\perp (Y(0), Y(1), S(0), S(1), X)$.

Covariates play two distinct roles. They change the point identifying assumption imposed at $\delta = 0$, and they may tighten the sensitivity bounds for $\delta \in (0, 1]$, hence deliver different nonparametric sharp bounds at $\delta = 1$. We now discuss each of the two phenomena in detail.

Change of identifying assumption. Covariates let us change the point identifying assumption that pins down the causal estimand at $\delta = 0$. In place of **MCAR**, we may ask that potential outcomes and attrition be independent only once we condition on X , that is

$$(S(0), S(1)) \perp\!\!\!\perp (Y(0), Y(1)) \mid X, \quad (\text{MAR})$$

the so called “missing at random” (MAR, henceforth) assumption. MAR identifies τ_0 through a covariate-weighted difference in means,

$$\tau_{\text{MAR}}(P_0) := \mathbb{E}_{P_0}[\mathbb{E}_{P_0}[Y \mid X, D = 1, S = 1] - \mathbb{E}_{P_0}[Y \mid X, D = 0, S = 1] \mid G = \text{AO}].$$

Under **MAR**, the inner difference equals $\mathbb{E}_{P_0}[Y(1) - Y(0) \mid X, G = \text{AO}]$, so that $\tau_{\text{MAR}}(P_0) = \tau_0$ by the law of iterated expectations. The difference-in-means estimand

$$\tau_{\text{MCAR}}(P_0) := \tau(P_{\text{T}}) = \mathbb{E}_{P_0}[Y \mid D = 1, S = 1] - \mathbb{E}_{P_0}[Y \mid D = 0, S = 1],$$

with $\tau(\cdot)$ the map (45), identifies the same target τ_0 , but under **MCAR** in place of **MAR**. Let $P_{X|C}$, $P_{X|\text{AO}}$, and $P_{X|D=1, S=1}$ denote the conditional distributions of X given $G = C$, given $G = \text{AO}$, and given $D = 1, S = 1$ under P_0 . Any discrepancy between the two estimands comes from the treated arm, as they differ only through the covariate distribution against which the treated stratum means $\mathbb{E}_{P_0}[Y \mid X = x, D = 1, S = 1]$ are averaged, namely $P_{X|D=1, S=1} = \pi P_{X|C} + (1 - \pi) P_{X|\text{AO}}$ for $\tau_{\text{MCAR}}(P_0)$ and $P_{X|\text{AO}}$ for $\tau_{\text{MAR}}(P_0)$. They therefore coincide when compliers and always observed share the same covariate distribution, as is the case when the share of compliers does not depend on x , or when the treated stratum means do not depend on x . Under **MAR** their difference equals $\pi(\mathbb{E}_{P_0}[Y(1) \mid G = C] - \mathbb{E}_{P_0}[Y(1) \mid G = \text{AO}])$, which is nonzero only when $Y(1)$ has a different mean for compliers and always observed, itself a violation of **MCAR**, and in this case only $\tau_{\text{MAR}}(P_0)$ recovers τ_0 .

Bounds for $\delta > 0$. In the literature on RCTs with attrition it is known that covariates can be used to tighten bounds at $\delta = 1$ (see Proposition 1b in Lee (2009) and Lemma 3 in Semenova (2025)). Whether covariates also tighten the bounds at interior budgets $\delta \in (0, 1)$ is a more delicate matter and it is not true in general, because the addition of covariates changes the point identifying assumption maintained at $\delta = 0$, thus precluding a direct comparison of the two sets of bounds. We also note that the distribution of X among the always observed is identified from the control group by that of $X \mid S = 1, D = 0$, so that only the distribution of the potential outcomes given X is left unrestricted by the data.

There are at least two ways to introduce covariates. To state them, fix $x \in \mathcal{X}$, let $P_C(x)$ and $P_{\text{AO}}(x)$ denote the conditional distributions of $Y(1)$ given $X = x$ and $G = C$ and given $X = x$ and $G = \text{AO}$ under P_0 , and let $P_{\text{T}}(x)$ denote the conditional distribution of Y given $D = 1, S = 1$, and $X = x$ under P_0 , that is, the outcome distribution of the observed treated units within the stratum. With a slight abuse of notation, in this section P_{T} denotes the joint distribution of (Y, X) given $D = 1$ and $S = 1$ under P_0 , of which $P_{\text{T}}(x)$ is the conditional given $X = x$. We maintain that $P_{\text{T}}(x)$ has a continuous distribution function for every stratum under consideration. Repeating the argument leading to (44) conditionally on $X = x$ gives

$$\begin{aligned} P_{\text{T}}(x) &= \pi(x) P_C(x) + (1 - \pi(x)) P_{\text{AO}}(x), \\ \pi(x) &= \frac{P_0(S = 1 \mid D = 1, X = x) - P_0(S = 1 \mid D = 0, X = x)}{P_0(S = 1 \mid D = 1, X = x)}. \end{aligned} \quad (48)$$

The within-stratum complier share $\pi(x)$ is identified from the data exactly as π is. We maintain that $\pi(x) \in (0, 1)$ for every $x \in \mathcal{X}$, so that every stratum contains both compliers and always observed units, and that $\mathbb{E}_{P_{\text{T}}}[\|Y\|]$ and $\mathbb{E}_{P_0}[\|Y\| \mid D = 0, S = 1]$ are finite. On the one hand, we can replicate the exercise above and impose the sensitivity

restriction within each stratum $x \in \mathcal{X}$ as

$$\text{TV}(P_C(x) \parallel P_{AO}(x)) \leq \delta. \quad (49)$$

This set of restrictions induces the collection of within-stratum robustness sets on the conditional distributions of the outcome given the covariates

$$\mathcal{P}_{\text{RCT}}^{\text{pw}}(\delta; x) := \{Q \in \Delta(\mathcal{Y}) : \text{TV}(Q \parallel P_T(x)) \leq \delta(1 - \pi(x)), \pi(x)Q \leq P_T(x)\}.$$

On the other hand, we can impose

$$\int_{\mathcal{X}} \text{TV}(P_C(x) \parallel P_{AO}(x)) dP_{X|D=0, S=1}(x) = \int_{\mathcal{X}} \text{TV}(P_C(x) \parallel P_{AO}(x)) dP_{X|AO}(x) \leq \delta, \quad (50)$$

where $P_{X|D=0, S=1}$ denotes the conditional distribution of X given $D = 0$ and $S = 1$ under P_0 , and where the first equality comes from the fact that observed control units are always observed units under **Mono**. For $Q \in \Delta(\mathcal{Y} \times \mathcal{X})$, we write Q_X for its marginal on $\mathcal{X}$ and $Q(\cdot | x)$ for the conditional distribution of the outcome given $X = x$, and likewise for R . The restriction (50) induces a single pooled robustness set on the joint distributions of outcome and covariates,

$$\begin{aligned} \mathcal{P}_{\text{RCT}, X}(\delta) &:= \left\{ Q \in \Delta(\mathcal{Y} \times \mathcal{X}) : \exists R \in \Delta(\mathcal{Y} \times \mathcal{X}) \text{ s.t. } P_T = \pi Q + (1 - \pi)R, \right. \\ &\quad \left. R_X = P_{X|D=0, S=1}, \int_{\mathcal{X}} \text{TV}(Q(\cdot | x) \parallel R(\cdot | x)) dP_{X|AO}(x) \leq \delta \right\} \\ &= \left\{ Q \in \Delta(\mathcal{Y} \times \mathcal{X}) : \pi Q \leq P_T, \mathbb{E}_Q[\mathbb{1}\{X = x\}] = \frac{\pi(x)}{\pi} \mathbb{E}_{P_T}[\mathbb{1}\{X = x\}], x \in \mathcal{X}, \right. \\ &\quad \left. \int_{\mathcal{X}} \text{TV}(Q(\cdot | x) \parallel P_T(x)) dP_{X|D=1, S=1}(x) \leq (1 - \pi)\delta \right\}, \end{aligned} \quad (51)$$

where the second equality uses the fact that X has discrete support and Lemma SA-12(i). The mixture constraint and the $|\mathcal{X}|$ constraints on the marginal in X encode what the data identify, whereas the budget constraint involves the only component the data leave unrestricted, the conditional distribution of $Y(1)$ given X . Unlike (49), the restriction in (50) allows deviations from the conditional treated outcome equality implied by **MAR** to be heterogeneous across strata and involves one robustness set rather than $|\mathcal{X}|$. It is therefore weaker, hence the bounds it induces are weakly wider. Under (49) the population reading of δ holds stratum by stratum, whereas under (50) it applies to the always observed as a whole. For instance, $\delta = 0.3$ then reads as “at most 30% of the always observed depart from **MAR**”, a share that can be allocated freely across strata, so that the share within a stratum may exceed δ as long as it falls below δ in other strata.

Sensitivity bounds. Even without **MAR**, $\mu^{\text{AO}}(0)$ remains identified as in (43). For the other mean potential outcome, we apply the results above stratum by stratum, for a fixed $x \in \mathcal{X}$. Since (48) and (49) reproduce (44) and the restriction underlying $\mathcal{P}_{\text{RCT}}(\delta)$ upon replacing (P_T, π) with $(P_T(x), \pi(x))$, the within-stratum robustness set is $\mathcal{P}_{\text{RCT}}^{\text{pw}}(\delta; x)$, and Theorem SA-34 applies within each stratum and yields the bounds $\underline{\mu}^{\text{AO}, \text{pw}}(1; \delta, x)$ and $\bar{\mu}^{\text{AO}, \text{pw}}(1; \delta, x)$, defined as in (46) upon the same replacement, on $\mathbb{E}_{P_0}[Y(1) | X = x, G = \text{AO}]$.

We then aggregate the above bounds using the distribution of X given $G = \text{AO}$, which, as noted above, is identified from the observed control group, and so

$$\begin{aligned} \underline{\tau}_X^{\text{pw}}(\delta) &:= \mathbb{E}_{P_0} [\underline{\mu}^{\text{AO}, \text{pw}}(1; \delta, X) | G = \text{AO}] - \mu^{\text{AO}}(0), \\ \bar{\tau}_X^{\text{pw}}(\delta) &:= \mathbb{E}_{P_0} [\bar{\mu}^{\text{AO}, \text{pw}}(1; \delta, X) | G = \text{AO}] - \mu^{\text{AO}}(0). \end{aligned} \quad (52)$$

At $\delta = 0$ the budget forces $P_C(x) = P_{AO}(x) = P_T(x)$ in every stratum, so the two bounds collapse to $\tau_{\text{MAR}}(P_0)$. At $\delta = 1$ the sensitivity bounds with covariates coincide with the Lee bounds (Lee, 2009, Proposition 1b), which are no wider than the bounds without covariates of Theorem SA-34 and are given by

$$\begin{aligned} \underline{\tau}_X^{\text{pw}}(1) &= \mathbb{E}_{P_0} [\mathbb{E}_{P_0} [Y | X, D = 1, S = 1, Y \leq y_{1-\pi(X)}(X)] | D = 0, S = 1] - \mu^{\text{AO}}(0) =: \underline{\tau}_{\text{Lee}, X}, \\ \bar{\tau}_X^{\text{pw}}(1) &= \mathbb{E}_{P_0} [\mathbb{E}_{P_0} [Y | X, D = 1, S = 1, Y \geq y_{\pi(X)}(X)] | D = 0, S = 1] - \mu^{\text{AO}}(0) =: \bar{\tau}_{\text{Lee}, X}, \end{aligned}$$

where, for $u \in (0, 1)$ and $x \in \mathcal{X}$, $y_u(x) := F_{P_T(x)}^{-1}(u)$ denotes the u -quantile of $P_T(x)$.

If we instead use the joint robustness set of (51) the sensitivity bounds for the complier mean are defined as

$$\underline{\mu}_X^{\text{C}}(1; \delta) := \inf_{Q \in \mathcal{P}_{\text{RCT}, X}(\delta)} \mathbb{E}_Q[Y], \quad \bar{\mu}_X^{\text{C}}(1; \delta) := \sup_{Q \in \mathcal{P}_{\text{RCT}, X}(\delta)} \mathbb{E}_Q[Y]. \quad (53)$$

Sensitivity bounds on τ_0 can then be formed as before through (46) and (47), with $\underline{\mu}_X^{\text{C}}(1; \delta)$ and $\bar{\mu}_X^{\text{C}}(1; \delta)$ in place of $\underline{\mu}^{\text{C}}(1; \delta)$ and $\bar{\mu}^{\text{C}}(1; \delta)$. We denote such bounds with $\underline{\tau}_X(\delta)$ and $\bar{\tau}_X(\delta)$. We stress that these bounds share the endpoints of the pointwise ones defined in (52). At $\delta = 0$ they also collapse on $\tau_{\text{MAR}}(P_0)$, whereas at $\delta = 1$ they coincide with the covariate based Lee bounds $\underline{\tau}_{\text{Lee}, X}$ and $\bar{\tau}_{\text{Lee}, X}$ because the total variation constraint is implied by the other constraints. For $\delta \in (0, 1)$, the joint bounds are weakly wider than the pointwise one in (52) because (49) implies (50) but not vice versa.

Obtaining the dual of the optimization problems in (53) is not immediate, but the results in Section SA2 still help us reduce them to a single finite-dimensional convex program, as we now show. In what follows, we maintain the assumptions that $\mathcal{X}$ is finite and that every stratum satisfies $\mathbb{E}_{P_T}[\mathbb{1}\{X = x\}] > 0$.

Because of the constraints on the marginal of X , a candidate $Q \in \mathcal{P}_{\text{RCT},X}(\delta)$ is determined entirely by its conditional distribution. Furthermore, we can rewrite the estimand as

$$\mathbb{E}_Q[Y] = \mathbb{E}_{P_T} \left[\frac{\pi(X)}{\pi} \mathbb{E}_{Q(\cdot|X)}[Y] \right].$$

These two facts suggest that we can focus on each stratum first. In this spirit, define the value function

$$V(t; x) := \sup \{ \mathbb{E}_Q[Y] : Q \in \Delta(\mathcal{Y}), \pi(x)Q \leq P_T(x), \text{TV}(Q \| P_T(x)) \leq t \}.$$

By Lemma SA-12(i), the feasible set of $V(t; x)$ is the mixture-restricted neighborhood $\mathcal{P}_{\text{mix}}^{\leq}(\rho, t)$ of (7), evaluated at baseline $P_T(x)$ and mixing weight $\kappa = \pi(x)$, with vacuous moment restriction. Moreover, the hypotheses of Theorem SA-14(i) hold here, since $\mathbb{E}_{P_T(x)}[|Y|] < \infty$ follows from $\mathbb{E}_{P_T}[|Y|] < \infty$ and because each stratum has positive mass, while with $d_m = 0$, both $\mathcal{M}(\rho) = \{0\}$ and $\Gamma_{\bar{\phi}_{\text{TV},x}}^-(t) = \{0\}$ per our convention, the latter nonempty because the baseline itself is feasible. Therefore, for every $t > 0$,

$$V(t; x) = \inf_{\eta \geq 0} \{ \eta t + \Upsilon(\eta; x) \}, \quad \Upsilon(\eta; x) := \inf_{\zeta \in \mathbb{R}} \left\{ \zeta + \mathbb{E}_{P_T(x)} [(\bar{\phi}_{\text{TV},x}^*)^\pi(Y - \zeta, \eta)] \right\}, \quad (54)$$

where $\bar{\phi}_{\text{TV},x} := \phi_{\text{TV}} + \iota_{[0,1/\pi(x)]}$ is the induced entropy of Proposition SA-13(i) with $\kappa = \pi(x)$. Specializing Example SA-3 to $\kappa = \pi(x)$, the perspective appearing in (54) is

$$(\bar{\phi}_{\text{TV},x}^*)^\pi(z, \eta) = \max \left\{ -\frac{\eta}{2}, z, \frac{z}{\pi(x)} - \eta \frac{1 - \pi(x)}{2\pi(x)} \right\}, \quad z \in \mathbb{R}, \eta \geq 0,$$

which at $\eta = 0$ reduces to the recession value $\max\{z, 0\}/\pi(x)$.

With these ingredients, the upper bound in (53) becomes an exact budget-allocation program across strata,

$$\bar{\mu}_X^C(1; \delta) = \sup \left\{ \mathbb{E}_{P_T} \left[\frac{\pi(X)}{\pi} V(t(X); X) \right] : t : \mathcal{X} \rightarrow \mathbb{R}_+, \mathbb{E}_{P_T}[t(X)] \leq (1 - \pi)\delta \right\}. \quad (55)$$

The program (55) is finite-dimensional and concave: each $V(\cdot; x)$ is nondecreasing and concave, because if Q_0 and Q_1 are feasible at budgets t_0 and t_1 then any mixture of the two is feasible at the mixed budget, by convexity of $\text{TV}(\cdot \| P_T(x))$, while the objective is linear. Moreover, each $V(\cdot; x)$ is constant on $[1 - \pi(x), \infty)$, because the ceiling alone forces $\text{TV}(Q \| P_T(x)) \leq 1 - \pi(x)$, in view of the rescaling identity of Lemma SA-17(i).

Since $\delta > 0$, the allocation $t \equiv (1 - \pi)\delta/2$ is strictly feasible, so Slater's condition holds and we can introduce a multiplier for the budget constraint to get

$$\bar{\mu}_X^C(1; \delta) = \inf_{\eta \geq 0} \left\{ \eta(1 - \pi)\delta + \mathbb{E}_{P_T} \left[\frac{\pi(X)}{\pi} \sup_{t \geq 0} \left\{ V(t; X) - \frac{\pi \eta}{\pi(X)} t \right\} \right] \right\}.$$

For each $x \in \mathcal{X}$ and $c \geq 0$, the inner supremum equals $\Upsilon(c; x)$. Indeed, $\Upsilon(\cdot; x)$ is convex on $\mathbb{R}_+$, as the partial minimum over ζ of a jointly convex function, nonincreasing, because the perspective displayed above is nonincreasing in η , and finite, as $\mathbb{E}_{P_T(x)}[|Y|] < \infty$. It is also lower semicontinuous at 0, because the perspective exceeds its value at $\eta = 0$ minus $\eta/(2\pi(x))$, so that $\Upsilon(\eta; x) \geq \Upsilon(0; x) - \eta/(2\pi(x))$. The extension of $\Upsilon(\cdot; x)$ by $+\infty$ to $(-\infty, 0)$ is therefore closed, proper, and convex. Writing $\Upsilon^*(\cdot; x)$ for its convex conjugate, Theorem 12.2 in Rockafellar (1970) gives $\Upsilon(c; x) = \sup_{s \in \mathbb{R}} \{cs - \Upsilon^*(s; x)\}$. For $t > 0$, $-\Upsilon^*(-t; x) = \inf_{\eta \geq 0} \{\eta t + \Upsilon(\eta; x)\} = V(t; x)$ by (54). For $t < 0$, the same infimum is $-\infty$, as $\Upsilon(\cdot; x) \leq \Upsilon(0; x)$. The term at $s = 0$ equals $\inf_{\eta \geq 0} \Upsilon(\eta; x)$, which is at most $V(t; x)$ for every $t > 0$ and hence, letting $t \downarrow 0$, at most $\sup_{t > 0} \{V(t; x) - ct\}$. Since $V(\cdot; x)$ is nondecreasing, the value at $t = 0$ does not affect the supremum either, so that $\sup_{t \geq 0} \{V(t; x) - ct\} = \Upsilon(c; x)$. Substituting $c = \pi\eta/\pi(x)$ in each stratum and minimizing jointly over the intercepts of the strata, we obtain

$$\bar{\mu}_X^C(1; \delta) = \inf_{\substack{\eta \geq 0, \\ (\zeta_x)_{x \in \mathcal{X}} \in \mathbb{R}^{|\mathcal{X}|}}} \left\{ \eta(1 - \pi)\delta + \sum_{x \in \mathcal{X}} \mathbb{E}_{P_T}[\mathbb{1}\{X = x\}] \frac{\pi(x)}{\pi} \left(\zeta_x + \mathbb{E}_{P_T(x)} \left[(\bar{\phi}_{\text{TV},x}^*)^\pi \left(Y - \zeta_x, \frac{\pi \eta}{\pi(x)} \right) \right] \right) \right\},$$

which is a single finite-dimensional convex program in the $|\mathcal{X}| + 1$ variables $(\eta, (\zeta_x)_{x \in \mathcal{X}})$, as the perspective is jointly convex by Lemma SA-3 and its arguments are affine in these variables.

SA6.1.2 Inference on the sensitivity bounds

We now quantify the sampling uncertainty in the estimated bounds. The large sample theory of Section SA4 holds nuisance parameters such as the complier share π fixed (Remark SA-10). Here no structural parameter enters, the bound is an explicit functional of $(P_T, \pi, \mu^{\text{AO}}(0))$, and the dual optimizer is unique whenever the density of P_T is continuous and strictly positive at the trimming quantiles. We therefore work directly with the trimmed-mean representation of the bound: conditional on π , standard quantile and truncated-moment expansions deliver its influence function, to which we add the first-order contribution of estimating π from the same sample. This route does not impose compactness of the outcome support. By Remark SA-9, the bootstrap of Algorithm SA-1 reduces here to the ordinary nonparametric bootstrap, in which π , the control-observed mean, and the trimming quantiles are recomputed in each resample.

Throughout, the data $\{(Y_i S_i, S_i, D_i)\}_{i=1}^n$ are an i.i.d. sample generated under P_0 and we only discuss the upper bound. Write $p := P_0(D = 1)$, $r_1 := P_0(S = 1 | D = 1)$, and $r_0 := P_0(S = 1 | D = 0)$ for the assignment and arm-specific response probabilities, so that the complier share of (44) is

$$\pi = 1 - \frac{r_0}{r_1}.$$

The bound on τ_0 is an explicit functional of $(P_T, \pi, \mu^{\text{AO}}(0))$ as shown above, with upper-tail trimming points $s_L(\delta) = F_{P_T}^{-1}(\pi\delta)$ and $s_U(\delta) = F_{P_T}^{-1}(1 - (1 - \pi)\delta)$. Its plug-in estimator replaces r_1, r_0, p by the corresponding sample frequencies, P_T by the empirical distribution $\hat{P}_{T,n}$ of the treated observed outcomes $\{Y_i : D_i = 1, S_i = 1\}$, the trimming points by the empirical quantiles $\hat{s}_L(\delta), \hat{s}_U(\delta)$ of $\hat{P}_{T,n}$, and $\mu^{\text{AO}}(0)$ by the control observed sample mean. Write $\hat{\tau}(\delta)$ for this plug-in estimator of the bound $\bar{\tau}(\delta)$, and let $\{\hat{\tau}^{(b)}(\delta)\}_{b=1}^B$ be its values on $B = 1000$ nonparametric bootstrap samples, with mean

$$\hat{\tau}_{\text{boot}}(\delta) = \frac{1}{B} \sum_{b=1}^B \hat{\tau}^{(b)}(\delta).$$

The limiting law of $\sqrt{n}(\hat{\tau}(\delta) - \bar{\tau}(\delta))$ is governed by three first-order influence functions. Writing $W = (YS, S, D)$, $\bar{\mu}(\delta) := \bar{\mu}^{\text{AO}}(1; \delta)$, and $\vartheta_+(y; \delta) := y \mathbf{1}\{s_L(\delta) < y < s_U(\delta)\} + \frac{1}{1-\pi} y \mathbf{1}\{y \geq s_U(\delta)\}$ so that $\mathbb{E}_{P_T}[\vartheta_+(Y; \delta)] = \bar{\mu}(\delta)$, these influence functions are

$$\begin{aligned} \psi_T(W; \delta) &= \frac{DS}{p r_1} \left\{ \vartheta_+(Y; \delta) - \bar{\mu}(\delta) + s_L(\delta) (\mathbf{1}\{Y \leq s_L(\delta)\} - \pi\delta) + \frac{\pi}{1-\pi} s_U(\delta) (\mathbf{1}\{Y \leq s_U(\delta)\} - 1 + (1 - \pi)\delta) \right\}, \\ \psi_\pi(W) &= \frac{r_0}{r_1^2} \frac{D}{p} (S - r_1) - \frac{1}{r_1} \frac{1-D}{1-p} (S - r_0), \quad \psi_0(W) = \frac{(1-D)S}{(1-p)r_0} (Y - \mu^{\text{AO}}(0)). \end{aligned}$$

Only $\psi_T(\cdot; \delta)$ varies with the sensitivity level. The other two, the influence functions of the estimated complier share $\hat{\pi} = 1 - \hat{r}_0/\hat{r}_1$ and of the control mean $\mu^{\text{AO}}(0)$, are free of δ . Denote the total derivative of the upper bound with respect to the complier share as

$$\varkappa(\delta) := -\delta \left(s_L(\delta) + \frac{\pi}{1-\pi} s_U(\delta) \right) + \frac{1}{(1-\pi)^2} \mathbb{E}_{P_T}[Y \mathbf{1}\{Y \geq s_U(\delta)\}]$$

and define the second-step influence function

$$\psi_2(W; \delta) := \psi_T(W; \delta) - \psi_0(W),$$

which holds π fixed, and the full influence function

$$\psi_{\text{full}}(W; \delta) := \psi_T(W; \delta) + \varkappa(\delta) \psi_\pi(W) - \psi_0(W),$$

which also takes $\hat{\pi}$ into account. Because ψ_{full} is the influence function of the upper bound functional, the moment function $\bar{\tau}(\delta) + \psi_{\text{full}}(W; \delta)$ has zero pathwise derivative with respect to $(P_T, \pi, \mu^{\text{AO}}(0))$ at their population values, that is, it is Neyman orthogonal. We form four estimators of the sampling variability of $\hat{\tau}(\delta)$

$$\hat{\sigma}_{\text{naive}}^2(\delta) = \hat{V}_2(\delta), \quad \hat{\sigma}_{\text{NO}}^2(\delta) = \hat{V}_{\text{full}}(\delta), \quad \hat{\sigma}_{\text{boot}}^2(\delta) = \frac{n}{B-1} \sum_{b=1}^B (\hat{\tau}^{(b)}(\delta) - \hat{\tau}_{\text{boot}}(\delta))^2, \quad \hat{\sigma}_{\text{Lee}}^2 = \hat{V}_{\text{Lee,U}} + \hat{V}_{\text{Lee,C}},$$

where $\hat{V}_2(\delta) := n^{-1} \sum_i \hat{\psi}_2(W_i; \delta)^2$ and $\hat{V}_{\text{full}}(\delta) := n^{-1} \sum_i \hat{\psi}_{\text{full}}(W_i; \delta)^2$ are the influence-function variances. The remaining two are the ingredients of the trimming bound variance of Lee (2009, Proposition 3): $V_{\text{Lee,U}}$ is the asymptotic variance of the estimated treated trimmed mean $\mathbb{E}_{P_T}[Y \mid Y \geq F_{P_T}^{-1}(\pi)]$ and $V_{\text{Lee,C}}$ that of the estimated control mean $\mu^{\text{AO}}(0)$, with sample analogs $\hat{V}_{\text{Lee,U}}$ and $\hat{V}_{\text{Lee,C}}$. The naive estimator $\hat{\sigma}_{\text{naive}}^2(\delta)$ treats the estimated complier share $\hat{\pi}$ as known, whereas the variance estimator $\hat{\sigma}_{\text{NO}}^2(\delta)$ is based on the Neyman-orthogonal score and therefore uses the full influence function, including the channel through which $\hat{\pi}$ contributes to sampling uncertainty.

We now state some standard regularity conditions before presenting the main inferential result of this section.

Assumption SA-7 (Inference regularity). *The sample $\{(Y_i S_i, S_i, D_i)\}_{i=1}^n$ is i.i.d. under P_0 , $p \in (0, 1)$, $r_0, r_1 \in (0, 1]$ with $\pi = 1 - r_0/r_1 \in (0, 1)$, $\delta \in (0, 1]$, $\mathbb{E}_{P_0}[Y^2 \mid D = d, S = 1] < \infty$, $d \in \{0, 1\}$, and the baseline P_T admits a density f_{P_T} that is continuous and strictly positive at the trimming quantiles $s_L(\delta)$ and $s_U(\delta)$.*

Proposition SA-35 (Nuisance parameter uncertainty quantification). *Let Assumption SA-7 hold. Then:*

(i) *The plug-in is asymptotically normal,*

$$\sqrt{n}(\hat{\tau}(\delta) - \bar{\tau}(\delta)) \rightsquigarrow \mathbf{N}(0, \sigma_{\text{full}}^2(\delta)), \quad \sigma_{\text{full}}^2(\delta) = \mathbb{E}[\psi_{\text{full}}(W; \delta)^2].$$

(ii) *Writing $\sigma_2^2(\delta) := \mathbb{E}[\psi_2(W; \delta)^2]$ for the asymptotic variance when π is known and $V_\pi := \mathbb{E}[\psi_\pi^2]$, as $n \rightarrow \infty$*

$$\hat{\sigma}_{\text{NO}}^2(\delta) \xrightarrow{\mathbb{P}} \sigma_{\text{full}}^2(\delta), \quad \hat{\sigma}_{\text{boot}}^2(\delta) \xrightarrow{\mathbb{P}} \sigma_{\text{full}}^2(\delta), \quad \hat{\sigma}_{\text{naive}}^2(\delta) \xrightarrow{\mathbb{P}} \sigma_2^2(\delta),$$

where $\sigma_2^2(\delta) = \sigma_{\text{full}}^2(\delta) - \varkappa(\delta)^2 V_\pi \leq \sigma_{\text{full}}^2(\delta)$. If, in addition, $r_1 < 1$ and f_{P_T} is continuous and strictly positive at $F_{P_T}^{-1}(\pi)$, then $\hat{\sigma}_{\text{Lee}}^2 \xrightarrow{\mathbb{P}} \sigma_{\text{full}}^2(1)$.

[Proof]

The term $\varkappa \psi_\pi$ is exactly the contribution of estimating π . Discarding it, that is, treating π as known as the general theory does, understates the variance. The understatement vanishes at $\delta = 0$, where $\varkappa \rightarrow 0$ and the bound is the difference in means, which does not require estimation of π . Because the plug-in is asymptotically normal (part (i)) with a variance that $\hat{\sigma}_{\text{NO}}(\delta)$ and $\hat{\sigma}_{\text{boot}}(\delta)$ estimate consistently (part (ii)), the Wald intervals based on these two

standard errors attain asymptotic coverage, whereas the naive interval under-covers whenever $\varkappa(\delta) \neq 0$. At the endpoint $\delta = 1$, the bound reduces to Lee’s trimmed mean and $\sigma_{\text{full}}^2(1) = V_{\text{Lee},U} + V_{\text{Lee},C}$ is the trimming bound variance of Lee (2009, Proposition 3).

SA6.2 Counterfactuals in Structural Models

Counterfactual predictions from structural models depend on the distribution specified for latent variables such as taste shocks, unobserved heterogeneity, or productivity shocks, which is typically not nonparametrically identifiable and is therefore chosen for tractability. Christensen and Connault (2023) (CC23) propose to assess how sensitive counterfactuals are to this choice by letting the latent distribution range over a ϕ -divergence neighborhood of the parametric specification. We conduct the same exercise with a sensitivity parameter that admits a population share interpretation, a case that is not covered by the theory of CC23, whose entropy functions are superlinear (see the discussion of the pointwise dual constraint at the beginning of Section SA2.5).

A generic structural model. Let $Z \in \mathcal{Z}$ collect the latent variables, let $P_* \in \mathcal{P}(\mathcal{Z})$ be the distribution postulated by the econometrician, which is known, and let $\theta \in \Theta$ collect the structural parameters. The model rationalizes the observed data through the moment conditions $\mathbb{E}_P[m(Z; \theta)] \in \mathcal{M}(\rho_0)$, where ρ_0 collects the moments that are estimated from the data, and the counterfactual of interest is the expectation of a known function, $\mathbb{E}_P[g(Z; \theta)]$. The relevant robustness set is the contamination robustness set of the main text, recalled in (20),

$$\mathcal{P}_C(\theta; \delta) = \{P \in \mathcal{P}(\mathcal{Z}) : P = (1 - \delta)P_* + \delta R \text{ for some } R \in \mathcal{P}(\mathcal{Z}), \mathbb{E}_P[m(Z; \theta)] \in \mathcal{M}(\rho_0)\},$$

so that δ is the share of the population whose latent variables are not drawn from P_* . This set is the member $\mathbb{U} = [1 - \delta, +\infty)$ of the family of Section SA4.1, that is, the contamination neighborhood $\mathcal{C}_{1-\delta}(P_*) \cap \mathcal{E}(\rho_0)$ of Section SA2.6. The sensitivity bounds on the counterfactual are

$$\underline{b}_C(\delta) := \inf_{\theta \in \Theta} \inf_{P \in \mathcal{P}_C(\theta; \delta)} \mathbb{E}_P[g(Z; \theta)], \quad \bar{b}_C(\delta) := \sup_{\theta \in \Theta} \sup_{P \in \mathcal{P}_C(\theta; \delta)} \mathbb{E}_P[g(Z; \theta)],$$

with the understanding that the inner infimum and supremum are $+\infty$ and $-\infty$, respectively, whenever $\mathcal{P}_C(\theta; \delta)$ is empty. Their plug-in estimators replace ρ_0 by an estimator $\hat{\rho}_n$, and, as the baseline is known, their large sample properties are those of Remark SA-11.

SA6.2.1 The marital college premium

We borrow the setting from Example 2.2 of the main text, which revisits one of the leading empirical applications of CC23: the marital college premium in the marriage-matching model that Chiappori, Salanié and Weiss (2017) estimate on American Community Survey data. Women belong to one of $J \in \mathbb{N}$ education groups, which we term “types,” $j = 0$ denotes the outside option of remaining unmatched, and a type- $b \in [J]$ woman receives utility ν_{0b} if unmatched and $t_{jb} + \nu_{jb}$ if matched with a type- $j \in [J]$ man, where we set $t_{0b} := 0$ for the outside option. As in the main text, the distribution of the preference shocks does not depend on the type of the woman (see also Remark SA-16), so that we consider a single random vector $U = (\nu_0, \dots, \nu_J)^\top$, whose postulated distribution $\tilde{P}$ makes $\nu_0, \dots, \nu_J$ independent Gumbel random variables. We fix two female types $b, b' \in [J]$, let $\varrho \in \mathbb{R}^{2J}$ collect their observed marriage probabilities with each male type, and let $\vartheta := (t_{jb}, t_{jb'}, j \in [J])$ range over a parameter space $\Theta \subseteq \mathbb{R}^{2J}$. The moment function is $s(U; \vartheta) = (s_1(U; \vartheta)^\top, s_2(U; \vartheta)^\top)^\top$, where

$$s_1(U; \vartheta) = \begin{bmatrix} \left(\mathbb{1} \left\{ j = \arg \max_{j' \in \{0, \dots, J\}} (t_{j'b} + \nu_{j'}) \right\} \right)_{j=1}^J \\ \left(\mathbb{1} \left\{ j = \arg \max_{j' \in \{0, \dots, J\}} (t_{j'b'} + \nu_{j'}) \right\} \right)_{j=1}^J \end{bmatrix}, \quad s_2(U; \vartheta) = (\nu_j, \nu_j^2 - \pi^2/6)_{j=0}^J,$$

with ties broken toward the lowest indexed alternative. The moment conditions $\mathbb{E}_P[s_1(U; \vartheta)] = \varrho$ equate the observed marriage probabilities to the model implied choice probabilities, and $\mathbb{E}_P[s_2(U; \vartheta)] = 0_{2(J+1)}$ imposes that each shock has mean zero and variance $\pi^2/6$, the variance of a standard Gumbel distribution. The counterfactual is the marital college premium for female types b and b' , $\mathbb{E}_P[f(U; \vartheta)]$, where

$$f(U; \vartheta) = \max_{j \in \{0, \dots, J\}} \{t_{jb} + \nu_j\} - \max_{j \in \{0, \dots, J\}} \{t_{jb'} + \nu_j\}.$$

The robustness set is $\mathcal{P}_{\text{STR}}(\vartheta; \delta)$ of (2.7) in Example 2.2 of the main text, so that δ is the share of women whose shocks are not drawn from the Gumbel distribution, and the sensitivity bounds $\underline{k}(\delta)$ and $\bar{k}(\delta)$ on the premium are the infimum and the supremum of $\mathbb{E}_P[f(U; \vartheta)]$ over $\vartheta \in \Theta$ and $P \in \mathcal{P}_{\text{STR}}(\vartheta; \delta)$. This is the generic exercise described above with $Z = U$, $\mathcal{Z} = \mathbb{R}^{J+1}$, $\theta = \vartheta$, $g = f$, $m = s$, $\rho_0 = \varrho$, the block s_2 having a known target, and $P_* = \tilde{P}$, so that $\mathcal{P}_{\text{STR}}(\vartheta; \delta) = \mathcal{P}_C(\vartheta; \delta)$, $\underline{k}(\delta) = \underline{b}_C(\delta)$, and $\bar{k}(\delta) = \bar{b}_C(\delta)$. The first identity uses that every $P \in \mathcal{P}_{\text{STR}}(\vartheta; \delta)$ lies in the admissible class $\mathcal{P}(\mathbb{R}^{J+1})$ of (1). Indeed, $|f(u; \vartheta)| \leq \max_{j \in \{0, \dots, J\}} |t_{jb} - t_{jb'}|$ and s_1 takes values in $\{0, 1\}^{2J}$, while the normalization $\mathbb{E}_P[s_2(U; \vartheta)] = 0_{2(J+1)}$ makes $s_2(U; \vartheta)$ integrable under P . The observed marriage probabilities are estimated by the corresponding empirical frequencies $\hat{\varrho}_n$. In the remainder of this section we state the results in the notation of the framework.

Remark SA-16 (Vacuous bounds). If g is unbounded above or below, then the robustness set alone is too large, as the contaminating distribution R can place its mass on regions where g is arbitrarily large or arbitrarily negative,

implying that

$$\bar{k}(\delta) = +\infty \quad \text{and} \quad \underline{k}(\delta) = -\infty$$

for every $\delta > 0$. This issue arises, for example, if one writes the premium using two different collections of shocks for types b and b' , and the shocks ν_{jb} and $\nu_{jb'}$ have unbounded support, as under Gaussian or Gumbel specifications. There are at least two ways to make such bounds non-trivial:

1. impose a common-shock representation, so that the premium becomes bounded;
2. intersect the robustness set with moment normalizations which impose structure on the tails of the distribution.

In this section we adopt both for simplicity, but the moment normalizations contained in s_2 would have sufficed to bound the premium uniformly over the robustness set. ♣

Throughout the remainder of this application, we consider only distributions that are absolutely continuous with respect to the baseline. This is without loss of generality for the sensitivity bounds, because condition (b) of Theorem SA-15 holds for the entropy $\phi_{\text{TV}} = \phi_{\text{TV}} + \iota_{[1-\delta, \infty)}$ of Proposition SA-13(ii), whose neighborhood is the contamination set, as we show in the next lemma. Moreover, Proposition SA-16 allows us to remove the pointwise constraint from the dual.

For the next result, define $\vec{\rho}_0 := (\rho_0^\top, 0_{2(J+1)}^\top)^\top$ and let $\Gamma_{\phi_{\text{TV}}}^{\ll}(\theta; \delta)$ be the attainable moment set of (3) for this entropy, restricted to distributions dominated by P_* , that is, by Proposition SA-13(ii), the set of the $\mathbb{E}_P[m(U; \theta)]$ with $P \in \mathcal{P}(\mathbb{R}^{J+1})$, $P \ll P_*$, and $P = (1 - \delta)P_* + \delta R$ for some $R \in \mathcal{P}(\mathbb{R}^{J+1})$. At every θ in a neighborhood of the population optimizers, we maintain the interior condition $\vec{\rho}_0 \in \text{ri}(\Gamma_{\phi_{\text{TV}}}^{\ll}(\theta; \delta))$. A primitive sufficient condition is obtained as in Lemma SA-20(ii): it suffices that the moment restrictions be satisfied by some $\bar{P} \ll P_*$ whose density with respect to P_* is bounded below by a constant strictly larger than $1 - \delta$ and under which $m(U; \theta)$ has a finite, full rank covariance matrix (see the discussion following Lemma SA-24).

Lemma SA-36 (Absolutely continuous Gumbel formulation). *For every $\theta \in \Theta$ and every $\lambda \in \mathcal{R}_\lambda$,*

$$\sup_{u \in \mathbb{R}^{J+1}} \{g(u; \theta) - \lambda^\top m(u; \theta)\} = P_*\text{-ess sup}_{u \in \mathbb{R}^{J+1}} \{g(u; \theta) - \lambda^\top m(u; \theta)\},$$

and the same identity holds with $-g$ in place of g . Consequently, for every (θ, δ) such that $\delta > 0$ and $\vec{\rho}_0 \in \text{ri}(\Gamma_{\phi_{\text{TV}}}^{\ll}(\theta; \delta))$, condition (b) of Theorem SA-15 holds for the entropy $\phi_{\text{TV}} = \phi_{\text{TV}} + \iota_{[1-\delta, \infty)}$, for $g(\cdot; \theta)$ and for $-g(\cdot; \theta)$.

[Proof]

SA6.3 Formula and Shift-Share Instruments

A growing body of empirical work identifies causal effects by means of formula instruments, that is, instruments obtained by applying a known formula to a vector of exogenous shocks and to a vector of predetermined covariates. We refer the reader to Borusyak, Hull and Jaravel (2025) for a review. Borusyak and Hull (2023) (BH23, henceforth) point out that exogeneity of the shocks alone does not deliver a consistent estimator. In the main text we show that this scenario fits into the framework of Section SA1.2. In this subsection we redefine all the relevant quantities and explain in detail how the sensitivity bounds can be computed.

We follow the notation of BH23 throughout. Let $i \in [n], n \in \mathbb{N}$ index the units of the sample and consider the causal or structural model

$$y_i = \beta x_i + \varepsilon_i,$$

where $y_i \in \mathbb{R}$ is a demeaned outcome, $x_i \in \mathbb{R}$ is a demeaned endogenous regressor, $\varepsilon_i \in \mathbb{R}$ is an unobserved residual, and $\beta \in \mathbb{R}$ is the parameter of interest. Let $v = (v_1, \dots, v_K) \in \mathcal{V}$, $K \in \mathbb{N}$, denote a vector of exogenous shocks, where $\mathcal{V}$ is the set of shock configurations the design allows for, a subset of $\mathbb{R}^K$ when the shocks are real valued, and let w collect predetermined covariates of unrestricted dimension. The candidate instrument of unit $i \in [n]$ is

$$z_i := f_i(v; w),$$

where $f_i(\cdot; w) : \mathcal{V} \rightarrow \mathbb{R}$ are known nonstochastic functions on $\mathcal{V}$. BH23 assume that the shocks v are exogenous, that is

$$v \perp\!\!\!\perp (\varepsilon_1, \dots, \varepsilon_n) \mid w,$$

and show that this alone does not make z_i orthogonal to ε_i . Their proposed solution is to recenter z_i at its expected value under the assignment process of the shocks. Computing that expected value requires the researcher to specify a distribution for the shocks, an assumption that economic theory rarely pins down. Exogeneity is maintained throughout the sensitivity exercise. What the sensitivity exercise relaxes is the assumption that the law of v given w is known. We let $\tilde{P} \in \Delta(\mathcal{V})$ denote the assignment process postulated by the econometrician, and we assume for simplicity that the distribution of v given w does not depend on w , as is the case in the application we consider.

The postulated assignment process defines the expected instruments $\mu_i \equiv \mu_i(\tilde{P})$, where for $i \in [n]$

$$\mu_i(P) := \mathbb{E}_P[f_i(v; w) \mid w] = \int_{\mathcal{V}} f_i(v'; w) dP(v'),$$

and the recentered instrument of unit i is $\tilde{z}_i(P) := z_i - \mu_i(P)$, with $\tilde{z}_i \equiv \tilde{z}_i(\tilde{P})$. The recentered instrumental variable estimator of BH23 is

$$\hat{\beta}_* := \frac{\sum_{i=1}^n \tilde{z}_i y_i}{\sum_{i=1}^n \tilde{z}_i x_i}. \quad (56)$$

We use a “design based” setup in which we hold the realized sample fixed (outcomes, treatments, covariates, and the realized shocks $\tilde{v}$, at which $z_i = f_i(\tilde{v}; w)$ for every $i \in [n]$), so that the recentered estimate is a function of the assignment process alone, and we ask how much it moves as that process ranges over a neighborhood of $\tilde{P}$. Because (56) depends on the postulated assignment process $\tilde{P}$ only through the expected instruments, the sensitivity exercise amounts to replacing $\tilde{P}$ by candidate assignment processes in some robustness set. We accordingly define

$$\hat{\beta}(P) := \frac{\sum_{i=1}^n \tilde{z}_i(P) y_i}{\sum_{i=1}^n \tilde{z}_i(P) x_i} = \frac{G_y(P)}{G_x(P)}, \quad \begin{aligned} G_y(P) &:= g_y(\tilde{v}) - \mathbb{E}_P[g_y], \\ G_x(P) &:= g_x(\tilde{v}) - \mathbb{E}_P[g_x], \end{aligned} \quad (57)$$

where

$$g_y(\cdot) := \sum_{i=1}^n y_i f_i(\cdot; w), \quad g_x(\cdot) := \sum_{i=1}^n x_i f_i(\cdot; w).$$

By construction $\hat{\beta}_* = \hat{\beta}(\tilde{P})$, so the exercise is centered around the reported estimate.

We can conduct a sensitivity exercise on $\hat{\beta}_*$ by letting the assignment process of v vary over the robustness set

$$\mathcal{P}_{\text{FI}}(\delta) := \{P \in \Delta(\mathcal{V}) : P = (1 - \delta)\tilde{P} + \delta R \text{ for some } R \in \Delta(\mathcal{V})\}.$$

Also, define the set

$$\mathcal{P}_{\neq 0} := \{P \in \Delta(\mathcal{V}) : G_x(P) \neq 0\},$$

which excludes the distributions in the robustness set at which the recentered first stage equals zero and the estimator is therefore undefined.

We assume throughout that the recentered first stage is positive at the postulated law, $G_x(\tilde{P}) > 0$, so that the reported estimate in (56) is well defined. This is without loss of generality, because replacing (g_y, g_x) by $(-g_y, -g_x)$ leaves $\hat{\beta}(\cdot)$ unchanged and reverses the sign of $G_x(\tilde{P})$. The space $\mathcal{V}$ of shock realizations is otherwise arbitrary, and we maintain instead a condition on the formula itself, namely that it is bounded uniformly over the sample,

$$\|f\|_\infty := \max_{i \in [n]} \sup_{v' \in \mathcal{V}} |f_i(v'; w)| < \infty. \quad (58)$$

Since the exercise is conducted conditionally on the realized sample, (58) implies

$$\sup_{v' \in \mathcal{V}} |g_y(v')| \leq \|f\|_\infty \sum_{i=1}^n |y_i|, \quad \sup_{v' \in \mathcal{V}} |g_x(v')| \leq \|f\|_\infty \sum_{i=1}^n |x_i|.$$

This condition keeps the sensitivity bounds informative. Without it, the robustness set can contain assignment processes that place mass where g_x or g_y is arbitrarily large, and the bounds can be infinite (Remark SA-17). Condition (58) does not restrict $\mathcal{V}$, and it holds in each of the three examples that motivate BH23, which we discuss next.

Example SA-4 (BH23 motivating examples). In their network spillover example, a stylized version of the setting of Miguel and Kremer (2004), the treatment of student i counts the number of her neighbors who have been dewormed in an intervention, so that

$$f_i(v'; w) = \sum_{k=1}^K \text{Neighbor}_{ik} v'_k,$$

where $v'_k \in \{0, 1\}$ indicates that student k is selected for the deworming intervention, $\text{Neighbor}_{ik} \in \{0, 1\}$ indicates that i and k are connected by an observed network link, and the shocks are assigned at the level of the $K = n$ students. Then, the formula is bounded by the largest number of neighbors that a student has. In their market access example, which follows Donaldson and Hornbeck (2016), the treatment of region i is the growth of its market access between two dates due to improvements to the interregional railroad network, that is

$$f_i(v'; w) = \log \text{MA}_i(\text{Network}_1(v')) - \log \text{MA}_i(\text{Network}_0), \quad \text{MA}_i(\text{Network}) := \sum_{j=1}^n \frac{\text{Pop}_j}{\text{Time}(\text{Network}, \text{Loc}_i, \text{Loc}_j)},$$

where Pop_j is the time invariant population of region j , Loc_j is its location, $\text{Time}(\cdot)$ gives the travel time between two regions, and Network is the set of railway lines and other types of transit in operation. The shocks enter through the network in operation at the later date: $v'_k \in \{0, 1\}$ indicates that planned line k faces no delays, so that

$$\text{Network}_1(v') = \text{Network}_0 \cup \{k \in [K] : v'_k = 1\},$$

where Network_0 denotes the initial network. Since the shocks are indicators, the formula takes finitely many values and is therefore bounded. The application in Section 6.3 of the main text takes up this design, in the version that BH23 apply to the construction of the Chinese high-speed rail network. In their Medicaid eligibility example, which

comes from Currie and Gruber (1996), the treatment of individual i indicates her eligibility for Medicaid, so that

$$f_i(v'; w) = \mathbf{1}\{\text{IncDem}_i \in v'_{\text{State}_i}\},$$

where IncDem_i is a vector of individual income and demographics, $\text{State}_i \in [K]$ indexes i 's state of residence, and v'_k is state k 's eligibility policy, that is, the set of income and demographic groups eligible for Medicaid in that state. Here $\|f\|_\infty = 1$ whatever the space of policy profiles.

Condition (58) fails, on the other hand, for a shift-share formula

$$f_i(v'; w) = \sum_{k \in [K]} w_{ik} v'_k$$

with shocks ranging over an unbounded set, such as $\mathcal{V} = \mathbb{R}^K$, and Remark SA-17 describes the two ways of rescuing the exercise in that case. $\blacklozenge$

The objects of interest are the largest and the smallest value that the recentered estimator takes as the law of v ranges over the robustness set, that is

$$\underline{\beta}(\delta) := \inf_{P \in \mathcal{P}_{\text{FI}}(\delta) \cap \mathcal{P}_{\neq 0}} \widehat{\beta}(P), \quad \overline{\beta}(\delta) := \sup_{P \in \mathcal{P}_{\text{FI}}(\delta) \cap \mathcal{P}_{\neq 0}} \widehat{\beta}(P). \quad (59)$$

The estimator in (57) is not linear in the candidate distribution P , so we cannot apply our framework directly to the two optimization programs above. Its numerator and denominator are, however, both affine in P , so that the estimator depends on P only through the two scalars $\mathbb{E}_P[g_y]$ and $\mathbb{E}_P[g_x]$, so that the exercise takes place in a two-dimensional set no matter the dimensionality of $\mathcal{V}$. Moreover, given $b \in \mathbb{R}$, the requirement that the candidate estimate equals b reduces to a single linear restriction on the candidate $P \in \mathcal{P}_{\neq 0}$. Indeed, $\widehat{\beta}(P) = b$ if and only if $G_y(P) - b G_x(P) = 0$ and

$$G_y(P) - b G_x(P) = \{g_y(\tilde{v}) - b g_x(\tilde{v})\} - \mathbb{E}_P[g_y - b g_x] = g_b(\tilde{v}) - \mathbb{E}_P[g_b],$$

where the first equality follows from the definitions of G_y and G_x in (57) and the second from setting $g_b := g_y - b g_x$. That is, the estimator attains the value b under P exactly when the sample orthogonality condition

$$\sum_{i=1}^n \tilde{z}_i(P) (y_i - b x_i) = 0$$

of the recentered instrument holds at b under P . Once b is fixed, g_b is a known function on $\mathcal{V}$ that does not vary with P , so that the exercise fits the framework of Section SA1.2 with $g = g_b$ and no moment restriction, as explained in the main text. The next proposition collects the properties of the sensitivity exercise.

Proposition SA-37 (Sensitivity bounds under a contamination neighborhood). *Let $\delta \in [0, 1]$. Write $\bar{g}_y^* := \mathbb{E}_{\tilde{P}}[g_y]$ and $\bar{g}_x^* := \mathbb{E}_{\tilde{P}}[g_x]$, so that $G_y(\tilde{P}) = g_y(\tilde{v}) - \bar{g}_y^*$ and $G_x(\tilde{P}) = g_x(\tilde{v}) - \bar{g}_x^*$. Also, define*

$$\begin{aligned} G_y(v'; \delta) &:= G_y((1 - \delta)\tilde{P} + \delta \delta_{v'}^{\text{Dirac}}) = G_y(\tilde{P}) - \delta \{g_y(v') - \bar{g}_y^*\}, \\ G_x(v'; \delta) &:= G_x((1 - \delta)\tilde{P} + \delta \delta_{v'}^{\text{Dirac}}) = G_x(\tilde{P}) - \delta \{g_x(v') - \bar{g}_x^*\}, \end{aligned}$$

for $v' \in \mathcal{V}$, the values of G_y and G_x when the postulated law is contaminated by a point mass at v' . Then the following statements hold.

- (i) *If $\inf_{v' \in \mathcal{V}} G_x(v'; \delta) > 0$, then the estimator (57) is well defined at every $P \in \mathcal{P}_{\text{FI}}(\delta)$ and both bounds in (59) are finite. If instead $\inf_{v' \in \mathcal{V}} G_x(v'; \delta) < 0$, the neighborhood contains assignment laws at which the recentered first stage vanishes. The former holds exactly on*

$$\mathcal{I}_{\text{FS}} := \left\{ \delta \in [0, 1] : \delta \sup_{v' \in \mathcal{V}} \{g_x(v') - \bar{g}_x^*\} < G_x(\tilde{P}) \right\},$$

an interval containing the origin whose right endpoint is the first-stage breakdown budget

$$\delta_{\text{FS}} := \min \left\{ 1, \frac{G_x(\tilde{P})}{\sup_{v' \in \mathcal{V}} \{g_x(v') - \bar{g}_x^*\}} \right\},$$

with the convention that the ratio equals $+\infty$ when the supremum in its denominator vanishes.

- (ii) *For every $\delta \in \mathcal{I}_{\text{FS}}$ the bounds in (59) are given in closed form by*

$$\underline{\beta}(\delta) = \inf_{v' \in \mathcal{V}} \frac{G_y(v'; \delta)}{G_x(v'; \delta)}, \quad \overline{\beta}(\delta) = \sup_{v' \in \mathcal{V}} \frac{G_y(v'; \delta)}{G_x(v'; \delta)}.$$

Every value strictly between the two bounds is taken by the estimator at some $P \in \mathcal{P}_{\text{FI}}(\delta)$.

[Proof]

Remark SA-17 (Unbounded formulas). When g_x is unbounded above on $\mathcal{V}$, as it is for a shift-share formula with shocks ranging over $\mathcal{V} = \mathbb{R}^K$ whenever $\sum_{i=1}^n x_i w_{ik} \neq 0$ for some $k \in [K]$, the exercise as stated is vacuous. Indeed, $\sup_{v' \in \mathcal{V}} \{g_x(v') - \bar{g}_x^*\}$ is then $+\infty$, so that at every $\delta > 0$ the robustness set contains assignment processes that

place a share δ of mass where g_x is arbitrarily large, the first-stage breakdown budget of part SA-37(i) equals zero, and both bounds on the estimate are infinite. Remark SA-16 records the same phenomenon in the marital college premium of Section SA6.2.1. Here we propose two alternatives to make the sensitivity exercise informative again.

Normalizing the moments of the shocks. A first solution is to leave $\mathcal{V}$ unrestricted and impose that the candidate distributions must match the second moments under $\tilde{P}$, a moment restriction of the kind the framework of Section SA1.2 accommodates. The closed forms of Proposition SA-37 do not survive the restriction, and the bounds are then computed from the dual representations of Section SA2.

Restricting the configuration space. The second route keeps the formula as it is and takes $\mathcal{V}$ to be a finite set of possible shock configurations that the design allows for. For example, for the shift-share formula of the BH23 examples above one may take $\mathcal{V}$ to be the finite set of permutations of the realized shocks $\tilde{v}$, on which the formula takes finitely many values and is therefore bounded. In the application of Section 6.3 of the main text, the permutations of line completion status of BH23 likewise define a finite $\mathcal{V}$, although there the formula is bounded on the whole of $\{0, 1\}^K$. ♣

Remark SA-18 (Inference). The bounds of this section carry no sampling uncertainty. As in the design-based setup above, the sample is held fixed and only the postulated assignment process varies. Unlike in Sections SA6.1 and SA6.2, where the baseline or the moment targets are estimated, the bounds of this section are therefore bounds for the estimate $\hat{\beta}_*$ rather than for an estimand. ♣

SA6.4 Other Applications

Our framework accommodates a considerably wider range of exercises than the RCTs with attrition of Section SA6.1, the counterfactuals in structural models of Section SA6.2, and the formula instruments of Section SA6.3. This section collects several other examples that are relevant for economic applications. We then close by discussing how the framework recovers exact nonparametric sharp bounds in all these applications, beyond sensitivity analysis. For expositional purposes, we avoid including moment conditions in the majority of these applications, but they can be incorporated in either of the two ways described in Section SA6.1.1.

Before turning to the individual applications, we present a lemma that we invoke repeatedly in what follows.

Lemma SA-38. *Let $(\mathcal{Y}, \mathcal{F})$ be a measurable space, and let Q be a probability measure on $(\mathcal{Y}, \mathcal{F})$. Then:*

- (i) *if P is a probability measure on $(\mathcal{Y}, \mathcal{F})$ such that $\text{TV}(P \parallel Q) \leq \delta \in [0, 1]$, there exist two mutually singular nonnegative measures ν^+ and ν^- with $P - Q = \nu^+ - \nu^-$, $\nu^+ \leq P$, and $\nu^- \leq Q$, together with a set $\mathcal{Y}^+ \in \mathcal{F}$ carrying ν^+ in the sense that $\nu^+(\mathcal{Y} \setminus \mathcal{Y}^+) = \nu^-(\mathcal{Y}^+) = 0$, such that*

$$\nu^+(\mathcal{Y}) = \nu^-(\mathcal{Y}) = \text{TV}(P \parallel Q) \leq \delta.$$

Moreover, the supremum defining the total variation is attained at $\mathcal{Y}^+$, that is,

$$\text{TV}(P \parallel Q) = \sup_{A \in \mathcal{F}} |P(A) - Q(A)| = |P(\mathcal{Y}^+) - Q(\mathcal{Y}^+)|;$$

- (ii) *if in addition $\mathcal{Y} = [\underline{Y}, \bar{Y}]$, then $\mathbb{E}_P[Y] \in [\underline{\mu}_Q(\delta), \bar{\mu}_Q(\delta)]$, where Y denotes the identity map on $\mathcal{Y}$, $q := F_Q^{-1}$ denotes the quantile function of Q , and*

$$\begin{aligned} \bar{\mu}_Q(\delta) &= \mathbb{E}_Q[Y] + \delta \bar{Y} - \int_0^\delta q(u) du = \delta \bar{Y} + \int_\delta^1 q(u) du, \\ \underline{\mu}_Q(\delta) &= \mathbb{E}_Q[Y] + \delta \underline{Y} - \int_{1-\delta}^1 q(u) du = \delta \underline{Y} + \int_0^{1-\delta} q(u) du. \end{aligned}$$

Moreover, both endpoints are attained in the sense that there exist probability measures P_+ and P_- on $(\mathcal{Y}, \mathcal{F})$ with $\text{TV}(P_\pm \parallel Q) \leq \delta$, $\mathbb{E}_{P_+}[Y] = \bar{\mu}_Q(\delta)$, and $\mathbb{E}_{P_-}[Y] = \underline{\mu}_Q(\delta)$;

- (iii) *if instead $\mathcal{Y} \subseteq \mathbb{R}$ and $\mathbb{E}_Q[|Y|] < \infty$, then, for any $\kappa \in (0, 1]$, the worst case means over the decomposition neighborhood $C_\kappa^{\leq}(Q)$ are*

$$\sup_{\tilde{P} \in C_\kappa^{\leq}(Q)} \mathbb{E}_{\tilde{P}}[Y] = \frac{1}{\kappa} \int_{1-\kappa}^1 q(u) du, \quad \inf_{\tilde{P} \in C_\kappa^{\leq}(Q)} \mathbb{E}_{\tilde{P}}[Y] = \frac{1}{\kappa} \int_0^\kappa q(u) du,$$

with Y and q as in part (ii). Moreover, both the supremum and the infimum are attained by distributions in $C_\kappa^{\leq}(Q)$.

[Proof]

Each part of Lemma SA-38 is put to work at a specific point of this section. Part (i) underlies the equivalence between corrupted sampling and the total variation neighborhood established in Section SA6.4.2. Part (ii) delivers the closed form sensitivity bounds of the external validity exercise of Section SA6.4.1, the missing-data bounds of Section SA6.4.3, and the closed forms of the discrete choice exercise of Section SA6.4.5. Part (iii) delivers the trimmed-mean bounds of the departures-from-randomization exercise of Section SA6.4.4.

SA6.4.1 RCT external validity

Suppose an RCT is conducted in one location, indexed by $S = 1$, and the object of interest is the average treatment effect in a different location, indexed by $S = 0$. This is a common situation in which an experiment run in one site, say a microcredit program evaluated in one country or a cash transfer piloted in selected municipalities, is used to inform policy elsewhere. Let $Y(0), Y(1) \in [\underline{Y}, \bar{Y}] =: \mathcal{Y}$ denote potential outcomes, let $D \in \{0, 1\}$ denote treatment assignment, let $Y = DY(1) + (1 - D)Y(0)$ be the observed outcome, and let P_0 denote the population distribution of $(Y(0), Y(1), D, S)$. Treatment is randomly assigned in the experimental location, that is $D \perp\!\!\!\perp (Y(0), Y(1))$ given $S = 1$, with $0 < P_0(D = 1 | S = 1) < 1$. Therefore, the source-site effect

$$\tau_1(P_0) := \mathbb{E}_{P_0}[Y(1) - Y(0) | S = 1] = \mathbb{E}_{P_0}[Y | D = 1, S = 1] - \mathbb{E}_{P_0}[Y | D = 0, S = 1]$$

is identified, whereas the target parameter is $\tau_0(P_0) = \mathbb{E}_{P_0}[Y(1) - Y(0) | S = 0]$, which need not equal $\tau_1(P_0)$ if the distribution of potential outcomes differs across locations. The standard assumption equating the two causal estimands is transportability, that is

$$(Y(0), Y(1)) | S = 0 \stackrel{d}{=} (Y(0), Y(1)) | S = 1.$$

This restriction is strong and untestable, so we treat it as a reference point and ask how departures from this assumption affect the target estimand.

Because only the marginals of $(Y(0), Y(1)) | S = 1$ are identified under random assignment and the average treatment effect depends only on those marginals, we conduct a separate sensitivity exercise for each arm. For $d \in \{0, 1\}$, let $P_{*,d}(\cdot) := P_0(Y(d) \in \cdot | S = 1)$ denote the baseline marginal distribution implied by transportability, and define

$$\mathcal{P}_{\text{EV}}^d(\delta, P_{*,d}) := \{P \in \Delta(\mathcal{Y}) : \text{TV}(P \| P_{*,d}) \leq \delta\},$$

and

$$\underline{\mu}_d(\delta) := \inf_{P \in \mathcal{P}_{\text{EV}}^d(\delta, P_{*,d})} \mathbb{E}_P[Y] = \delta \underline{Y} + \int_0^{1-\delta} q_d(u) du, \quad \bar{\mu}_d(\delta) := \sup_{P \in \mathcal{P}_{\text{EV}}^d(\delta, P_{*,d})} \mathbb{E}_P[Y] = \delta \bar{Y} + \int_\delta^1 q_d(u) du,$$

where the second equalities follow from Lemma SA-38(ii) and q_d denotes the quantile function of $P_{*,d}$. The sensitivity bounds for $\tau_0(P_0)$ are then

$$\underline{\tau}_{\text{TV}}(\delta) = \underline{\mu}_1(\delta) - \bar{\mu}_0(\delta), \quad \bar{\tau}_{\text{TV}}(\delta) = \bar{\mu}_1(\delta) - \underline{\mu}_0(\delta).$$

At $\delta = 0$ both bounds equal $\tau_1(P_0)$, whereas at $\delta = 1$ the closed forms return $\underline{\tau}_{\text{TV}}(1) = \underline{Y} - \bar{Y}$ and $\bar{\tau}_{\text{TV}}(1) = \bar{Y} - \underline{Y}$, the sharp nonparametric bounds implied by the bounded support alone (Manski, 2003). The exercise thus interpolates between exact transportability and the fully agnostic bounds, with δ measuring how much probability mass of the target site distribution differs from that of the experimental site. In Section SA6.4.7 we show that the recovery of exact nonparametric sharp bounds is not specific to this application but a feature of our framework.

SA6.4.2 Contaminated and corrupted data

Suppose we are interested in a linear functional of the distribution of a random variable $Y_1 \in \mathcal{Y}$, but we are concerned that the population we sample from might contain data errors, that is, we observe

$$Y = (1 - E)Y_0 + EY_1,$$

where $E \in \{0, 1\}$ is unobserved and indicates whether an observation is error free ($E = 1$) or not ($E = 0$). Realizations of Y corresponding to Y_0 are termed “contaminated” or “corrupted”. A leading occurrence is survey data in which a share of the responses is misreported or miscoded, as with income measures in household surveys. Let P denote the law of the observed outcome Y , let P_1 denote the marginal law of Y_1 , our object of interest, let $\tilde{P}$ denote the conditional law of Y_0 given $E = 0$, and write $\delta := \mathbb{P}(E = 0) \in (0, 1)$ for the error probability. Horowitz and Manski (1995) study this model under two sampling schemes. First, under “contaminated sampling” the occurrence of errors is independent of the error-free outcome, $E \perp\!\!\!\perp Y_1$. Second, under “corrupted sampling” the dependence between E and Y_1 is left unrestricted. By setting the baseline distribution as $P_* = P$, our framework covers both settings described above. Indeed, contaminated sampling is based on the decomposition neighborhood of Section SA2.6, whereas corrupted sampling is based on the pure total variation neighborhood.

Contaminated sampling. Under independence the distribution of Y_1 among error-free observations coincides with P_1 , so the sampling distribution is the mixture

$$P = \delta \tilde{P} + (1 - \delta)P_1.$$

Therefore, the candidate distributions are the distributions that can arise as the error-free component of the observed mixture, namely those $Q \in \Delta(\mathcal{Y})$ for which $P = \delta R + (1 - \delta)Q$ for some $R \in \Delta(\mathcal{Y})$ or, equivalently by Lemma SA-12(i) with $\kappa = 1 - \delta$, those satisfying the domination constraint $(1 - \delta)Q \leq P$. The robustness set is thus

$$\mathcal{P}_{\text{CON}}(\delta, P) := \{Q \in \Delta(\mathcal{Y}) : (1 - \delta)Q \leq P\}.$$

Sensitivity bounds can then be obtained from the dual of Theorem SA-14(i).

Corrupted sampling. Suppose now that the independence assumption is dropped, so that errors may occur selectively. Writing $P_{1|e}$ for the conditional distribution of Y_1 given $E = e$, the observed distribution and the

distribution of interest decompose as

$$P = \delta \tilde{P} + (1 - \delta)P_{1|1}, \quad P_1 = \delta P_{1|0} + (1 - \delta)P_{1|1},$$

where $P_{1|1}$ need no longer coincide with P_1 and $P_{1|0}$ is entirely unrestricted. In this case the robustness set coincides with the total variation neighborhood of Section SA2.6 centered at the baseline distribution P , namely

$$\mathcal{P}_{\text{COR}}(\delta, P) := \{Q \in \Delta(\mathcal{Y}) : \text{TV}(Q \| P) \leq \delta\}.$$

To see that P_1 must satisfy $\text{TV}(P_1 \| P) \leq \delta$, fix a measurable event A and decompose both $P_1(A)$ and $P(A)$ over $\{E = 1\}$ and $\{E = 0\}$. The contributions on $\{E = 1\}$ cancel, because conditionally on this event we get $Y = Y_1$, and so

$$P_1(A) - P(A) = \mathbb{P}(Y_1 \in A, E = 0) - \mathbb{P}(Y_0 \in A, E = 0),$$

which is a difference of two numbers lying in $[0, \delta]$, so that $|P_1(A) - P(A)| \leq \delta$. Taking the supremum over A yields $\text{TV}(P_1 \| P) \leq \delta$.

Conversely, every distribution in the total variation neighborhood arises as the distribution of Y_1 under some corrupted-sampling process with error probability δ , so the inclusion just derived above is in fact an equality. To see it, fix $Q \in \mathcal{P}_{\text{COR}}(\delta, P)$ and set $t := \text{TV}(Q \| P) \leq \delta$. In view of Lemma SA-38(i) and writing $Q - P = \nu^+ - \nu^-$ with $\nu^+ \leq Q$, $\nu^- \leq P$, and $\nu^+(\mathcal{Y}) = \nu^-(\mathcal{Y}) = t$, the two distributions share the common component $C := P - \nu^- = Q - \nu^+$, which is a nonnegative measure of mass $1 - t \geq 1 - \delta$. Let a unit be error free ($E = 1$) with probability $1 - \delta$, in which case its true and recorded outcomes coincide and are drawn from the normalized common component $C/(1 - t)$. Instead, with probability δ the unit is erroneous ($E = 0$), its recorded outcome Y_0 is drawn from $(P - \frac{1-\delta}{1-t}C)/\delta$ and its true outcome Y_1 from $(Q - \frac{1-\delta}{1-t}C)/\delta$. Both numerators are nonnegative measures of total mass δ , because $\frac{1-\delta}{1-t} \leq 1$ and C is dominated by both P and Q . Hence, after division by δ , the displayed measures are probability laws. By construction $\mathbb{P}(E = 0) = \delta$, the recorded outcome has distribution P , and the true outcome has distribution Q , as desired. When $t < \delta$, part of the erroneous units simply carry correct values, reflecting that the corruption budget need not be exhausted. Sensitivity bounds can thus be obtained via Theorem SA-11.

SA6.4.3 Missing data

Consider an outcome Y with support $\mathcal{Y} = [\underline{Y}, \bar{Y}]$ that is observed only when an indicator $S \in \{0, 1\}$ equals one – as with item nonresponse in a survey or dropout in a panel. Write $p := \mathbb{P}(S = 1)$, with $p \in (0, 1)$, and let $P_*(\cdot) := \mathbb{P}(Y \in \cdot | S = 1)$ and $R(\cdot) := \mathbb{P}(Y \in \cdot | S = 0)$ denote the outcome distributions of the observed and of the missing units, respectively. The object of interest is the linear functional of the whole population

$$\mu_{\text{MD}}(P) := \mathbb{E}_P[g(Y)], \quad P = p P_* + (1 - p) R.$$

While p and P_* are identified from the data, R , the law of Y among the missing, is not. The main difference from the sensitivity exercise we conducted in Section SA6.1 is that the baseline distribution P_* has to be a mixing component of any candidate distribution P , rather than being the mixture itself.

We can point identify $\mu_{\text{MD}}(P)$ under an MCAR-type assumption, which would imply that $R = P_*$. The sensitivity exercise would then relax this benchmark by imposing that $\text{TV}(R \| P_*) \leq \delta$. The mixture displays P as a contamination of P_* with mixing weight p and mixing distribution R , in the sense of Section SA2.6, so, in view of the rescaling identity $\text{TV}(P \| P_*) = (1 - p)\text{TV}(R \| P_*)$ of Lemma SA-17(ii), the requirement is equivalent to the pair of constraints $\text{TV}(P \| P_*) \leq (1 - p)\delta$ and $p P_* \leq P$. Therefore, the robustness set is

$$\mathcal{P}_{\text{MD}}(\delta, P_*) := \{P \in \Delta(\mathcal{Y}) : \text{TV}(P \| P_*) \leq (1 - p)\delta, p P_* \leq P\},$$

with δ bounding the fraction of the missing subpopulation that does not satisfy the missing completely at random assumption. The sensitivity bounds for $\delta \in [0, 1]$ are then defined as

$$\underline{\mu}_{\text{MD}}(\delta) := \inf_{P \in \mathcal{P}_{\text{MD}}(\delta, P_*)} \mathbb{E}_P[g(Y)], \quad \bar{\mu}_{\text{MD}}(\delta) := \sup_{P \in \mathcal{P}_{\text{MD}}(\delta, P_*)} \mathbb{E}_P[g(Y)],$$

and their duals are readily obtained from Theorem SA-14(ii). Were the support unbounded, the moment restrictions $\mathbb{E}_P[m(Y)] \in \mathcal{M}(\rho)$ would again be what keeps the bounds finite (cf. Section SA6.2 and Remark SA-16). The closest exercise in the literature is that of Ober-Reynolds (2026), who studies GMM breakdown points in the presence of missing data under Hellinger neighborhoods. The neighborhoods there are defined through densities with respect to the baseline, which restricts the candidate distributions to be absolutely continuous with respect to it, and the results require strict convexity of the entropy function, so they do not cover TV. We stress that our framework covers the Hellinger divergence.

The same setting described above can also be used to perform sensitivity analysis around a causal estimand that the randomized experiment with attrition of Section SA6.1 set aside, namely the ATE for the whole population,

$$\tau_{\text{MB}}(P_0) := \mathbb{E}_{P_0}[Y(1) - Y(0)],$$

where, as in Section SA6.1, P_0 is the distribution of the population vector $(Y(0), Y(1), S(0), S(1), D)$. Formally, we set $g(y) = y$ in the above setting and focus on the treated arm. The law of $Y(1)$ is the mixture

$$P_{Y(1)} = \pi P_{Y(1)|\text{NA}} + (1 - \pi)P_{Y(1)|\text{A}},$$

where $P_{Y(1)|\text{NA}}$ and $P_{Y(1)|\text{A}}$ denote the distribution of $Y(1)$ for the non-attriters ($S(1) = 1$) and the attriters

($S(1) = 0$), respectively, and $\pi := P_0(S(1) = 1) \in (0, 1)$ is the non-attritor share, identified as $P_0(S = 1 \mid D = 1)$ by the random assignment of D . Under **MCAR**, $P_{Y(1)|\text{NA}} = P_{Y(1)|A}$. Absent **MCAR**, the two laws need not coincide, and δ controls their distance. The baseline of this exercise is the identified outcome distribution of the non-attritors in the treated arm, $P_T = P_0(Y \in \cdot \mid D = 1, S = 1)$, which equals $P_{Y(1)|\text{NA}}$ by the random assignment of D . The robustness set is thus

$$\mathcal{P}_{\text{MB}}(\delta, P_T) := \{P \in \Delta(\mathcal{Y}) : \text{TV}(P \parallel P_T) \leq (1 - \pi)\delta, \pi P_T \leq P\}.$$

The upper sensitivity bound $\bar{\mu}_1^{\text{MB}}(\delta)$ is then the supremum of $\mathbb{E}_P[Y(1)]$ over $P \in \mathcal{P}_{\text{MB}}(\delta, P_T)$.

Applying Lemma **SA-38(ii)** to the attritors' distribution, with baseline P_T and budget $\delta \in [0, 1]$, we get

$$\bar{\mu}_1^{\text{MB}}(\delta) = \pi \mathbb{E}_{P_T}[Y] + (1 - \pi) \left(\mathbb{E}_{P_T}[Y] + \delta \bar{Y} - \int_0^\delta q(u) du \right),$$

where q denotes the quantile function of P_T . In words, the upper bound is obtained by trimming from the identified average the low δ -mass tail of P_T and relocating it to $\bar{Y}$. At $\delta = 0$ the two subpopulation distribution are forced to coincide, as they would under **MCAR**, and the display reduces to $\mathbb{E}_{P_T}[Y]$, so that $\mathbb{E}_{P_0}[Y(1)]$ is point identified. At $\delta = 1$ the attritors' distribution is unrestricted, and the display allocates their entire mass $1 - \pi$ to $\bar{Y}$, which delivers the worst case bounds of Manski (2003). In contrast with the Lee bounds case, the mixture constraint is reversed, so that $P_T \ll P$ but not conversely, and the elements of $\mathcal{P}_{\text{MB}}(\delta, P_T)$ need not be dominated by P_T , so that the relocated mass may be singular with respect to P_T .

The same construction applies to $Y(0)$ upon replacing π with $P_0(S(0) = 1)$ and P_T with $P_0(Y \in \cdot \mid D = 0, S = 1)$, and the lower bounds $\underline{\mu}_d^{\text{MB}}(\delta)$ follow from the symmetric closed form $\underline{\mu}_Q(\delta)$ of Lemma **SA-38(ii)**, so that

$$\tau_{\text{MB}}(P_0) \in \left[\underline{\mu}_1^{\text{MB}}(\delta) - \bar{\mu}_0^{\text{MB}}(\delta), \bar{\mu}_1^{\text{MB}}(\delta) - \underline{\mu}_0^{\text{MB}}(\delta) \right].$$

The interval is sharp because no restriction links the law of $Y(1)$ on $\{S(1) = 0\}$ to that of $Y(0)$ on $\{S(0) = 0\}$.

SA6.4.4 Departures from randomization

Consider an observational study in which treatment effects are identified by assuming that assignment is as good as random, that is $D \perp\!\!\!\perp (Y(0), Y(1))$, where $Y(d) \in [\underline{Y}, \bar{Y}] =: \mathcal{Y}$ almost surely. In some situations the assumption may fail for part of the population. For example, in a job-training evaluation, some individuals may select into the program on unobserved motivation, while for the remainder enrollment is as good as random. Suppose that the distribution P_0 from which the data are sampled is the mixture

$$P_0 = \delta P_C + (1 - \delta) P_U,$$

where P_U is a distribution in which assignment is random, P_C describes the “confounded” population in which $D \not\perp\!\!\!\perp Y(d)$, and δ is the proportion of confounded units. The causal estimand of interest is

$$\tau_U(P_U) := \mathbb{E}_{P_U}[Y(1) - Y(0)],$$

the average treatment effect on the unconfounded subpopulation.

Let $\varpi_d := P_0(D = d)$, $d \in \{0, 1\}$, and $\pi := P_U(D = 1)$. Among treated units, a fraction $\alpha_1 := (1 - \delta)\pi/\varpi_1$ comes from the unconfounded component, and similarly a fraction $\alpha_0 := (1 - \delta)(1 - \pi)/\varpi_0$ among untreated units, so the observed outcome distributions $P_{*,d}(\cdot) := P_0(Y \in \cdot \mid D = d)$ are the mixtures

$$P_{*,d} = \alpha_d P_{U,d} + (1 - \alpha_d) P_{C,d}, \quad d \in \{0, 1\},$$

with $P_{U,d}$ and $P_{C,d}$ the conditional laws of Y given $D = d$ under P_U and P_C , respectively. The baseline distribution of the exercise is thus $P_{*,d}$, which is identified from the data and plays the role of the mixture as in the RCT with attrition example of Section **SA6.1**.

The candidate distributions are the possible unconfounded components $P_{U,d}$, which satisfy the domination constraint $\alpha_d P_{U,d} \leq P_{*,d}$ by Lemma **SA-12(i)**. The domination constraint already implies $\text{TV}(P_{U,d} \parallel P_{*,d}) \leq 1 - \alpha_d$, in view of the rescaling identity of Lemma **SA-17(i)**, so the robustness set of this exercise is, for $d \in \{0, 1\}$, the set $\mathcal{P}_{\text{DR}}^d(1 - \alpha_d, P_{*,d})$, where, for $\delta \in [0, 1]$,

$$\mathcal{P}_{\text{DR}}^d(\delta, P_{*,d}) := \{P \in \Delta(\mathcal{Y}) : P_{*,d} = (1 - \delta)P + \delta R \text{ for some } R \in \Delta(\mathcal{Y})\}.$$

We can obtain sensitivity bounds by means of the dual in Theorem **SA-14(i)**, applied arm by arm. One caveat is in order: the data identify ϖ_d but not π , so the weights α_d are not identified. We sidestep this issue by imposing equal treatment rates across the two components, $P_C(D = 1) = \pi$, which yields $\pi = \varpi_1$ and $\alpha_d = 1 - \delta$ in both arms, so that the robustness set of each arm becomes $\mathcal{P}_{\text{DR}}^d(\delta, P_{*,d})$. Alternatively, one could profile the bounds over the values of π for which $\alpha_0, \alpha_1 \in [0, 1]$.

Since assignment is random under P_U , we have $\mathbb{E}_{P_U}[Y(d)] = \mathbb{E}_{P_{U,d}}[Y]$ for $d \in \{0, 1\}$, so it suffices to bound each arm mean as $P_{U,d}$ ranges over the robustness set $\mathcal{P}_{\text{DR}}^d(\delta, P_{*,d})$ defined above. Fix $\delta \in [0, 1)$ and let q_d denote the quantile function of $P_{*,d}$. In view of Lemma **SA-38(iii)**, applied with $\kappa = 1 - \delta$ and baseline $P_{*,d}$, the sensitivity bounds for

the mean of each arm $d \in \{0, 1\}$ are

$$\underline{\mu}_d^U(\delta) := \inf_{P \in \mathcal{P}_{\text{DR}}^d(\delta, P_{*,d})} \mathbb{E}_P[Y] = \frac{1}{1-\delta} \int_0^{1-\delta} q_d(u) du, \quad \bar{\mu}_d^U(\delta) := \sup_{P \in \mathcal{P}_{\text{DR}}^d(\delta, P_{*,d})} \mathbb{E}_P[Y] = \frac{1}{1-\delta} \int_\delta^1 q_d(u) du,$$

the averages of Y over the lower and upper $(1-\delta)$ -fraction of each arm. Combining the two arms then yields

$$\underline{\mu}_1^U(\delta) - \bar{\mu}_0^U(\delta) \leq \tau_U(P_U) \leq \bar{\mu}_1^U(\delta) - \underline{\mu}_0^U(\delta).$$

In this application, δ thus traces out the interpolation between point identification and worst case selection and bounds the fraction of the population that does not satisfy unconfoundedness. At $\delta = 0$ the trimmed means reduce to the arm means and both bounds collapse to the difference in means of the two arms, whereas as $\delta \rightarrow 1$ each arm is trimmed down to its most extreme $(1-\delta)$ -fraction, so that $\bar{\mu}_d^U(\delta)$ and $\underline{\mu}_d^U(\delta)$ converge to the essential supremum and the essential infimum of the observed distribution $P_{*,d}$.

The exercise accommodates the more general case of unconfoundedness or selection on observables, which relaxes randomization as

$$D \perp\!\!\!\perp (Y(0), Y(1)) \mid X.$$

It suffices to posit the mixture decomposition, with the same confounded share δ , within each covariate cell and repeat the argument conditionally on $X = x$, which bounds $\mathbb{E}_{P_U}[Y(d) \mid X = x]$. The robustness set of this extension is, within the covariate cell $X = x$, the set $\mathcal{P}_{\text{DR}}^d(\delta, P_{*,d}(x))$, where $P_{*,d}(x) := P_0(Y \in \cdot \mid D = d, X = x)$ denotes the observed outcome distribution of arm d within the cell. Since a constant share implies that X has the same distribution under P_U and P_0 , averaging these bounds over the observed distribution of X then bounds the average treatment effect for the unconfounded subpopulation, in the spirit of Bonvini and Kennedy (2022). We stress that our target parameter differs from theirs, as they bound the average treatment effect for the entire population, which corresponds to P_0 in our notation, whereas we target the unconfounded subpopulation P_U .

SA6.4.5 Discrete choice models

The linearity of the estimand $\mathbb{E}_P[g(Z; \theta)]$ in P is what makes the dualization behind Theorem SA-9 possible. The estimands of interest in discrete choice models do not possess this structure. In this section, we explain why discrete choice models cannot, in general, be accommodated within our framework and propose three alternatives. For simplicity, we illustrate the argument on the canonical binary logit, but the same reasoning extends to multinomial logit, probit, and ordered choice.

Let $X \in \mathbb{R}^k$ be a vector of regressors and $Y \in \{0, 1\}$ a binary outcome, whose conditional choice probability we assume to be $\mathbb{P}(Y = 1 \mid X = x) = \Lambda(x^\top \beta)$, where $\Lambda(t) := (1 + e^{-t})^{-1}$ is the logistic link function. The population analog of the MLE is the solution $\beta(P)$ of the nonlinear system of moment equations

$$\mathbb{E}_P[(Y - \Lambda(X^\top \beta(P)))X] = 0. \quad (60)$$

The functional $\beta(P)$ is not of the form $\mathbb{E}_P[f(Z)]$ for any known f . In this particular case the link is nonlinear in β , so (60) cannot be solved in closed form. The coefficient $\beta(P)$ is, however, seldom the target of applied work. Researchers are typically interested, for example, in the average marginal effect

$$\text{AME}_j(P) := \mathbb{E}_{P_X}[\lambda(X^\top \beta(P))\beta_j(P)], \quad \beta_j(P) = e_j^\top \beta(P),$$

where $\lambda(t) := \Lambda'(t) = \Lambda(t)(1 - \Lambda(t))$. Unlike in a linear model, where fixing the regressor marginal renders the estimand an expectation under the conditional law of the error, conditioning on X does not restore linearity here. Although the framework does not apply to $\beta(P)$ directly, we propose three ways in which some version of the sensitivity analysis exercise can still be performed in a discrete choice model.

Transportability of P_X . Fix β at the population value $\beta(P_0)$ solving (60) and the link to be the logit one, and let the covariate distribution vary. The AME reduces to the linear functional

$$\text{AME}_j(P_X) = \beta_j(P_0) \mathbb{E}_{P_X}[\lambda(X^\top \beta(P_0))],$$

which is bounded because $\lambda \in [0, 1/4]$. In this case, sensitivity to contamination or corruption in P_X can be readily assessed following Section SA6.4.2. Under corrupted sampling the robustness set is $\mathcal{P}_{\text{COR}}(\delta, P_X)$, whereas under contaminated sampling it is $\mathcal{P}_{\text{CON}}(\delta, P_X)$, the two sets of Section SA6.4.2 evaluated at baseline P_X .

Relaxation of the logit link. This route holds the covariate marginal P_X fixed and perturbs the conditional distribution of the outcome, retaining the logit specification as the center of the robustness set. Because Y is binary, for any measurable conditional choice probability $P : \mathbb{R}^k \rightarrow [0, 1]$ let $P_X \otimes \text{Be}(P)$ denote the joint law under which $X \sim P_X$ and Y given $X = x$ is distributed as the Bernoulli law $\text{Be}(P(x))$. Writing $P_*(x) := \Lambda(x^\top \beta_0)$ for the logit specification, with $\beta_0 := \beta(P_0)$ solving (60) at the law P_0 from which the data are sampled, so that $\text{Be}(P_*(x))$ is the implied baseline conditional, the robustness set for each covariate cell x reads

$$\mathcal{P}_{\text{DC}}(\delta, \text{Be}(P_*(x))) := \{Q \in \Delta(\{0, 1\}) : \text{TV}(Q \parallel \text{Be}(P_*(x))) \leq \delta\} \quad \text{for } P_X\text{-almost every } x,$$

which contains all conditional distributions that are at most δ away from the logit specification in total variation.

The within-cell sets do not restrict the derivatives of P , so derivative-based functionals such as the AME_j admit no informative bounds and the natural targets are averaged discrete contrasts. For a binary regressor of interest $X_j \in \{0, 1\}$, write $X = (X_j, X_{-j})$ and $x = (x_j, x_{-j})$, where X_{-j} collects the remaining covariates, with marginal

$P_{X_{-j}}$ fixed by P_X , and, with a slight abuse of notation, let $P(t, x_{-j})$ denote the value of P when the regressor of interest equals $t \in \{0, 1\}$. The target estimand is the average effect of switching X_j from 0 to 1,

$$\tau_{\text{DC}}(P) := \mathbb{E}_{P_{X_{-j}}} [P(1, X_{-j}) - P(0, X_{-j})],$$

which is linear in P . Sensitivity bounds can be obtained cell by cell. Since $\text{TV}(Q \parallel \text{Be}(p)) = |Q(\{1\}) - p|$ for every $Q \in \Delta(\{0, 1\})$ and $p \in [0, 1]$, the choice probabilities $Q(\{1\})$ of the distributions $Q \in \mathcal{P}_{\text{DC}}(\delta, \text{Be}(P_*(x)))$ form exactly the interval $[\underline{P}(x), \overline{P}(x)]$, where $\underline{P}(x) := \max\{P_*(x) - \delta, 0\}$ and $\overline{P}(x) := \min\{P_*(x) + \delta, 1\}$. Aggregating over the fixed marginal $P_{X_{-j}}$ yields

$$\begin{aligned} \tau_{\text{DC}}(\delta) &:= \mathbb{E}_{P_{X_{-j}}} [\underline{P}(1, X_{-j})] - \mathbb{E}_{P_{X_{-j}}} [\overline{P}(0, X_{-j})], \\ \bar{\tau}_{\text{DC}}(\delta) &:= \mathbb{E}_{P_{X_{-j}}} [\overline{P}(1, X_{-j})] - \mathbb{E}_{P_{X_{-j}}} [\underline{P}(0, X_{-j})]. \end{aligned}$$

At $\delta = 0$ the interval collapses to the logit implied contrast $\tau_{\text{DC}}(P_*)$, whereas at $\delta = 1$ it widens to the trivial bounds $[-1, 1]$.

An alternative robustness set is worth recording. It replaces the within-cell restriction with its integrated version, $\mathbb{E}_{P_X}[|P(X) - P_*(X)|] \leq \delta$. Since two joint distributions sharing a marginal have total variation equal to the average of their conditional total variations (see, e.g., Polyanskiy and Wu, 2025, Proposition 7.2), this restriction is equivalent to

$\mathcal{P}_{\text{DC}}^{\text{int}}(\delta, P_*) := \{P_X \otimes \text{Be}(P') \in \Delta(\mathbb{R}^k \times \{0, 1\}) : P' : \mathbb{R}^k \rightarrow [0, 1] \text{ measurable, } \text{TV}(P_X \otimes \text{Be}(P') \parallel P_X \otimes \text{Be}(P_*)) \leq \delta\}$, so that the budget δ is allocated across covariate cells rather than imposed cell by cell. Provided the propensity score $\pi(x_{-j}) := \mathbb{P}(X_j = 1 \mid X_{-j} = x_{-j})$ lies in $(0, 1)$ for $P_{X_{-j}}$ -almost every x_{-j} , the law of iterated expectations gives $\tau_{\text{DC}}(P) = \mathbb{E}_{P_X \otimes \text{Be}(P)}[\omega(X)Y]$ for the inverse-probability weight $\omega(x) := x_j/\pi(x_{-j}) - (1 - x_j)/(1 - \pi(x_{-j}))$, so that, when X takes finitely many values, bounds are delivered by the dual of Theorem SA-11 applied with baseline $P_X \otimes \text{Be}(P_*)$ and with the fixed covariate marginal imposed through one moment equality per covariate cell.

Unobserved heterogeneity. Keep the logit specification at the individual level and augment the latent utility index with a unit-level intercept

$$Y = \mathbf{1}\{X^\top \beta + A + \varepsilon > 0\}$$

with $\varepsilon \sim \text{Logistic}(0, 1)$ independent of (X, A) and $A \sim Q \in \Delta(\mathbb{R})$ independent of X . The average marginal effect is thus

$$\text{AME}_j(Q) = \beta_j \int \psi(a) dQ(a), \quad \psi(a) := \mathbb{E}_{P_X}[\lambda(X^\top \beta + a)],$$

which is bounded since $\lambda \leq 1/4$ and linear in Q . We center the robustness set at a Gaussian benchmark, $Q_* := \mathcal{N}(\mu_0, \sigma_0^2)$ for given $\mu_0 \in \mathbb{R}$ and $\sigma_0 > 0$, a classic random effects logit specification, and take as robustness set the total variation ball

$$\mathcal{P}_{\text{DC}}^{\text{het}}(\delta, Q_*) := \{Q \in \Delta(\mathbb{R}) : \text{TV}(Q \parallel Q_*) \leq \delta\},$$

the set of heterogeneity distributions deviating from normality by at most δ in total variation. Let q_ψ denote the quantile function of $\psi(A)$ when $A \sim Q_*$, and set

$$\underline{\psi} := \inf_{a \in \mathbb{R}} \psi(a), \quad \bar{\psi} := \sup_{a \in \mathbb{R}} \psi(a).$$

The lower and upper values of the average derivative are

$$\begin{aligned} \underline{\mu}_\psi(\delta) &:= \inf_{Q \in \mathcal{P}_{\text{DC}}^{\text{het}}(\delta, Q_*)} \int \psi(a) dQ(a) = \delta \underline{\psi} + \int_0^{1-\delta} q_\psi(u) du, \\ \bar{\mu}_\psi(\delta) &:= \sup_{Q \in \mathcal{P}_{\text{DC}}^{\text{het}}(\delta, Q_*)} \int \psi(a) dQ(a) = \delta \bar{\psi} + \int_\delta^1 q_\psi(u) du. \end{aligned}$$

These closed forms follow from Lemma SA-38(ii) applied to the law of $\psi(A)$ under Q_* . Therefore, the sensitivity bounds on the average marginal effect are

$$\begin{aligned} \underline{\text{AME}}_j(\delta) &:= \inf_{Q \in \mathcal{P}_{\text{DC}}^{\text{het}}(\delta, Q_*)} \text{AME}_j(Q) = \min\{\beta_j \underline{\mu}_\psi(\delta), \beta_j \bar{\mu}_\psi(\delta)\}, \\ \bar{\text{AME}}_j(\delta) &:= \sup_{Q \in \mathcal{P}_{\text{DC}}^{\text{het}}(\delta, Q_*)} \text{AME}_j(Q) = \max\{\beta_j \underline{\mu}_\psi(\delta), \beta_j \bar{\mu}_\psi(\delta)\}. \end{aligned}$$

The upper value $\bar{\mu}_\psi(\delta)$ is attained whenever there exists $a_+ \in \mathbb{R}$ with $\psi(a_+) = \bar{\psi}$, and the lower value $\underline{\mu}_\psi(\delta)$ is attained whenever there exists $a_- \in \mathbb{R}$ with $\psi(a_-) = \underline{\psi}$. By appropriately changing the baseline distribution, we can also study sensitivity to departures from a no heterogeneity benchmark. In particular, we can choose the baseline distribution to be $\delta_{a_0}^{\text{Dirac}}$, the Dirac measure at some $a_0 \in \mathbb{R}$, so that the robustness set is

$$\mathcal{P}_{\text{DC}}^{\text{het}}(\delta, \delta_{a_0}^{\text{Dirac}}) = \{Q \in \Delta(\mathbb{R}) : \text{TV}(Q \parallel \delta_{a_0}^{\text{Dirac}}) \leq \delta\}.$$

Finally, we can also allow A to depend on X by constructing robustness sets that vary the conditional distribution of A given X while keeping the identified marginal of X fixed.

SA6.4.6 Just-identified method of moments

Let $Z \sim P_0 \in \Delta(\mathcal{Z})$ denote the law from which the data are drawn, let $\Theta \subseteq \mathbb{R}^{d_\theta}$, and let $m : \mathcal{Z} \times \Theta \rightarrow \mathbb{R}^{d_m}$ be a moment function with as many elements as parameters ($d_m = d_\theta$) that defines the moment condition

$$\mathbb{E}_{P_0}[m(Z; \theta)] = 0. \quad (61)$$

We assume that the moment condition in (61) has a unique solution $\theta \in \Theta$. This assumption is standard in much of the literature on generalized method of moments estimators (see, e.g., Newey and McFadden, 1994, Section 2.2.3). The formulation covers just-identified moment estimators and the first-order conditions of M -estimators. We consider the case in which $d_m > d_\theta$ at the end of the section.

Suppose that an econometrician has as target estimand $\theta_j = e_j^\top \theta$ and is concerned that the sampling distribution P_0 differs from the population of interest P_1 because a share δ of observations is erroneous. Under contaminated sampling,

$$P_0 = \delta \tilde{P} + (1 - \delta)P_1,$$

where $\tilde{P}$ is the distribution of the erroneous observations, whereas under corrupted sampling P_1 need not coincide with the error-free mixture component. As established in Section SA6.4.2, the two sampling schemes instead imply, respectively, the domination restriction $(1 - \delta)P_1 \leq P_0$ and the total variation restriction $\text{TV}(P_1 \| P_0) \leq \delta$. In both settings, the baseline distribution is $P_* = P_0$, whereas the distribution of interest is P_1 . Accordingly, the robustness set can be defined as either

$$\mathcal{P}_{\text{MM}}^{\text{con}}(\theta; \delta, P_*) := \{P \in \Delta(\mathcal{Z}) : (1 - \delta)P \leq P_*, \mathbb{E}_P[m(Z; \theta)] = 0\}$$

in the case of contamination, or, in the case of corruption, as

$$\mathcal{P}_{\text{MM}}^{\text{cor}}(\theta; \delta, P_*) := \{P \in \Delta(\mathcal{Z}) : \text{TV}(P \| P_*) \leq \delta, \mathbb{E}_P[m(Z; \theta)] = 0\}.$$

Sensitivity bounds can then be obtained by optimizing over the parameter values that are consistent with at least one distribution in the robustness set, that is,

$$\bar{\theta}_j(\delta) = \sup_{\theta \in \Theta_{\text{MM}}(\delta)} e_j^\top \theta, \quad \underline{\theta}_j(\delta) = \inf_{\theta \in \Theta_{\text{MM}}(\delta)} e_j^\top \theta, \quad \Theta_{\text{MM}}(\delta) := \{\theta \in \Theta : \mathcal{P}_{\text{MM}}^\bullet(\theta; \delta, P_*) \neq \emptyset\},$$

where $\bullet \in \{\text{con}, \text{cor}\}$.

If instead the model is over-identified, i.e., $d_m > d_\theta$, there might not exist a $\theta \in \Theta$ for which the sample analog $\mathbb{E}_{\hat{P}_n}[m(Z; \theta)] = 0$ of (61) holds, even when (61) itself has a solution. In this case, we refer the reader to the two solutions proposed to circumvent this issue in Section 3.3 of Christensen and Connault (2023).

SA6.4.7 Nonparametric sharp bounds

A recurring feature of the applications of this section is the behavior of the bounds at the endpoint $\delta = 1$. For this value of the sensitivity parameter, the total variation constraint becomes vacuous, and the sensitivity bounds coincide with the sharp nonparametric bounds implied by whatever support and moment restrictions are maintained. So far this coincidence was verified case by case. The next result, together with the discussion that follows it, shows that this is a general feature of our framework. In particular, we show that it holds for every entropy function that is finite at the endpoints of its effective domain, including the induced entropies of Proposition SA-13, for which the sharp bounds are those implied by the maintained mixture restriction. Throughout, $\mathcal{P}_\phi(\theta; \rho, P_*, \delta)$ denotes the robustness set of (2) with entropy ϕ , and $\mathcal{E}(\theta; \rho) := \{P \in \mathcal{P}(\mathcal{Z}) : \mathbb{E}_P[m(Z; \theta)] \in \mathcal{M}(\rho)\}$ denotes the collection of distributions compatible with the moment constraints at θ , as in Section SA2.6.

Theorem SA-39 (Exact nonparametric sharp bounds). *Let $P_* \in \Delta(\mathcal{Z})$, and let ϕ be an entropy function in the sense of Definition SA-4 whose effective domain has endpoints $\underline{x} := \inf \text{dom}(\phi)$ and $\bar{x} := \sup \text{dom}(\phi)$, with $\underline{x} < \bar{x}$. Assume that $\phi(\underline{x}) < \infty$, that $\phi(\bar{x}) < \infty$ if $\bar{x} < \infty$, and that $\phi^\infty(1) < \infty$ if $\bar{x} = \infty$. Define the domination set and the threshold*

$$\mathcal{D}_\phi(P_*) := \{P \in \Delta(\mathcal{Z}) : \underline{x} P_* \leq P \leq \bar{x} P_*\}, \quad \bar{\delta}_\phi := \frac{(\bar{x} - 1)\phi(\underline{x}) + (1 - \underline{x})\phi(\bar{x})}{\bar{x} - \underline{x}},$$

if $\bar{x} < \infty$, and

$$\mathcal{D}_\phi(P_*) := \{P \in \Delta(\mathcal{Z}) : \underline{x} P_* \leq P\}, \quad \bar{\delta}_\phi := \phi(\underline{x}) + (1 - \underline{x})\phi^\infty(1),$$

if $\bar{x} = \infty$. Then, for every $\delta \geq \bar{\delta}_\phi$ and every $\theta \in \Theta$,

$$\mathcal{P}_\phi(\theta; \rho, P_*, \delta) = \mathcal{D}_\phi(P_*) \cap \mathcal{E}(\theta; \rho),$$

and consequently the sensitivity bounds of (4) satisfy

$$\begin{aligned} \inf_{\theta \in \Theta} \inf_{P \in \mathcal{P}_\phi(\theta; \rho, P_*, \delta)} \mathbb{E}_P[g(Z; \theta)] &= \inf_{\theta \in \Theta} \inf_{P \in \mathcal{D}_\phi(P_*) \cap \mathcal{E}(\theta; \rho)} \mathbb{E}_P[g(Z; \theta)], \\ \sup_{\theta \in \Theta} \sup_{P \in \mathcal{P}_\phi(\theta; \rho, P_*, \delta)} \mathbb{E}_P[g(Z; \theta)] &= \sup_{\theta \in \Theta} \sup_{P \in \mathcal{D}_\phi(P_*) \cap \mathcal{E}(\theta; \rho)} \mathbb{E}_P[g(Z; \theta)]. \end{aligned}$$

[Proof]

Three cases deserve to be singled out. First, when $\text{dom}(\phi) = [0, \infty)$ the domination set is the whole of $\Delta(\mathcal{Z})$, the

threshold is $\bar{\delta}_\phi = \phi(0) + \phi^\infty(1)$, and the theorem delivers exactly the nonparametric sharp bounds implied by the moment constraints. In terms of Table SA-2, this covers the total variation entropy, for which $\phi_{\text{TV}}(0) = \phi_{\text{TV}}^\infty(1) = 1/2$ and thus $\bar{\delta}_{\phi_{\text{TV}}} = 1$, as well as the squared Hellinger entropy ($\bar{\delta}_\phi = 2$) and the Jensen–Shannon entropy ($\bar{\delta}_\phi = 2 \log 2$). It excludes the Kullback–Leibler and Pearson χ^2 entropies, for which $\phi^\infty(1) = +\infty$, and the likelihood and Neyman χ^2 entropies, for which $\phi(0) = +\infty$. Second, for the upper-truncated entropy $\bar{\phi} = \phi + \iota_{[0, 1/\kappa]}$ of Proposition SA-13(i) we have $\underline{x} = 0$ and $\bar{x} = 1/\kappa$, so that the domination set is the baseline-as-mixture restriction $\{P \in \Delta(\mathcal{Z}) : \kappa P \leq P_*\}$ and $\bar{\delta}_{\bar{\phi}} = (1 - \kappa)\phi(0) + \kappa\phi(1/\kappa)$, provided $\phi(0)$ and $\phi(1/\kappa)$ are finite. Third, for the lower-truncated entropy $\underline{\phi} = \phi + \iota_{[\kappa, \infty)}$ of Proposition SA-13(ii) we have $\underline{x} = \kappa$ and $\bar{x} = \infty$, so that the domination set is the baseline-as-mixing-component restriction $\{P \in \Delta(\mathcal{Z}) : \kappa P_* \leq P\}$ and $\bar{\delta}_{\underline{\phi}} = \phi(\kappa) + (1 - \kappa)\phi^\infty(1)$, provided $\phi^\infty(1)$ is finite. In the last two cases the theorem delivers the sharp bounds implied by the mixture restriction together with the moment constraints.

A similar result was derived in Theorem 2.1 in Christensen and Connault (2023), which shows that divergence-based sensitivity bounds *approximate* the nonparametric sharp bounds as the sensitivity parameter diverges. However, three differences are worth highlighting. First, the assumptions of CC23’s Theorem 2.1 do not allow for $\phi = \phi_{\text{TV}}$. Second, the entropy functions CC23 consider are superlinear, that is, $\phi^\infty(1) = +\infty$, so the associated divergences have unbounded range and no finite upper bound for the budget exists. Therefore, the sharp bounds are recovered only in the limit as $\delta \rightarrow +\infty$. On the contrary, our result shows that the framework presented here allows us to construct the *exact* nonparametric sharp bounds. Third, CC23 recover the nonparametric sharp bounds when only distributions that are absolutely continuous with respect to the baseline are considered. This is worrisome, for example, when the latter has a positive Lebesgue density, because distributions with mass points are automatically discarded. Our nonparametric sharp bounds are instead constructed by considering all these distributions as well. We recall that our Theorem SA-15 gives sufficient conditions for when ignoring such distributions is allowed.

Finally, in some applications the nonparametric sharp bounds have known closed forms, such as the Lee bounds, to which the trimming bounds of Theorem SA-34 reduce at $\delta = 1$, and the Manski bounds of Sections SA6.4.1 and SA6.4.3, whereas in others, such as the marital college premium of Section 6.2 of the main text, they do not. Theorem SA-39 delivers them in either case.

SA7 Additional Empirical Applications

This section collects two empirical exercises that take the attrition framework of Section SA6.1 to the data and complement the one in Section 6.1 of the main text, which revisits the online tutoring program of Carlana and La Ferrara (2025), a randomized experiment in which attrition is moderate in level but strongly differential across arms. Differential attrition is ultimately what threatens a design that aims at identifying the ATE for the always observed. Heuristically, when the two arms respond at nearly the same rate, the complier share among treated respondents, π , of Section SA6.1, is close to zero, there are few compliers “contaminating” the treatment group, and their impact on the causal estimand is limited. When treatment assignment itself induces response, π is large, and compliers can substantially move the estimand of interest. Section SA7.1 revisits the large experiment of Christensen et al. (2026), which randomized price discounts for transport services and sits at the opposite end of the spectrum relative to the tutoring program: attrition is severe in level yet similar across arms, so that the sensitivity paths widen slowly. Section SA7.2 runs the same sensitivity exercise on every eligible trial in a public collection of clinical randomized experiments and asks how often attrition threatens the sign of the estimated effect.

SA7.1 RCT with Attrition I: Christensen, Nino and Osman (2026)

Setup. The experiment recruited 1,373 existing Uber riders in Cairo, Egypt, and randomized them into a 25% price reduction (457 participants), a 50% price reduction (461), or standard prices (455), applied directly in the app for twelve weeks. The outcome we consider is the headline measure of total mobility of Christensen et al. (2026): kilometers traveled per week, recorded by the Google Maps Timeline application and collected in four follow up phone surveys. For this particular outcome variable, the authors report intention-to-treat effects on the inverse hyperbolic sine of weekly kilometers of 0.10 (0.10) for the 25% reduction and 0.40 (0.09) for the 50% reduction, against a control mean of 205 kilometers per week, in column (1) of their Table 2. Attrition operates at the level of the individual survey round, as we report in Table SA-3: response rates in the control group fall from 82% and 78% in the first two rounds to 40% and 38% in the last two. Treatment assignment plausibly induces differential attrition, as subsidized riders remain engaged with the app, and the authors themselves report the bounds of Lee (2009) in their Table B11.

Table SA-3: Attrition by survey round in the Uber price-reduction experiment.

	Control ($n = 455$)	25% reduction ($n = 457$)	50% reduction ($n = 461$)
Round 1	18.0%	18.8%	15.8%
Round 2	22.2%	17.5%	14.1%
Round 3	59.8%	56.0%	53.8%
Round 4	62.2%	60.2%	54.0%
All rounds	40.5%	38.1%	34.4%

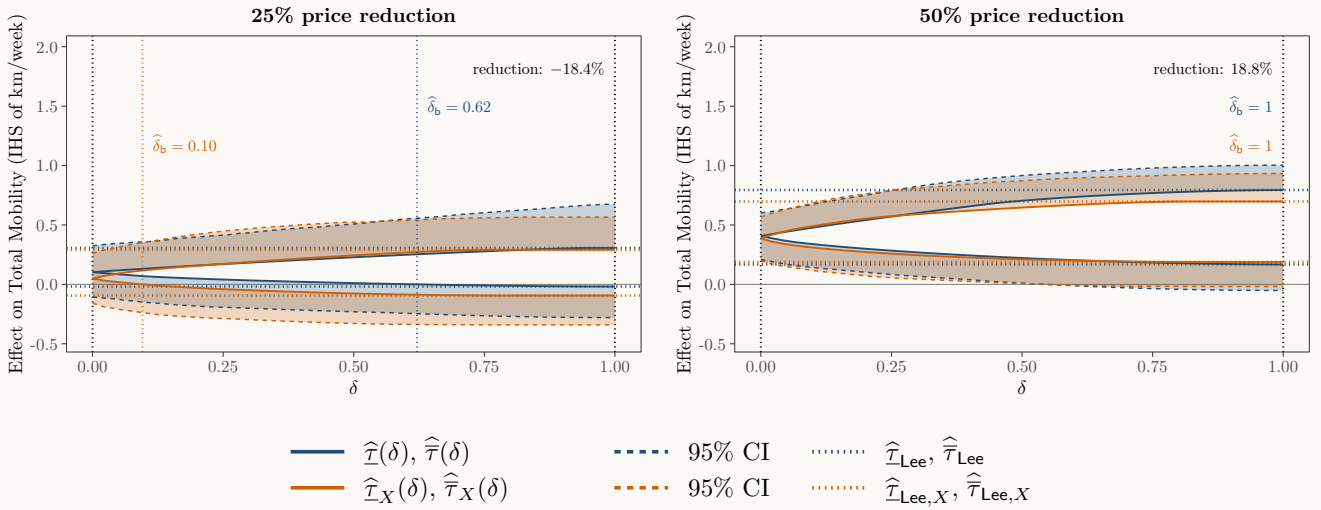
Notes: Computed from the replication data of Christensen et al. (2026). Columns are the three randomized arms, with n the number of participants assigned to each. Cells are the share of that arm’s participants whose weekly-kilometers outcome is unobserved.

We take the independent sampling unit to be the participant, of whom there are 1,373, and treat the four survey rounds as repeated measurements within each participant. The sensitivity bounds pool the resulting participant–round observations, while inference resamples participants and thereby preserves dependence across rounds within each participant. We let Y be the inverse hyperbolic sine of weekly kilometers, S the indicator that the participant responded in the round in question, and D the treatment indicator of the contrast analyzed, namely each price reduction (25% or 50%) against control (no reduction). The baseline P_T is estimated by the empirical distribution of the observed treated units, while $\mu^{\text{AO}}(0)$ is identified by (43) and estimated by the mean among the observed controls. The estimated complier shares are $\hat{\pi} = 0.039$ for the 25% arm and 0.093 for the 50% arm. Both estimates are an order of magnitude smaller than the complier shares of the tutoring program, which are 0.47 and 0.22 in its two waves.

Sensitivity bounds. Figure SA-1 traces $(\hat{\tau}(\delta), \hat{\tau}(\delta))$ and $(\hat{\tau}_X(\delta), \hat{\tau}_X(\delta))$ over a grid of 51 values of δ , separately for the two comparisons. We stress two features of the analysis without covariates. First, the sensitivity paths originate at the sample analog of $\tau_{\text{MCAR}}(P_0)$, which is 0.103 for the 25% arm and 0.407 for the 50% arm, and end at $\delta = 1$ on the Lee bounds, namely $[-0.018, 0.307]$ and $[0.168, 0.794]$, respectively. Second, the small complier shares make the paths fan out slowly, as the breakdown budgets reflect. Indeed, $\hat{\tau}(\delta)$ stays above zero at every budget the exercise considers in the 50% arm, so that $\hat{\delta}_b$ attains its largest possible value, $\hat{\delta}_b = 1$. In the 25% arm the estimated lower bound reaches zero at $\hat{\delta}_b = 0.62$. When sampling uncertainty is taken into account, the 95% confidence interval of $[\hat{\tau}(\delta), \hat{\tau}(\delta)]$ contains zero from $\delta = 0.54$ onwards in the 50% arm and already at $\delta = 0$ in the 25% arm, in line with the original point estimates in Table 2 of Christensen et al. (2026).

We now comment on the sensitivity exercise which includes covariates. Formally, we impose the aggregate sensitivity constraint across covariate strata as in (50) and compute the joint bounds through the program in (55). We take X to be the full interaction of an indicator for above median baseline travel distance and the two randomization stratifiers, gender and an indicator for whether the participant was actively looking for a job. We retain a cell only if it contains at least five treated and five control respondents, and all eight cells satisfy this rule. The control respondents keep the plug-in $\hat{\pi}(x)$ strictly below one, as Section SA6.1.1 maintains for $\pi(x)$.³ We recall that the covariate-augmented paths originate at the sample analog of $\tau_{\text{MAR}}(P_0)$ rather than at the sample analog of $\tau_{\text{MCAR}}(P_0)$. The introduction of covariates narrows the bounds at $\delta = 1$ by 18.8% in the 50% arm, but widens them by 18.4% in the 25% arm. Conditioning leaves the breakdown budget of the 50% arm at $\hat{\delta}_b = 1$, but cuts that of the 25% arm from 0.62 to 0.10. Covariates thus need not raise the breakdown budget, because they also move the origin of the path. Here the covariate-augmented path of that arm starts closer to zero, at 0.050 against 0.103, and its bounds are in fact wider than the unconditional ones at every $\delta > 0$, so that no tightening is available to offset the shift in the origin. Finally, we stress that differences between the estimated sensitivity bounds can also stem from differences in the underlying samples, as the covariates are not observed for every unit for which Y and D are observed.

Fig. SA-1: TV sensitivity bounds for the effect of Uber price reductions on total mobility.



Notes: Computed from the replication data of Christensen et al. (2026). Each panel plots the estimated $\hat{\tau}(\delta)$ and $\hat{\tau}_X(\delta)$ of (47) (solid blue lines) over $\delta \in \{0, 0.02, \dots, 1\}$, together with their joint covariate-augmented counterparts $\hat{\tau}_X(\delta)$ and $\hat{\tau}_X(\delta)$ obtained from (53) (solid orange lines), and the corresponding bounds of Lee (2009) (dotted horizontal lines). The dashed lines are the outer 95% confidence bands of the bounds, pointwise in δ , formed by the 2.5% quantile of the bootstrapped lower bound and the 97.5% quantile of the bootstrapped upper bound. The colored vertical dotted lines mark the estimated breakdown budgets $\hat{\delta}_b$ of Section SA5 and the annotations beside them report their values. The remaining annotation reports the reduction in the width of the sensitivity bounds at $\delta = 1$ achieved by the use of additional covariates, so that a negative value means that the covariates widen the bounds. The confidence intervals are computed using 1,000 iterations of a nonparametric percentile bootstrap clustered at the individual level. The corresponding standard errors are reported in Table SA-4.

Summary measures. Table SA-5 reports the measures of Section SA5 for the two comparisons. The shadow prices are small, 0.08 at $\hat{\delta}_b = 0.62$ in the 25% arm and 0.18 at $\hat{\delta}_{cb} = 0.54$ in the 50% arm, which is what the small complier shares lead one to expect in view of the trimming quantiles of Theorem SA-34.

The flatness of the bounds, which is what makes the present application robust, also makes certifying additional robustness cheap at the margin, because by (42) the marginal sample requirement is governed mainly by the ratio of the shadow price to the slack $\tau(\delta) - \tau_*$, with reference value $\tau_* = 0$, and this slack is large in the 50% arm. The precision correction σ'/σ in (42) is comparatively small. Certifying five additional percentage points would require 167 additional participants in the 50% contrast. At fifty dollars per participant, this requirement amounts to approximately \$1,670 per additional percentage point of certified robustness, well below the \$12k to \$80k per percentage point of the tutoring program reported in Section 6.1 of the main text. On the other hand, the 25% contrast illustrates a case in which the certified breakdown budget is 0: signing the effect at $\delta = 0$ alone would require 2,941 additional participants, and certifying a budget of five percentage points 4,702.

Discussion. The exercise is the mirror image of the tutoring application of Section 6.1 of the main text. There, attrition is moderate in level but strongly differential across arms, and a small share of treated respondents departing from MCAR suffices to make the effect vanish. Here, attrition is severe in level but similar across arms. In the 50% arm no departure from MCAR makes the effect vanish, and in the 25% arm close to two thirds of the treated

³When $\hat{\pi}(x) < 0$, we set it to 0, since Mono implies that the two arms' response rates are ordered within each cell and thus we attribute a violation of that ordering in the sample to noise.

Table SA-4: TV sensitivity bounds for the effect of Uber price reductions.

δ	25% reduction			50% reduction		
	$\widehat{\tau}(\delta)$	$\widehat{\tau}(\delta)$	Width	$\widehat{\tau}(\delta)$	$\widehat{\tau}(\delta)$	Width
<i>Panel A: Without covariates</i>						
0.00	0.103 (0.108)	0.103 (0.108)	0.000 (0.000)	0.407 (0.101)	0.407 (0.101)	0.000 (0.000)
0.25	0.041 (0.115)	0.171 (0.117)	0.130 (0.084)	0.284 (0.107)	0.571 (0.113)	0.287 (0.080)
0.50	0.011 (0.122)	0.228 (0.137)	0.217 (0.139)	0.222 (0.112)	0.704 (0.121)	0.482 (0.117)
0.75	-0.009 (0.128)	0.278 (0.159)	0.287 (0.183)	0.183 (0.116)	0.769 (0.120)	0.585 (0.128)
1.00	-0.018 (0.130)	0.307 (0.178)	0.325 (0.210)	0.168 (0.117)	0.794 (0.118)	0.626 (0.129)
<i>Panel B: With covariates</i>						
0.00	0.050 (0.109)	0.050 (0.109)	0.000 (0.000)	0.388 (0.097)	0.388 (0.097)	0.000 (0.000)
0.25	-0.040 (0.113)	0.173 (0.117)	0.213 (0.070)	0.258 (0.100)	0.574 (0.102)	0.316 (0.057)
0.50	-0.077 (0.115)	0.244 (0.128)	0.321 (0.106)	0.211 (0.103)	0.646 (0.109)	0.435 (0.086)
0.75	-0.093 (0.117)	0.291 (0.137)	0.384 (0.126)	0.190 (0.105)	0.693 (0.117)	0.503 (0.105)
1.00	-0.094 (0.117)	0.291 (0.138)	0.385 (0.128)	0.189 (0.105)	0.697 (0.119)	0.509 (0.110)

Notes: Plug-in estimates of the bounds $\widehat{\tau}(\delta)$ and $\widehat{\tau}(\delta)$ of (47) (Panel A) and of their joint covariate-augmented counterparts $\widehat{\tau}_X(\delta)$ and $\widehat{\tau}_X(\delta)$ of (53) (Panel B) at five budgets. Width is the difference between the upper and the lower bound. Bootstrap standard errors are in parentheses, where the standard error of the width is computed from the bootstrap distribution of the width itself. All other technical details can be found in the notes of Figure SA-1.

Table SA-5: Summary measures for the effect of Uber price reductions on total mobility.

	Breakdown		Price of robustness	Certification frontier			
	$\widehat{\delta}_b$	$\widehat{\delta}_{cb}$	$\widehat{\eta}_b$	n	n^*	Δn	\$/pp
25% reduction	0.621	0.000	0.08	912	5,614	+4,702	47k
50% reduction	1	0.543	0.18	916	1,083	+167	2k

Notes: All measures are computed without covariates and with reference value $\tau_\star = 0$. $\widehat{\delta}_b$ is the plug-in breakdown budget of (37), and $\widehat{\delta}_{cb}$ is the certified breakdown budget of (39) at $\alpha = 0.05$, computed from the percentile bootstrap band as in Remark SA-14. $\widehat{\eta}_b$ is the plug-in shadow price, evaluated at $\widehat{\delta}_b$, or at $\widehat{\delta}_{cb}$ when $\widehat{\delta}_b = 1$, n^* is the certification frontier (41) at $\widehat{\delta}_{cb} + 0.05$ and rounded up to the nearest integer, and $\Delta n = n^* - n$ is the additional sample the study would have to collect, measured against the realized sample size n rather than against the frontier at the certified budget. The last column converts Δn into dollars per additional percentage point of certified budget at a marginal cost of fifty dollars per participant enrolled and surveyed, the midpoint of the range implied by the survey costs of the impact-evaluation portfolio of Puri and Rathinam (2019).

respondents must depart from MCAR before the effect loses its sign. The sensitivity exercise thus prices the differential in response across arms, not its level, exactly as the trimming quantiles of Theorem SA-34 predict.

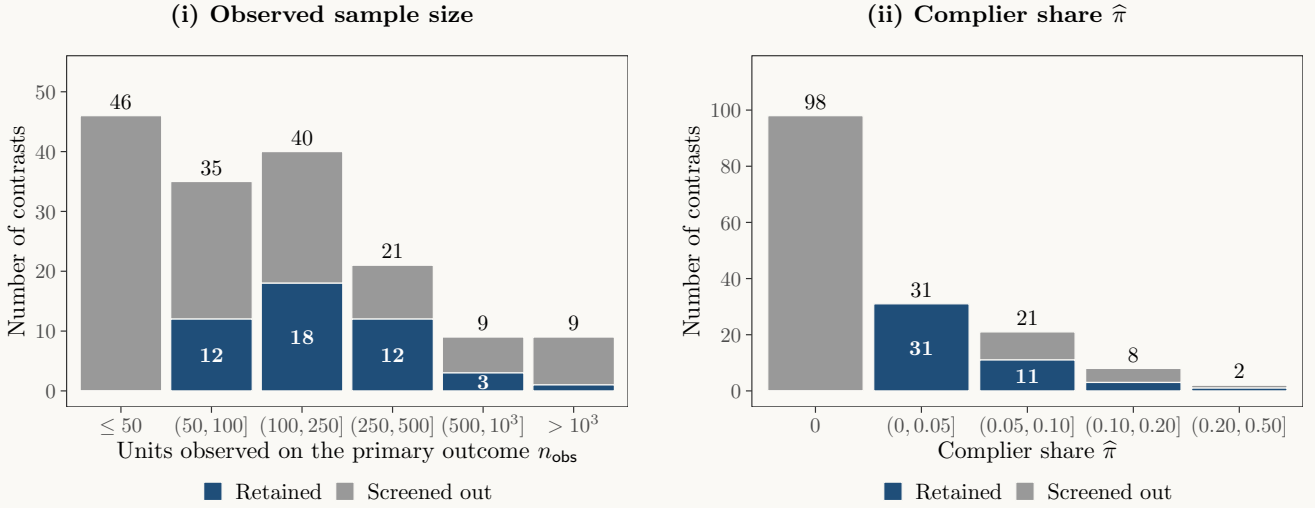
SA7.2 RCT with Attrition II: Shao, Lyu, Yu and Wang (2026)

The case study of the previous subsection and the one in Section 6.1 of the main text answer the question “how fragile is this particular conclusion” for one experiment at a time. A different question is how often attrition is a concern for robustness across a population of published trials. To answer this question, we deploy our framework on every eligible trial in RCT-Bench, a public collection of 125 individually randomized clinical trials with participant-level data – 54,258 participants in all, published between 1991 and 2026 – assembled by Shao et al. (2026) to study covariate

adjustment.⁴ Clinical trials are a natural population for the exercise as attrition is endemic and pre-analysis plans typically commit researchers to use a specific outcome.

Study selection. The unit of analysis is a treatment-control contrast rather than a trial: in trials with more than two arms, we compare each treated arm with the common control group. The 125 trials yield 160 contrasts, 97 from two-arm trials and 63 from the 28 trials with three or more arms. We retain the contrasts with more than fifty units observed on the primary outcome and with a strictly positive complier share $\hat{\pi} > 0$, that is, with some differential attrition. The size screen leaves 114 contrasts, of which 68 exhibit no differential attrition at all ($\hat{\pi} = 0$), leaving 46 contrasts drawn from 37 distinct trials. These are the units displayed in Figure SA-3. They are small to medium size studies. Across the 46 contrasts, the median number of participants randomized to the two arms is 232, and the median number observed on the primary outcome is 195. The largest contrast has 2,207 randomized participants and the smallest 56. Figure SA-2 displays the histograms for the observed sample size and the estimated complier share $\hat{\pi}$ for the whole corpus of 160 contrasts.

Fig. SA-2: Sample size and differential attrition across the clinical corpus.



Notes: Both panels use data from the 160 treatment-control contrasts of the corpus described in the text, shading in blue the 46 contrasts the exercise retains and in gray the remainder. Retention requires (panel (i)) more than fifty units observed on the primary outcome and a strictly positive estimated complier share $\hat{\pi}$ (panel (ii)). The number atop each bar is the bin total. The number inside the blue block is the count of retained contrasts.

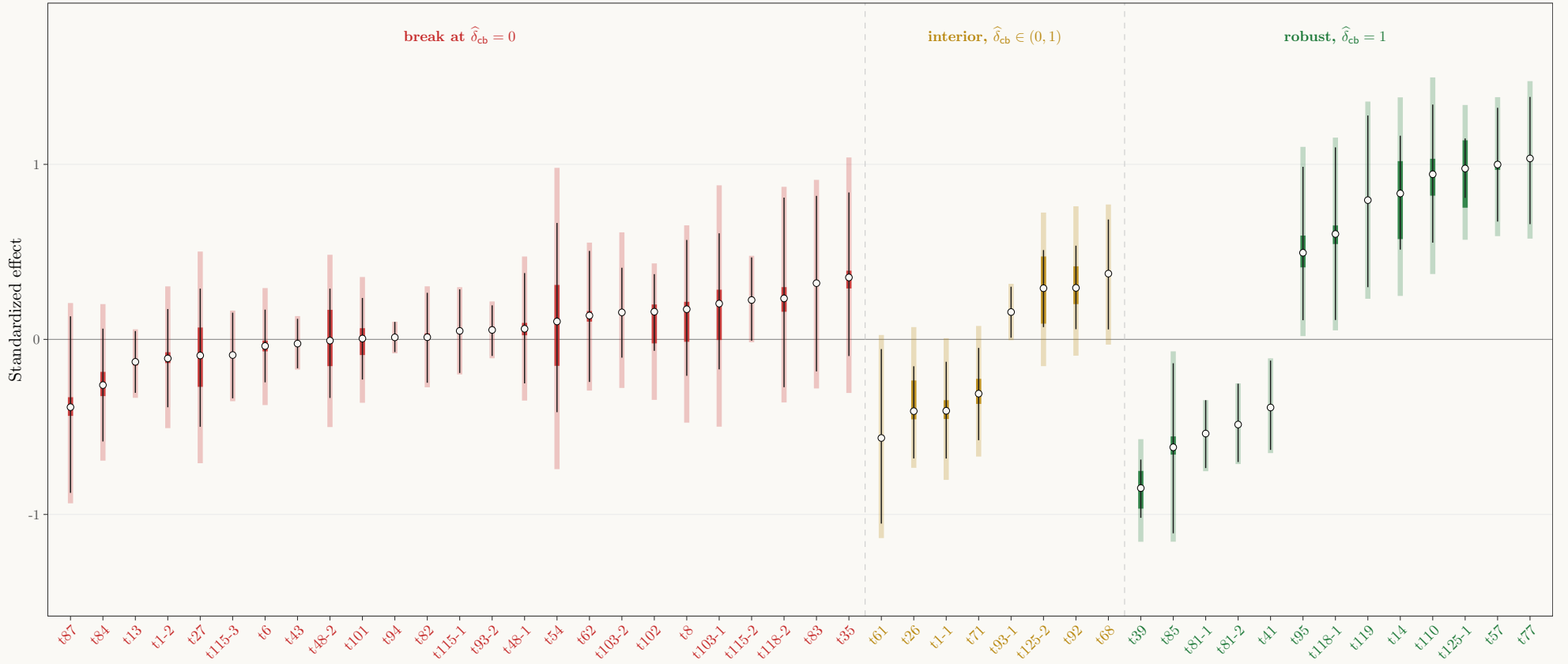
Results. For each contrast we trace $\hat{\tau}(\delta)$ and $\hat{\pi}(\delta)$ on a grid with step size 0.005, with $B = 2,000$ nonparametric bootstrap iterations. The estimated complier shares are small for most contrasts: the median $\hat{\pi}$ is 0.018, the maximum is 0.371, and the smallest positive value is 7×10^{-5} . We summarize each contrast by the certified breakdown budget $\hat{\delta}_{\text{cb}}$ of (39), computed from the percentile bootstrap band as in Remark SA-14. We set $\hat{\delta}_{\text{cb}} = 1$ when zero is never reached on $[0, 1]$.

For each contrast, Figure SA-3 reports on a common standardized scale the difference in means among respondents, the sample analog of $\tau_{\text{MCAR}}(P_0)$ (white dots), and its 95% confidence intervals (black lines), together with $[\hat{\tau}(1), \hat{\pi}(1)]$ (non-transparent bands) and their 95% confidence intervals (shaded bands). First, 25 of the 46 contrasts have $\hat{\delta}_{\text{cb}} = 0$, meaning that the 95% confidence interval for $[\tau(0), \pi(0)]$ already contains 0. These are drawn in red in Figure SA-3. Put differently, the estimated treatment effect is not signed even under MCAR, so that no positive budget is needed to overturn the conclusion. Second, 13 contrasts, drawn in green in Figure SA-3, are robust at every budget, meaning that the confidence intervals for $[\tau(1), \pi(1)]$ do not contain 0. In other words, even if the whole population of treated respondents departs from MCAR and does so adversarially, the treatment effect would still be statistically significant. Third, the remaining 8 contrasts, depicted in amber in Figure SA-3, have a certified breakdown budget in the interior of $[0, 1]$, that is, $\hat{\delta}_{\text{cb}} \in (0, 1)$.

Before concluding, we stress that we did not use covariate-adjusted sensitivity bounds to prioritize comparability across studies. Therefore, a trial with $\hat{\delta}_{\text{cb}} = 0$ may well report a significant covariate-adjusted treatment effect in the primary analysis.

⁴The collection is available at https://bingkaiwang.github.io/RCT_Bench-main/website/. For each trial we take the first listed primary outcome (endpoint) as the outcome and treat a unit as attrited when that outcome is missing in the participant-level file.

Fig. SA-3: TV sensitivity bounds, contrast by contrast.



Notes: The 46 treatment-control contrasts that the screens of this subsection admit, drawn from 37 distinct trials, one vertical bar each, grouped by the value of the certified breakdown budget $\hat{\delta}_{cb}$ of (39), computed as in Remark SA-14, namely $\hat{\delta}_{cb} = 0$ (left), $\hat{\delta}_{cb} \in (0, 1)$ (middle), and $\hat{\delta}_{cb} = 1$ (right), the latter under the convention that the infimum over an empty set equals one, drawn in red, amber, and green, respectively, and ordered within each group by increasing standardized effect. A label “ $tj-k$ ” denotes the k -th treated arm of trial j , indexed in the alphabetical order of the arm names recorded in the metadata, whereas a simple “ tj ” is a two-arm trial. Because the primary outcomes are on incomparable scales, every quantity is divided by $\hat{\sigma}$, the contrast’s pooled-respondent outcome standard deviation, so that the vertical axis is in standard-deviation units. Each bar nests four quantities: the outer 95% confidence band of the bounds at $\delta = 1$ (shaded bands); the identified set at $\delta = 1$, whose endpoints are the Lee bounds (non-transparent bands); the 95% confidence interval of the difference in means among respondents (black lines); and the difference in means among respondents itself, the sample analog of $\tau_{MCAR}(P_0)$ (white dots). Confidence intervals use $B = 2,000$ nonparametric bootstrap iterations per contrast.

SA8 Proofs

SA8.1 Proof of Lemma SA-1

Proof. Choose $\varrho = P_* + \nu^\perp$ as the dominating measure and let $\mathcal{N} \in \mathcal{F}$ be such that $P_*(\mathcal{N}) = 0$ and $\nu^\perp(\mathcal{Z} \setminus \mathcal{N}) = 0$. Such $\mathcal{N}$ exists because $\nu^\perp \perp P_*$. On $\mathcal{Z} \setminus \mathcal{N}$ we have $dP_*/d\varrho = 1$ and $d\nu/d\varrho = w$ ϱ -almost everywhere, so the integrand in Definition SA-5 equals $\phi^\pi(w, 1) = \phi(w)$, whose integral may be taken over all of $\mathcal{Z}$ because $P_*(\mathcal{N}) = 0$. On $\mathcal{N}$ we have $dP_*/d\varrho = 0$ and $d\nu/d\varrho = 1$ ϱ -almost everywhere, since ϱ and ν both restrict to $\nu^\perp$ on $\mathcal{N}$, so the integrand equals $\phi^\pi(1, 0) = \phi^\infty(1)$ by definition of the perspective function, and it integrates to $\phi^\infty(1) \varrho(\mathcal{N}) = \phi^\infty(1) \nu^\perp(\mathcal{Z})$. Summing the two contributions yields the claim. Finally, the argument does not depend on the choice of the dominating measure, in view of the positive homogeneity of the perspective function ϕ^π . ■

SA8.2 Proof of Lemma SA-2

Proof. Since $\phi_{\text{TV}}(s) = \frac{1}{2}|s - 1|$ for $s \geq 0$ and $\phi_{\text{TV}}^\infty(s) = s/2$ for $s \geq 0$, the perspective function satisfies $\phi_{\text{TV}}^\pi(s, t) = \frac{1}{2}|s - t|$ for all $s, t \geq 0$: for $t > 0$ this follows from $t\phi_{\text{TV}}(s/t) = \frac{t}{2}|s/t - 1| = \frac{1}{2}|s - t|$, and for $t = 0$ from $\phi_{\text{TV}}^\pi(s, 0) = \phi_{\text{TV}}^\infty(s) = s/2$. Hence, by Definition SA-5,

$$D_{\phi_{\text{TV}}}(P \| Q) = \int \phi_{\text{TV}}^\pi(p, q) d\varrho = \frac{1}{2} \int |p - q| d\varrho.$$

Next, for any $\mathcal{A} \in \mathcal{F}$, since $P(\mathcal{A}) - Q(\mathcal{A}) = \int_{\mathcal{A}} (p - q) d\varrho$, dropping the negative part of the integrand and enlarging the domain of integration yields $P(\mathcal{A}) - Q(\mathcal{A}) \leq \int [p - q]_+ d\varrho$, with equality at $\mathcal{A} = \{p \geq q\}$. Symmetrically, $Q(\mathcal{A}) - P(\mathcal{A}) \leq \int [q - p]_+ d\varrho$. Since $\int (p - q) d\varrho = 0$ in view of $P(\mathcal{Z}) = Q(\mathcal{Z}) = 1$, the two bounds coincide and equal $\frac{1}{2} \int |p - q| d\varrho$, whence

$$\text{TV}(P \| Q) = \sup_{\mathcal{A} \in \mathcal{F}} |P(\mathcal{A}) - Q(\mathcal{A})| = \frac{1}{2} \int |p - q| d\varrho,$$

attained at $\mathcal{A} = \{p \geq q\}$. ■

SA8.3 Proof of Lemma SA-3

Proof. By Definition SA-4, ϕ is a convex function. Furthermore, ϕ is proper: it is not identically $+\infty$ because $\phi(1) = 0$, and it never takes the value $-\infty$ because $\phi \geq 0$. Thus, by Theorem 12.2 in Rockafellar (1970), ϕ^* is proper, convex, and closed. Closedness and lower semicontinuity are equivalent for any proper convex function (Rockafellar, 1970, p. 52).

Regarding the second part, take any proper closed convex function $f : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$ and define the function $\tilde{f} : \mathbb{R}^k \times \mathbb{R} \rightarrow \overline{\mathbb{R}}$

$$\tilde{f}(z, t) = \begin{cases} tf(z/t), & t > 0, \\ \iota_{\{0\}}(z), & t = 0, \\ +\infty, & t < 0. \end{cases}$$

Then, $\tilde{f}$ is also proper and convex (Rockafellar, 1970, p. 35). The closure of $\tilde{f}$ is equivalent to

$$(\text{cl } \tilde{f})(z, t) = \begin{cases} tf(z/t), & t > 0, \\ f^\infty(z), & t = 0, \\ +\infty, & t < 0, \end{cases}$$

as highlighted on p. 67 in Rockafellar (1970). Note that $\text{cl } \tilde{f}$ is closed and convex by construction and it is proper by Theorem 7.4 in Rockafellar (1970), because $\tilde{f}$ is proper. Hence $\text{cl } \tilde{f}$ is also lower semicontinuous. Finally, by definition of perspective function (Definition SA-3), we have $f^\pi(z, t) = (\text{cl } \tilde{f})(z, t)$. Taking $f = \phi$, which is proper and convex as argued above and lower semicontinuous by Definition SA-4, and $f = \phi^*$ concludes. ■

SA8.4 Proof of Lemma SA-4

Proof. Fix $z \in \mathbb{R}$. If $t < 0$, the statement follows from the definition of perspective function as $(\phi^*)^\pi(z, t) = +\infty$. If $t > 0$, then

$$(\phi^*)^\pi(z, t) = t\phi^*(z/t) \geq \tilde{x}z - t\phi(\tilde{x}),$$

for all $\tilde{x} \in \text{dom}(\phi)$, where the inequality follows by multiplying the Young–Fenchel inequality (Rockafellar, 1970, p. 105),

$$\phi^*(y) + \phi(x) \geq xy \quad \forall x \in \text{dom}(\phi), y \in \text{dom}(\phi^*),$$

taken at $y = z/t$ and $x = \tilde{x}$, by $t > 0$. If $z/t \notin \text{dom}(\phi^*)$, then $(\phi^*)^\pi(z, t) = t\phi^*(z/t) = +\infty$ and the claimed inequality holds trivially. If $t = 0$, then, since ϕ is closed, proper, and convex, it coincides with its biconjugate,

$\phi = \phi^{**}$, so the second assertion of Theorem 13.3 in Rockafellar (1970), applied with $f = \phi^*$, gives that

$$(\phi^*)^\infty(z) = \iota_{\text{dom}(\phi^{**})}^*(z) = \iota_{\text{dom}(\phi)}^*(z) = \sup_{x \in \text{dom}(\phi)} xz \geq \tilde{x}z = \tilde{x}z - t\phi(\tilde{x}).$$

■

SA8.5 Proof of Corollary SA-7

Proof. Lemma SA-6 gives $\text{sup dom}(\phi^*) = \phi^\infty(1)$. Hence (1) and (2) are equivalent. Condition (2) is equivalent to the existence of an $M \in \mathbb{R}$ such that $\text{dom}(\phi^*) \subseteq (-\infty, M]$, which is precisely condition (3). When these conditions hold, the same identity permits the choice $M = \phi^\infty(1)$. ■

SA8.6 Proof of Theorem SA-9

Proof. The proof follows the logic of Kuhn, Shafiee and Wiesemann (2025) and proceeds in four steps:

1. Split $\mathcal{P}_\phi(\rho, \delta)$ along its fibers over $\mathcal{M}(\rho)$.
2. Dualize the inner maximization problem. This step transforms the infinite-dimensional problem over $\mathcal{P}_\phi(\rho, \delta)$ into a finite-dimensional problem in a Euclidean space.
3. Using classical results from convex analysis, the optimal value function h of the primal of the inner maximization problem turns out to be convex. Its biconjugate then satisfies the inequality $h^{**} \leq h$.
4. By the last step, it is possible to upper bound the value of the inner maximization problem.

Fiber decomposition and value function of the primal. Fix $\delta \in \mathbb{R}_+$ and let ϕ be an entropy function. Then, note that we can separate the set of probability distributions $\mathcal{P}_\phi(\rho, \delta)$ along its fibers as follows

$$\begin{aligned} \mathcal{P}_\phi(\rho, \delta) &= \{P \in \mathcal{P}(\mathcal{Z}) : \mathbb{E}_P[m(Z)] \in \mathcal{M}(\rho), D_\phi(P \| P_*) \leq \delta\} \\ &= \bigcup_{\bar{m} \in \mathcal{M}(\rho)} \{P \in \mathcal{P}(\mathcal{Z}) : \mathbb{E}_P[m(Z)] = \bar{m}, D_\phi(P \| P_*) \leq \delta\}. \end{aligned}$$

The inner problem in (4) can then be rewritten as

$$\sup_{P \in \mathcal{P}_\phi(\rho, \delta)} \mathbb{E}_P[g(Z)] = \sup_{\bar{m} \in \mathcal{M}(\rho)} \sup_{\substack{P \in \mathcal{P}(\mathcal{Z}) : \mathbb{E}_P[m(Z)] = \bar{m}, \\ D_\phi(P \| P_*) \leq \delta}} \mathbb{E}_P[g(Z)] = \sup_{\bar{m} \in \mathcal{M}(\rho)} -h(1, \delta, \bar{m}), \quad (62)$$

where h is the value function (6).

Convexity of the value function. Consider the function on $\mathbb{R}^{d_m+2} \times \mathcal{S}$, where $\mathcal{S}$ denotes the vector space of signed measures on $\mathcal{Z}$ with finite m -moment, obtained from (6) by adding to the objective $-\int g d\nu$ the convex indicators of the constraints, namely those of the sets $\{(\nu, x) : \int m d\nu = x\}$, $\{(\nu, a) : \int d\nu = a\}$, $\{(\nu, b) : D_\phi(\nu \| P_*) \leq b\}$, and $\mathcal{M}_{+,m}(\mathcal{Z})$. This function is jointly convex in (a, b, x, ν) : the objective is linear in ν ; the first two indicator sets are linear subspaces; the third is convex because $\nu \mapsto D_\phi(\nu \| P_*)$ is convex on $\mathcal{M}_+(\mathcal{Z})$ – for any two measures, fix a common dominating measure and use the joint convexity of the perspective ϕ^π (see Theorem 7.5(b) in Polyanskiy and Wu (2025) for the case of probability measures) – and $\mathcal{M}_{+,m}(\mathcal{Z})$ is a convex cone. As h is the infimum of this function over $\nu \in \mathcal{S}$, Theorem 1 in Rockafellar (1974) shows that h is jointly convex in its three arguments. Its biconjugate h^{**} is therefore convex and closed and satisfies $h \geq h^{**}$. We first work out the convex conjugate of h , h^* , and then the biconjugate h^{**} .

The convex conjugate. The convex conjugate of h can be obtained as

$$\begin{aligned} h^*(-\zeta, -\eta, -\lambda) &= \sup_{(a, b, x) \in \mathbb{R}^{d_m+2}} \left\{ -\lambda^\top x - \zeta a - \eta b - h(a, b, x) \right\} \\ &= \sup_{\substack{(\nu, x, a, b) \in \\ \mathcal{M}_{+,m}(\mathcal{Z}) \times \mathbb{R}^{d_m+2}}} \left\{ -\lambda^\top x - \zeta a - \eta b + \int g(z) d\nu(z) : \right. \\ &\quad \left. \int m(z) d\nu(z) = x, \int d\nu(z) = a, D_\phi(\nu \| P_*) \leq b \right\} \\ &= \sup_{\substack{(\nu, b) \in \\ \mathcal{M}_{+,m}(\mathcal{Z}) \times \mathbb{R}}} \left\{ \int (g(z) - \lambda^\top m(z) - \zeta) d\nu(z) - \eta b : D_\phi(\nu \| P_*) \leq b \right\}. \end{aligned}$$

First note that $D_\phi(\nu \| P_*) \geq 0$ for every $\nu \in \mathcal{M}_{+,m}(\mathcal{Z})$, as the integrand ϕ^π in Definition SA-5 is nonnegative because $\phi \geq 0$. It must thus be that $b \geq 0$. If $\eta < 0$, the supremum is $+\infty$, as letting $b \rightarrow +\infty$ at $\nu = P_*$, for which $D_\phi(P_* \| P_*) = 0 \leq b$ and the integrals of g and m are finite, sends the objective to $+\infty$. If $\eta \geq 0$, the objective is nonincreasing in b , so that, for each ν with $D_\phi(\nu \| P_*) < \infty$, it suffices to take $b = D_\phi(\nu \| P_*)$. By plugging the

constraint into the objective, the optimization problem becomes

$$\begin{aligned} & \sup_{\nu \in \mathcal{M}_{+,m}(\mathcal{Z})} \int (g(z) - \lambda^\top m(z) - \zeta) d\nu(z) - \eta D_\phi(\nu \| P_*) \\ &= \sup_{\nu \in \mathcal{M}_{+,m}(\mathcal{Z})} \int \left\{ (g(z) - \lambda^\top m(z) - \zeta) \frac{d\nu}{d\varrho}(z) - \eta \phi^\pi \left(\frac{d\nu}{d\varrho}(z), \frac{dP_*}{d\varrho}(z) \right) \right\} d\varrho(z) \end{aligned} \quad (63)$$

where, for a given ν , $\varrho \in \mathcal{M}_{+,m}(\mathcal{Z})$ is any measure with $\nu \ll \varrho$ and $P_* \ll \varrho$, and the equality exploits the definition of (generalized) ϕ -divergence and of perspective function (see Definition SA-3 and Definition SA-5). The supremum may be restricted to the ν with $D_\phi(\nu \| P_*) < \infty$, which we call feasible, since for every other ν the constraint $D_\phi(\nu \| P_*) \leq b$ fails for all real b . For each feasible ν , the function g is ν -integrable by assumption. In particular, P_* is feasible, as $P_* \in \mathcal{M}_{+,m}(\mathcal{Z})$ and $D_\phi(P_* \| P_*) = 0$, so that $g(Z) - \lambda^\top m(Z) - \zeta$ is P_* -integrable.

Every feasible ν is dominated by $\nu + P_*$, which lies in $\mathcal{M}_{+,m}(\mathcal{Z})$, and the objective in (63) does not depend on the choice of ϱ by Remark SA-2. Hence (63) is the supremum, over the $\varrho \in \mathcal{M}_{+,m}(\mathcal{Z})$ with $P_* \ll \varrho$, of the supremum of the same objective over the feasible $\nu \ll \varrho$. Fix such a ϱ and let $\beta \in L_1(\varrho)$ be a version of $dP_*/d\varrho$. For $\alpha \in L_1(\varrho)$, define

$$u_\alpha(z) := (g(z) - \lambda^\top m(z) - \zeta) \alpha(z) - \eta \phi^\pi(\alpha(z), \beta(z)) \quad \text{and} \quad I(\alpha) := \int u_\alpha(z) d\varrho(z),$$

where $I(\alpha)$ is assigned the value $-\infty$ whenever the negative part of u_α fails to be ϱ -integrable. We claim that the supremum over the feasible $\nu \ll \varrho$ equals $\sup_{\alpha \in L_1(\varrho)} I(\alpha)$. One inequality is immediate, as the density $d\nu/d\varrho$ of a feasible $\nu \ll \varrho$ lies in $L_1(\varrho)$ and $I(d\nu/d\varrho)$ equals the objective at ν . The reverse inequality is not immediate, because for $\alpha \in L_1(\varrho)$ the measure $\alpha\varrho$ need not have finite moment $\int \|m\| \alpha d\varrho$ and finite divergence from P_* . Fix $\alpha \in L_1(\varrho)$ with $I(\alpha) > -\infty$. Then $u_\alpha > -\infty$ holds ϱ -almost everywhere, and since $(g - \lambda^\top m - \zeta) \alpha$ is finite ϱ -almost everywhere, so is $\phi^\pi(\alpha, \beta)$. As $\phi(s) = +\infty$ for $s < 0$, Definition SA-3 gives $\phi^\pi(s, t) = +\infty$ whenever $s < 0$ and $t \geq 0$, so that $\alpha \geq 0$ holds ϱ -almost everywhere. For $k \in \mathbb{N}$, let

$$A_k := \{z \in \mathcal{Z} : \|m(z)\| \leq k, \phi^\pi(\alpha(z), \beta(z)) \leq k\} \quad \text{and} \quad \alpha_k := \alpha \mathbb{1}_{A_k} + \beta \mathbb{1}_{\mathcal{Z} \setminus A_k}.$$

The measure $\nu_k := \alpha_k \varrho$ is nonnegative and finite. It has finite moment, as

$$\int \|m(z)\| d\nu_k(z) \leq k \int \alpha(z) d\varrho(z) + \mathbb{E}_{P_*}[\|m(Z)\|] < \infty,$$

and it has finite divergence, as

$$D_\phi(\nu_k \| P_*) = \int_{A_k} \phi^\pi(\alpha(z), \beta(z)) d\varrho(z) \leq k \varrho(\mathcal{Z}) < \infty,$$

where the equality holds because $\phi^\pi(t, t) = 0$ for every $t \geq 0$, by $\phi(1) = 0$ when $t > 0$ and by $\phi^\infty(0) = 0$ when $t = 0$. Hence ν_k is feasible, and, by the same property and $\beta d\varrho = dP_*$, its objective value is

$$I(\alpha_k) = \int_{A_k} u_\alpha(z) d\varrho(z) + \mathbb{E}_{P_*}[(g(Z) - \lambda^\top m(Z) - \zeta) \mathbb{1}_{\{Z \notin A_k\}}].$$

The sets A_k increase to $\{z \in \mathcal{Z} : \phi^\pi(\alpha(z), \beta(z)) < \infty\}$, which has full ϱ -measure, and hence full P_* -measure as $P_* \ll \varrho$. The second term thus vanishes as $k \rightarrow \infty$ by dominated convergence, since $g(Z) - \lambda^\top m(Z) - \zeta$ is P_* -integrable, while monotone convergence applied separately to u_α^+ and u_α^- gives $\int_{A_k} u_\alpha d\varrho \rightarrow I(\alpha)$. Therefore $I(\alpha_k) \rightarrow I(\alpha)$, which proves the claim. Consequently, for each such ϱ and β , the supremum of the objective over the feasible $\nu \ll \varrho$ equals the left-hand side of

$$\begin{aligned} & \sup_{\alpha \in L_1(\varrho)} \int \left\{ (g(z) - \lambda^\top m(z) - \zeta) \alpha(z) - \eta \phi^\pi(\alpha(z), \beta(z)) \right\} d\varrho(z) \\ &= \int \sup_{\alpha \in \mathbb{R}} \left\{ (g(z) - \lambda^\top m(z) - \zeta) \alpha - \eta \phi^\pi(\alpha, \beta(z)) \right\} d\varrho(z), \end{aligned} \quad (64)$$

where the equality is the maximization form of Theorem 14.60 in Rockafellar and Wets (2009), which applies because for each z the integrand is upper semicontinuous in α (the affine term is continuous and $-\eta \phi^\pi(\cdot, \beta(z))$ is upper semicontinuous), so its negative is a normal integrand on the decomposable space $L_1(\varrho)$. Then, recalling that the integral runs over $\mathcal{Z}$, it is possible to split $\mathcal{Z}$ into the set where β vanishes and the set where it is positive. Formally:

$$\mathcal{Z}_+(\beta) := \{z \in \mathcal{Z} : \beta(z) > 0\} \quad \text{and} \quad \mathcal{Z}_0(\beta) := \{z \in \mathcal{Z} : \beta(z) = 0\}.$$

Using Lemma SA-5, (64) can be rewritten as

$$\begin{aligned} & \int_{\mathcal{Z}_+(\beta)} \eta \phi^* \left(\frac{g(z) - \lambda^\top m(z) - \zeta}{\eta} \right) \beta(z) d\varrho(z) + \\ & \int_{\mathcal{Z}_0(\beta)} \eta \iota_{\text{cl}(\text{dom}(\phi^*))} \left(\frac{g(z) - \lambda^\top m(z) - \zeta}{\eta} \right) d\varrho(z). \end{aligned} \quad (65)$$

The first integral is an expectation over P_* because $P_*(\mathcal{Z}_+(\beta)) = \int_{\mathcal{Z}_+(\beta)} dP_*(z) = \int_{\mathcal{Z}_+(\beta)} \beta(z) d\varrho(z) = 1$, where the last equality follows from the definition of the partition $\mathcal{Z} = \mathcal{Z}_+(\beta) \cup \mathcal{Z}_0(\beta)$.

Putting everything together, the conjugate of h can be rewritten as

$$h^*(-\zeta, -\eta, -\lambda) = \mathbb{E}_{P_*} \left[\eta \phi^* \left(\frac{g(Z) - \lambda^\top m(Z) - \zeta}{\eta} \right) \right] + \sup_{\substack{\varrho \in \mathcal{M}_{+,m}(\mathcal{Z}): P_* \ll \varrho, \\ \beta = dP_*/d\varrho \text{ a.s.}}} \left\{ \int_{\mathcal{Z}_0(\beta)} \eta \iota_{\text{cl}(\text{dom}(\phi^*))} \left(\frac{g(z) - \lambda^\top m(z) - \zeta}{\eta} \right) d\varrho(z) \right\}. \quad (66)$$

If there exists $z_0 \in \mathcal{Z}$ such that $\frac{g(z_0) - \lambda^\top m(z_0) - \zeta}{\eta} \notin \text{cl}(\text{dom}(\phi^*))$, then the objective evaluates to $+\infty$. To see this, there are two possibilities:

- if z_0 is an atom of P_* , then the first expectation evaluates to $+\infty$, since $\phi^* = +\infty$ strictly outside $\text{cl}(\text{dom}(\phi^*))$ and the integrand is bounded below by the P_* -integrable function $g(Z) - \lambda^\top m(Z) - \zeta$;
- if z_0 is not an atom of P_* , then one can always choose $\varrho = P_* + \delta_{z_0}^{\text{Dirac}}$ and set $\beta(z) = \mathbf{1}\{z \neq z_0\}$ so that the second integral picks up the Dirac mass at z_0 , where the integrand evaluates to $+\infty$ by definition of the indicator function $\iota_{\text{cl}(\text{dom}(\phi^*))}$.

Lemma SA-6 shows that $\text{cl}(\text{dom}(\phi^*)) = (-\infty, \phi^\infty(1)]$ if $\phi^\infty(1) < \infty$ and $\text{cl}(\text{dom}(\phi^*)) = \mathbb{R}$ otherwise. It follows that $\frac{g(z) - \lambda^\top m(z) - \zeta}{\eta} \in \text{cl}(\text{dom}(\phi^*))$ is equivalent to

$$\zeta + \eta \phi^\infty(1) \geq \sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\}. \quad (67)$$

Whenever the above condition is satisfied, the second integral in (66) is identically 0 by definition of the indicator function $\iota_{\text{cl}(\text{dom}(\phi^*))}$. Thus, the convex conjugate of h is

$$h^*(-\zeta, -\eta, -\lambda) = \begin{cases} \mathbb{E}_{P_*} \left[\eta \phi^* \left(\frac{g(Z) - \lambda^\top m(Z) - \zeta}{\eta} \right) \right] & \text{if (67) holds,} \\ +\infty & \text{otherwise.} \end{cases}$$

Extension to $\eta = 0$. Recall that we derived the above expression under the assumption that $\eta > 0$. We now claim that

$$h^*(-\zeta, -\eta, -\lambda) = \begin{cases} \mathbb{E}_{P_*} \left[(\phi^*)^\pi (g(Z) - \lambda^\top m(Z) - \zeta, \eta) \right] & \text{if } \eta \geq 0 \text{ and (67) holds,} \\ +\infty & \text{otherwise,} \end{cases}$$

holds. To see why, note that h^* is convex and closed (Rockafellar, 1970, Theorem 12.2). Furthermore, Lemma SA-3 gives that ϕ^* and $(\phi^*)^\pi$ are both convex, proper, and lower semicontinuous. By Lemma SA-4, $(\phi^*)^\pi(z, \eta) \geq z$ for every $\eta \geq 0$. Along any convergent sequence of multipliers $(\zeta_n, \eta_n, \lambda_n) \rightarrow (\zeta, \eta, \lambda)$ the (ζ_n, λ_n) are bounded, so $g(z) - \lambda_n^\top m(z) - \zeta_n \geq g(z) - C(1 + \|m(z)\|)$ for a finite constant C independent of n , and this lower bound is P_* -integrable because $\mathbb{E}_{P_*}[g(Z)]$ and $\mathbb{E}_{P_*}[\|m(Z)\|]$ are finite (as $P_* \in \mathcal{M}_{+,m}(\mathcal{Z})$ with $D_\phi(P_* \| P_*) = 0$). Hence the integrands $(\phi^*)^\pi(g(Z) - \lambda_n^\top m(Z) - \zeta_n, \eta_n)$ are bounded below by this single integrable minorant, and the lower semicontinuity of $(\phi^*)^\pi$ together with Fatou's lemma gives that $(\zeta, \eta, \lambda) \mapsto \mathbb{E}_{P_*}[(\phi^*)^\pi(g(Z) - \lambda^\top m(Z) - \zeta, \eta)]$ is lower semicontinuous, while joint convexity of $(\phi^*)^\pi$ makes it convex, hence it is closed and convex.

Denote by $F(\zeta, \eta, \lambda)$ the right-hand side of the claimed display. We have just shown that its expectation component is closed and convex. The constraint (67) defines, in turn, a closed convex subset of $\mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}$, because $\lambda \mapsto \sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\}$ is a supremum of affine functions of λ , hence convex and lower semicontinuous, while (ζ, η) enter the constraint linearly. When $\phi^\infty(1) = +\infty$, the convention of the statement renders the constraint vacuous and the set is the entire space. The convention is essential here: under the alternative reading $0 \cdot (+\infty) = 0$, the constraint would be vacuous for $\eta > 0$ yet active on the slice $\{\eta = 0\}$, and the resulting set would fail to be closed. As the sum of a closed convex function and the indicator of a closed convex set, F is closed and convex, and it coincides with h^* on $\{\eta > 0\}$ by the previous step.

Agreement on $\{\eta > 0\}$ alone does not determine the values of two closed convex functions on the boundary slice $\{\eta = 0\}$, so we further verify that their domains share a relative interior point with $\eta > 0$. Both functions are nonincreasing in η : for h^* , this follows from its representation as a supremum over pairs (ν, b) with $b \geq D_\phi(\nu \| P_*) \geq 0$ obtained at the beginning of the previous step, whose objective is nonincreasing in η pointwise in (ν, b) . For F , it follows from the identity $(\phi^*)^\pi(x, \eta) = \sup_{u \in \text{dom}(\phi)} \{ux - \eta \phi(u)\}$ for $\eta \geq 0$ together with $\phi \geq 0$, and from the fact that the left-hand side of (67) is nondecreasing in η , so that feasibility at $(\zeta, 0, \lambda)$ implies feasibility at (ζ, η, λ) for every $\eta > 0$. Neither function takes the value $-\infty$: F is bounded below by the integrable minorant of the previous paragraph, and $h^*(-\zeta, -\eta, -\lambda) \geq \mathbb{E}_{P_*}[g(Z) - \lambda^\top m(Z)] - \zeta > -\infty$, as follows from evaluating the supremum defining it at $(\nu, b) = (P_*, 0)$. Consequently, finiteness of either function at a point of the slice propagates to every point vertically above it.

If $h^* = +\infty$ throughout $\{\eta > 0\}$, then so is F , and the propagation property forces both functions to equal $+\infty$ on the slice as well, so that they agree everywhere. Otherwise, the set $D := \text{dom}(h^*) \cap \{\eta > 0\} = \text{dom}(F) \cap \{\eta > 0\}$ is nonempty and convex, and the propagation property yields $D \subseteq \text{dom}(h^*) \subseteq \text{cl}(D)$, and likewise for F . Since

a convex set and its closure have the same relative interior (Rockafellar, 1970, Theorem 6.3), it follows that $\text{ri}(\text{dom}(h^*)) = \text{ri}(\text{dom}(F)) = \text{ri}(D) \subseteq \{\eta > 0\}$. Fix any $y \in \text{ri}(D)$ and any point x on the slice $\{\eta = 0\}$. For a closed proper convex function f , the value at x is recovered as the limit along the segment joining x to any point of $\text{ri}(\text{dom}(f))$, that is, $f(x) = \lim_{t \rightarrow 0} f((1-t)x + ty)$ (Rockafellar, 1970, Corollary 7.5.1). Every point of the open segment has a strictly positive η -coordinate, where h^* and F coincide, so the two limits coincide, and hence so do the values at x . This proves the claim.

The biconjugate and the bound. Using the definition, the biconjugate of h is then immediate to construct:

$$\begin{aligned} h^{**}(a, b, x) &= \sup_{\substack{(\zeta, \eta, \lambda) \in \\ \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{dm}}} -\lambda^\top x - \zeta a - \eta b - h^*(-\zeta, -\eta, -\lambda) \\ &= \begin{cases} \sup_{\substack{(\zeta, \eta, \lambda) \in \\ \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{dm}}} -\lambda^\top x - \zeta a - \eta b - \mathbb{E}_{P_*} \left[(\phi^*)^\pi(g(Z) - \lambda^\top m(Z) - \zeta, \eta) \right] \\ \text{s.t. } \zeta + \eta \phi^\infty(1) \geq \sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\}. \end{cases} \end{aligned} \quad (68)$$

Let $\mathcal{F}$ denote the set of (ζ, η, λ) satisfying $\eta \geq 0$ and the pointwise constraint in (68). Therefore, by (62) and the biconjugate inequality $h^{**} \leq h$,

$$\begin{aligned} \sup_{P \in \mathcal{P}_\phi(\rho, \delta)} \mathbb{E}_P[g(Z)] &\leq \sup_{\bar{m} \in \mathcal{M}(\rho)} \inf_{(\zeta, \eta, \lambda) \in \mathcal{F}} \left\{ \lambda^\top \bar{m} + \zeta + \eta \delta + \mathbb{E}_{P_*} \left[(\phi^*)^\pi(g(Z) - \lambda^\top m(Z) - \zeta, \eta) \right] \right\} \\ &\leq \inf_{(\zeta, \eta, \lambda) \in \mathcal{F}} \left\{ \iota_{\mathcal{M}(\rho)}^*(\lambda) + \zeta + \eta \delta + \mathbb{E}_{P_*} \left[(\phi^*)^\pi(g(Z) - \lambda^\top m(Z) - \zeta, \eta) \right] \right\}, \end{aligned}$$

where the second inequality is the minimax inequality. The support function identity established in Section SA1.2 restricts λ to $\mathcal{R}_\lambda$ and replaces $\iota_{\mathcal{M}(\rho)}^*(\lambda)$ with $\lambda_{12}^\top \rho$, which gives the stated weak dual bound. ■

SA8.7 Proof of Lemma SA-10

Proof. Relative-interior qualification. Write $c := \mathbb{E}_{P_*}[m(Z)]$. Recall that $(a, b, x) \in \text{dom}(h)$ exactly when some $\nu \in \mathcal{M}_{+,m}(\mathcal{Z})$ has mass $\int d\nu = a$, moment $\int m d\nu = x$ and $D_\phi(\nu \| P_*) \leq b$, in which case the value of (6) is smaller than $+\infty$, since $-\int g(z) d\nu(z) < \infty$ by hypothesis.

Uniform strict feasibility near $\bar{m}$. Since $c \in \Gamma(\delta)$, because $D_\phi(P_* \| P_*) = 0 \leq \delta$, and $\bar{m} \in \text{ri}(\Gamma(\delta))$, Theorem 6.4 in Rockafellar (1970) yields a $\mu_0 > 1$ such that $c + \mu_0(\bar{m} - c) \in \Gamma(\delta)$. Fixing any $\mu \in (1, \mu_0)$, the point

$$c + \mu(\bar{m} - c) = \left(1 - \frac{\mu-1}{\mu_0-1}\right)\bar{m} + \frac{\mu-1}{\mu_0-1}(c + \mu_0(\bar{m} - c))$$

is a convex combination of $\bar{m} \in \text{ri}(\Gamma(\delta))$ and $c + \mu_0(\bar{m} - c) \in \Gamma(\delta)$, with weight $1 - \frac{\mu-1}{\mu_0-1} > 0$ on $\bar{m}$. Therefore, $c + \mu(\bar{m} - c) \in \text{ri}(\Gamma(\delta))$ by Theorem 6.1 in Rockafellar (1970). By continuity of $y \mapsto c + \mu(y - c)$, there is $r > 0$ such that the relative neighborhood $V := \mathcal{B}_r(\bar{m}) \cap \text{aff}(\Gamma(\delta))$ of $\bar{m}$ satisfies $V \subseteq \Gamma(\delta)$ and $c + \mu(y - c) \in \Gamma(\delta)$ for all $y \in V$. For $y \in V$, fix $P_y \in \mathcal{P}(\mathcal{Z})$ so that $\mathbb{E}_{P_y}[m(Z)] = c + \mu(y - c)$ and $D_\phi(P_y \| P_*) \leq \delta$. Set $Q_y := (1 - \frac{1}{\mu})P_* + \frac{1}{\mu}P_y$. Then $\mathbb{E}_{Q_y}[m(Z)] = y$ and, by convexity of $D_\phi(\cdot \| P_*)$ with $D_\phi(P_* \| P_*) = 0$, we get

$$D_\phi(Q_y \| P_*) \leq \frac{1}{\mu} D_\phi(P_y \| P_*) \leq \frac{\delta}{\mu} < \delta.$$

A relative neighborhood in $\text{dom}(h)$. Consider the affine set

$$\mathbb{A} := \{(a, b, x) \in \mathbb{R}^{d_m+2} : x - (a-1)c \in \text{aff}(\Gamma(\delta))\}.$$

For a target $(a, b, x) \in \mathbb{A}$ in a small neighborhood of $(1, \delta, \bar{m})$, let $y := x - (a-1)c$, which lies in $\text{aff}(\Gamma(\delta))$ and hence in V for the neighborhood small enough. Writing the Lebesgue decomposition of P_y with respect to P_* as $P_y = w'_y P_* + P_y^\perp$, where $w'_y := d(P_y)_{\text{ac}}/dP_*$ is the density of its absolutely continuous part and $P_y^\perp \perp P_*$ its singular part, set

$$\nu := (a-1)P_* + Q_y = \left(a - \frac{1}{\mu} + \frac{1}{\mu}w'_y\right)P_* + \frac{1}{\mu}P_y^\perp \in \mathcal{M}_{+,m}(\mathcal{Z}).$$

This ν is nonnegative if $a \geq \frac{1}{\mu}$, and has mass $\int d\nu = a$ and moment $\int m d\nu = (a-1)c + y = x$. Once the divergence bound below is established, the hypothesis also gives $\int g(z) d\nu(z) > -\infty$. Set $w_a := \frac{a-1/\mu}{1-1/\mu}$, so the absolutely continuous density of ν equals $(1 - \frac{1}{\mu})w_a + \frac{1}{\mu}w'_y$. By convexity of ϕ , P_* -a.e.,

$$\phi\left(\left(1 - \frac{1}{\mu}\right)w_a + \frac{1}{\mu}w'_y\right) \leq \left(1 - \frac{1}{\mu}\right)\phi(w_a) + \frac{1}{\mu}\phi(w'_y),$$

so, adding the singular contribution $\frac{1}{\mu}\phi^\infty(1)P_y^\perp(\mathcal{Z})$ and applying Lemma SA-1 twice,

$$\begin{aligned} D_\phi(\nu \| P_*) &\leq \left(1 - \frac{1}{\mu}\right)\phi(w_a) + \frac{1}{\mu} \left[\int \phi(w'_y) dP_* + \phi^\infty(1)P_y^\perp(\mathcal{Z}) \right] \\ &= \left(1 - \frac{1}{\mu}\right)\phi(w_a) + \frac{1}{\mu} D_\phi(P_y \| P_*) \leq \left(1 - \frac{1}{\mu}\right)\phi(w_a) + \frac{\delta}{\mu}, \end{aligned}$$

which holds uniformly in $y \in V$, the right-hand side depending on the target only through a . As $a \rightarrow 1$ we have $w_a \rightarrow 1$ and, since ϕ is continuous at 1 with $\phi(1) = 0$, $\phi(w_a) \rightarrow 0$. Therefore, the upper bound above tends to $\frac{\delta}{\mu} < \delta$,

because $\mu > 1$. Hence, for (a, b) close enough to $(1, \delta)$, the upper bound is smaller than b , so that $D_\phi(\nu \| P_*) \leq b$. This fact, together with $\int d\nu = a$ and $\int m d\nu = x$, gives $(a, b, x) \in \text{dom}(h)$ throughout a relative neighborhood of $(1, \delta, \bar{m})$ within $\mathbb{A}$.

The affine hull of $\text{dom}(h)$. To conclude that $(1, \delta, \bar{m}) \in \text{ri}(\text{dom}(h))$, it remains to verify that $\text{dom}(h) \subseteq \mathbb{A}$, so that the neighborhood just constructed is a neighborhood relative to $\text{aff}(\text{dom}(h))$. Take $(a, b, x) \in \text{dom}(h)$ and let $\nu \in \mathcal{M}_{+,m}(\mathcal{Z})$ have mass a , moment x , and $D_\phi(\nu \| P_*) \leq b < \infty$. For $s \in (0, 1)$ small enough so that $1 - sa > 0$, the probability measure $Q_s := s\nu + (1 - sa)P_*$ can be written as the convex combination $Q_s = s\nu + (1 - s)t_s P_*$ with $t_s := (1 - sa)/(1 - s)$, so that, by convexity of $D_\phi(\cdot \| P_*)$,

$$D_\phi(Q_s \| P_*) \leq s D_\phi(\nu \| P_*) + (1 - s) \phi(t_s) \leq s b + (1 - s) \phi(t_s),$$

where $D_\phi(t_s P_* \| P_*) = \phi(t_s)$ by Definition SA-5 with $\varrho = P_*$. Since $t_s \rightarrow 1$ as $s \downarrow 0$ and ϕ is continuous at 1 with $\phi(1) = 0$, the right-hand side falls below δ for s small enough and $\mathbb{E}_{Q_s}[m(Z)] = s x + (1 - sa) c \in \Gamma(\delta)$. Rearranging, $x - (a - 1)c = c + s^{-1}(\mathbb{E}_{Q_s}[m(Z)] - c)$, and since $\mathbb{E}_{Q_s}[m(Z)] - c$ belongs to the subspace $\text{aff}(\Gamma(\delta)) - c$, which is closed under scaling, we conclude $x - (a - 1)c \in \text{aff}(\Gamma(\delta))$, i.e. $(a, b, x) \in \mathbb{A}$. Hence $\text{dom}(h) \subseteq \mathbb{A}$, while the neighborhood constructed above is open relative to $\mathbb{A}$, so $\text{aff}(\text{dom}(h)) = \mathbb{A}$ and $(1, \delta, \bar{m}) \in \text{ri}(\text{dom}(h))$.

Properness. We argue by contradiction. Suppose that $h(a, b, x) = -\infty$ for some $(a, b, x) \in \mathbb{R}^{d_m+2}$. Any $\nu \in \mathcal{M}_{+,m}(\mathcal{Z})$ that is feasible for (6) at (a, b, x) satisfies $D_\phi(\nu \| P_*) \leq b < \infty$, so that $\int |g| d\nu < \infty$ by hypothesis and $\int g d\nu$ is well defined and finite. Since $h(a, b, x) = -\infty$, the supremum of $\int g d\nu$ over the feasible measures equals $+\infty$, so we may select a sequence $\{\nu_k\}_{k \geq 1}$ of feasible measures with $\int g d\nu_k \geq 4^k$ for every $k \geq 1$. In particular, $\int [g]_+ d\nu_k \geq 4^k$, because $\int g d\nu_k \leq \int [g]_+ d\nu_k$.

Let $\nu^\dagger := \sum_{k \geq 1} 2^{-k} \nu_k$, a nonnegative measure of mass a , where here and below every interchange of a sum and an integral follows from Tonelli's theorem. Write $\nu_k = w_k P_* + \nu_k^\perp$ for the Lebesgue decomposition of ν_k with respect to P_* , as in Lemma SA-1, where $\nu_k^\perp$ is concentrated on a P_* -null set N_k . The function $w^\dagger := \sum_{k \geq 1} 2^{-k} w_k$ is finite P_* -almost everywhere, as $\int w^\dagger dP_* \leq \sum_{k \geq 1} 2^{-k} \nu_k(\mathcal{Z}) = a$, and $\nu^{\dagger \perp} := \sum_{k \geq 1} 2^{-k} \nu_k^\perp$ is concentrated on the P_* -null set $\cup_{k \geq 1} N_k$. By uniqueness of the Lebesgue decomposition, $\nu^\dagger = w^\dagger P_* + \nu^{\dagger \perp}$ is thus the Lebesgue decomposition of $\nu^\dagger$ with respect to P_* , and Lemma SA-1 gives

$$D_\phi(\nu^\dagger \| P_*) = \int_{\mathcal{Z}} \phi(w^\dagger(z)) dP_*(z) + \phi^\infty(1) \nu^{\dagger \perp}(\mathcal{Z}).$$

We bound the two terms in turn. As the entropy function ϕ is convex, lower semicontinuous, and proper, it is the pointwise supremum of the affine functions it majorizes (Rockafellar, 1970, Theorem 12.1). For any affine $\ell(t) = \alpha t + \beta$ with $\ell \leq \phi$ and any $z \in \mathcal{Z}$ with $w^\dagger(z) < \infty$,

$$\ell(w^\dagger(z)) = \alpha \sum_{k \geq 1} 2^{-k} w_k(z) + \beta \sum_{k \geq 1} 2^{-k} = \sum_{k \geq 1} 2^{-k} \ell(w_k(z)) \leq \sum_{k \geq 1} 2^{-k} \phi(w_k(z)),$$

where the first equality uses $\sum_{k \geq 1} 2^{-k} = 1$ and the convergence of $\sum_{k \geq 1} 2^{-k} w_k(z)$. Taking the supremum over such ℓ yields $\phi(w^\dagger) \leq \sum_{k \geq 1} 2^{-k} \phi(w_k)$ so that

$$\int_{\mathcal{Z}} \phi(w^\dagger(z)) dP_*(z) \leq \sum_{k \geq 1} 2^{-k} \int_{\mathcal{Z}} \phi(w_k(z)) dP_*(z).$$

For the singular term, $\nu^{\dagger \perp}(\mathcal{Z}) = \sum_{k \geq 1} 2^{-k} \nu_k^\perp(\mathcal{Z})$, so that $\phi^\infty(1) \nu^{\dagger \perp}(\mathcal{Z}) = \sum_{k \geq 1} 2^{-k} \phi^\infty(1) \nu_k^\perp(\mathcal{Z})$ under the convention that the product evaluates to 0 when the singular mass vanishes. Combining the two bounds and applying Lemma SA-1 to each ν_k , gives

$$D_\phi(\nu^\dagger \| P_*) \leq \sum_{k \geq 1} 2^{-k} \left[\int_{\mathcal{Z}} \phi(w_k(z)) dP_*(z) + \phi^\infty(1) \nu_k^\perp(\mathcal{Z}) \right] = \sum_{k \geq 1} 2^{-k} D_\phi(\nu_k \| P_*) \leq b.$$

Finally, we get

$$\int_{\mathcal{Z}} [g(z)]_+ d\nu^\dagger(z) = \sum_{k \geq 1} 2^{-k} \int_{\mathcal{Z}} [g(z)]_+ d\nu_k(z) \geq \sum_{k \geq 1} 2^{-k} 4^k = +\infty.$$

The measure $\nu^\dagger$ thus belongs to $\mathcal{M}_+(\mathcal{Z})$ and satisfies $D_\phi(\nu^\dagger \| P_*) \leq b < \infty$, yet $\int_{\mathcal{Z}} |g| d\nu^\dagger \geq \int_{\mathcal{Z}} [g]_+ d\nu^\dagger = +\infty$, contradicting the hypothesis. Therefore, $h(a, b, x) > -\infty$ for every $(a, b, x) \in \mathbb{R}^{d_m+2}$. Moreover, P_* is feasible in the definition of $h(1, 0, \mathbb{E}_{P_*}[m(Z)])$, and $\mathbb{E}_{P_*}[|g(Z)|] < \infty$ by hypothesis. Hence this value of h is finite, so h is proper. $\blacksquare$

SA8.8 Proof of Theorem SA-11

Proof. Let K_ρ be the set defined before Lemma SA-10, and let $q := \iota_{K_\rho}$. The set K_ρ is nonempty and polyhedral, so q is a proper polyhedral convex function. By the fiber decomposition (62),

$$-\sup_{P \in \mathcal{P}_\phi(\rho, \delta)} \mathbb{E}_P[g(Z)] = \inf_{y \in \mathbb{R}^{d_m+2}} \{h(y) + q(y)\}.$$

The function h is proper and convex, where we established convexity in the proof of Theorem SA-9, and properness follows from Lemma SA-10.

Choose $\bar{m}_0 \in \mathcal{M}(\rho) \cap \text{ri}(\Gamma(\delta))$. By Lemma SA-10, $(1, \delta, \bar{m}_0) \in \text{ri}(\text{dom}(h))$, while $(1, \delta, \bar{m}_0) \in \text{dom}(q) = K_\rho$. Lemma SA-8 therefore gives

$$\inf_{y \in \mathbb{R}^{d_m+2}} \{h(y) + q(y)\} = \sup_{y^* \in \mathbb{R}^{d_m+2}} \{-h^*(y^*) - q^*(-y^*)\}.$$

Write $y^* = (-\zeta, -\eta, -\lambda)$. The conjugate of q is

$$\begin{aligned} q^*(\zeta, \eta, \lambda) &= \sup_{(1, \delta, \bar{m}) \in K_\rho} \{\zeta + \eta\delta + \lambda^\top \bar{m}\} \\ &= \zeta + \eta\delta + \iota_{\mathcal{M}(\rho)}^*(\lambda). \end{aligned}$$

The conjugate calculation in the proof of Theorem SA-9 gives

$$h^*(-\zeta, -\eta, -\lambda) = \mathbb{E}_{P_*} \left[(\phi^*)^\pi (g(Z) - \lambda^\top m(Z) - \zeta, \eta) \right]$$

when $\eta \geq 0$ and

$$\zeta + \eta\phi^\infty(1) \geq \sup_{z \in \mathcal{Z}} \{g(z) - \lambda^\top m(z)\},$$

and gives $+\infty$ otherwise. Negating the Fenchel identity and using

$$\iota_{\mathcal{M}(\rho)}^*(\lambda) = \begin{cases} \lambda_{12}^\top \rho, & \lambda \in \mathcal{R}_\lambda, \\ +\infty, & \lambda \notin \mathcal{R}_\lambda, \end{cases}$$

yields exactly the dual program in the statement. ■

SA8.9 Proof of Lemma SA-12

Proof. We prove each part separately.

Part (i): baseline as the mixture. The implication $(1) \Rightarrow (2)$ is immediate. For $(2) \Rightarrow (1)$, define

$$R(A) := \frac{P_*(A) - \kappa P(A)}{1 - \kappa}.$$

Then R is a probability measure and $P_* = \kappa P + (1 - \kappa)R$.

Next, $(2) \Rightarrow (3)$: if $P_*(A) = 0$, then $\kappa P(A) \leq P_*(A) = 0$, hence $P(A) = 0$, so $P \ll P_*$. Let $u := dP/dP_*$. Then for every $A \in \mathcal{F}$,

$$\int_A \kappa u dP_* = \kappa P(A) \leq P_*(A) = \int_A 1 dP_*,$$

so $\kappa u \leq 1$ P_* -a.s., that is $u \leq 1/\kappa$.

Finally, $(3) \Rightarrow (2)$ follows by integrating the bound $0 \leq dP/dP_* \leq 1/\kappa$. Indeed, for every $A \in \mathcal{F}$,

$$\kappa P(A) = \int_A \kappa \frac{dP}{dP_*} dP_* \leq \int_A dP_* = P_*(A),$$

which is precisely (2).

Part (ii): baseline as a mixing component. The implication $(1) \Rightarrow (2)$ is immediate. For $(2) \Rightarrow (1)$, define

$$R(A) := \frac{P(A) - \kappa P_*(A)}{1 - \kappa}.$$

Then R is a probability measure and $P = \kappa P_* + (1 - \kappa)R$.

Next, $(2) \Rightarrow (3)$. Consider the Lebesgue decomposition of P with respect to P_* ,

$$P = uP_* + P^\perp, \quad P^\perp \perp P_*,$$

where $u := dP_{ac}/dP_*$ is the Radon–Nikodym derivative of the absolutely continuous part $P_{ac} \ll P_*$ of P with respect to P_* , and $P^\perp$ is its singular part. Choose a measurable set N such that $P_*(N) = 0$ and $P^\perp(N^c) = 0$. Note that such N exists by the Lebesgue decomposition. Define

$$B := \{z \in \mathcal{Z} : u(z) < \kappa\}.$$

Then

$$P(B \cap N^c) = \int_{B \cap N^c} u dP_* < \kappa P_*(B \cap N^c),$$

unless $P_*(B \cap N^c) = 0$. But by (2) we must have

$$P(B \cap N^c) \geq \kappa P_*(B \cap N^c).$$

Hence $P_*(B \cap N^c) = 0$, and since $P_*(N) = 0$, it follows that $P_*(B) = 0$. Therefore, $u \geq \kappa$ P_* -a.s.

Finally, $(3) \Rightarrow (2)$ follows by integrating:

$$P(A) = \int_A u dP_* + P^\perp(A) \geq \int_A \kappa dP_* = \kappa P_*(A).$$

When $P \ll P_*$, we have $P^\perp = 0$ and $u = dP/dP_*$, which gives the final claim of part (ii). ■

SA8.10 Proof of Proposition SA-13

Proof. We prove each part separately, writing throughout the Lebesgue decomposition $P = wP_* + P^\perp$ of a candidate $P \in \Delta(\mathcal{Z})$ with respect to P_* , with $P^\perp \perp P_*$ and $w := dP_{ac}/dP_*$, where w is the Radon–Nikodym derivative of the absolutely continuous part $P_{ac} \ll P_*$ of P with respect to P_* .

Part (i): baseline as the mixture. The function $\bar{\phi} = \phi + \iota_{[0, 1/\kappa]}$ is convex and lower semicontinuous as the sum of a convex lower semicontinuous function and the indicator of the closed convex set $[0, 1/\kappa]$. Moreover, $\bar{\phi}(1) = \phi(1) + \iota_{[0, 1/\kappa]}(1) = 0$ because $1 \in [0, 1/\kappa]$ (as $\kappa < 1$), and $\bar{\phi}(s) = +\infty$ for $s < 0$. Hence $\bar{\phi}$ is an entropy function in the sense of Definition SA-4. Its effective domain $\text{dom}(\bar{\phi}) = \text{dom}(\phi) \cap [0, 1/\kappa]$ is bounded, so $\bar{\phi}^\infty(1) = +\infty$. By Lemma SA-1, applied to the entropy function $\bar{\phi}$,

$$D_{\bar{\phi}}(P \| P_*) = \int \bar{\phi}(w) dP_* + \bar{\phi}^\infty(1) P^\perp(\mathcal{Z}).$$

By Lemma SA-12(i), the mixture restriction (7) is equivalent to $P \ll P_*$ and $0 \leq w \leq 1/\kappa$ P_* -a.s. If it holds, then $P^\perp(\mathcal{Z}) = 0$ and $\iota_{[0, 1/\kappa]}(w) = 0$ P_* -a.s., whence $D_{\bar{\phi}}(P \| P_*) = \int \bar{\phi}(w) dP_* = D_\phi(P \| P_*)$. If it fails, then either $P^\perp(\mathcal{Z}) > 0$, so the second term is $+\infty$ since $\bar{\phi}^\infty(1) = +\infty$, or $w > 1/\kappa$ on a set of positive P_* -measure, so $\iota_{[0, 1/\kappa]}(w) = +\infty$ there and the first term is $+\infty$. Either way, $D_{\bar{\phi}}(P \| P_*) = +\infty$. As the mixture restriction is equivalent to $\kappa P \leq P_*$ by the same lemma, this proves the stated identity for $\bar{\phi}$. Therefore $D_{\bar{\phi}}(P \| P_*) \leq \delta$ holds if and only if the mixture restriction holds and $D_\phi(P \| P_*) \leq \delta$, which by (7) is exactly membership in $\mathcal{P}_{\text{mix}}^{\leq}(\rho, \delta)$. This proves $\mathcal{P}_{\text{mix}}^{\leq}(\rho, \delta) = \mathcal{P}_{\bar{\phi}}(\rho, \delta)$.

Part (ii): baseline as a mixing component. The function $\underline{\phi} = \phi + \iota_{[\kappa, \infty)}$ is convex and lower semicontinuous as the sum of a convex lower semicontinuous function and the indicator of the closed convex set $[\kappa, \infty)$. Moreover, $\underline{\phi}(1) = \phi(1) + \iota_{[\kappa, \infty)}(1) = 0$ because $1 \in [\kappa, \infty)$ (as $\kappa < 1$), and $\underline{\phi}(s) = +\infty$ for $s < 0$. Hence $\underline{\phi}$ is an entropy function in the sense of Definition SA-4. Since $\text{dom}(\underline{\phi}) = \text{dom}(\phi) \cap [\kappa, \infty)$ coincides with $\text{dom}(\phi)$ for large arguments, the truncation leaves the upper-tail recession unchanged, so $\underline{\phi}^\infty(1) = \phi^\infty(1)$. By Lemma SA-1, applied to the entropy function $\underline{\phi}$, together with $\underline{\phi}^\infty(1) = \phi^\infty(1)$,

$$D_{\underline{\phi}}(P \| P_*) = \int \underline{\phi}(w) dP_* + \underline{\phi}^\infty(1) P^\perp(\mathcal{Z}) = \begin{cases} D_\phi(P \| P_*), & w \geq \kappa \text{ } P_*$$
-a.s., \\ +\infty, & \text{otherwise,} \end{cases}

since when $w \geq \kappa$ P_* -a.s. we have $\iota_{[\kappa, \infty)}(w) = 0$ P_* -a.s., so $\int \underline{\phi}(w) dP_* = \int \phi(w) dP_*$ and the singular term equals $\phi^\infty(1) P^\perp(\mathcal{Z})$, whereas otherwise $\iota_{[\kappa, \infty)}(w) = +\infty$ on a set of positive P_* -measure. By Lemma SA-12(ii), the condition $w \geq \kappa$ P_* -a.s. is exactly the mixture restriction (8), as well as the dominance $\kappa P_* \leq P$, which proves the stated identity for $\underline{\phi}$. Hence $D_{\underline{\phi}}(P \| P_*) \leq \delta$ holds if and only if the mixture restriction holds and $D_\phi(P \| P_*) \leq \delta$, which is exactly membership in $\mathcal{P}_{\text{mix}}(\rho, \delta)$. This proves $\mathcal{P}_{\text{mix}}(\rho, \delta) = \mathcal{P}_{\underline{\phi}}(\rho, \delta)$. ■

SA8.11 Proof of Theorem SA-14

Proof. By Proposition SA-13, each mixture-constrained robustness set is an ordinary robustness set generated by an induced entropy. Both parts therefore follow from Theorem SA-11 once its hypotheses are verified for the corresponding entropy.

Part (i): baseline as the mixture. Proposition SA-13(i) gives

$$\mathcal{P}_{\text{mix}}^{\leq}(\rho, \delta) = \mathcal{P}_{\bar{\phi}}(\rho, \delta), \quad \bar{\phi} = \phi + \iota_{[0, 1/\kappa]}.$$

The induced entropy is continuous at 1 because it coincides with ϕ on a neighborhood of 1, and $\bar{\phi}^\infty(1) = +\infty$. Moreover, every $\nu \in \mathcal{M}_+(\mathcal{Z})$ with $D_{\bar{\phi}}(\nu \| P_*) < \infty$ satisfies $\nu \ll P_*$ and $d\nu/dP_* \leq 1/\kappa$ P_* -almost surely. Hence

$$\int_{\mathcal{Z}} |g(z)| d\nu(z) \leq \frac{1}{\kappa} \mathbb{E}_{P_*}[|g(Z)|] < \infty,$$

where the last inequality holds as $P_* \in \mathcal{P}(\mathcal{Z})$. The compatibility condition is imposed in the statement, so Theorem SA-11 applies with $\bar{\phi}$ in place of ϕ . Its pointwise constraint is vacuous because the recession is infinite, and the resulting dual is the one displayed in part (i).

Part (ii): baseline as a mixing component. Proposition SA-13(ii) gives

$$\mathcal{P}_{\text{mix}}(\rho, \delta) = \mathcal{P}_{\underline{\phi}}(\rho, \delta), \quad \underline{\phi} = \phi + \iota_{[\kappa, \infty)}.$$

Again, the induced entropy is continuous at 1 because it coincides with ϕ on a neighborhood of 1, while $\underline{\phi}^\infty(1) = \phi^\infty(1)$. The integrability and compatibility conditions required by Theorem SA-11 are imposed directly in the statement. Applying that theorem with $\underline{\phi}$ in place of ϕ therefore gives the dual in part (ii). ■

SA8.12 Proofs of Theorem SA-15 and Proposition SA-16

Proof. We first establish a preliminary fact about the deletion of the constraint, then prove the proposition, and finally prove the theorem. Throughout, we write $y_\lambda(z) := g(z) - \lambda^\top m(z)$, $s(\lambda) := \sup_{z \in \mathcal{Z}} y_\lambda(z)$, and $e(\lambda) := P_*\text{-ess sup}_{z \in \mathcal{Z}} y_\lambda(z)$. As $P_* \in \mathcal{P}(\mathcal{Z})$, $y_\lambda \in L_1(P_*)$, so $e(\lambda) > -\infty$. Recall from (11) the dual objective f of

Theorem SA-9, whose expectation term has the integrand $(\phi^*)^\pi(y_\lambda(Z) - \zeta, \eta)$. Lemma SA-4 bounds the integrand from below by $y_\lambda - \zeta$, so this expectation is well defined, possibly with value $+\infty$. Recall also from (12) the dual feasibility set $\mathcal{F} = \{(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda : \zeta + \eta \phi^\infty(1) \geq s(\lambda)\}$, and define its essential-supremum counterpart

$$\mathcal{F}^\ll := \{(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda : \zeta + \eta \phi^\infty(1) \geq e(\lambda)\}, \quad (69)$$

noting that $\mathcal{F} \subseteq \mathcal{F}^\ll \subseteq \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$, as the essential supremum never exceeds the supremum. Throughout this proof the budget δ is held fixed, and we suppress it from the notation, writing f for the dual objective at this budget and V_P for $V_P(\delta)$, and likewise for the three remaining values of (10) and (13). To prove the theorem we need to show that $V_P = V_P^\ll$, whereas to prove the proposition we need to show that $V_P^\ll = V_D^\circ$.

The essential-supremum constraint. We first show that $f(\zeta, \eta, \lambda) = +\infty$ whenever $(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$ lies outside $\mathcal{F}^\ll$. Consequently,

$$V_D^\circ = \inf_{(\zeta, \eta, \lambda) \in \mathcal{F}^\ll} f(\zeta, \eta, \lambda). \quad (70)$$

In this sense, deleting the constraint (9) amounts to replacing the supremum $s(\lambda)$ in (12) with the essential supremum $e(\lambda)$ in (69). To accomplish this, we prove two claims. Claim 1: if $\phi^\infty(1) < \infty$, then $[1, +\infty) \subseteq \text{dom}(\phi)$. Claim 2: the recession function of the conjugate satisfies $(\phi^*)^\infty(x) = +\infty$ for every $x > 0$. Claim 1 follows because, by Definition SA-2 with base point $z_0 = 1$, the quantity $\alpha^{-1}\{\phi(1+\alpha) - \phi(1)\}$ is nondecreasing in $\alpha > 0$ with limit $\phi^\infty(1)$, so that $\phi(1+\alpha) \leq \alpha \phi^\infty(1) < \infty$ for every $\alpha > 0$. For Claim 2, ϕ is closed, proper, and convex, so $\phi^{**} = \phi$, and ϕ^* is closed, proper, and convex by Lemma SA-3. Then Theorem 13.3 in Rockafellar (1970), applied with $f = \phi^*$, identifies $(\phi^*)^\infty$ with the support function of $\text{dom}(\phi^{**}) = \text{dom}(\phi)$, hence $(\phi^*)^\infty(x) = \sup_{s \in \text{dom}(\phi)} sx \geq \sup_{s \geq 1} sx = +\infty$ for every $x > 0$, in view of Claim 1.

Fix $(\zeta, \eta, \lambda) \in (\mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda) \setminus \mathcal{F}^\ll$, so that $\zeta + \eta \phi^\infty(1) < e(\lambda)$. This forces $\phi^\infty(1) < \infty$, because if $\phi^\infty(1) = +\infty$, the left-hand side would equal $+\infty$ for $\eta > 0$ and, under the convention stated after (9), for $\eta = 0$ as well, a contradiction. Write $c := \zeta + \eta \phi^\infty(1)$ and note that $P_*\{y_\lambda > c\} > 0$, because if this probability were zero, c would be an essential upper bound for y_λ and $e(\lambda) \leq c$, which is again a contradiction. Consider two cases. If $\eta > 0$, then on the event $\{y_\lambda > c\}$ we have $(y_\lambda - \zeta)/\eta > \phi^\infty(1)$, which lies strictly above $\sup \text{dom}(\phi^*)$ in view of Corollary SA-7, and so $\phi^*((y_\lambda - \zeta)/\eta) = +\infty$ there and $(\phi^*)^\pi(y_\lambda - \zeta, \eta) = \eta \phi^*((y_\lambda - \zeta)/\eta) = +\infty$ on a set of positive P_* -probability. If $\eta = 0$, the constraint reads $\zeta < e(\lambda)$, so $P_*\{y_\lambda - \zeta > 0\} > 0$ by the same argument, and on that event $(\phi^*)^\pi(y_\lambda - \zeta, 0) = (\phi^*)^\infty(y_\lambda - \zeta) = +\infty$ by the consequence recorded above. In either case the integrand equals $+\infty$ on a set of positive P_* -probability, so the expectation in (11) equals $+\infty$. Hence $f(\zeta, \eta, \lambda) = +\infty$. The infimum of f over $\mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$ therefore coincides with its infimum over $\mathcal{F}^\ll$, which proves (70).

Proof of Proposition SA-16. Define the convex cone $\mathcal{M}_{+,m}^\ll(\mathcal{Z}) := \{\nu \in \mathcal{M}_{+,m}(\mathcal{Z}) : \nu \ll P_*\}$ and the function

$$h^\ll(a, b, x) := \inf_{\nu \in \mathcal{M}_{+,m}^\ll(\mathcal{Z})} \left\{ - \int g(z) d\nu(z) : \int m(z) d\nu(z) = x, \int d\nu(z) = a, D_\phi(\nu \| P_*) \leq b \right\},$$

so that, by (62), $V_P^\ll = \sup_{\bar{m} \in \mathcal{M}(\rho)} -h^\ll(1, \delta, \bar{m})$. We compute its conjugate and then apply Lemma SA-8, as in Theorem SA-11.

Convexity and the conjugate on $\{\eta > 0\}$. Joint convexity of $h^\ll$ follows exactly as in the proof of Theorem SA-9. For the conjugate, fix $\eta > 0$. The computation leading to (63) is unchanged, with the supremum now running over the feasible $\nu \ll P_*$, that is, the $\nu \in \mathcal{M}_{+,m}^\ll(\mathcal{Z})$ with $D_\phi(\nu \| P_*) < \infty$. We may thus take $\varrho = P_*$, so that $\beta \equiv 1$ holds P_* -almost everywhere and $\phi^\pi(\alpha, \beta) = \phi(\alpha)$. By the claim established in the proof of Theorem SA-9, the supremum over the feasible $\nu \ll P_*$ then equals, writing $y := y_\lambda$,

$$\sup_{\alpha \in L_1(P_*)} \int [(y(z) - \zeta)\alpha(z) - \eta \phi(\alpha(z))] dP_*(z), \quad (71)$$

where the integral is assigned the value $-\infty$ whenever the negative part of its integrand fails to be integrable. In particular, neither the moment restriction nor the finiteness of the divergence needs to be imposed on α . The interchange of supremum and integral is granted by Theorem 14.60 in Rockafellar and Wets (2009), and the definition of the conjugate, $\sup_{u \in \mathbb{R}} \{tu - \eta \phi(u)\} = \eta \phi^*(t/\eta)$ for $t \in \mathbb{R}$ and $\eta > 0$, yields

$$h^{\ll,*}(-\zeta, -\eta, -\lambda) = \mathbb{E}_{P_*} \left[\eta \phi^* \left(\frac{y(Z) - \zeta}{\eta} \right) \right], \quad \eta > 0.$$

Extension to $\eta = 0$. Let $F^\circ(\zeta, \eta, \lambda) := \mathbb{E}_{P_*}[(\phi^*)^\pi(y(Z) - \zeta, \eta)]$. The corresponding step of the proof of Theorem SA-9 shows that this expectation is closed and convex in (ζ, η, λ) , while the conjugate $h^{\ll,*}$ is closed and convex. Both functions are nonincreasing in η : $h^{\ll,*}$ through its representation as a supremum over pairs (ν, b) with $\nu \in \mathcal{M}_{+,m}^\ll(\mathcal{Z})$ and $b \geq D_\phi(\nu \| P_*) \geq 0$, whose objective is nonincreasing in η pointwise in (ν, b) , and F° through the identity $(\phi^*)^\pi(x, \eta) = \sup_{u \in \text{dom}(\phi)} \{ux - \eta \phi(u)\}$ for $\eta \geq 0$ together with $\phi \geq 0$. Neither takes the value $-\infty$: $F^\circ \geq \mathbb{E}_{P_*}[y(Z)] - \zeta > -\infty$ by Lemma SA-4, and $h^{\ll,*}(-\zeta, -\eta, -\lambda) \geq \mathbb{E}_{P_*}[y(Z)] - \zeta > -\infty$ by evaluating its defining supremum at $(\nu, b) = (P_*, 0)$, which is admissible because $P_* \in \mathcal{M}_{+,m}^\ll(\mathcal{Z})$. Since the two functions agree on $\{\eta > 0\}$ by the previous step, an argument identical to the one in the proof of Theorem SA-9 yields $h^{\ll,*}(-\zeta, -\eta, -\lambda) = F^\circ(\zeta, \eta, \lambda)$ for every $(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}$.

The domain of $h^\ll$ and the assembly. The proof of Lemma SA-10 applies step by step upon replacing $\Gamma(\delta)$

with $\Gamma^{\ll}(\delta)$ and choosing every auxiliary measure to be absolutely continuous. Hence $(1, \delta, \omega) \in \text{ri}(\text{dom}(h^{\ll}))$ for every $\omega \in \text{ri}(\Gamma^{\ll}(\delta))$ and $h^{\ll}$ is proper. The set K_ρ defined before Lemma SA-10 is nonempty and polyhedral. By the compatibility condition, there is an $\omega_0 \in \mathcal{M}(\rho) \cap \text{ri}(\Gamma^{\ll}(\delta))$, and therefore $K_\rho \cap \text{ri}(\text{dom}(h^{\ll})) \neq \emptyset$. Applying Lemma SA-8 to $h^{\ll}$ and ι_{K_ρ} , and using the conjugate computed above together with the support function of $\mathcal{M}(\rho)$ recorded in Section SA1.2 gives

$$V_P^{\ll} = \inf_{(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda} f(\zeta, \eta, \lambda) = V_D^\circ,$$

which proves the proposition.

Proof of Theorem SA-15. Suppose condition (a) holds, that is, $\phi^\infty(1) = +\infty$. Then, by Lemma SA-1 and $\phi \geq 0$, every P with $D_\phi(P \| P_*) < \infty$ has a singular part of zero mass, that is, $P \ll P_*$. Hence the two robustness sets coincide and $V_P = V_P^{\ll}$ follows immediately. Suppose instead that condition (b) holds. Then the equality $e(\lambda) = s(\lambda)$ for every $\lambda \in \mathcal{R}_\lambda$, assumed in condition (b), makes the defining inequalities of (12) and (69) identical, so $\mathcal{F} = \mathcal{F}^{\ll}$, and (70) gives $V_D^\circ = \inf_{\mathcal{F}^{\ll}} f = \inf_{\mathcal{F}} f = V_D$. Because Proposition SA-16 holds under the conditions of the theorem, we get $V_P^{\ll} = V_D^\circ$. Moreover, $V_P \leq V_D$ by Theorem SA-9, whose integrability hypothesis is implied by condition (b), as $\mathcal{M}_{+,m}(\mathcal{Z}) \subseteq \mathcal{M}_+(\mathcal{Z})$, while $V_P^{\ll} \leq V_P$, as the supremum defining V_P runs over a larger robustness set. Using all these facts, we obtain

$$V_P^{\ll} \leq V_P \leq V_D = V_D^\circ = V_P^{\ll},$$

so that the four values coincide. In particular $V_P = V_P^{\ll}$, which was to be shown. $\blacksquare$

SA8.13 Proof of Lemma SA-17

Proof. Recall that $\mathcal{C}_\kappa(P_*) = \{P \in \Delta(\mathcal{Z}) : P = \kappa P_* + (1 - \kappa)R, R \in \Delta(\mathcal{Z})\}$ and $\mathcal{C}_\kappa^{\ll}(P_*) = \{P \in \Delta(\mathcal{Z}) : P_* = \kappa P + (1 - \kappa)R, R \in \Delta(\mathcal{Z})\}$, and that $\mathcal{C}_{\kappa,\delta}(P_*)$ and $\mathcal{C}_{\kappa,\delta}^{\ll}(P_*)$ are defined in (14) by intersecting these two sets with $\mathcal{T}_\delta(P_*)$. Fix $\kappa \in (0, 1)$ and $\delta \in [0, 1]$. The endpoints $\kappa \in \{0, 1\}$ are treated at the end.

Step 1: domination constraints and uniqueness of the mixing distribution. By construction, $P \in \mathcal{C}_\kappa(P_*)$ if and only if there exists $R \in \Delta(\mathcal{Z})$ with $P = \kappa P_* + (1 - \kappa)R$, which is condition (1) of Lemma SA-12(ii). By the equivalence (1) $\Leftrightarrow$ (2) of that lemma this holds if and only if $P \geq \kappa P_*$ as measures, thus

$$\mathcal{C}_\kappa(P_*) = \{P \in \Delta(\mathcal{Z}) : P \geq \kappa P_*\}. \quad (72)$$

Symmetrically, $P \in \mathcal{C}_\kappa^{\ll}(P_*)$ if and only if there exists $R \in \Delta(\mathcal{Z})$ with $P_* = \kappa P + (1 - \kappa)R$, which is condition (1) of Lemma SA-12(i). By the equivalence (1) $\Leftrightarrow$ (2) of that lemma this holds if and only if $\kappa P \leq P_*$ as measures, so that

$$\mathcal{C}_\kappa^{\ll}(P_*) = \{P \in \Delta(\mathcal{Z}) : \kappa P \leq P_*\}. \quad (73)$$

Moreover, since $\kappa < 1$, the two mixture equations can be solved for the mixing distribution, giving $R = (P - \kappa P_*)/(1 - \kappa)$ and $R = (P_* - \kappa P)/(1 - \kappa)$, respectively, so that in both cases the mixing distribution is uniquely determined by P and P_* , as claimed in the statement.

Step 2: the rescaling identities. Let $P \in \mathcal{C}_\kappa(P_*)$ with mixing distribution R . Substituting the mixture equation gives $P(A) - P_*(A) = (1 - \kappa)(R(A) - P_*(A))$ for every $A \in \mathcal{F}$, and taking absolute values and suprema over $A \in \mathcal{F}$ yields $\text{TV}(P \| P_*) = (1 - \kappa)\text{TV}(R \| P_*)$. Let instead $P \in \mathcal{C}_\kappa^{\ll}(P_*)$ with mixing distribution R . The same substitution gives $P_*(A) - P(A) = (1 - \kappa)(R(A) - P(A))$ and $P_*(A) - R(A) = \kappa(P(A) - R(A))$ for every $A \in \mathcal{F}$, whence $\text{TV}(P \| P_*) = (1 - \kappa)\text{TV}(P \| R)$ and $\text{TV}(R \| P_*) = \kappa \text{TV}(P \| R)$. Since the total variation distance never exceeds one, these identities imply in particular that every distribution in $\mathcal{C}_\kappa(P_*)$ or in $\mathcal{C}_\kappa^{\ll}(P_*)$ lies within total variation distance $1 - \kappa$ of the baseline.

Step 3: the two characterizations. Consider part (ii) first. By definition, $\mathcal{C}_{\kappa,\delta}(P_*)$ collects the distributions $P = \kappa P_* + (1 - \kappa)R$ with $R \in \Delta(\mathcal{Z})$ and $\text{TV}(P \| P_*) \leq \delta$. By Step 2, the latter constraint holds if and only if $\text{TV}(R \| P_*) \leq \delta/(1 - \kappa)$, which proves the first equality in (17), while the second follows upon intersecting both sides of (72) with $\mathcal{T}_\delta(P_*)$. The argument for part (i) is identical, with (73) in place of (72) and with the identity $\text{TV}(P \| P_*) = (1 - \kappa)\text{TV}(P \| R)$ turning the constraint $\text{TV}(P \| P_*) \leq \delta$ into $\text{TV}(P \| R) \leq \delta/(1 - \kappa)$, which proves (16). Next, the containment $\mathcal{C}_{\kappa,\delta}(P_*) \cup \mathcal{C}_{\kappa,\delta}^{\ll}(P_*) \subseteq \mathcal{T}_{\min\{\delta, 1-\kappa\}}(P_*)$ follows by combining the constraint $\text{TV}(P \| P_*) \leq \delta$, valid on both sets by definition, with the bound $\text{TV}(P \| P_*) \leq 1 - \kappa$ obtained at the end of Step 2. Finally, if $\delta \geq 1 - \kappa$, the same bound gives $\mathcal{C}_\kappa(P_*) \subseteq \mathcal{T}_{1-\kappa}(P_*) \subseteq \mathcal{T}_\delta(P_*)$ and $\mathcal{C}_\kappa^{\ll}(P_*) \subseteq \mathcal{T}_\delta(P_*)$, so intersecting either contamination neighborhood with $\mathcal{T}_\delta(P_*)$ leaves it unchanged, that is, $\mathcal{C}_{\kappa,\delta}(P_*) = \mathcal{C}_\kappa(P_*)$ and $\mathcal{C}_{\kappa,\delta}^{\ll}(P_*) = \mathcal{C}_\kappa^{\ll}(P_*)$.

Endpoints. If $\kappa = 0$, both mixture equations hold with the unique choices $R = P$ and $R = P_*$, respectively, so $\mathcal{C}_0(P_*) = \mathcal{C}_0^{\ll}(P_*) = \Delta(\mathcal{Z})$, matching the domination sets $\{P \geq 0\}$ and $\{0 \leq P_*\}$, and the rescaling identities hold by direct substitution. Steps 2 and 3 then apply verbatim, and both characterizations reduce to $\mathcal{C}_{0,\delta}(P_*) = \mathcal{C}_{0,\delta}^{\ll}(P_*) = \mathcal{T}_\delta(P_*)$. If $\kappa = 1$, both mixture equations force $P = P_*$, so $\mathcal{C}_1(P_*) = \mathcal{C}_1^{\ll}(P_*) = \{P_*\}$. Since $\text{TV}(P_* \| P_*) = 0 \leq \delta$, intersecting with $\mathcal{T}_\delta(P_*)$ gives $\mathcal{C}_{1,\delta}(P_*) = \mathcal{C}_{1,\delta}^{\ll}(P_*) = \{P_*\}$ for every $\delta \in [0, 1]$. $\blacksquare$

SA8.14 Proof of Lemma SA-18

Proof. We prove each part separately. Throughout, we fix $\theta \in \Theta$ and recall the image $\mathcal{N}(\theta)$ of the moment function from Section SA3.1. For this proof it is convenient to permute the coordinates of $\mathcal{M}(\rho)$ so that the $\mathbf{d}_{m,\leq}$

inequality constraints precede the $\mathbf{d}_{m,=}$ equality constraints, where $\mathbf{d}_{m,\leq} := \mathbf{d}_1 + \mathbf{d}_3$ and $\mathbf{d}_{m,=} := \mathbf{d}_2 + \mathbf{d}_4$, and to write $\lambda_{\leq} \in \mathbb{R}^{\mathbf{d}_{m,\leq}}$ for the inequality multipliers and $\lambda_{=} \in \mathbb{R}^{\mathbf{d}_{m,=}}$ for the equality multipliers, so that $\lambda = (\lambda_{\leq}, \lambda_{=})$. The mapping with the previous notation is thus $\lambda_{\leq} = (\lambda_1, \lambda_3)$ and $\lambda_{=} = (\lambda_2, \lambda_4)$, and the vector $\vec{\rho}$ is permuted accordingly, so that its first $\mathbf{d}_{m,\leq}$ components are $(\rho_1^\top, 0_{\mathbf{d}_3}^\top)^\top$ and its last $\mathbf{d}_{m,=}$ components are $(\rho_2^\top, 0_{\mathbf{d}_4}^\top)^\top$. Finally, we write $\lambda_{12} = (\lambda_1^\top, \lambda_2^\top)^\top \in \mathbb{R}^{\mathbf{d}_e}$ and use the notation $\xi = (\zeta, \eta, \lambda^\top)^\top$ to gather all multipliers together.

Part (i). Under Assumption SA-1, and under Assumption SA-2 in case (a) of Assumption SA-2(i), the Slater condition of Assumption SA-1(iv) or of Assumption SA-2(vi) places the origin in the interior of $\text{co}(\mathcal{N}(\theta)) - \vec{\rho}$. In case (b), every distribution Q in the definition of $\mathcal{N}_{\bar{x}}(\theta)$ satisfies $\mathbb{E}_Q[\|m(Z; \theta)\|] \leq \bar{x} \mathbb{E}_{P_*}[\|m(Z; \theta)\|] < \infty$, as $P_* \in \mathcal{P}(\mathcal{Z})$, so that the separation argument in the proof of Lemma SA-20(i) gives $\mathcal{N}_{\bar{x}}(\theta) \subseteq \text{cl}(\text{co}(\mathcal{N}(\theta)))$. The Slater condition of Assumption SA-2(vi) then places the origin in the interior of $\text{cl}(\text{co}(\mathcal{N}(\theta))) - \vec{\rho}$, and hence in the interior of $\text{co}(\mathcal{N}(\theta)) - \vec{\rho}$, as a convex set whose closure has nonempty interior has the same interior as its closure (Rockafellar, 1970, Theorem 6.3). In either case, the origin 0 lies in the interior of $\text{co}(\mathcal{N}(\theta)) - \vec{\rho} = \text{co}(\mathcal{N}(\theta) - \vec{\rho}) \subseteq \mathbb{R}^{\mathbf{d}_m}$, where the equality follows as the convex hull is translation invariant. By definition of the interior in $\mathbb{R}^{\mathbf{d}_m}$, there exists $r > 0$ such that $\mathcal{B}_r(0) \subseteq \text{co}(\mathcal{N}(\theta) - \vec{\rho})$. Set $\varepsilon := r/(2\sqrt{\mathbf{d}_m})$ and consider the point $v_+ \in \mathbb{R}^{\mathbf{d}_m}$ with components $v_{+,j} := -\varepsilon$ for $j = 1, \dots, \mathbf{d}_{m,\leq}$ and $v_{+,j} := +\varepsilon$ for $j = \mathbf{d}_{m,\leq} + 1, \dots, \mathbf{d}_m$. Since $\|v_+\| = \varepsilon\sqrt{\mathbf{d}_m} = r/2 < r$, we have $v_+ \in \mathcal{B}_r(0) \subseteq \text{co}(\mathcal{N}(\theta) - \vec{\rho})$. By Carathéodory's theorem in $\mathbb{R}^{\mathbf{d}_m}$, there exist points $z_1, \dots, z_{\mathbf{d}_m+1} \in \mathcal{Z}$ and weights $\omega_1, \dots, \omega_{\mathbf{d}_m+1} \geq 0$ with $\sum_{i=1}^{\mathbf{d}_m+1} \omega_i = 1$ such that

$$\sum_{i=1}^{\mathbf{d}_m+1} \omega_i (m(z_i; \theta) - \vec{\rho}) = v_+.$$

Use these weights to construct the finitely supported distribution

$$Q_+ := \sum_{i=1}^{\mathbf{d}_m+1} \omega_i \delta_{z_i}^{\text{Dirac}} \in \Delta(\mathcal{Z}).$$

Then, $\mathbb{E}_{Q_+}[m_j(Z; \theta)] = \vec{\rho}_j - \varepsilon < \vec{\rho}_j$ for $j \leq \mathbf{d}_{m,\leq}$ and $\mathbb{E}_{Q_+}[m_j(Z; \theta)] = \vec{\rho}_j + \varepsilon$ for $j > \mathbf{d}_{m,\leq}$, as required. The distribution Q_- is constructed symmetrically by replacing v_+ with $v_- := -\varepsilon \mathbf{1}_{\mathbf{d}_m} \in \mathbb{R}^{\mathbf{d}_m}$.

Part (ii). Under Assumption SA-2 the constraint is vacuous, hence $\mathcal{F}(\theta) = \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$ and there is nothing more to prove. Consider now Assumption SA-1, so that $\phi^\infty(1) \in (0, +\infty)$ by Assumption SA-1(ii). Set $\lambda^\circ := 0 \in \mathcal{R}_\lambda$, $\eta^\circ := 1$, and $\zeta^\circ := \bar{g}$, where $\sigma(0, \theta) = \bar{g} < \infty$ by Assumption SA-1(i). Then $\xi^\circ \in \mathbb{R} \times \mathbb{R}_{++} \times \mathcal{R}_\lambda$ and $\zeta^\circ + \eta^\circ \phi^\infty(1) = \bar{g} + \phi^\infty(1) > \bar{g} = \sigma(\lambda^\circ, \theta)$, where the inequality follows from $\phi^\infty(1) > 0$, so ξ° is a strictly feasible point of $\mathcal{F}(\theta)$, which was to be shown.

Part (iii). For a set S and a level $\alpha \in \mathbb{R}$, we write $\text{lev}_{\leq \alpha}(f, S) := \{x \in S : f(x) \leq \alpha\}$ for the sublevel set of f relative to S . The argument rests on the following recession criterion for a closed proper convex function on $\mathbb{R}^n$: it has a nonempty bounded sublevel set if and only if its recession function is strictly positive at every nonzero direction, in which case all its nonempty sublevel sets are bounded. This follows from Theorems 8.4 and 8.7 and Corollary 8.7.1 in Rockafellar (1970), since all nonempty sublevel sets of such a function share the recession cone formed by the directions at which its recession function is nonpositive, a nonempty closed convex set is bounded if and only if its recession cone is $\{0\}$, and the recession function vanishes at the origin. We apply this criterion to the extension of $f(\cdot, \theta)$ by $+\infty$ outside $\mathcal{F}(\theta)$, which we now show is closed, proper, and convex on $\mathbb{R}^{\mathbf{d}_m+2}$.

Properties of f . By Lemma SA-3, ϕ^* and its perspective $(\phi^*)^\pi$ are both convex, proper, and lower semicontinuous. The term within the expectation admits an integrable lower bound that is uniform along any convergent sequence of multipliers $\xi_s \rightarrow \xi$. Indeed, under Assumption SA-1, Lemma SA-4 at $\tilde{x} = 0$ gives $(\phi^*)^\pi(x, \eta) \geq -\eta \phi(0)$ with $\phi(0) < \infty$ by Assumption SA-1(ii), whereas under Assumption SA-2, the same lemma at $\tilde{x} = 1$ gives

$$(\phi^*)^\pi(g(z; \theta) - \lambda^\top m(z; \theta) - \zeta, \eta) \geq g(z; \theta) - \lambda^\top m(z; \theta) - \zeta \geq -(|g(z; \theta)| + \sup_s \|\lambda_s\| \|m(z; \theta)\| + \sup_s |\zeta_s|),$$

which is a P_* -integrable bound as $P_* \in \mathcal{P}(\mathcal{Z})$. Therefore, the expectation $\mathbb{E}_{P_*}[(\phi^*)^\pi(g(Z; \theta) - \lambda^\top m(Z; \theta) - \zeta, \eta)]$ is convex, because integration preserves convexity, and lower semicontinuous. Adding the affine continuous term $\lambda_{12}^\top \rho + \zeta + \eta \delta$ preserves both properties, so $f(\cdot, \theta)$ is convex and lower semicontinuous on $\mathbb{R}^{\mathbf{d}_m+2}$, hence closed.

Properties of the extension of f . If $\phi^\infty(1) = +\infty$, then the dual constraint is vacuous for every $\eta \geq 0$ – its left-hand side $\zeta + \eta \phi^\infty(1)$ equals $+\infty$, for $\eta > 0$ because $\eta \phi^\infty(1) = +\infty$ and for $\eta = 0$ under the convention $0 \cdot \phi^\infty(1) = +\infty$ – so that $\mathcal{F}(\theta)$ coincides with the ambient cone $\mathcal{L}$ of Section SA3.1, which after the permutation reads $\mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+^{\mathbf{d}_{m,\leq}} \times \mathbb{R}^{\mathbf{d}_{m,=}}$ and is closed and convex. The closedness and convexity of the extension of $f(\cdot, \theta)$ by $+\infty$ outside $\mathcal{F}(\theta)$ then follow directly from the lower semicontinuity and convexity of f established above. If instead $\phi^\infty(1) < \infty$, rewrite

$$\mathcal{F}(\theta) = \{\xi \in \mathbb{R}^{\mathbf{d}_m+2} : \lambda_{\leq} \geq 0, \eta \geq 0, \sigma(\lambda, \theta) \leq \zeta + \eta \phi^\infty(1)\},$$

with $\sigma(\lambda, \theta)$ the dual support function of Section SA3.1, so that $\mathcal{F}(\theta)$ is the intersection of the polyhedral cone $\{\lambda_{\leq} \geq 0, \eta \geq 0\}$ with the set $\{\xi : \sigma(\lambda, \theta) \leq \zeta + \eta \phi^\infty(1)\}$, the preimage of the epigraph of σ under the linear map $\xi \mapsto (\lambda, \zeta + \eta \phi^\infty(1))$. Note that since σ is a supremum of affine functions in λ , it is convex and its epigraph is convex. Moreover, it is the pointwise supremum of continuous functions of λ , thus it is lower semicontinuous and

its epigraph is closed. As the preimage of a closed convex set under a linear map, $\{\xi : \sigma(\lambda, \theta) \leq \zeta + \eta\phi^\infty(1)\}$ is closed and convex, and therefore $\mathcal{F}(\theta)$ is the intersection of two closed convex sets, hence itself closed and convex. The extension is then the sum of f and the indicator function of the closed convex set $\mathcal{F}(\theta)$, so its closedness and convexity again follow from the corresponding properties of f established above.

Properness of f . The function $f(\cdot, \theta)$ never takes the value $-\infty$ because at each ξ the term within the expectation admits an integrable lower bound, as shown in the previous paragraph. Moreover, f is not identically $+\infty$. Under Assumption SA-1, it is finite at the point $\xi_0 = (\bar{g}, 1, 0)$ with $\bar{g} = \sup_{z \in \mathcal{Z}} g(z; \theta) < \infty$, where the finiteness follows by Assumption SA-1(i) and which is feasible since $\bar{g} + \phi^\infty(1) > \bar{g} = \sigma(0, \theta)$ in view of $\phi^\infty(1) > 0$. At this value the integrand equals $\phi^*(g(Z; \theta) - \bar{g})$, because the perspective at $\eta = 1$ reduces to ϕ^* itself, and $g(z; \theta) - \bar{g} \leq 0$ on $\mathcal{Z}$. Since ϕ^* is nondecreasing with $\phi^*(0) = -\inf \phi = 0$ and $\phi^*(x) \geq x$ by Lemma SA-4, the integrand lies between $g(Z; \theta) - \bar{g}$ and 0. Both bounds are P_* -integrable as $P_* \in \mathcal{P}(\mathcal{Z})$, implying that $f(\xi_0, \theta) < \infty$. Under Assumption SA-2, it is finite at the point $\xi_0 := \xi^*$ of Assumption SA-2(ii), which is feasible since $\mathcal{F}(\theta)$ is unrestricted. We conclude that f is proper.

Weak duality with slack. We next record an inequality that is used in several steps below and in the proof of Lemma SA-21. Fix $\xi = (\zeta, \eta, \lambda) \in \mathcal{F}(\theta)$ and $P \in \mathcal{P}(\mathcal{Z})$ with $D_\phi(P \| P_*) < \infty$. Write $w(z) := g(z; \theta) - \lambda^\top m(z; \theta)$ and let $P = rP_* + P_\perp$ be the Lebesgue decomposition of P with respect to P_* , so that $D_\phi(P \| P_*) = \mathbb{E}_{P_*}[\phi(r(Z))] + \phi^\infty(1)P_\perp(\mathcal{Z})$ by Lemma SA-1. In particular, $0 \leq \mathbb{E}_{P_*}[\phi(r(Z))] \leq D_\phi(P \| P_*) < \infty$, so that $r(z) \in \text{dom}(\phi)$ for P_* -almost every $z \in \mathcal{Z}$, and Lemma SA-4 at $\tilde{x} = r(z)$ gives

$$(\phi^*)^\pi(w(z) - \zeta, \eta) \geq (w(z) - \zeta)r(z) - \eta\phi(r(z))$$

for P_* -almost every $z \in \mathcal{Z}$. As $P \in \mathcal{P}(\mathcal{Z})$, the function $(w - \zeta)r$ belongs to $L_1(P_*)$, and so does $\phi(r)$, so integration yields

$$\mathbb{E}_{P_*}[(\phi^*)^\pi(w(Z) - \zeta, \eta)] \geq \int_{\mathcal{Z}} (w(z) - \zeta)r(z) dP_*(z) - \eta \mathbb{E}_{P_*}[\phi(r(Z))]. \quad (74)$$

As for the singular part, if $\phi^\infty(1) = +\infty$, then $D_\phi(P \| P_*) < \infty$ forces $P_\perp(\mathcal{Z}) = 0$, whereas if $\phi^\infty(1) < \infty$, the constraint defining $\mathcal{F}(\theta)$ gives $w(z) - \zeta \leq \eta\phi^\infty(1)$ for every $z \in \mathcal{Z}$. In either case

$$\int_{\mathcal{Z}} (w(z) - \zeta) dP_\perp(z) \leq \eta\phi^\infty(1)P_\perp(\mathcal{Z}), \quad (75)$$

under the convention that the right-hand side evaluates to 0 when $P_\perp(\mathcal{Z}) = 0$. It then follows that

$$\begin{aligned} f(\xi, \theta) &= \lambda_{12}^\top \rho + \zeta + \eta\delta + \mathbb{E}_{P_*}[(\phi^*)^\pi(w(Z) - \zeta, \eta)] \\ &\geq \lambda_{12}^\top \rho + \zeta + \eta\delta + \mathbb{E}_P[w(Z) - \zeta] - \int_{\mathcal{Z}} (w(z) - \zeta) dP_\perp(z) - \eta \mathbb{E}_{P_*}[\phi(r(Z))] \\ &\geq \lambda_{12}^\top \rho + \zeta + \eta\delta + \mathbb{E}_P[w(Z) - \zeta] - \eta D_\phi(P \| P_*) \\ &= \mathbb{E}_P[g(Z; \theta)] + \eta(\delta - D_\phi(P \| P_*)) + (\lambda_{12}^\top \rho - \lambda^\top \mathbb{E}_P[m(Z; \theta)]), \end{aligned} \quad (76)$$

where the equality in the first line is the definition of f , the first inequality combines (74) with the decomposition $P = rP_* + P_\perp$, the second inequality follows from (75) and the decomposition of $D_\phi(P \| P_*)$ displayed above, and the last equality holds because P is a probability measure. If, in addition, P is primal feasible, that is, $D_\phi(P \| P_*) \leq \delta$ and $\mathbb{E}_P[m(Z; \theta)] \in \mathcal{M}(\rho)$, then both terms after $\mathbb{E}_P[g(Z; \theta)]$ in (76) are nonnegative, the second because $\lambda \in \mathcal{R}_\lambda$ and, by the support function of $\mathcal{M}(\rho)$ recorded in Section SA1.2, $\lambda^\top x \leq \lambda_{12}^\top \rho$ for every $x \in \mathcal{M}(\rho)$. In this case $f(\xi, \theta) \geq \mathbb{E}_P[g(Z; \theta)]$.

Choice of the candidate sublevel set. By the recession criterion above, applied to the extension of $f(\cdot, \theta)$ by $+\infty$ outside $\mathcal{F}(\theta)$, it suffices to exhibit one nonempty bounded sublevel set $\text{lev}_{\leq \alpha}(f, \mathcal{F}(\theta))$. Indeed, the recession function of the extension is then strictly positive at every $d \neq 0$. At every $d \in \text{rec}(\mathcal{F}(\theta))$ it coincides with $f^\infty(d, \theta)$, because, by Definition SA-2, both recession functions can be computed along the ray $\xi_0 + \tau d$, which stays in $\mathcal{F}(\theta)$, where the two functions agree. This is the claim of part (iii). The rest of the proof does so using the distributions $Q_\pm$ from part (i), the distribution P° of Assumption SA-1(iii) or Assumption SA-2(iii), and, when $\phi^\infty(1) = +\infty$, the full-support condition imposed in case (a) of Assumption SA-2(v).

To avoid clutter, we present the argument for $\mathbf{d}_{m,=} = \mathbf{d}_{m,\leq} = 1$. The result readily generalizes to higher (but finite) dimensions: for every unit vector $u = (u_\leq, u_>=)$ with $u_\leq \geq 0$, Carathéodory's theorem furnishes, exactly as in the proof of part (i), a distribution $Q_u \in \Delta(\mathcal{Z})$ with $\mathbb{E}_{Q_u}[m(Z; \theta)] = \bar{\rho} - \varepsilon u$, hence with $\mathbb{E}_{Q_u}[u^\top m(Z; \theta)] = -\varepsilon$ after the recentering below, and the diverging-multiplier arguments then apply with Q_u , for u the limiting direction of the multipliers, in place of Q_+ or Q_- . Without loss of generality, let $\bar{\rho} = 0$. Otherwise, redefine the moment function by setting $\tilde{m}_j(Z; \theta) := m_j(Z; \theta) - \bar{\rho}_j$ for each $j = 1, \dots, \mathbf{d}_m$, so that the inequality constraints $\mathbb{E}_P[m_j(Z; \theta)] \leq \bar{\rho}_j$ become $\mathbb{E}_P[\tilde{m}_j(Z; \theta)] \leq 0$ and the equality constraints $\mathbb{E}_P[m_j(Z; \theta)] = \bar{\rho}_j$ become $\mathbb{E}_P[\tilde{m}_j(Z; \theta)] = 0$, and replace m_j with $\tilde{m}_j$ throughout. Since $\lambda^\top \bar{\rho} = \lambda_{12}^\top \rho$, the invertible affine change of variables $\tilde{\zeta} := \zeta + \lambda_{12}^\top \rho$ maps the dual objective and the pointwise constraint onto those of the recentered problem, and it carries bounded sublevel sets onto bounded sublevel sets. As shown in the properness paragraph, $f(\cdot, \theta)$ is finite at the feasible point ξ_0 identified there so we may pick $\alpha^* := f(\xi_0, \theta) + 1$ which makes $\text{lev}_{\leq \alpha^*}(f, \mathcal{F}(\theta))$ nonempty.

Case I: $\eta = 0$. Define

$$\mathcal{L}_{|0} := \left\{ (\zeta, \eta, \lambda_{\leq}, \lambda_{=}) \in \mathbb{R} \times \{0\} \times \mathbb{R}_+ \times \mathbb{R} : \zeta \geq \sup_{z \in \mathcal{Z}} \{g(z; \theta) - \lambda^\top m(z; \theta)\} \right\}.$$

We now show that (a) the portion of $\text{lev}_{\leq \alpha^*}(f, \mathcal{F}(\theta))$ on the slice $\{\eta = 0\}$ is contained in $\mathcal{L}_{|0}$ and (b) that $f(\xi, \theta) \geq \zeta$ on this slice.

We start with (a). Indeed, note that when $\phi^\infty(1) < \infty$, the slice of $\mathcal{F}(\theta)$ is $\mathcal{L}_{|0}$ itself. When instead $\phi^\infty(1) = +\infty$, the slice of $\mathcal{F}(\theta) = \mathcal{L}$ is unconstrained, but every point of it outside $\mathcal{L}_{|0}$ has $f(\xi, \theta) = +\infty$ and so belongs to no sublevel set. To see this, consider any $\xi = (\zeta, 0, \lambda) \in \mathcal{L}$ with $\zeta < \sup_{z \in \mathcal{Z}} \{g(z; \theta) - \lambda^\top m(z; \theta)\}$. The set $\{z \in \mathcal{Z} : g(z; \theta) - \lambda^\top m(z; \theta) > \zeta\}$ is nonempty and open, by continuity of g and m , and hence has positive P_* -measure, since under Assumption SA-2(v) every nonempty open subset of $\mathcal{Z} = \text{supp}(P_*)$ has positive mass. At each such z the term within the expectation in f equals

$$(\phi^*)^\pi(g(z; \theta) - \lambda^\top m(z; \theta) - \zeta, 0) = (\phi^*)^\infty(g(z; \theta) - \lambda^\top m(z; \theta) - \zeta) = +\infty.$$

To see this, first note that Definition SA-4 and the domain condition $[0, \infty) \subseteq \text{dom}(\phi)$ (imposed in case (a) of Assumption SA-2(i), or implied by Assumption SA-1(ii)) imply that $\text{dom}(\phi) = [0, \infty)$, and then

$$(\phi^*)^\infty(x) = \iota_{\text{dom}(\phi)}^*(x) = \sup_{y \in \text{dom}(\phi)} \{xy\} = \sup_{y \geq 0} \{xy\} = \begin{cases} 0, & x \leq 0, \\ +\infty, & x > 0, \end{cases} \quad (77)$$

where the first equality follows from Theorem 13.3 in Rockafellar (1970). It follows that $(\phi^*)^\infty$ equals $+\infty$ whenever its argument is positive. This happens whenever $g(z; \theta) - \lambda^\top m(z; \theta) > \zeta$, which is satisfied for the particular choice of z above. Since the set of z for which this happens has positive P_* -measure, the expectation in f is $+\infty$ as claimed.

We now turn to (b). Since $(\phi^*)^\pi(x, 0) = (\phi^*)^\infty(x) \geq 0$ for every x by (77), the expectation in f is nonnegative on the slice $\{\eta = 0\}$, hence $f(\xi, \theta) \geq \zeta$. The same conclusion holds under option (b) of Assumption SA-2(i), where $\text{dom}(\phi) = [0, \bar{x}]$ with $\bar{x} = \sup \text{dom}(\phi) \in (1, \infty)$: there $(\phi^*)^\infty(x) = \sup_{y \in [0, \bar{x}]} xy = \bar{x} [x]_+ \geq 0$, so $f(\xi, \theta) \geq \zeta$ on the slice as well.

Option (b) of Assumption SA-2(i) is treated at the end of this proof. The relaxed-set analysis that follows therefore concerns Assumption SA-1 and option (a) of Assumption SA-2 (and applies verbatim, with the λ -coordinates deleted, when those assumptions are invoked with $\mathbf{d}_m = 0$).

We now construct two relaxed constraint sets containing $\mathcal{L}_{|0}$, whose intersection with the sublevel set we then show to be bounded. Because $\sup_{z \in \mathcal{Z}} h(z) \geq \mathbb{E}_Q[h(Z)]$ for any h whenever $\text{supp}(Q) \subseteq \mathcal{Z}$, replacing the supremum in the constraint of $\mathcal{L}_{|0}$ with an expectation under Q_- , respectively Q_+ , can only enlarge the set. In particular, we use $\mathbb{E}_{Q_-}[m_2(Z; \theta)] = -\varepsilon$ and $\mathbb{E}_{Q_+}[m_2(Z; \theta)] = +\varepsilon$ from Part (i). This yields the following two sets

$$\overline{\mathcal{L}}_{|0,-} := \{(\zeta, \eta, \lambda_{\leq}, \lambda_{=}) \in \mathbb{R} \times \{0\} \times \mathbb{R}_+ \times \mathbb{R} : \zeta \geq \mathbb{E}_{Q_-}[g(Z; \theta)] - \lambda_{\leq} \mathbb{E}_{Q_-}[m_1(Z; \theta)] + \lambda_{=}\varepsilon\} \supseteq \mathcal{L}_{|0}$$

and

$$\overline{\mathcal{L}}_{|0,+} := \{(\zeta, \eta, \lambda_{\leq}, \lambda_{=}) \in \mathbb{R} \times \{0\} \times \mathbb{R}_+ \times \mathbb{R} : \zeta \geq \mathbb{E}_{Q_+}[g(Z; \theta)] - \lambda_{\leq} \mathbb{E}_{Q_+}[m_1(Z; \theta)] - \lambda_{=}\varepsilon\} \supseteq \mathcal{L}_{|0}.$$

We build two such sets because the equality multiplier $\lambda_{=}$ is unconstrained in sign: the term $+\lambda_{=}\varepsilon$ in $\overline{\mathcal{L}}_{|0,-}$ makes its constraint fail as $\lambda_{=} \rightarrow +\infty$, while the term $-\lambda_{=}\varepsilon$ in $\overline{\mathcal{L}}_{|0,+}$ makes its constraint fail as $\lambda_{=} \rightarrow -\infty$. The inequality multiplier $\lambda_{\leq}$ is nonnegative and is handled by either set, since $\mathbb{E}_{Q_{\pm}}[m_1(Z; \theta)] < 0$ makes $-\lambda_{\leq} \mathbb{E}_{Q_{\pm}}[m_1(Z; \theta)] \rightarrow +\infty$ as $\lambda_{\leq} \rightarrow +\infty$.

Consider any sequence with $\|(\zeta, \lambda_{\leq}, \lambda_{=})\| \rightarrow +\infty$ lying in $\text{lev}_{\leq \alpha^*}(f, \overline{\mathcal{L}}_{|0,-} \cap \overline{\mathcal{L}}_{|0,+})$. By definition of sublevel set, it satisfies $f(\xi, \theta) \leq \alpha^*$ together with the defining constraints of $\overline{\mathcal{L}}_{|0,-}$ and $\overline{\mathcal{L}}_{|0,+}$. We show that each such sequence must eventually violate one of these requirements thus contradicting its membership in the sublevel set. Thus, no diverging sequence can lie in $\text{lev}_{\leq \alpha^*}(f, \overline{\mathcal{L}}_{|0,-} \cap \overline{\mathcal{L}}_{|0,+})$. This implies that the set is bounded, and, a fortiori, that $\text{lev}_{\leq \alpha^*}(f, \mathcal{L}_{|0})$ is bounded as well.

The different diverging sequences we must consider are:

- if $|\zeta| \rightarrow \infty$, then either f diverges ($\zeta \rightarrow \infty$) or the constraint is violated ($\zeta \rightarrow -\infty$);
- if $\lambda_{\leq} \rightarrow \infty$, then the constraint is violated because $\mathbb{E}_{Q_-}[m_1(Z; \theta)] < 0$ and ζ is bounded along this sequence;
- if $|\lambda_{=}| \rightarrow \infty$, then either the constraint in $\overline{\mathcal{L}}_{|0,-}$ ($\lambda_{=} \rightarrow \infty$) or the one in $\overline{\mathcal{L}}_{|0,+}$ ($\lambda_{=} \rightarrow -\infty$) is violated;
- if $|\zeta| \rightarrow \infty$ and $\lambda_{\leq} \rightarrow \infty$, then either f diverges ($\zeta \rightarrow \infty$) or the constraints in $\overline{\mathcal{L}}_{|0,-}$ and $\overline{\mathcal{L}}_{|0,+}$ are violated ($\zeta \rightarrow -\infty$);
- if $|\zeta| \rightarrow \infty$ and $|\lambda_{=}| \rightarrow \infty$, then when $\zeta \rightarrow \infty$ the function diverges, whereas when $\zeta \rightarrow -\infty$, either the constraint in $\overline{\mathcal{L}}_{|0,-}$ is violated ($\lambda_{=} \rightarrow \infty$) or the one in $\overline{\mathcal{L}}_{|0,+}$ is violated ($\lambda_{=} \rightarrow -\infty$);
- if $\lambda_{\leq} \rightarrow \infty$ and $|\lambda_{=}| \rightarrow \infty$, then either the constraint in $\overline{\mathcal{L}}_{|0,-}$ is violated ($\lambda_{=} \rightarrow \infty$) or the one in $\overline{\mathcal{L}}_{|0,+}$ is violated ($\lambda_{=} \rightarrow -\infty$) because ζ is bounded along this sequence;

- if $|\zeta| \rightarrow \infty$, $\lambda_{\leq} \rightarrow \infty$, and $|\lambda_{=}| \rightarrow \infty$, then whenever $\zeta \rightarrow \infty$ the function diverges, whereas if $\zeta \rightarrow -\infty$ either the constraint in $\overline{\mathcal{L}}_{|0,-}$ is violated ($\lambda_{=} \rightarrow \infty$) or the one in $\overline{\mathcal{L}}_{|0,+}$ is violated ($\lambda_{=} \rightarrow -\infty$).

Therefore $\text{lev}_{\leq \alpha^*}(f, \overline{\mathcal{L}}_{|0,-} \cap \overline{\mathcal{L}}_{|0,+})$ is bounded and so is $\text{lev}_{\leq \alpha^*}(f, \mathcal{L}_{|0})$.

Case II: $\eta > 0$. Denote by

$$\mathcal{L}_{>0} := \left\{ (\zeta, \eta, \lambda_{\leq}, \lambda_{=}) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R}_+ \times \mathbb{R} : \zeta + \eta\phi^\infty(1) \geq \sup_{z \in \mathcal{Z}} \{g(z; \theta) - \lambda^\top m(z; \theta)\} \right\}$$

the feasible set at $\eta > 0$. On $\mathcal{L}_{>0}$ the distributions Q_+, Q_- from part (i) induce the two relaxed inequalities

$$\begin{aligned} \zeta + \eta\phi^\infty(1) &\geq \mathbb{E}_{Q_-}[g(Z; \theta)] - \lambda_{\leq} \mathbb{E}_{Q_-}[m_1(Z; \theta)] + \lambda_{=}\varepsilon, \\ \zeta + \eta\phi^\infty(1) &\geq \mathbb{E}_{Q_+}[g(Z; \theta)] - \lambda_{\leq} \mathbb{E}_{Q_+}[m_1(Z; \theta)] - \lambda_{=}\varepsilon. \end{aligned}$$

We first derive two lower bounds on f over $\mathcal{L}_{>0}$.

First bound. The first is obtained via the distribution P° provided by Assumption SA-1(iii) or Assumption SA-2(iii), which satisfies the moment restrictions together with $\delta_\circ := D_\phi(P^\circ \| P_*) < \delta$. The distribution P° belongs to $\mathcal{P}(\mathcal{Z})$ by Assumption SA-1(iii), respectively by Assumption SA-2(iii). Fix $\xi \in \mathcal{L}_{>0} \subseteq \mathcal{F}(\theta)$. Inequality (76) at $P = P^\circ$, in the recentered problem, gives

$$f(\xi, \theta) \geq \mathbb{E}_{P^\circ}[g(Z; \theta)] + \eta(\delta - \delta_\circ) - \lambda^\top \mathbb{E}_{P^\circ}[m(Z; \theta)] \geq \mathbb{E}_{P^\circ}[g(Z; \theta)] + \eta(\delta - \delta_\circ),$$

where the second inequality holds because $\lambda^\top \mathbb{E}_{P^\circ}[m(Z; \theta)]$ equals $\lambda_{\leq} \mathbb{E}_{P^\circ}[m_1(Z; \theta)] \leq 0$, in view of $\lambda_{\leq} \geq 0$, $\mathbb{E}_{P^\circ}[m_1(Z; \theta)] \leq 0$, and $\mathbb{E}_{P^\circ}[m_2(Z; \theta)] = 0$. Because we assumed that $\delta > \delta_\circ$, the bound above shows that f diverges along any sequence in $\mathcal{L}_{>0}$ with $\eta \rightarrow \infty$, regardless of the behavior of $(\zeta, \lambda_{\leq}, \lambda_{=})$.

Second bound. The second bound handles diverging ζ : since $0 \in \text{dom}(\phi)$, Lemma SA-4 at $\tilde{x} = 0$ gives $(\phi^*)^\pi(w(z) - \zeta, \eta) \geq -\eta\phi(0)$, implying that $f(\xi, \theta) \geq \zeta + \eta(\delta - \phi(0))$ on $\mathcal{L}_{>0}$. We argue by contradiction. Consider any diverging sequence in $\text{lev}_{\leq \alpha^*}(f, \mathcal{L}_{>0})$ such that $\|(\zeta, \eta, \lambda_{\leq}, \lambda_{=})\| \rightarrow +\infty$. If $\eta \rightarrow \infty$ along a subsequence, the first bound gives $f(\xi, \theta) \geq \mathbb{E}_{P^\circ}[g(Z; \theta)] + \eta(\delta - \delta_\circ) \rightarrow +\infty$, already contradicting $f \leq \alpha^*$. It therefore suffices to treat sequences along which η remains in a bounded subset of $\mathbb{R}_{++}$, so that only $(\zeta, \lambda_{\leq}, \lambda_{=})$ diverges.

Suppose first that $\phi^\infty(1) < \infty$. These cases reduce to the $\eta = 0$ analysis up to bounded scalars: (a) if $\zeta \rightarrow +\infty$, the second bound gives $f(\xi, \theta) \geq \zeta + \eta(\delta - \phi(0))$, and the right-hand side diverges since $\eta(\delta - \phi(0))$ is bounded along the sequence; (b) if $\zeta \rightarrow -\infty$, the pointwise constraint is eventually violated uniformly in λ . To see this, the two relaxed inequalities read

$$\zeta + \eta\phi^\infty(1) \geq \mathbb{E}_{Q_-}[g(Z; \theta)] + \varepsilon\lambda_{\leq} + \varepsilon\lambda_{=}$$

and

$$\zeta + \eta\phi^\infty(1) \geq \mathbb{E}_{Q_+}[g(Z; \theta)] + \varepsilon\lambda_{\leq} - \varepsilon\lambda_{=},$$

since $\mathbb{E}_{Q_\pm}[m_1(Z; \theta)] = -\varepsilon$ after the recentering, and taking the larger of the two right-hand sides gives

$$\zeta + \eta\phi^\infty(1) \geq \min_{\pm} \mathbb{E}_{Q_\pm}[g(Z; \theta)] + \varepsilon\lambda_{\leq} + \varepsilon|\lambda_{=}| \geq \min_{\pm} \mathbb{E}_{Q_\pm}[g(Z; \theta)],$$

a finite lower bound that holds for every $\lambda \in \mathbb{R}_+ \times \mathbb{R}$. Along the sequence the left-hand side diverges to $-\infty$, so the constraint eventually fails regardless of the behavior of λ ; (c) if $\lambda_{\leq} \rightarrow +\infty$ or $|\lambda_{=}| \rightarrow \infty$ with ζ bounded, the same arguments based on the $Q_\pm$ distributions used for $\eta = 0$ apply. These cases are sufficient because the case for ζ does not depend on λ .

Suppose instead that $\phi^\infty(1) = +\infty$, so that, by the restriction of this analysis to option (a) of Assumption SA-2(i), $[0, \infty) \subseteq \text{dom}(\phi)$ and g, m are continuous. The constraint defining $\mathcal{L}_{>0}$ and the two relaxed inequalities are then vacuous. Consider the following cases. If $\zeta \rightarrow +\infty$, the argument in (a) applies verbatim, as it does not use the constraint. If $\zeta \rightarrow -\infty$ with λ bounded, then Assumption SA-2(i) allows us to use Lemma SA-4 at some $\tilde{x} > 1$ in $\text{dom}(\phi)$ to get $(\phi^*)^\pi(w(z) - \zeta, \eta) \geq \tilde{x}(w(z) - \zeta) - \eta\phi(\tilde{x})$, implying that

$$f(\xi, \theta) \geq \zeta(1 - \tilde{x}) + \tilde{x} \mathbb{E}_{P_*}[w(Z)] + \eta(\delta - \phi(\tilde{x})) \rightarrow +\infty.$$

If instead $s := \|\lambda\| \rightarrow \infty$ with ζ bounded above, write $\lambda = s u_s$ and extract a further subsequence along which $u_s \rightarrow u = (u_{\leq}, u_{=}) \in \mathbb{R}_+ \times \mathbb{R}$ with $\|u\| = 1$, using compactness of the unit sphere and closedness of the constraint $\lambda_{\leq} \geq 0$. We claim that $P_*(u^\top m(Z; \theta) < 0) > 0$. To see this, let $Q = Q_-$ if $u_{=} \geq 0$ and $Q = Q_+$ if $u_{=} < 0$, with $Q_\pm$ and ε as in part (i). Then

$$\mathbb{E}_Q[u^\top m(Z; \theta)] = u_{\leq} \mathbb{E}_Q[m_1(Z; \theta)] + u_{=} \mathbb{E}_Q[m_2(Z; \theta)] < 0,$$

where the first term is nonpositive, since $u_{\leq} \geq 0$ and $\mathbb{E}_{Q_\pm}[m_1(Z; \theta)] < 0$, and strictly negative when $u_{=} = 0$, as then $u_{\leq} = 1$, while the second term equals $-|u_{=}| \varepsilon \leq 0$ by the choice of Q , with strict inequality when $u_{=} \neq 0$. A random variable with strictly negative mean is strictly negative with positive probability, so $Q(u^\top m(Z; \theta) < 0) > 0$ and the set $\{z \in \mathcal{Z} : u^\top m(z; \theta) < 0\}$ is nonempty. Moreover, this set is open by continuity of m , and therefore has positive P_* -measure by the full-support condition of Assumption SA-2(v). By continuity of P_* from below there then exists $c > 0$ such that the event $\mathcal{A} := \{z \in \mathcal{Z} : u^\top m(z; \theta) < -2c\}$ has $p := P_*(\mathcal{A}) > 0$. As $P_* \in \mathcal{P}(\mathcal{Z})$ and $u_s \rightarrow u$, for all sufficiently large s

$$\int_{\mathcal{A}} |(u_s - u)^\top m(z; \theta)| dP_*(z) \leq \|u_s - u\| \mathbb{E}_{P_*} \|m(Z; \theta)\| \leq cp,$$

so that, using $u^\top m(z; \theta) < -2c$ on $\mathcal{A}$,

$$\int_{\mathcal{A}} u_s^\top m(z; \theta) dP_*(z) \leq \int_{\mathcal{A}} u^\top m(z; \theta) dP_*(z) + cp \leq -2cp + cp = -cp. \quad (78)$$

Set $\tilde{K} := 2/p \in \text{dom}(\phi)$. It then follows that

$$\begin{aligned} f(\xi, \theta) &\stackrel{(i)}{=} \zeta + \eta\delta + \mathbb{E}_{P_*}[(\phi^*)^\pi(w(Z) - \zeta, \eta)] \\ &\stackrel{(ii)}{\geq} \zeta + \eta\delta + \tilde{K} \int_{\mathcal{A}} (w(z) - \zeta) dP_*(z) - \eta(p\phi(\tilde{K}) + (1-p)\phi(0)) \\ &\stackrel{(iii)}{\geq} \zeta(1 - \tilde{K}p) - \tilde{K} \mathbb{E}_{P_*}[[g(Z; \theta)]_-] + \tilde{K}cps + \eta(\delta - p\phi(\tilde{K}) - (1-p)\phi(0)) \\ &\stackrel{(iv)}{=} -\zeta - \frac{2}{p} \mathbb{E}_{P_*}[[g(Z; \theta)]_-] + 2cs + \eta(\delta - p\phi(2/p) - (1-p)\phi(0)) \rightarrow +\infty, \end{aligned}$$

where (i) is the definition of f under the centering $\rho = 0$; (ii) applies Lemma SA-4 pointwise at $\tilde{x} = \tilde{K}$ on $\mathcal{A}$ and at $\tilde{x} = 0$ on $\mathcal{Z} \setminus \mathcal{A}$ and integrates against P_* , using $P_*(\mathcal{A}) = p$ and $P_*(\mathcal{Z} \setminus \mathcal{A}) = 1 - p$; (iii) expands $w = g - s u_s^\top m$, bounds $\int_{\mathcal{A}} g dP_* \geq -\mathbb{E}_{P_*}[[g(Z; \theta)]_-]$, which is finite by Assumption SA-2(iv), and uses (78) multiplied by $-s$; and (iv) substitutes $\tilde{K} = 2/p$, so that $\tilde{K}p = 2$ and hence $\zeta(1 - \tilde{K}p) = -\zeta$. The limit is $+\infty$ because $-\zeta$ is bounded below along this subsequence (recall ζ is bounded above), η is bounded, and $cs \rightarrow +\infty$. In every case $f \rightarrow +\infty$ along a subsequence, contradicting $f \leq \alpha^*$.

Therefore, we can conclude that $\text{lev}_{\leq \alpha^*}(f, \mathcal{L}_{>0})$ is bounded.

Combining the two cases. Under Assumption SA-1 or option (a) of Assumption SA-2(i), we now show that $\text{lev}_{\leq \alpha^*}(f, \mathcal{F}(\theta)) = \text{lev}_{\leq \alpha^*}(f, \mathcal{L}_{|0}) \cup \text{lev}_{\leq \alpha^*}(f, \mathcal{L}_{>0})$. For the inclusion $\subseteq$, note that every point of $\text{lev}_{\leq \alpha^*}(f, \mathcal{F}(\theta))$ has either $\eta = 0$, and then it lies in $\mathcal{L}_{|0}$ – since any point of the $\eta = 0$ slice outside $\mathcal{L}_{|0}$ has $f = +\infty$, as shown at the start of the $\eta = 0$ case – or $\eta > 0$, in which case it lies in $\mathcal{L}_{>0}$ by construction. For the reverse inclusion $\supseteq$, both $\mathcal{L}_{|0} \subseteq \mathcal{F}(\theta)$ and $\mathcal{L}_{>0} \subseteq \mathcal{F}(\theta)$ by construction, so the two sublevel sets on the right are contained in $\text{lev}_{\leq \alpha^*}(f, \mathcal{F}(\theta))$. As the union of two bounded sets, $\text{lev}_{\leq \alpha^*}(f, \mathcal{F}(\theta))$ is bounded. By the reduction in the choice of the candidate sublevel set, this proves part (iii) under Assumption SA-1 and under option (a) of Assumption SA-2(i).

Option (b) of Assumption SA-2(i) for arbitrary $d_m \geq 0$. It remains to establish part (iii) under option (b) of Assumption SA-2(i). The relaxed-set analysis above does not apply in this case because it relies on $\text{dom}(\phi) = [0, \infty)$ and on the continuity and full-support condition of Assumption SA-2(v), both available only under option (a). We instead verify the claim of part (iii) directly. Since $\phi^\infty(1) = +\infty$ under option (b), the feasible set is the cone $\mathcal{F}(\theta) = \mathcal{L}$, so that $\text{rec}(\mathcal{F}(\theta)) = \mathcal{L}$ and the claim reads: $f^\infty(d, \theta) > 0$ for every nonzero $d = (d_\zeta, d_\eta, d_\lambda) \in \mathcal{L}$. As before, we work in the recentered problem, $\vec{\rho} = 0$. The verification has two steps. We first exploit the boundedness of $\text{dom}(\phi)$ to show that $f(\cdot, \theta)$ is finite on $\mathcal{L}$ and to obtain $f^\infty(\cdot, \theta)$ in closed form. We then show that the closed form is strictly positive along each applicable class of directions. When $d_m = 0$, the λ -block is absent, we set $d_\lambda = 0$, and only the first two classes below arise. When $d_m \geq 1$, all three classes arise: those with $d_\eta > 0$, where the distribution P° of Assumption SA-2(iii) provides the required slack; those with $d_\eta = 0$ and $d_\lambda = 0$, where positivity is a consequence of $\bar{x} > 1$; and those with $d_\eta = 0$ and $d_\lambda \neq 0$, where the condition of Assumption SA-2(vi) takes over the role that the full-support condition played in the analysis above.

We begin with the closed form. Note first that $\text{dom}(\phi) = [0, \bar{x}]$. Write $h := (\phi^*)^\pi$ and $C_1 := \max\{\phi(0), \phi(\bar{x})\}$, which bounds ϕ on $[0, \bar{x}]$ by convexity and the nonnegativity of ϕ . Evaluating the supremum defining ϕ^* at the points 0 and $\bar{x}$ of $\text{dom}(\phi)$ gives the lower bound in

$$\max\{-\phi(0), \bar{x}y - \phi(\bar{x})\} \leq \phi^*(y) \leq \bar{x}[y]_+,$$

while the upper bound uses $\phi \geq 0$ together with $xy \leq \bar{x}[y]_+$ for $x \in [0, \bar{x}]$ and every $y \in \mathbb{R}$. Consequently $|h(x, \eta)| \leq \bar{x}|x| + C_1\eta$ for every $\eta \geq 0$, and when $\eta = 0$ we have $h(x, 0) = (\phi^*)^\infty(x) = \bar{x}[x]_+$, as computed in Case I. In particular, $\text{dom}(h) = \mathbb{R} \times \mathbb{R}_+$. The displayed linear growth bound, together with $P_* \in \mathcal{P}(\mathcal{Z})$, shows that $f(\cdot, \theta)$ is finite on $\mathcal{L}$. The arguments at the beginning of this part show that it is closed, proper, and convex. Moreover, h is closed, proper, and convex by Lemma SA-3, positively homogeneous by construction, and satisfies $h(0, 0) = 0$, so Theorem 8.5 in Rockafellar (1970) gives $h^\infty = h$. By the same theorem, for every $x_0 \in \text{dom}(h)$ and every direction v , the difference quotients $[h(x_0 + \tau v) - h(x_0)]/\tau$ are nondecreasing in $\tau > 0$ and increase to $h^\infty(v) = h(v)$. Fix any $\xi_0 = (\zeta_0, \eta_0, \lambda_0) \in \mathcal{L}$. Since $\mathcal{L}$ is a convex cone containing d , the ray $\xi_0 + \tau d$ belongs to $\mathcal{L}$. At each $z \in \mathcal{Z}$ the term within the expectation in f is h composed with an affine map of ξ , whose value at ξ_0 lies in $\text{dom}(h)$, so the difference quotients of this term along d increase pointwise to $h(-d_\lambda^\top m(z; \theta) - d_\zeta, d_\eta)$. Monotone convergence therefore yields

$$f^\infty(d, \theta) = d_\zeta + d_\eta\delta + \mathbb{E}_{P_*} [h(-d_\lambda^\top m(Z; \theta) - d_\zeta, d_\eta)].$$

We now show that the right-hand side above is strictly positive in each of the applicable cases described at the beginning of this part. (1) Directions with $d_\eta > 0$. Let $\delta_\circ := D_\phi(P^\circ \| P_*) < \delta$, with P° the distribution of Assumption SA-2(iii), which belongs to $\mathcal{P}(\mathcal{Z})$ by Assumption SA-2(iii). For every $\tau > 0$ the point $\xi_0 + \tau d$ lies in

$\mathcal{L} = \mathcal{F}(\theta)$, so (76) at $P = P^\circ$, in the recentered problem, gives

$$f(\xi_0 + \tau d, \theta) \geq \mathbb{E}_{P^\circ}[g(Z; \theta)] + (\eta_0 + \tau d_\eta)(\delta - \delta_\circ) - (\lambda_0 + \tau d_\lambda)^\top \mathbb{E}_{P^\circ}[m(Z; \theta)].$$

Subtracting the finite value $f(\xi_0, \theta)$, dividing by τ , and letting $\tau \rightarrow \infty$, Definition SA-2 yields

$$f^\infty(d, \theta) \geq d_\eta(\delta - \delta_\circ) - d_\lambda^\top \mathbb{E}_{P^\circ}[m(Z; \theta)] \geq d_\eta(\delta - \delta_\circ) > 0,$$

where the second inequality holds because $d_\lambda \in \mathcal{R}_\lambda$ and $\mathbb{E}_{P^\circ}[m(Z; \theta)]$ lies in the recentered moment constraint set, so that $d_\lambda^\top \mathbb{E}_{P^\circ}[m(Z; \theta)] \leq 0$ by the support function of $\mathcal{M}(\rho)$ recorded in Section SA1.2, evaluated at $\rho = 0$. (2) Directions with $d_\eta = 0$ and $d_\lambda = 0$, so that $d_\zeta \neq 0$. In this case $f^\infty(d, \theta) = d_\zeta + \bar{x}[-d_\zeta]_+$, which equals $d_\zeta > 0$ when $d_\zeta > 0$ and $(1 - \bar{x})d_\zeta > 0$ when $d_\zeta < 0$, in view of $\bar{x} > 1$. (3) When $d_m \geq 1$, directions with $d_\eta = 0$ and $d_\lambda \neq 0$. Write $Y := -d_\lambda^\top m(Z; \theta)$, so that $f^\infty(d, \theta) = d_\zeta + \bar{x} \mathbb{E}_{P_*}[[Y - d_\zeta]_+]$. We say that a distribution $Q \in \Delta(\mathcal{Z})$ is “ceiling-feasible” if $Q \ll P_*$ with $dQ/dP_* \leq \bar{x}$ P_* -almost surely, as in the definition of $\mathcal{N}_{\bar{x}}(\theta)$ in Assumption SA-2(vi). For every such Q ,

$$\mathbb{E}_Q[Y] \leq d_\zeta + \mathbb{E}_Q[[Y - d_\zeta]_+] \leq d_\zeta + \bar{x} \mathbb{E}_{P_*}[[Y - d_\zeta]_+] = f^\infty(d, \theta),$$

where the first inequality applies $y \leq t + [y - t]_+$ pointwise at $t = d_\zeta$ and the second bounds the density of Q by $\bar{x}$. Since Assumption SA-2(vi) places 0 in the interior of $\mathcal{N}_{\bar{x}}(\theta) - \bar{\rho}$, there exists $r_0 > 0$ with $-r_0 d_\lambda / \|d_\lambda\| \in \mathcal{N}_{\bar{x}}(\theta) - \bar{\rho}$. In the recentered problem this says that some ceiling-feasible Q satisfies $\mathbb{E}_Q[m(Z; \theta)] = -r_0 d_\lambda / \|d_\lambda\|$, and hence $\mathbb{E}_Q[Y] = r_0 \|d_\lambda\|$. Therefore $f^\infty(d, \theta) \geq r_0 \|d_\lambda\| > 0$, for any value of d_ζ .

In every applicable case $f^\infty(d, \theta) > 0$, which establishes part (iii) under option (b) for every $d_m \geq 0$. Parts (i)–(ii) require no separate argument: part (ii) is the vacuity statement already proven, and part (i) was proven under both options above. ■

SA8.15 Proof of Theorem SA-19

Proof. Fix $\theta \in \Theta$, and let $\mathcal{F}(\theta)$ and $f(\cdot, \theta)$ be the dual feasible set and the dual objective of Section SA3.1.

Properties of the feasible set and of the objective. By the properties established in the proof of Lemma SA-18(iii), $\mathcal{F}(\theta)$ is closed and convex, $f(\cdot, \theta)$ is a closed proper convex function, and $f(\xi_0, \theta) < \infty$ at the feasible point ξ_0 exhibited there, so that $\mathcal{F}(\theta) \cap \text{dom}(f(\cdot, \theta)) \neq \emptyset$.

Reduction to a recession condition. Since $\mathcal{F}(\theta)$ is not compact, the Weierstrass theorem does not apply directly, and we rely instead on Proposition 3.2.2 in Bertsekas (2009). This result concerns the minimization of a closed convex function over a closed convex set. It states that the set of minimizers is nonempty and compact if and only if the set and the function have no nonzero direction of recession in common. Now, a direction d is a direction of recession of $f(\cdot, \theta)$ exactly when $f^\infty(d, \theta) \leq 0$, whereas Lemma SA-18(iii) shows that $f^\infty(d, \theta) > 0$ for every nonzero direction of recession d of $\mathcal{F}(\theta)$. Hence $\mathcal{F}(\theta)$ and $f(\cdot, \theta)$ have no nonzero direction of recession in common, and the set of minimizers of (21) is nonempty and compact. Convexity of the set of minimizers follows from the convexity of $f(\cdot, \theta)$ and $\mathcal{F}(\theta)$, since the minimizers of a convex function over a convex set form a convex set. ■

SA8.16 Proof of Lemma SA-20

Proof. Throughout, $\Gamma(\delta)$ is convex and contains $c := \mathbb{E}_{P_*}[m(Z; \theta)]$, as recalled in Section SA2.3.

Part (i). For (a), if $\bar{\rho} = c$ we take $P^\circ := P_*$. Otherwise, since $\bar{\rho} \in \text{int}(\Gamma(\delta))$ and $c \in \Gamma(\delta)$, Theorem 6.4 in Rockafellar (1970) yields $\mu_0 > 1$ such that $c + \mu_0(\bar{\rho} - c) \in \Gamma(\delta)$, that is, some $P_1 \in \mathcal{P}(\mathcal{Z})$ with $\mathbb{E}_{P_1}[m(Z; \theta)] = c + \mu_0(\bar{\rho} - c)$ and $D_\phi(P_1 \| P_*) \leq \delta$. Set $P^\circ := (1 - \mu_0^{-1})P_* + \mu_0^{-1}P_1$, which belongs to $\mathcal{P}(\mathcal{Z})$. Then $\mathbb{E}_{P^\circ}[m(Z; \theta)] = (1 - \mu_0^{-1})c + \mu_0^{-1}(c + \mu_0(\bar{\rho} - c)) = \bar{\rho}$, and, since $D_\phi(\cdot \| P_*)$ is convex, as recalled in Section SA2.3, $D_\phi(P^\circ \| P_*) \leq (1 - \mu_0^{-1}) \cdot 0 + \mu_0^{-1} \delta < \delta$. For (b), let $P \in \mathcal{P}(\mathcal{Z})$ and suppose that $\mathbb{E}_P[m(Z; \theta)]$ does not belong to the closed convex set $\text{cl}(\text{co}(\mathcal{N}(\theta)))$. By the separating hyperplane theorem there exist $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$ such that $a^\top \mathbb{E}_P[m(Z; \theta)] > b \geq a^\top m(z; \theta)$ for every $z \in \mathcal{Z}$. Integrating the second inequality under P gives $a^\top \mathbb{E}_P[m(Z; \theta)] \leq b$, a contradiction. Hence $\Gamma(\delta) \subseteq \text{cl}(\text{co}(\mathcal{N}(\theta)))$, so that $\bar{\rho}$ lies in the interior of $\text{cl}(\text{co}(\mathcal{N}(\theta)))$. As this interior is nonempty, the affine hull of $\text{co}(\mathcal{N}(\theta))$ is the whole of $\mathbb{R}^d$, and, since a convex set and its closure have the same relative interior (Rockafellar, 1970, Theorem 6.3), $\bar{\rho} \in \text{int}(\text{co}(\mathcal{N}(\theta)))$, which is (b). For (c), let $P \in \mathcal{P}(\mathcal{Z})$ with $D_\phi(P \| P_*) \leq \delta$. As $\bar{x} < \infty$ forces $\phi^\infty(1) = +\infty$, Lemma SA-1 gives $P \ll P_*$ and $\int \phi(dP/dP_*)dP_* < \infty$, so that $dP/dP_* \in \text{dom}(\phi) \subseteq [0, \bar{x}]$ P_* -almost surely. Thus $\mathbb{E}_P[m(Z; \theta)] \in \mathcal{N}_{\bar{x}}(\theta)$, whence $\Gamma(\delta) \subseteq \mathcal{N}_{\bar{x}}(\theta)$ and $\bar{\rho} \in \text{int}(\mathcal{N}_{\bar{x}}(\theta))$, which is (c).

Part (ii). For $K > 0$ let $m_K(z) := m(z; \theta) \mathbf{1}\{\|m(z; \theta)\| \leq K\}$, let $\mu_K := \mathbb{E}_{P^\circ}[m_K(Z)]$, and let $\Sigma_K := \mathbb{E}_{P^\circ}[(m(Z; \theta) - \bar{\rho})(m_K(Z) - \mu_K)^\top]$, which is finite because $\|m_K\| \leq K$ and $m(\cdot; \theta)$ is P° -integrable. As $K \rightarrow \infty$, $m_K \rightarrow m(\cdot; \theta)$ pointwise with $\|m_K\| \leq \|m(\cdot; \theta)\|$ and $\mathbb{E}_{P^\circ} \|m(Z; \theta)\|^2 < \infty$, so the dominated convergence theorem gives $\Sigma_K \rightarrow \mathbb{V}_{P^\circ}[m(Z; \theta)]$. As the latter matrix is nonsingular and the nonsingular matrices form an open set, Σ_K is nonsingular for some K , which we fix from now on, and we write $\sigma_{\min}(\Sigma_K) > 0$ for its smallest singular value. Let $B := 2K$, so that $\|m_K(z) - \mu_K\| \leq B$ for every z , fix $v \in \mathbb{R}^d$ with $\|v\| \leq 1$ and $t \in \mathbb{R}$ with $|t| \leq 1/(2B)$, and define $r_t(z) := 1 + t v^\top (m_K(z) - \mu_K)$, which takes values in $[\frac{1}{2}, \frac{3}{2}]$, and $P_t := r_t P^\circ$. Since $\int r_t dP^\circ = 1 + t v^\top (\mu_K - \mu_K) = 1$, the measure P_t is a probability distribution. It belongs to $\mathcal{P}(\mathcal{Z})$ because $r_t \leq \frac{3}{2}$ and $P^\circ \in \mathcal{P}(\mathcal{Z})$, and

$$\mathbb{E}_{P_t}[m(Z; \theta)] = \mathbb{E}_{P^\circ}[m(Z; \theta) r_t(Z)] = \bar{\rho} + t \mathbb{E}_{P^\circ}[m(Z; \theta)(m_K(Z) - \mu_K)^\top] v = \bar{\rho} + t \Sigma_K v,$$

where the last equality uses $\mathbb{E}_{P^\circ}[m_K(Z) - \mu_K] = 0$ and $\mathbb{E}_{P^\circ}[m(Z; \theta)] = \bar{\rho}$. As Σ_K is nonsingular, the set $\{\bar{\rho} + t\Sigma_K v : \|v\| \leq 1\}$ contains the closed ball $\bar{B}_{|t|\sigma_{\min}(\Sigma_K)}(\bar{\rho})$. It remains to show that $D_\phi(P_t \| P_*) < \delta$ whenever $|t|$ is small enough. Write $P^\circ = w^\circ P_* + P^{\circ\perp}$ for the Lebesgue decomposition of P° with respect to P_* . Then $P_t = r_t w^\circ P_* + r_t P^{\circ\perp}$ is that of P_t , and Lemma SA-1 gives

$$D_\phi(P_t \| P_*) = \int \phi(r_t w^\circ) dP_* + \phi^\infty(1) \int r_t dP^{\circ\perp}.$$

For $r \geq 1$ and $w \geq 0$, Theorem 8.5 in Rockafellar (1970) gives $\phi(rw) \leq \phi(w) + (r-1)w\phi^\infty(1)$, as the recession function of ϕ is $y \mapsto y\phi^\infty(1)$ on $[0, \infty)$. For $r \in [\frac{1}{2}, 1)$, convexity of ϕ and $\phi \geq 0$ give $\phi(rw) \leq r\phi(w) + (1-r)\phi(0) \leq \phi(w) + (1-r)\phi(0)$. Since $|r_t - 1| \leq |t|B$, in both cases $\phi(r_t w^\circ) \leq \phi(w^\circ) + |t|B(w^\circ\phi^\infty(1) + \phi(0))$ pointwise, and integrating under P_* , using $\int w^\circ dP_* \leq 1$ and $\int r_t dP^{\circ\perp} \leq (1 + |t|B)P^{\circ\perp}(\mathcal{Z})$, yields

$$D_\phi(P_t \| P_*) \leq D_\phi(P^\circ \| P_*) + |t|B(\phi(0) + 2\phi^\infty(1)).$$

The right-hand side is smaller than δ for $|t|$ small enough, because $D_\phi(P^\circ \| P_*) < \delta$, and for every such t the closed ball $\bar{B}_{|t|\sigma_{\min}(\Sigma_K)}(\bar{\rho})$ is contained in $\Gamma(\delta)$, which proves the first claim of (ii). If $P^\circ \ll P_*$, then $P_t = r_t P^\circ \ll P_*$ for every t , so the same ball is contained in the attainable set of the distributions dominated by P_* , which proves the second claim. ■

SA8.17 Proof of Lemma SA-21

Proof. Throughout the proof, we suppress the arguments ρ and δ and thus write $\mathcal{P}_\phi(\theta)$ for the primal feasible set and $V_P(\theta)$ and $V_D(\theta)$ for the value functions. We write P° for the strictly feasible distribution of Assumption SA-1(iii) or Assumption SA-2(iii).

Part (i): Finiteness. We first bound the two value functions from below. By either Assumption SA-1(iii) or Assumption SA-2(iii), $P^\circ \in \mathcal{P}_\phi(\theta)$. Therefore $V_P(\theta) \geq \mathbb{E}_{P^\circ}[g(Z; \theta)] > -\infty$. Moreover, for every primal-feasible P and dual-feasible ξ , inequality (76) in the proof of Lemma SA-18(iii), whose two slack terms are nonnegative at primal-feasible P , gives $\mathbb{E}_P[g(Z; \theta)] \leq f(\xi, \theta)$. Taking the supremum over P and the infimum over ξ therefore gives $V_D(\theta) \geq V_P(\theta) > -\infty$.

We next bound the two value functions from above. Since $V_P(\theta) \leq V_D(\theta)$, it suffices to find a feasible point at which the dual objective is finite. Such a point is provided by the properness paragraph in the proof of Lemma SA-18(iii), which shows that $f(\xi_0, \theta) < \infty$ at the feasible point $\xi_0 = (\bar{g}, 1, 0)$, with $\bar{g} = \sup_{z \in \mathcal{Z}} g(z; \theta)$, under Assumption SA-1, and at $\xi_0 = \xi^*$ of Assumption SA-2(ii) under Assumption SA-2. Under Assumption SA-1, the point ξ_0 is also the strictly feasible point constructed in the proof of Lemma SA-18(ii). Therefore $V_P(\theta) \leq V_D(\theta) \leq f(\xi_0, \theta) < \infty$, which together with the lower bound establishes part (i).

Part (ii): Upper semicontinuity. We introduce the notation

$$\Psi(\xi, \theta) := \mathbb{E}_{P_*}[(\phi^*)^\pi(g(Z; \theta) - \lambda^\top m(Z; \theta) - \zeta, \eta)],$$

so that $f(\xi, \theta) = \lambda_{12}^\top \rho + \zeta + \eta\delta + \Psi(\xi, \theta)$. Since V_D is real valued on Θ follows from part SA-21(i), whose hypotheses hold at every $\theta \in \Theta$ under Assumption SA-3. Fix $\theta_0 \in \Theta$ and let $(\theta_n)_{n \geq 1} \subseteq \Theta$ be an arbitrary sequence with $\theta_n \rightarrow \theta_0$. We now show that $\limsup_{n \rightarrow \infty} V_D(\theta_n) \leq V_D(\theta_0)$.

Step 1: strategy and two preliminary facts. We start showing two facts that are used repeatedly below. First, for every $\xi = (\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$, every $\theta \in \Theta$, and every $z \in \mathcal{Z}$,

$$(\phi^*)^\pi(g(z; \theta) - \lambda^\top m(z; \theta) - \zeta, \eta) \geq g(z; \theta) - \lambda^\top m(z; \theta) - \zeta. \quad (79)$$

Indeed, this is Lemma SA-4 with $\tilde{x} = 1$. Second, the conjugate ϕ^* is nondecreasing and finite on $(-\infty, \phi^\infty(1))$, hence continuous there, as a finite convex function on an open interval.

The remainder of the proof implements the following strategy in the finite-recession and superlinear cases. Since $V_D(\theta_n) \leq f(\xi, \theta_n)$ for every $\xi \in \mathcal{F}(\theta_n)$, it suffices to construct, for every $\varepsilon > 0$, a multiplier ξ_ε such that (a) $f(\xi_\varepsilon, \theta_0) \leq V_D(\theta_0) + 2\varepsilon$, (b) $\xi_\varepsilon \in \mathcal{F}(\theta_n)$ for all n large, and (c) $\limsup_{n \rightarrow \infty} f(\xi_\varepsilon, \theta_n) \leq f(\xi_\varepsilon, \theta_0)$. Indeed, (a)–(c) give $\limsup_{n \rightarrow \infty} V_D(\theta_n) \leq V_D(\theta_0) + 2\varepsilon$, and letting $\varepsilon \rightarrow 0$ concludes.

Step 2: the finite-recession case. Suppose Assumption SA-1 holds at every $\theta \in \Theta$ and fix $\varepsilon > 0$.

Since $V_D(\theta_0) \in \mathbb{R}$, there exists $\xi = (\zeta, \eta, \lambda) \in \mathcal{F}(\theta_0)$ with $f(\xi, \theta_0) \leq V_D(\theta_0) + \varepsilon$. By Step 1, $f(\xi, \theta_0)$ is finite. Feasibility of ξ and $\phi^\infty(1) < \infty$ imply $\sigma(\lambda, \theta_0) \leq \zeta + \eta\phi^\infty(1) < \infty$, while $\sigma(\lambda, \theta_0) > -\infty$ because $\mathcal{Z}$ is nonempty. Let $\xi^\circ = (\zeta^\circ, \eta^\circ, \lambda^\circ)$ be the strictly feasible point of Lemma SA-18(ii), with slack $s^\circ := \zeta^\circ + \eta^\circ\phi^\infty(1) - \sigma(\lambda^\circ, \theta_0) > 0$ and $\eta^\circ > 0$. We take for ξ° the point constructed in the proof of that lemma, which is the point ξ_0 of part (i) at θ_0 , so that $f(\xi^\circ, \theta_0) < \infty$, while $f(\xi^\circ, \theta_0) > -\infty$ by Step 1. For $t \in (0, 1]$, define $\xi_t := (1-t)\xi + t\xi^\circ = (\zeta_t, \eta_t, \lambda_t)$, which belongs to the convex set $\mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$. Since $\sigma(\cdot, \theta_0)$ is convex on $\mathcal{R}_\lambda$, as a pointwise supremum of functions affine in λ , the slack at ξ_t satisfies

$$\zeta_t + \eta_t\phi^\infty(1) - \sigma(\lambda_t, \theta_0) \geq (1-t)\{\zeta + \eta\phi^\infty(1) - \sigma(\lambda, \theta_0)\} + t s^\circ \geq t s^\circ > 0,$$

so ξ_t is strictly feasible at θ_0 , with $\eta_t \geq t\eta^\circ > 0$. Moreover, $f(\cdot, \theta_0)$ is convex on $\mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$, hence

$$f(\xi_t, \theta_0) \leq (1-t)f(\xi, \theta_0) + t f(\xi^\circ, \theta_0).$$

Since both values on the right-hand side are finite, choosing t small enough yields $f(\xi_t, \theta_0) \leq V_D(\theta_0) + 2\varepsilon$. We fix such

a t for the remainder of this step, abbreviate $\xi' := \xi_t = (\zeta', \eta', \lambda')$, and write $s := \zeta' + \eta' \phi^\infty(1) - \sigma(\lambda', \theta_0) \geq t s^\circ > 0$ for the associated slack.

Since $\sigma(\lambda', \theta_0) < \infty$, Assumption SA-3(ii) applies at θ_0 and yields a radius $r > 0$ such that $\sigma(\lambda', \theta) \leq \sigma(\lambda', \theta_0) + s/2$ for every $\theta \in \mathcal{B}_r(\theta_0) \cap \Theta$. Then, for every $\theta \in \mathcal{B}_r(\theta_0) \cap \Theta$ we have:

$$\zeta' + \eta' \phi^\infty(1) = \sigma(\lambda', \theta_0) + s \geq \sigma(\lambda', \theta) + s/2,$$

so that $\xi' \in \mathcal{F}(\theta)$ and

$$\frac{g(z; \theta) - \lambda'^\top m(z; \theta) - \zeta'}{\eta'} \leq \frac{\sigma(\lambda', \theta_0) + s/2 - \zeta'}{\eta'} = \phi^\infty(1) - \frac{s}{2\eta'} < \phi^\infty(1),$$

for every $z \in \mathcal{Z}$. It follows that, since ϕ^* is nondecreasing and finite on $(-\infty, \phi^\infty(1))$ by Step 1, the integrand of $\Psi(\xi', \theta)$ is bounded above, uniformly over $z \in \mathcal{Z}$ and $\theta \in \mathcal{B}_r(\theta_0) \cap \Theta$, by the finite constant $C := \eta' \phi^*(\phi^\infty(1) - s/(2\eta'))$.

By Assumption SA-3(i) applied at θ_0 , the maps $\theta' \mapsto g(z; \theta')$ and $\theta' \mapsto m(z; \theta')$ are continuous at θ_0 for P_* -almost every z , the exceptional null set possibly depending on θ_0 . Therefore, writing $y_n(z) := g(z; \theta_n) - \lambda'^\top m(z; \theta_n) - \zeta'$ and $y_0(z) := g(z; \theta_0) - \lambda'^\top m(z; \theta_0) - \zeta'$, we have $y_n(z) \rightarrow y_0(z)$. Moreover, $y_0(z) \leq \sigma(\lambda', \theta_0) - \zeta' = \eta' \phi^\infty(1) - s$, so that $y_0(z)/\eta'$ lies in the open interval $(-\infty, \phi^\infty(1))$, on which ϕ^* is continuous by Step 1. Hence $\eta' \phi^*(y_n(z)/\eta') \rightarrow \eta' \phi^*(y_0(z)/\eta')$ for P_* -almost every z . For all n large enough that $\theta_n \in \mathcal{B}_r(\theta_0)$, the functions $z \mapsto C - \eta' \phi^*(y_n(z)/\eta')$ are nonnegative, so Fatou's lemma gives

$$\liminf_{n \rightarrow \infty} \mathbb{E}_{P_*} \left[C - \eta' \phi^*(y_n(Z)/\eta') \right] \geq \mathbb{E}_{P_*} \left[C - \eta' \phi^*(y_0(Z)/\eta') \right].$$

Since $\Psi(\xi', \theta_0) > -\infty$ by Step 1, we get $\limsup_{n \rightarrow \infty} \Psi(\xi', \theta_n) \leq \Psi(\xi', \theta_0)$. As the remaining terms of $f(\xi', \cdot)$ do not depend on θ , we conclude that $\limsup_{n \rightarrow \infty} f(\xi', \theta_n) \leq f(\xi', \theta_0) \leq V_D(\theta_0) + 2\varepsilon$, so that requirements (a)–(c) of Step 1 are met by $\xi_\varepsilon = \xi'$, which shows upper semicontinuity.

Step 3: the superlinear case. Suppose Assumption SA-2 holds at every $\theta \in \Theta$, so that $\mathcal{F}(\theta) = \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$ for every $\theta \in \Theta$ by Lemma SA-18(ii). Requirement (b) of Step 1 therefore holds for every multiplier, and it remains to show (a) and (c).

We claim that

$$V_D(\theta_0) = \inf \{ f(\xi, \theta_0) : \xi \in \mathbb{R} \times \mathbb{R}_{++} \times \mathcal{R}_\lambda \}, \quad (80)$$

which excludes the set of multipliers with $\eta = 0$. To see this, fix $(\zeta, \lambda) \in \mathbb{R} \times \mathcal{R}_\lambda$ and write $y(z) := g(z; \theta_0) - \lambda^\top m(z; \theta_0) - \zeta$. For every $y \in \mathbb{R}$, the map $\eta \mapsto \eta \phi^*(y/\eta)$ is nonincreasing on $(0, \infty)$ and increases to $(\phi^*)^\pi(y, 0) = (\phi^*)^\infty(y)$ as $\eta \rightarrow 0$. Indeed, with $s = 1/\eta$ and $\phi^*(0) = -\inf \phi = 0$, it equals the difference quotient $\{\phi^*(sy) - \phi^*(0)\}/s$, which is nondecreasing in s by convexity of ϕ^* and converges to $(\phi^*)^\infty(y)$ as $s \rightarrow \infty$ (Rockafellar, 1970, Theorem 8.5). Hence, as $\eta \rightarrow 0$, the integrand of $\Psi(\zeta, \eta, \lambda, \theta_0)$ converges pointwise to the integrand of $\Psi(\zeta, 0, \lambda, \theta_0)$. Moreover, the minorant in (79) implies that the shifted integrands

$$\eta \phi^*(y(Z)/\eta) - y(Z)$$

are nonnegative and increase pointwise to $(\phi^*)^\infty(y(Z)) - y(Z)$ as $\eta \downarrow 0$. Since $y(Z)$ is integrable under the maintained conditions, the monotone convergence theorem applied to these shifted integrands gives $\Psi(\zeta, \eta, \lambda, \theta_0) \rightarrow \Psi(\zeta, 0, \lambda, \theta_0)$ as $\eta \rightarrow 0$. Since also $\eta \delta \rightarrow 0$, we conclude that $f(\zeta, \eta, \lambda, \theta_0) \rightarrow f(\zeta, 0, \lambda, \theta_0)$ as $\eta \rightarrow 0$, so the infimum over multipliers with $\eta > 0$ is not larger than the value of $f(\cdot, \theta_0)$ at any multiplier with $\eta = 0$, which proves (80).

Fix $\varepsilon > 0$ and pick $\xi = (\zeta, \eta, \lambda)$ with $\eta > 0$ and $f(\xi, \theta_0) \leq V_D(\theta_0) + 2\varepsilon$, which is possible in view of (80), so that requirement (a) of Step 1 holds. It remains to verify requirement (c). By Assumption SA-3(i), for P_* -almost every z , $y_n(z) := g(z; \theta_n) - \lambda^\top m(z; \theta_n) - \zeta \rightarrow y_0(z) := g(z; \theta_0) - \lambda^\top m(z; \theta_0) - \zeta$ along the given sequence. We now distinguish the two cases of Assumption SA-2(i).

In case (a), the conjugate ϕ^* is finite and convex, thus continuous, so $\eta \phi^*(y_n(z)/\eta) \rightarrow \eta \phi^*(y_0(z)/\eta)$ for P_* -almost every z . Assumption SA-3(iii) provides a radius $r > 0$ and a function $G \in L_1(P_*)$ dominating the integrand of $\Psi(\xi, \cdot)$ on $\mathcal{B}_r(\theta_0) \cap \Theta$. Applying Fatou's lemma to the nonnegative functions $z \mapsto G(z) - \eta \phi^*(y_n(z)/\eta)$ yields $\limsup_{n \rightarrow \infty} \Psi(\xi, \theta_n) \leq \Psi(\xi, \theta_0)$, and requirement (c) follows because the remaining terms of $f(\xi, \cdot)$ do not depend on θ .

In case (b), every subgradient of ϕ^* lies in $\text{dom}(\phi) \subseteq [0, \bar{x}]$, because ϕ^* is a pointwise supremum of the affine maps $y \mapsto ty - \phi(t)$ indexed by $t \in \text{dom}(\phi)$. Hence ϕ^* is finite, nondecreasing, and Lipschitz with constant $\bar{x}$. In particular, because $\phi^*(0) = 0$ and the Lipschitz property,

$$\eta \phi^*(y_n(z)/\eta) \leq \bar{x} |y_n(z)| \leq \bar{x} (G_g(z) + \|\lambda\| G_m(z) + |\zeta|)$$

for all n large and P_* -almost every z , by Assumption SA-3(iii). As the integrand converges P_* -almost surely by continuity of ϕ^* , requirement (c) follows as in case (a).

The requirements (a) and (c) of Step 1 are thus satisfied in both cases of Assumption SA-2(i), which shows upper semicontinuity. $\blacksquare$

SA8.18 Proof of Theorem SA-22

Proof. By Lemma SA-21, $V_D(\cdot; \delta)$ is real valued and upper semicontinuous on Θ . Since Θ is nonempty and compact, the extreme value theorem for real valued upper semicontinuous functions on compact sets yields a point $\theta^* \in \Theta$ such that

$$V_D(\theta^*; \delta) = \sup_{\theta \in \Theta} V_D(\theta; \delta).$$

Writing $V_D^* := V_D(\theta^*; \delta)$, the set of maximizers is

$$\arg \sup_{\theta \in \Theta} V_D(\theta; \delta) = \{\theta \in \Theta : V_D(\theta; \delta) \geq V_D^*\}.$$

It is nonempty and closed by upper semicontinuity, and therefore compact as a closed subset of Θ . Moreover, Assumption SA-3 supplies, at every $\theta \in \Theta$, the hypotheses of Theorem SA-19. Hence, at every outer maximizer, the set of minimizers of (21) is nonempty, compact, and convex. This proves part SA-22(i).

Compatibility for strong duality. Fix an arbitrary $\theta \in \Theta$, let P° denote the strictly feasible distribution of Assumption SA-1(iii) or Assumption SA-2(iii), and let $\Gamma(\theta; b)$ be the attainable moment set of (3). The entropy function is continuous at 1: under Assumption SA-1(ii), $\phi(0) < \infty$ and $\phi^\infty(1) < \infty$ imply $\text{dom}(\phi) = [0, \infty)$, whereas under Assumption SA-2(i), convexity and the stated domain conditions place 1 in the interior of $\text{dom}(\phi)$. The feasibility margin also implies $\delta > 0$. It remains only to verify $\mathcal{M}(\rho) \cap \text{ri}(\Gamma(\theta; \delta)) \neq \emptyset$.

When $\mathbf{d}_m = 0$ there is no moment constraint and nothing to verify. Suppose therefore that $\mathbf{d}_m \geq 1$, write $\bar{m}^\circ := \mathbb{E}_{P^\circ}[m(Z; \theta)] \in \mathcal{M}(\rho)$ and $\delta_\circ := D_\phi(P^\circ \| P_*) < \delta$. We first construct a constant $K < \infty$, a radius $r > 0$, and distributions $\{Q_v : \|v\| < r\} \subseteq \Delta(\mathcal{Z})$ with $\mathbb{E}_{Q_v}[\|m(Z; \theta)\|] < \infty$, $D_\phi(Q_v \| P_*) \leq K$, and $\mathbb{E}_{Q_v}[m(Z; \theta)] = \bar{\rho} + v$.

Under Assumption SA-1, every $Q \in \Delta(\mathcal{Z})$ satisfies $D_\phi(Q \| P_*) \leq \phi(0) + \phi^\infty(1) =: K$, which is finite by Assumption SA-1(ii). Indeed, writing $Q = wP_* + Q_\perp$ for the Lebesgue decomposition of Q with respect to P_* , convexity of ϕ gives $\phi(w) \leq (1 - w/T)\phi(0) + (w/T)\phi(T)$ for every $T \geq w$, and letting $T \rightarrow \infty$ yields $\phi(w) \leq \phi(0) + \phi^\infty(1)w$. Integrating against P_* and invoking Lemma SA-1 then delivers $D_\phi(Q \| P_*) \leq \phi(0) + \phi^\infty(1)\{\mathbb{E}_{P_*}[w(Z)] + Q_\perp(\mathcal{Z})\} = \phi(0) + \phi^\infty(1)$. The Slater condition of Assumption SA-1(iv) yields a radius $r > 0$ such that $\bar{\rho} + \mathcal{B}_r(0) \subseteq \text{co}(\mathcal{N}(\theta))$, and Carathéodory's theorem, exactly as in the proof of Lemma SA-18(i), implies that, for every $\|v\| < r$, there exists a finitely supported $Q_v \in \Delta(\mathcal{Z})$ with $\mathbb{E}_{Q_v}[m(Z; \theta)] = \bar{\rho} + v$.

Under Assumption SA-2, for $C > 0$ with $[0, C] \subseteq \text{dom}(\phi)$ let $\mathcal{Q}_C := \{Q \in \Delta(\mathcal{Z}) : Q \ll P_*, dQ/dP_* \leq C\}$. Every $Q \in \mathcal{Q}_C$ satisfies $\mathbb{E}_Q[\|m(Z; \theta)\|] \leq C \mathbb{E}_{P_*}[\|m(Z; \theta)\|] < \infty$, as $P_* \in \mathcal{P}(\mathcal{Z})$, and

$$D_\phi(Q \| P_*) = \mathbb{E}_{P_*}[\phi(\frac{dQ}{dP_*}(Z))] \leq \max\{\phi(0), \phi(C)\} < \infty,$$

because a convex function on an interval is bounded by its values at the endpoints. In case (b) of Assumption SA-2(i), Assumption SA-2(vi) gives $r > 0$ with $\bar{\rho} + \mathcal{B}_r(0) \subseteq \mathcal{N}_x(\theta)$, so each point $\bar{\rho} + v$ with $\|v\| < r$ is the moment under some $Q_v \in \mathcal{Q}_x$, and we may take $K := \max\{\phi(0), \phi(\bar{x})\}$, finite by Assumption SA-2(i). In case (a) of Assumption SA-2(i), let $\mathcal{N}_C(\theta) := \{\mathbb{E}_Q[m(Z; \theta)] : Q \in \mathcal{Q}_C\}$ and $\mathcal{N}_\infty(\theta) := \bigcup_{C \geq 1} \mathcal{N}_C(\theta)$, which is convex. We claim that $\text{co}(\mathcal{N}(\theta)) \subseteq \text{cl}(\mathcal{N}_\infty(\theta))$. Fix $z_0 \in \mathcal{Z}$ and $\varepsilon > 0$. The conditional distribution $P_*(\cdot | \mathcal{B}_\varepsilon(z_0))$ is well defined because $P_*(\mathcal{B}_\varepsilon(z_0)) > 0$ in view of the full-support condition of Assumption SA-2(v), belongs to $\mathcal{Q}_C$ for $C = 1/P_*(\mathcal{B}_\varepsilon(z_0))$, and its moment vector $\mathbb{E}_{P_*}[m(Z; \theta) | Z \in \mathcal{B}_\varepsilon(z_0)]$ converges to $m(z_0; \theta)$ as $\varepsilon \rightarrow 0$, by continuity of $m(\cdot; \theta)$. Finite mixtures of such conditional distributions, which again belong to some $\mathcal{Q}_C$, therefore approximate every convex combination of points of $\mathcal{N}(\theta)$, proving the claim. The Slater condition of Assumption SA-2(vi) then gives $r_1 > 0$ with $\bar{\rho} + \mathcal{B}_{r_1}(0) \subseteq \text{co}(\mathcal{N}(\theta)) \subseteq \text{cl}(\mathcal{N}_\infty(\theta))$, so that $\bar{\rho} \in \text{int}(\text{cl}(\mathcal{N}_\infty(\theta))) = \text{int}(\mathcal{N}_\infty(\theta))$, where the equality holds because a convex set and its closure share the same relative interior (Rockafellar, 1970, Theorem 6.3) and both sets are of full dimension. Since $\mathcal{N}_\infty(\theta)$ thus contains a neighborhood of $\bar{\rho}$, we may pick $\mathbf{d}_m + 1$ points of $\mathcal{N}_\infty(\theta)$ spanning a simplex whose interior contains $\bar{\rho}$, where the i -th point belongs to $\mathcal{N}_{C_i}(\theta)$ for some $C_i \geq 1$. Letting $C^* := \max_i C_i$, convexity of $\mathcal{N}_{C^*}(\theta)$ yields $r > 0$ with $\bar{\rho} + \mathcal{B}_r(0) \subseteq \mathcal{N}_{C^*}(\theta)$, and we take $K := \max\{\phi(0), \phi(C^*)\}$, finite as $[0, \infty) \subseteq \text{dom}(\phi)$.

With the family $\{Q_v\}$ at hand, fix $t \in (0, 1)$ small enough that $(1 - t)\delta_\circ + tK < \delta$ and define $P_v := (1 - t)P^\circ + tQ_v$ for $\|v\| < r$. The distribution P_v belongs to $\mathcal{P}(\mathcal{Z})$, since $P^\circ \in \mathcal{P}(\mathcal{Z})$ and Q_v is finitely supported under Assumption SA-1, whereas under Assumption SA-2 the density dQ_v/dP_* is bounded and $P_* \in \mathcal{P}(\mathcal{Z})$. Moreover, $D_\phi(P_v \| P_*) \leq (1 - t)\delta_\circ + tK < \delta$, by convexity of $D_\phi(\cdot \| P_*)$, and $\mathbb{E}_{P_v}[m(Z; \theta)] = (1 - t)\bar{m}^\circ + t(\bar{\rho} + v)$. Hence $\Gamma(\theta; \delta)$ contains $\mathcal{B}_{tr}(\bar{m}_c)$, where $\bar{m}_c := (1 - t)\bar{m}^\circ + t\bar{\rho}$, so that $\bar{m}_c \in \text{int}(\Gamma(\theta; \delta)) \subseteq \text{ri}(\Gamma(\theta; \delta))$, whereas $\bar{m}_c \in \mathcal{M}(\rho)$ as a convex combination of the points $\bar{m}^\circ$ and $\bar{\rho}$ of the convex set $\mathcal{M}(\rho)$. Since θ was arbitrary, the compatibility condition holds at every $\theta \in \Theta$.

Under the additional integrability condition in part SA-22(ii), all the conditions of Theorem SA-11 therefore hold at every $\theta \in \Theta$. Applying that theorem pointwise gives

$$V_D(\theta; \delta) = V_P(\theta; \delta) \quad \text{for every } \theta \in \Theta.$$

The two value functions therefore have exactly the same maximizers, which proves part SA-22(ii). $\blacksquare$

SA8.19 Proof of Lemma SA-23

Proof. Entropy function and divergence. We apply Proposition SA-13 twice, with P in place of P_* . First, let $\phi_{\underline{u}} := \phi_{\text{TV}} + \iota_{[\underline{u}, \infty)}$. If $\underline{u} > 0$, part (ii) of that proposition with $\kappa = \underline{u} \in (0, 1)$ shows that $\phi_{\underline{u}}$ is an entropy function with $\phi_{\underline{u}}^\infty(1) = \phi_{\text{TV}}^\infty(1) = 1/2$, and that $D_{\phi_{\underline{u}}}(Q \| P)$ equals $D_{\phi_{\text{TV}}}(Q \| P)$ if $\underline{u}P \leq Q$ and $+\infty$ otherwise. If $\underline{u} = 0$, then $\phi_{\underline{u}} = \phi_{\text{TV}}$, as $\phi_{\text{TV}} = +\infty$ on $(-\infty, 0)$, and the same conclusions hold trivially. Second, if $\bar{u} < \infty$, part (i) with $\kappa = 1/\bar{u} \in (0, 1)$, applied to $\phi_{\underline{u}}$, shows that $\phi_{\bar{u}} = \phi_{\underline{u}} + \iota_{[0, \bar{u}]}$ is an entropy function with $\phi_{\bar{u}}^\infty(1) = +\infty$, and that $D_{\phi_{\bar{u}}}(Q \| P)$ equals $D_{\phi_{\underline{u}}}(Q \| P)$ if $Q \leq \bar{u}P$ and $+\infty$ otherwise. If $\bar{u} = +\infty$, then $\phi_{\bar{u}} = \phi_{\underline{u}}$ and there is nothing to add. Combining the two steps with the definition of $\mathcal{P}_{\mathbb{U}}(P)$, we obtain $D_{\phi_{\bar{u}}}(Q \| P) = +\infty$ if $Q \notin \mathcal{P}_{\mathbb{U}}(P)$ and $D_{\phi_{\bar{u}}}(Q \| P) = D_{\phi_{\text{TV}}}(Q \| P) = \text{TV}(Q \| P)$ otherwise, where the second equality is Lemma SA-2. Finally, $\phi_{\bar{u}}$ is continuous at 1, as it coincides on $\mathbb{U}$ with the continuous function ϕ_{TV} and 1 is an interior point of $\mathbb{U}$. The identity between (22) and the divergence neighborhood follows, since every $Q \in \mathcal{P}_{\mathbb{U}}(P)$ belongs to $\mathcal{P}(\mathcal{Z})$ under either condition of the lemma. This is immediate when $g(\cdot; \theta)$ and $m(\cdot; \theta)$ are bounded. When $\bar{u} < \infty$ and $P \in \mathcal{P}(\mathcal{Z})$, the inequality $Q \leq \bar{u}P$ gives $\mathbb{E}_Q[|g(Z; \theta)|] \leq \bar{u} \mathbb{E}_P[|g(Z; \theta)|] < \infty$, and likewise for $\|m(Z; \theta)\|$.

Conjugate and perspective. For $y \in \mathbb{R}$, $\phi_{\bar{u}}^*(y) = \sup_{u \in \mathbb{U}} \{uy - \frac{1}{2}|u-1|\}$. The function $u \mapsto uy - \frac{1}{2}|u-1|$ is concave and piecewise linear on $\mathbb{U}$, with slope $y + \frac{1}{2}$ on $u < 1$ and slope $y - \frac{1}{2}$ on $u > 1$, so its supremum over the interval is attained at an endpoint of $\mathbb{U}$ or at the kink $u = 1$. When $\bar{u} < \infty$ this gives $\phi_{\bar{u}}^*(y) = \max_{u \in \{\underline{u}, 1, \bar{u}\}} \{uy - \frac{1}{2}|u-1|\}$, a finite function of y . When $\bar{u} = +\infty$, the supremum is $+\infty$ for $y > \frac{1}{2}$, as the slope on $u > 1$ is then positive, and equals $\max\{uy - \frac{1}{2}(1-\underline{u}), y\}$ for $y \leq \frac{1}{2}$, as that slope is then nonpositive. For $\eta > 0$, the perspective is $(\phi_{\bar{u}}^*)^\pi(x, \eta) = \eta \phi_{\bar{u}}^*(x/\eta)$, and multiplying each branch through by η yields (23), the domain restriction $x/\eta \leq \frac{1}{2}$ becoming $x \leq \eta/2$. At $\eta = 0$ the perspective is the recession function of $\phi_{\bar{u}}^*$, that is, $\lim_{t \rightarrow \infty} \phi_{\bar{u}}^*(y_0 + tx)/t$ for any y_0 in the domain of $\phi_{\bar{u}}^*$. When $\bar{u} < \infty$ this limit is $\max_{u \in \{\underline{u}, 1, \bar{u}\}} ux$, which coincides with the first branch of (23) at $\eta = 0$. When $\bar{u} = +\infty$ and $x > 0$ the argument $y_0 + tx$ exceeds $\frac{1}{2}$ for t large and the limit is $+\infty$, in agreement with the third branch at $\eta = 0$, whereas for $x \leq 0$ the limit is $\max\{\underline{u}x, x\}$, in agreement with the second branch. Hence $(\phi_{\bar{u}}^*)^\pi = r$ on $\mathbb{R} \times \mathbb{R}_+$. ■

SA8.20 Proof of Lemma SA-24

Proof. Throughout the proof we fix θ and abbreviate $g(\cdot) := g(\cdot; \theta)$, $m(\cdot) := m(\cdot; \theta)$, and $\mathcal{F} := \mathcal{F}(\theta)$. Both g and m are bounded, as they are continuous on the compact $\mathcal{Z}$.

Part (i). Let $Q \in \mathcal{P}_{\mathbb{U}}(P)$ have Lebesgue decomposition $Q = wP + Q^\perp$ with respect to P , so that $w \in \mathbb{U}$ P -almost surely and $Q^\perp = 0$ when $\bar{u} < \infty$ (Lemma SA-23), and let $\xi = (\zeta, \eta, \lambda) \in \mathcal{F}$. Put $x(z) := g(z) - \lambda^\top m(z) - \zeta$. By (r1), $w(z)x(z) \leq r(x(z), \eta) + \eta|w(z) - 1|/2$ for P -almost every z , whence, integrating against P ,

$$\int x \, dQ = \mathbb{E}_P[w x] + \int x \, dQ^\perp \leq \mathbb{E}_P[r(x, \eta)] + \frac{\eta}{2} \mathbb{E}_P|w - 1| + \int x \, dQ^\perp.$$

When $\bar{u} < \infty$ the last integral vanishes. When $\bar{u} = +\infty$, dual feasibility gives $x \leq \eta/2$ on $\mathcal{Z}$, so that $\int x \, dQ^\perp \leq \frac{\eta}{2} Q^\perp(\mathcal{Z})$. In both cases, since $\text{TV}(Q \| P) = \frac{1}{2} \mathbb{E}_P|w - 1| + \frac{1}{2} Q^\perp(\mathcal{Z})$ by Lemmas SA-1 and SA-2,

$$\mathbb{E}_Q[g] - \lambda^\top \mathbb{E}_Q[m] - \zeta = \int x \, dQ \leq P\ell_{\xi, \theta} + \eta \text{TV}(Q \| P).$$

Rearranging and using the definition (25) of J gives the displayed inequality. If $Q \in \mathcal{P}_{\mathbb{U}}(\theta; \rho, P, \delta)$, then $\text{TV}(Q \| P) \leq \delta$ and, since $\lambda \in \mathcal{R}_\lambda$ and $\mathbb{E}_Q[m] \in \mathcal{M}(\rho)$, the support function identity of Section SA1.2 gives $\lambda^\top \mathbb{E}_Q[m] \leq \iota_{\mathcal{M}(\rho)}^*(\lambda) = \lambda_{12}^\top \rho$, so both terms on the right are nonnegative. Taking the infimum over $\xi \in \mathcal{F}$ and the supremum over Q shows that $\bar{K}_\delta(\theta; \rho, P)$ dominates the upper bound. Moreover, when the robustness set is nonempty this bound exceeds $-\infty$, as g is bounded, while $\bar{K}_\delta(\theta; \rho, P) \leq \sup_{z \in \mathcal{Z}} g(z) < \infty$ by evaluating J at the feasible point $(\sup_z g(z), 0, 0)$, at which the integrand equals $\underline{u}(g - \sup_z g) \leq 0$ by (23).

Part (ii). We apply Theorem SA-11 with the entropy $\phi_{\bar{u}}$ of Lemma SA-23 and the baseline P , which is legitimate because that theorem is stated for an arbitrary baseline distribution and an arbitrary entropy function continuous at 1. Its hypotheses hold: $\delta > 0$; $\int_{\mathcal{Z}} |g| \, d\nu < \infty$ for every finite measure ν ; and the compatibility condition $\mathcal{M}(\rho) \cap \text{ri}(\Gamma(\delta)) \neq \emptyset$ holds because, by Lemma SA-23, the attainable set $\Gamma(\delta)$ of that theorem coincides with $\Gamma_{\mathbb{U}}(\theta; P, \delta)$, and the constraint qualification supplies a point of $\mathcal{M}(\rho)$ in its interior, hence in its relative interior. The theorem states that the upper bound equals the infimum over $(\zeta, \eta, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$ of $\lambda_{12}^\top \rho + \zeta + \eta\delta + \mathbb{E}_P[(\phi_{\bar{u}}^*)^\pi(g - \lambda^\top m - \zeta, \eta)]$ subject to $\zeta + \eta\phi_{\bar{u}}^\infty(1) \geq \sigma(\lambda, \theta)$. By Lemma SA-23, the perspective is r and the recession constant is $1/2$ when $\bar{u} = +\infty$, so that the program is $\inf_{\xi \in \mathcal{F}} J(\xi, \theta; \rho, P)$, and $+\infty$ when $\bar{u} < \infty$, so that the constraint is vacuous and the program is again $\inf_{\xi \in \mathcal{F}} J(\xi, \theta; \rho, P)$ with $\mathcal{F} = \mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$. This is (27).

Part (iii). Recession cone. When $\bar{u} < \infty$, $\mathcal{F}$ is the closed convex cone $\mathbb{R} \times \mathbb{R}_+ \times \mathcal{R}_\lambda$ and coincides with its recession cone. When $\bar{u} = +\infty$, $\mathcal{F}$ is the sublevel set $\{\xi : f(\xi) \leq 0\}$ of the closed proper convex function $f(\xi) := \sigma(\lambda, \theta) - \zeta - \eta/2 + \iota_{\mathbb{R}_+}(\eta) + \iota_{\mathcal{R}_\lambda}(\lambda)$, in which $\lambda \mapsto \sigma(\lambda, \theta)$ is finite, continuous, and convex on $\mathbb{R}^d$, as a supremum over the compact $\mathcal{Z}$ of functions affine in λ with bounded intercepts. Its recession function is $f^\infty(d) = \sigma^\infty(\lambda_d, \theta) - \zeta_d - \eta_d/2 + \iota_{\mathbb{R}_+}(\eta_d) + \iota_{\mathcal{R}_\lambda}(\lambda_d)$: indeed, for any λ° , $\sigma(\lambda^\circ + t\lambda_d, \theta)/t = \sup_{z \in \mathcal{Z}} \{(g(z) - \lambda^{\circ\top} m(z))/t - \lambda_d^\top m(z)\} \rightarrow \sigma^\infty(\lambda_d, \theta)$ as $t \rightarrow \infty$, the first term in the braces converging to zero uniformly in z , while

the recession function of the indicator of a closed convex cone is the indicator itself. The sublevel-set description of recession cones (Rockafellar, 1970, Theorem 8.7) then gives $0^+\mathcal{F} = \{d : f^\infty(d) \leq 0\}$, which is the set displayed in the statement.

Recession function of the objective. Fix $\xi^\circ \in \mathcal{F}$ and $d \in 0^+\mathcal{F}$, and write $x^\circ := g - \lambda^{\circ\top}m - \zeta^\circ$ and $x_d := -\lambda_d^\top m - \zeta_d$. For $t > 0$, positive homogeneity (r2) gives $r(x^\circ + tx_d, \eta^\circ + t\eta_d)/t = r(x^\circ/t + x_d, \eta^\circ/t + \eta_d)$, which converges to $r(x_d, \eta_d)$ pointwise on $\mathcal{Z}$ as $t \rightarrow \infty$: when $\bar{u} < \infty$ by continuity of r on $\mathbb{R} \times \mathbb{R}_+$, and when $\bar{u} = +\infty$ by continuity of r on its closed domain $\{x \leq \eta/2\}$, in which the points $(x^\circ/t + x_d, \eta^\circ/t + \eta_d)$ lie for every $t > 0$ because $\xi^\circ + td \in \mathcal{F}$, and to which the limit (x_d, η_d) belongs because $d \in 0^+\mathcal{F}$. The convergence is dominated, by (r3) and the boundedness of g and m , so the dominated convergence theorem yields $\lim_{t \rightarrow \infty} \{J(\xi^\circ + td, \theta; \rho, P) - J(\xi^\circ, \theta; \rho, P)\}/t = \zeta_d + \eta_d \delta + \lambda_{d,12}^\top \rho + P r(x_d, \eta_d)$, which is the recession function of the closed proper convex function $J(\cdot, \theta; \rho, P) + \iota_{\mathcal{F}}$ in the direction d (Definition SA-2). The function is convex because $\ell_{\xi, \theta}(z)$ is convex in ξ for every z by (r2), closed because $\mathcal{F}$ is closed and $\xi \mapsto J(\xi, \theta; \rho, P)$ is continuous on $\mathcal{F}$ by the dominated convergence theorem and (r3), and proper because it is finite on $\mathcal{F}$.

Positivity. Suppose the constraint qualification holds at (θ, ρ, P) with radius r and point $w_0 \in \mathcal{M}(\rho)$, and let $\bar{d} \in 0^+\mathcal{F}$ be nonzero. Applying part (i) to the program with g replaced by 0, for which d is a feasible point of its dual feasible set by the description of $0^+\mathcal{F}$ just obtained, we get, for every $Q \in \mathcal{P}_U(P)$,

$$J^\infty(d, \theta; \rho, P) \geq \eta_d(\delta - \text{TV}(Q \| P)) + (\lambda_{d,12}^\top - \lambda_d^\top \mathbb{E}_Q[m]). \quad (81)$$

We distinguish three cases. If $\eta_d > 0$, take $Q = Q^\circ$, a distribution of the family with $\mathbb{E}_{Q^\circ}[m] \in \mathcal{M}(\rho)$ and $\text{TV}(Q^\circ \| P) < \delta$, which exists by part (iv) below, whose proof does not use the present part. Then the first term in (81) is strictly positive and the second is nonnegative, as in part (i). If $\eta_d = 0$ and $\lambda_d \neq 0$, the point $w := w_0 - \frac{r}{2}\lambda_d/\|\lambda_d\|$ lies in $\mathcal{B}_r(w_0) \subseteq \Gamma_U(\theta; P, \delta)$, so there is $Q \in \mathcal{P}_U(P)$ with $\mathbb{E}_Q[m] = w$, and (81) gives $J^\infty(d, \theta; \rho, P) \geq \lambda_{d,12}^\top \rho - \lambda_d^\top w_0 + \frac{r}{2}\|\lambda_d\| \geq \frac{r}{2}\|\lambda_d\| > 0$, the second inequality by the support function identity since $w_0 \in \mathcal{M}(\rho)$ and $\lambda_d \in \mathcal{R}_\lambda$. If $\eta_d = 0$ and $\lambda_d = 0$, then $\zeta_d \neq 0$ and, by (23) at $\eta = 0$, $J^\infty(d, \theta; \rho, P) = \zeta_d + r(-\zeta_d, 0)$. When $\bar{u} = +\infty$ the recession-cone description forces $\zeta_d \geq \sigma^\infty(0, \theta) = 0$, hence $\zeta_d > 0$ and $r(-\zeta_d, 0) = -\underline{u}\zeta_d$, so that the recession function equals $(1 - \underline{u})\zeta_d > 0$. When $\bar{u} < \infty$ the same computation applies if $\zeta_d > 0$, whereas if $\zeta_d < 0$ then $r(-\zeta_d, 0) = -\bar{u}\zeta_d$ and the recession function equals $(\bar{u} - 1)|\zeta_d| > 0$, where the inequality follows from $\bar{u} > 1$.

Existence. The closed proper convex function $J(\cdot, \theta; \rho, P) + \iota_{\mathcal{F}}$ has no nonzero direction of recession, since its recession function is $+\infty$ off $0^+\mathcal{F}$ and strictly positive on $0^+\mathcal{F} \setminus \{0\}$. By Proposition 3.2.2 in Bertsekas (2009), its set of minimizers is nonempty and compact, and it is convex as the set of minimizers of a convex function. This set is $\Xi_\delta(\theta; \rho, P)$.

Part (iv). Suppose first that the constraint qualification holds with radius r and point $w_0 \in \mathcal{M}(\rho)$. Since $\Gamma_U(\theta; P, \delta) \subseteq \Gamma_U(\theta; P, 1)$, the second requirement holds. For the first, let $b_P := \mathbb{E}_P[m]$, which belongs to $\Gamma_U(\theta; P, \delta)$ because $P \in \mathcal{P}_U(P)$ and $\text{TV}(P \| P) = 0$. If $b_P = w_0$, then $Q^\circ := P$ satisfies the first requirement with $\text{TV}(Q^\circ \| P) = 0 < \delta$. Otherwise, choose $\varepsilon > 0$ so small that $w_1 := w_0 + \varepsilon(w_0 - b_P) \in \mathcal{B}_r(w_0)$, let $Q_1 \in \mathcal{P}_U(P)$ attain w_1 with $\text{TV}(Q_1 \| P) \leq \delta$, and set $Q^\circ := (1 + \varepsilon)^{-1}Q_1 + \varepsilon(1 + \varepsilon)^{-1}P$, which belongs to the convex set $\mathcal{P}_U(P)$ and satisfies $\mathbb{E}_{Q^\circ}[m] = (1 + \varepsilon)^{-1}(w_1 + \varepsilon b_P) = w_0 \in \mathcal{M}(\rho)$ and, by convexity of $\text{TV}(\cdot \| P)$ and $\text{TV}(P \| P) = 0$, $\text{TV}(Q^\circ \| P) \leq (1 + \varepsilon)^{-1}\text{TV}(Q_1 \| P) \leq \delta/(1 + \varepsilon) < \delta$.

Conversely, let Q° satisfy the first requirement, with $\delta^\circ := \text{TV}(Q^\circ \| P) < \delta$ and $w^\circ := \mathbb{E}_{Q^\circ}[m] \in \mathcal{M}(\rho)$, and let $w_1 \in \mathcal{M}(\rho)$ and $r_1 > 0$ be such that $\mathcal{B}_{r_1}(w_1) \subseteq \Gamma_U(\theta; P, 1)$. As $\delta^\circ < \delta \leq 1$, we may put $t := (\delta - \delta^\circ)/(1 - \delta^\circ) \in (0, 1]$. For any $w \in \mathcal{B}_{r_1}(w_1)$, attained by some $Q_w \in \mathcal{P}_U(P)$, the mixture $Q_t := (1 - t)Q^\circ + tQ_w$ belongs to $\mathcal{P}_U(P)$, has moment vector $(1 - t)w^\circ + tw$, and satisfies $\text{TV}(Q_t \| P) \leq (1 - t)\delta^\circ + t\text{TV}(Q_w \| P) \leq (1 - t)\delta^\circ + t = \delta$, using $\text{TV} \leq 1$. Hence $\Gamma_U(\theta; P, \delta)$ contains $\mathcal{B}_{tr_1}((1 - t)w^\circ + tw_1)$, whose center is a point of the convex set $\mathcal{M}(\rho)$, which is the constraint qualification with radius tr_1 . Finally, the first requirement says precisely that the infimum (30) is strictly below δ . ■

SA8.21 Proof of Lemma SA-25

Proof. Throughout the proof we fix θ and abbreviate $m(\cdot) := m(\cdot; \theta)$, $\mathcal{F}_D := \mathcal{F}_D(\theta)$, and $\sigma^\infty(\lambda) := \sigma^\infty(\lambda, \theta)$. We recall that $\lambda^\top \bar{\rho} = \lambda_{12}^\top \rho$ for every $\lambda \in \mathbb{R}^{d_m}$, and set $\bar{M} := \sup_{z \in \mathcal{Z}} \|m(z)\|$, which is finite because m is continuous on the compact $\mathcal{Z}$. For $\lambda \in \mathbb{R}^{d_m}$ we write $h_\lambda := -\lambda^\top m(\cdot)$, a continuous function on $\mathcal{Z}$ with $\sup_z h_\lambda(z) = \sigma^\infty(\lambda) \leq \|\lambda\| \bar{M}$, and we call ζ admissible for λ if $(\zeta, \lambda) \in \mathcal{F}_D$, that is, if $\zeta \geq \sigma^\infty(\lambda) - \frac{1}{2}$ when $\bar{u} = +\infty$ and always when $\bar{u} < \infty$.

A variational inequality. We first show that for every $(\zeta, \lambda) \in \mathcal{F}_D$ and every $Q \in \mathcal{P}_U(P)$,

$$\mathbb{E}_Q[h_\lambda] - \zeta \leq \mathbb{E}_P[h_\lambda - \zeta, 1] + \text{TV}(Q \| P). \quad (82)$$

This is the inequality of Lemma SA-24(i), applied with 0 in place of $g(\cdot; \theta)$ and with $\eta = 1$, once the terms δ and $\lambda_{12}^\top \rho$ common to both sides cancel. Indeed, for this choice the dual support function reduces to $\sigma^\infty(\lambda)$, so that $(\zeta, 1, \lambda)$ belongs to the dual feasible set (26) if and only if $(\zeta, \lambda) \in \mathcal{F}_D$.

Attainment of the primal projection. Let $\mathcal{K}_{\theta, \rho} := \{Q \in \mathcal{P}_U(P) : \mathbb{E}_Q[m(Z; \theta)] \in \mathcal{M}(\rho)\}$. Because $\mathcal{Z}$ is compact, $\Delta(\mathcal{Z})$ is weakly compact. The set $\mathcal{P}_U(P)$ is weakly closed: if $Q_n \Rightarrow Q$ with $Q_n \in \mathcal{P}_U(P)$, the portmanteau theorem gives $Q(U) \leq \liminf_n Q_n(U) \leq \bar{u}P(U)$ for every open U and $Q(C) \geq \limsup_n Q_n(C) \geq \underline{u}P(C)$ for every closed C .

Continuity of m and closedness of $\mathcal{M}(\rho)$ imply that $\mathcal{K}_{\theta,\rho}$ is weakly closed, hence weakly compact. Moreover, on the compact metric space $\mathcal{Z}$, $2\text{TV}(Q \| P) = \sup\{|\mathbb{E}_Q[a] - \mathbb{E}_P[a]| : a \in C(\mathcal{Z}), \|a\|_\infty \leq 1\}$, so $Q \mapsto \text{TV}(Q \| P)$ is weakly lower semicontinuous, as a supremum of weakly continuous functions. It follows that the infimum in (30) is attained whenever $\mathcal{K}_{\theta,\rho}$ is nonempty. The same argument shows that $\{Q \in \mathcal{P}_\mathbb{U}(P) : \text{TV}(Q \| P) \leq \delta'\}$ is weakly compact for every $\delta' \geq 0$, so that, as $Q \mapsto \mathbb{E}_Q[m]$ is weakly continuous and linear, the attainable sets $\Gamma_\mathbb{U}(\theta; P, \delta')$ are compact and convex.

Part (i). Suppose $\mathcal{B}_r(w_0) \subseteq \Gamma_\mathbb{U}(\theta; P, 1)$ with $w_0 \in \mathcal{M}(\rho)$. We prove the weak direction first, then the multiplier and intercept bounds, then the strong direction, and finally attainment.

Weak direction. Fix $(\zeta, \lambda) \in \mathcal{F}_\mathbb{D}$ and let Q be feasible for the projection problem, so that $w := \mathbb{E}_Q[m] \in \mathcal{M}(\rho)$. Since $\lambda \in \mathcal{R}_\lambda$, the support function of $\mathcal{M}(\rho)$ gives $\lambda^\top w \leq \lambda_{12}^\top \rho$, whence $\mathbb{E}_Q[h_\lambda] = -\lambda^\top w \geq -\lambda_{12}^\top \rho$. Combining with (82), $\text{TV}(Q \| P) \geq v_P(\zeta, \lambda)$. Taking the infimum over feasible Q and the supremum over $(\zeta, \lambda) \in \mathcal{F}_\mathbb{D}$ yields $\text{D}(\theta; \rho, P) \geq \sup_{\mathcal{F}_\mathbb{D}} v_P$.

Multiplier bound. For $(\zeta, \lambda) \in \mathcal{F}_\mathbb{D}$ with $\lambda \neq 0$ and any $r' < r$, the point $w := w_0 - r'\lambda/\|\lambda\|$ lies in $\mathcal{B}_r(w_0)$ and is therefore attained by some $Q_w \in \mathcal{P}_\mathbb{U}(P)$. Applying (82) to Q_w and using $\text{TV} \leq 1$,

$$1 \geq \text{TV}(Q_w \| P) \geq -\zeta - \lambda^\top w - \mathbb{E}_P[r(h_\lambda - \zeta, 1)] = v_P(\zeta, \lambda) + \lambda_{12}^\top \rho - \lambda^\top w \geq v_P(\zeta, \lambda) + r'\|\lambda\|,$$

the last inequality because $\lambda_{12}^\top \rho \geq \lambda^\top w_0$ and $\lambda^\top (w_0 - w) = r'\|\lambda\|$. The point $(\zeta, \lambda) = (-\frac{1}{2}, 0)$ is admissible and, by (23), $v_P(-\frac{1}{2}, 0) = -r(\frac{1}{2}, 1) + \frac{1}{2} = 0$, so $\sup_{\mathcal{F}_\mathbb{D}} v_P \geq 0$, and every (ζ, λ) with $v_P(\zeta, \lambda) \geq 0$, in particular every maximizer, satisfies $\|\lambda\| \leq 1/r'$ for every $r' < r$, hence $\|\lambda\| \leq 1/r$.

The intercept. Fix λ and consider $F_\lambda(\zeta) := \zeta + \mathbb{E}_P[r(h_\lambda - \zeta, 1)]$ on the set of admissible ζ , so that $v_P(\zeta, \lambda) = -F_\lambda(\zeta) - \lambda_{12}^\top \rho$. The function F_λ is finite and convex, as $r(\cdot, 1)$ is convex, and its one-sided derivatives are

$$F'_{\lambda,+}(\zeta) = 1 - \mathbb{E}_P[r'_-(h_\lambda - \zeta)], \quad F'_{\lambda,-}(\zeta) = 1 - \mathbb{E}_P[r'_+(h_\lambda - \zeta)],$$

where $r'_\pm$ denote the one-sided derivatives of $x \mapsto r(x, 1)$ on its domain, by the dominated convergence theorem applied to difference quotients bounded by $w^\dagger$ ((r3)). From (23), $r'_-(x)$ equals $\underline{u}$ for $x \leq -\frac{1}{2}$, 1 for $-\frac{1}{2} < x \leq \frac{1}{2}$, and $\bar{u}$ for $x > \frac{1}{2}$, while $r'_+(x)$ equals $\underline{u}$ for $x < -\frac{1}{2}$, 1 for $-\frac{1}{2} \leq x < \frac{1}{2}$, and $\bar{u}$ for $x \geq \frac{1}{2}$, where these last two values are relevant only when $\bar{u} < \infty$. In every case the one-sided derivatives take values in $\mathbb{U}$ and the subdifferential of $r(\cdot, 1)$ at x is the interval $[r'_-(x), r'_+(x)] \subseteq \mathbb{U}$. We claim that F_λ attains its minimum over the admissible set and that every minimizer ζ^* satisfies $|\zeta^*| \leq \|\lambda\|\bar{M} + \frac{1}{2}$. When $\bar{u} < \infty$ every $\zeta \in \mathbb{R}$ is admissible. For $\zeta > \sigma^\infty(\lambda) + \frac{1}{2}$ we have $h_\lambda - \zeta < -\frac{1}{2}$ on $\mathcal{Z}$, so $F'_{\lambda,+}(\zeta) = 1 - \underline{u} > 0$, and for $\zeta < \inf_z h_\lambda(z) - \frac{1}{2}$ we have $h_\lambda - \zeta > \frac{1}{2}$, so $F'_{\lambda,-}(\zeta) = 1 - \bar{u} < 0$, as $\bar{u} > 1$. The convex function F_λ is therefore increasing to the right of $\sigma^\infty(\lambda) + \frac{1}{2}$ and decreasing to the left of $\inf_z h_\lambda - \frac{1}{2}$, so it attains its minimum, and only there, on the compact interval between these two points, whose endpoints are bounded by $\|\lambda\|\bar{M} + \frac{1}{2}$ in absolute value. When $\bar{u} = +\infty$, the admissible set is the closed half line $[\zeta_\lambda, \infty)$ with $\zeta_\lambda := \sigma^\infty(\lambda) - \frac{1}{2}$, on which $h_\lambda - \zeta \leq \frac{1}{2}$ and hence $r'_-(h_\lambda - \zeta) \leq 1$, so that $F'_{\lambda,+} \geq 0$ throughout: F_λ is nondecreasing on the half line and the endpoint ζ_λ is a minimizer. Any other minimizer ζ^* makes F_λ constant on $[\zeta_\lambda, \zeta^*]$, hence $\mathbb{E}_P[r'_-(h_\lambda - \zeta')] = 1$, that is $P(h_\lambda - \zeta' \leq -\frac{1}{2}) = 0$, for every $\zeta' \in (\zeta_\lambda, \zeta^*)$, which forces $\zeta^* \leq \inf_z h_\lambda(z) + \frac{1}{2}$ up to P -null sets and in particular $|\zeta^*| \leq \|\lambda\|\bar{M} + \frac{1}{2}$. Combined with the multiplier bound, every maximizer of v_P over $\mathcal{F}_\mathbb{D}$ satisfies $|\zeta^*| \leq \bar{M}/r + \frac{1}{2}$, as claimed.

Strong direction. Define $T : \mathbb{R}^{\text{dm}} \rightarrow [0, +\infty]$ by

$$T(w) := \inf\{\text{TV}(Q \| P) : Q \in \mathcal{P}_\mathbb{U}(P), \mathbb{E}_Q[m] = w\},$$

with $T(w) = +\infty$ when no such Q exists, so that $\text{D}(\theta; \rho, P) = \inf_{w \in \mathbb{R}^{\text{dm}}} \{T(w) + \iota_{\mathcal{M}(\rho)}(w)\}$ by definition. The function T is proper, since $T \geq 0$ and $T(\mathbb{E}_P[m]) = 0$, its effective domain is $\Gamma_\mathbb{U}(\theta; P, 1)$, and it is convex: given w, w' with near-optimal Q, Q' and $t \in (0, 1)$, the mixture $tQ + (1-t)Q'$ belongs to the convex set $\mathcal{P}_\mathbb{U}(P)$, attains the moment vector $tw + (1-t)w'$, and, by convexity of $\text{TV}(\cdot \| P)$, satisfies $\text{TV}(tQ + (1-t)Q' \| P) \leq t\text{TV}(Q \| P) + (1-t)\text{TV}(Q' \| P)$. We apply Lemma SA-8 to the proper convex functions T and $\iota_{\mathcal{M}(\rho)}$, where the latter is polyhedral. The hypothesis of the lemma is that $\mathcal{M}(\rho)$ meets the relative interior of $\text{dom}(T)$, and $w_0 \in \mathcal{M}(\rho)$ is an interior point of $\text{dom}(T)$ by assumption. The lemma yields, with the supremum attained because the common value is finite,

$$\text{D}(\theta; \rho, P) = \sup_{x^* \in \mathbb{R}^{\text{dm}}} \{-T^*(x^*) - \iota_{\mathcal{M}(\rho)}^*(-x^*)\} = \max_{\lambda \in \mathcal{R}_\lambda} \{-\lambda_{12}^\top \rho - T^*(-\lambda)\},$$

where, at $\lambda := -x^*$, the support function $\iota_{\mathcal{M}(\rho)}^*(\lambda)$ equals $\lambda_{12}^\top \rho$ when $\lambda \in \mathcal{R}_\lambda$ and $+\infty$ otherwise. It remains to compute, for $\lambda \in \mathcal{R}_\lambda$,

$$T^*(-\lambda) = \sup_{w \in \mathbb{R}^{\text{dm}}} \{-\lambda^\top w - T(w)\} = \sup_{Q \in \mathcal{P}_\mathbb{U}(P)} \{\mathbb{E}_Q[h_\lambda] - \text{TV}(Q \| P)\},$$

where the second equality collapses the partial minimization inside T . We claim that

$$\sup_{Q \in \mathcal{P}_\mathbb{U}(P)} \{\mathbb{E}_Q[h_\lambda] - \text{TV}(Q \| P)\} = \min_{\zeta \text{ admissible for } \lambda} F_\lambda(\zeta). \quad (83)$$

The inequality “ $\leq$ ” is (82) rearranged, valid for every admissible ζ . For “ $\geq$ ” we exhibit, for a minimizer ζ^* of F_λ over the admissible set, a distribution $Q^* \in \mathcal{P}_\mathbb{U}(P)$ with $\mathbb{E}_{Q^*}[h_\lambda] - \text{TV}(Q^* \| P) = F_\lambda(\zeta^*)$, by matching the mass of the pointwise maximizers in (r1). Recall that $u \in [r'_-(x), r'_+(x)]$ if and only if $u \in \mathbb{U}$ attains the supremum in (r1) at $(x, 1)$, that is, $ux = r(x, 1) + \frac{1}{2}|u - 1|$, the standard characterization of the subdifferential of a conjugate function.

Suppose first that $\bar{u} < \infty$, so that every $\zeta \in \mathbb{R}$ is admissible and any minimizer ζ^* satisfies $F'_{\lambda,-}(\zeta^*) \leq 0 \leq F'_{\lambda,+}(\zeta^*)$. Then $\mathbb{E}_P[r'_-(h_\lambda - \zeta^*)] \leq 1 \leq \mathbb{E}_P[r'_+(h_\lambda - \zeta^*)]$, and there is $t \in [0, 1]$ such that $w^* := (1-t)r'_-(h_\lambda - \zeta^*) + tr'_+(h_\lambda - \zeta^*)$ satisfies $\mathbb{E}_P[w^*] = 1$. As r'_- and r'_+ take values in $\mathbb{U}$, so does w^* . Set $Q^* := w^*P \in \mathcal{P}_\mathbb{U}(P)$. Since $w^*(z)$ lies in the subdifferential of $r(\cdot, 1)$ at $h_\lambda(z) - \zeta^*$, $w^*(h_\lambda - \zeta^*) = r(h_\lambda - \zeta^*, 1) + \frac{1}{2}|w^* - 1|$ pointwise, and integrating against P gives $\mathbb{E}_{Q^*}[h_\lambda] - \zeta^* = \mathbb{E}_P[r(h_\lambda - \zeta^*, 1)] + \text{TV}(Q^* \| P)$, that is, $\mathbb{E}_{Q^*}[h_\lambda] - \text{TV}(Q^* \| P) = F_\lambda(\zeta^*)$. Suppose next that $\bar{u} = +\infty$, and take $\zeta^* = \zeta_\lambda$, which is a minimizer by the intercept step and at which $F'_{\lambda,+}(\zeta_\lambda) \geq 0$, that is, $\mathbb{E}_P[w^*] \leq 1$ for $w^* := r'_-(h_\lambda - \zeta_\lambda) \in \{u, 1\}$. Let $z^* \in \mathcal{Z}$ attain the supremum of the continuous h_λ on the compact $\mathcal{Z}$, set $s := 1 - \mathbb{E}_P[w^*] \geq 0$, and $Q^* := w^*P + s\delta_{z^*}^{\text{Dirac}}$, a distribution in $\mathcal{P}_\mathbb{U}(P)$ because $Q^* \geq w^*P \geq \underline{u}P$ and no cap is imposed. Then, using $h_\lambda(z^*) - \zeta_\lambda = \frac{1}{2}$ and the subdifferential identity for w^* ,

$$\mathbb{E}_{Q^*}[h_\lambda] - \zeta_\lambda = \mathbb{E}_P[w^*(h_\lambda - \zeta_\lambda)] + \frac{s}{2} = \mathbb{E}_P[r(h_\lambda - \zeta_\lambda, 1)] + \frac{1}{2}\mathbb{E}_P|w^* - 1| + \frac{s}{2} \geq \mathbb{E}_P[r(h_\lambda - \zeta_\lambda, 1)] + \text{TV}(Q^* \| P),$$
 the last inequality because $Q^* - P = (w^* - 1)P + s\delta_{z^*}^{\text{Dirac}}$ and the total variation norm is subadditive, so that $\text{TV}(Q^* \| P) \leq \frac{1}{2}\mathbb{E}_P|w^* - 1| + \frac{s}{2}$ by Lemma SA-2. Hence $\mathbb{E}_{Q^*}[h_\lambda] - \text{TV}(Q^* \| P) \geq F_\lambda(\zeta_\lambda)$, which proves (83). Combining (83) with the Fenchel display,

$$D(\theta; \rho, P) = \max_{\lambda \in \mathcal{R}_\lambda} \left\{ -\lambda_{12}^\top \rho - \min_{\zeta \text{ admissible}} F_\lambda(\zeta) \right\} = \sup_{(\zeta, \lambda) \in \mathcal{F}_\mathbb{D}} v_P(\zeta, \lambda),$$

which, together with the weak direction, proves the equality in (32).

Attainment. By the multiplier bound the supremum in (32) may be restricted to $\|\lambda\| \leq 1/r$, and by the intercept analysis the supremum over ζ at each such λ is attained on the compact interval $I_\lambda := [-\bar{M}/r - \frac{1}{2}, \bar{M}/r + \frac{1}{2}]$ intersected with the admissible set, which is nonempty. The map $(\zeta, \lambda) \mapsto v_P(\zeta, \lambda)$ is jointly continuous on $\{(\zeta, \lambda) \in \mathcal{F}_\mathbb{D} : \|\lambda\| \leq 1/r\}$ by the dominated convergence theorem and (r3), and the correspondence $\lambda \mapsto I_\lambda \cap \{\zeta : (\zeta, \lambda) \in \mathcal{F}_\mathbb{D}\}$ is compact-valued and continuous on the compact set $\{\lambda \in \mathcal{R}_\lambda : \|\lambda\| \leq 1/r\}$, since σ^∞ is continuous as the support function of the compact set $\{-m(z) : z \in \mathcal{Z}\}$. Berge's maximum theorem then makes $\lambda \mapsto \max_\zeta v_P(\zeta, \lambda)$ continuous on that compact set, so the supremum over λ is attained as well.

Part (ii). Finite side. If $D(\theta; \rho, P) \leq \delta$, then by the attainment of the primal projection there is $Q^\circ \in \mathcal{P}_\mathbb{U}(P)$ with $\mathbb{E}_{Q^\circ}[m] \in \mathcal{M}(\rho)$ and $\text{TV}(Q^\circ \| P) \leq \delta$, so the robustness set (22) is nonempty and $\bar{K}_\delta(\theta; \rho, P)$ is finite by Lemma SA-24(i).

Infinite side. Suppose the hypothesis of part (i) holds and $D(\theta; \rho, P) > \delta$. By part (i) there is $(\tilde{\zeta}, \tilde{\lambda}) \in \mathcal{F}_\mathbb{D}$ with $v_P(\tilde{\zeta}, \tilde{\lambda}) > \delta$. Fix any $\xi^0 \in \mathcal{F}(\theta)$, which exists as the set is nonempty, and consider the ray $\xi^0 + td$, $t \geq 0$, in the direction $d := (\tilde{\zeta}, 1, \tilde{\lambda})$. Each point of the ray is dual feasible: $\eta^0 + t \geq 0$, $\lambda^0 + t\tilde{\lambda} \in \mathcal{R}_\lambda$ because $\mathcal{R}_\lambda$ is a convex cone, and, when $\bar{u} = +\infty$, by subadditivity of the supremum together with the two feasibility constraints,

$$\sigma(\lambda^0 + t\tilde{\lambda}, \theta) \leq \left(\zeta^0 + \frac{\eta^0}{2} \right) + t \left(\tilde{\zeta} + \frac{1}{2} \right) = (\zeta^0 + t\tilde{\zeta}) + \frac{\eta^0 + t}{2}.$$

Along the ray, by positive homogeneity (r2) and the dominated convergence theorem, exactly as in the computation of the recession function in the proof of Lemma SA-24(iii),

$$\lim_{t \rightarrow \infty} \frac{J(\xi^0 + td, \theta; \rho, P)}{t} = \tilde{\zeta} + \delta + \tilde{\lambda}_{12}^\top \rho + \mathbb{E}_P[r(-\tilde{\zeta} - \tilde{\lambda}^\top m(Z), 1)] = \delta - v_P(\tilde{\zeta}, \tilde{\lambda}) < 0.$$

Hence $J(\xi^0 + td, \theta; \rho, P) \rightarrow -\infty$ and $\bar{K}_\delta(\theta; \rho, P) = -\infty$.

Part (iii). Write $\bar{M}_\Theta := \sup_{(z, \theta') \in \mathcal{Z} \times \Theta} \|m(z; \theta')\|$ and $d_m(\theta_1, \theta_2) := \sup_{z \in \mathcal{Z}} \|m(z; \theta_1) - m(z; \theta_2)\|$, which tends to zero as $\theta_2 \rightarrow \theta_1$ by the uniform continuity of m on the compact $\mathcal{Z} \times \Theta$.

Stability of the attainable set. For a unit vector $u \in \mathbb{R}^{d_m}$, $\theta' \in \Theta$, and $P' \in \Delta(\mathcal{Z})$, let $h_\Gamma(u, \theta'; P') := \sup\{u^\top w : w \in \Gamma_\mathbb{U}(\theta'; P', \delta)\}$ be the support function of the attainable set, a compact convex set by the attainment step. This is the upper bound over the family with integrand $u^\top m(\cdot; \theta')$ and no moment restrictions, so by Lemma SA-24(ii), whose constraint qualification is vacuous in the absence of moment restrictions,

$$h_\Gamma(u, \theta'; P') = \inf\{\zeta + \eta\delta + P'r(u^\top m(\cdot; \theta') - \zeta, \eta) : (\zeta, \eta) \in \mathcal{F}_0(u, \theta')\},$$

where $\mathcal{F}_0(u, \theta') := \{(\zeta, \eta) \in \mathbb{R} \times \mathbb{R}_+ : \zeta + \eta/2 \geq \sup_z u^\top m(z; \theta')\}$ when $\bar{u} = +\infty$ and $\mathcal{F}_0(u, \theta') := \mathbb{R} \times \mathbb{R}_+$ otherwise. The near-minimizers of this program lie in a fixed compact set. Indeed, the point $(\bar{M}_\Theta, 0)$ is feasible with objective at most $\bar{M}_\Theta$, as the integrand is $\underline{u}(u^\top m - \bar{M}_\Theta) \leq 0$ there by (23), so $h_\Gamma \leq \bar{M}_\Theta$. On the other hand, for any feasible (ζ, η) , Lemma SA-24(i) with $Q = P'$ gives an objective at least $-\bar{M}_\Theta + \eta\delta$, and (r4) gives an objective at least $(1 - \underline{u})\zeta - \underline{u}\bar{M}_\Theta - \frac{1-\underline{u}}{2}\eta$ and, when $\bar{u} < \infty$, at least $(1 - \bar{u})\zeta - \bar{u}\bar{M}_\Theta - \frac{\bar{u}-1}{2}\eta$, whereas when $\bar{u} = +\infty$ feasibility gives $\zeta \geq -\bar{M}_\Theta - \eta/2$. Hence every feasible (ζ, η) whose objective is at most $h_\Gamma + 1$ satisfies $0 \leq \eta \leq C_\eta := (2\bar{M}_\Theta + 1)/\delta$ and $|\zeta| \leq C_\zeta$ for a constant C_ζ depending only on $\bar{M}_\Theta$, δ , and $\mathbb{U}$. Let $\mathcal{A}_\Gamma := \{(\zeta, \eta, \lambda) : |\zeta| \leq C_\zeta, 0 \leq \eta \leq C_\eta, \|\lambda\| \leq 1\}$, a compact set, and note that the integrand $r(u^\top m(\cdot; \theta') - \zeta, \eta)$ of a feasible (ζ, η) with $(\zeta, \eta, -u) \in \mathcal{A}_\Gamma$ is the member of $\mathcal{H}_{\mathcal{A}_\Gamma}^\infty$ with index $(\zeta, \eta, -u)$, as, when $\bar{u} = +\infty$, feasibility is exactly the requirement $\zeta + \eta/2 \geq \sigma^\infty(-u, \theta')$ of (31), while this requirement is absent when $\bar{u} < \infty$. Now let (θ', P') and (θ'', P'') be given, fix $\varepsilon > 0$, and pick $(\zeta, \eta) \in \mathcal{F}_0(u, \theta')$ ε -optimal at (θ', P') with $\varepsilon \leq 1$, so that $(\zeta, \eta, -u) \in \mathcal{A}_\Gamma$. When $\bar{u} = +\infty$ the shift $s := [\sup_z u^\top m(z; \theta'') - \sup_z u^\top m(z; \theta')]_+ \leq d_m(\theta', \theta'')$ makes $(\zeta + s, \eta)$ feasible at θ'' , and we set $s := 0$ when $\bar{u} < \infty$. Then, by (r3),

$$h_\Gamma(u, \theta''; P'') \leq \zeta + s + \eta\delta + P''r(u^\top m(\cdot; \theta'') - \zeta - s, \eta)$$

$$\begin{aligned} &\leq \zeta + \eta\delta + P'r(u^\top m(\cdot; \theta') - \zeta, \eta) + \|P' - P''\|_{\mathcal{H}_{\mathcal{A}_\Gamma}^\infty} + s + u^\dagger(d_m(\theta', \theta'') + s) \\ &\leq h_\Gamma(u, \theta'; P') + \varepsilon + \|P' - P''\|_{\mathcal{H}_{\mathcal{A}_\Gamma}^\infty} + (1 + 2u^\dagger)d_m(\theta', \theta''), \end{aligned}$$

where the middle inequality splits the P'' -integral into the P' -integral of the same function, whose index lies in $\mathcal{A}_\Gamma$, plus the difference in baselines, plus the P'' -integral of the pointwise change in the integrand, which the Lipschitz property bounds. When $\bar{u} = +\infty$ both arguments of r lie in its domain by feasibility. Letting $\varepsilon \rightarrow 0$ and exchanging the roles of the two pairs,

$$\sup_{\|u\|=1} |h_\Gamma(u, \theta'; P') - h_\Gamma(u, \theta''; P'')| \leq (1 + 2u^\dagger)d_m(\theta', \theta'') + \|P' - P''\|_{\mathcal{H}_{\mathcal{A}_\Gamma}^\infty}.$$

Persistence. Let $w_0 \in \mathcal{M}(\rho)$ with $\mathcal{B}_r(w_0) \subseteq \Gamma_\mathbb{U}(\theta; P, \delta)$, and let (θ', ρ', P') satisfy $(1 + 2u^\dagger)d_m(\theta, \theta') + \|P - P'\|_{\mathcal{H}_{\mathcal{A}_\Gamma}^\infty} < r/3$ and $\|\rho' - \rho\| < r/3$. Writing h' for the support function of $\Gamma' := \Gamma_\mathbb{U}(\theta'; P', \delta)$ and h for that of $\Gamma := \Gamma_\mathbb{U}(\theta; P, \delta)$, the inclusion $\mathcal{B}_r(w_0) \subseteq \Gamma$ reads $h(u) \geq u^\top w_0 + r$ for every unit u , so the display gives $h'(u) \geq u^\top w_0 + 2r/3$ for every unit u , that is, $\mathcal{B}_{2r/3}(w_0) \subseteq \Gamma'$, as the set Γ' is closed and convex. Let $w'_0 := w_0 + \bar{\rho}' - \bar{\rho}$, where $\bar{\rho}'$ is the padded target of ρ' . Then $w'_0 \in \mathcal{M}(\rho')$, the shift acting on the first two blocks only and preserving the inequalities and equalities defining $\mathcal{M}(\rho)$, and $\|w'_0 - w_0\| = \|\rho' - \rho\| < r/3$, so that $\mathcal{B}_{r/3}(w'_0) \subseteq \mathcal{B}_{2r/3}(w_0) \subseteq \Gamma'$. This is the constraint qualification at (θ', ρ', P') with radius $r/3$. Choosing $\varepsilon > 0$ such that $\|\theta' - \theta\| < \varepsilon$ implies $(1 + 2u^\dagger)d_m(\theta, \theta') < r/6$, which is possible by uniform continuity, and $\varepsilon \leq r/6$, the persistence claim holds on the neighborhood of the statement for any compact $\mathcal{A}_\mathbb{D} \supseteq \mathcal{A}_\Gamma$, since $\|\cdot\|_{\mathcal{H}_{\mathcal{A}_\Gamma}^\infty} \leq \|\cdot\|_{\mathcal{H}_{\mathcal{A}_\mathbb{D}}^\infty}$.

Lipschitz continuity. Let (θ_1, ρ_1, P_1) and (θ_2, ρ_2, P_2) lie in the neighborhood just constructed, so that the constraint qualification holds at both with radius $r/3$, and write $\mathbf{D}_i := \mathbf{D}(\theta_i; \rho_i, P_i)$, $d_m := d_m(\theta_1, \theta_2)$, and $\sigma_i^\infty(\lambda) := \sigma^\infty(\lambda, \theta_i)$. By part (i) applied at (θ_1, ρ_1, P_1) with radius $r/3$, there is a maximizer (ζ_1, λ_1) of v_{P_1} with $\|\lambda_1\| \leq 3/r$ and $|\zeta_1| \leq \frac{1}{2} + 3\bar{M}_\Theta/r$. When $\bar{u} = +\infty$ the shift $s := [\sigma_2^\infty(\lambda_1) - \sigma_1^\infty(\lambda_1)]_+ \leq \|\lambda_1\|d_m$ makes $(\zeta_1 + s, \lambda_1)$ admissible at θ_2 , and we set $s := 0$ when $\bar{u} < \infty$. Set $C_2 := \frac{1}{2} + 3\bar{M}_\Theta/r + 6\bar{M}_\Theta/r$ and let $\mathcal{A}_\mathbb{D}$ be a compact set containing $\mathcal{A}_\Gamma$ and $\{(\zeta, 1, \lambda) : |\zeta| \leq C_2, \|\lambda\| \leq 3/r\}$, so that the integrands $f_1 := r(-\zeta_1 - \lambda_1^\top m(\cdot; \theta_1), 1)$ and $r(-\zeta_1 - s - \lambda_1^\top m(\cdot; \theta_2), 1)$ belong to $\mathcal{H}_{\mathcal{A}_\mathbb{D}}^\infty$, their indices satisfying the recession-cone requirement of (31) by admissibility when $\bar{u} = +\infty$, while this requirement is absent when $\bar{u} < \infty$. Evaluating the θ_2 -problem at $(\zeta_1 + s, \lambda_1)$, which bounds $\mathbf{D}_2$ from below by the weak direction of part (i), and using that $r(\cdot, 1)$ is $u^\dagger$ -Lipschitz on its domain,

$$\begin{aligned} \mathbf{D}_2 &\geq -\mathbb{E}_{P_2}[r(-\zeta_1 - s - \lambda_1^\top m(Z; \theta_2), 1)] - \zeta_1 - s - \lambda_{1,12}^\top \rho_2 \\ &\geq -\mathbb{E}_{P_1}[f_1] - \zeta_1 - \lambda_{1,12}^\top \rho_1 - \|P_1 - P_2\|_{\mathcal{H}_{\mathcal{A}_\mathbb{D}}^\infty} - u^\dagger(s + \|\lambda_1\|d_m) - s - \|\lambda_1\|\|\rho_1 - \rho_2\| \\ &\geq \mathbf{D}_1 - C(\|\rho_1 - \rho_2\| + d_m + \|P_1 - P_2\|_{\mathcal{H}_{\mathcal{A}_\mathbb{D}}^\infty}), \quad C := \max\left\{1, \frac{3(2u^\dagger + 1)}{r}\right\}, \end{aligned}$$

where the middle inequality splits $|\mathbb{E}_{P_2}[f_2] - \mathbb{E}_{P_1}[f_1]| \leq \|P_1 - P_2\|_{\mathcal{H}_{\mathcal{A}_\mathbb{D}}^\infty} + \mathbb{E}_{P_2}[|f_2 - f_1|]$ with f_2 the first integrand displayed, both of whose arguments lie in the domain of $r(\cdot, 1)$, and the last inequality uses $v_{P_1}(\zeta_1, \lambda_1) = \mathbf{D}_1$, the strong direction at (θ_1, ρ_1, P_1) , together with $s \leq (3/r)d_m$ and $\|\lambda_1\| \leq 3/r$. Exchanging the roles of the two problems gives the reverse bound, whence the asserted Lipschitz inequality. The constants C , C_2 , and the set $\mathcal{A}_\mathbb{D}$ depend only on r , δ , $\mathbb{U}$, and $\bar{M}_\Theta$, as claimed. $\blacksquare$

SA8.22 Proof of Lemma SA-26

Proof. Envelope. Recall a from Section SA4.3, so that $\ell_{\xi, \theta} = r(a(\cdot; \theta, \zeta, \lambda), \eta)$. Under Assumption SA-4(i), g and m are continuous on the compact $\mathcal{Z} \times \Theta$, hence bounded, $|g| \leq \bar{G}$ and $\|m\| \leq \bar{M}$. Over the compact index $\mathcal{A}$ one has $\|\lambda\| \leq L_0$, $|\zeta| \leq C_0$, and $\eta \leq \eta_0$, so $|a| \leq \bar{G} + L_0\bar{M} + C_0 =: R_0$ and therefore, by (r3), $|\ell_{\xi, \theta}| \leq F_0 := u^\dagger R_0 + \frac{u^\dagger}{2} \eta_0 < \infty$ on the index set of $\mathcal{H}_{\mathcal{A}}^J$. As a constant, F_0 is a P_* -square-integrable envelope with $\|F_0\|_{Q,2} = F_0$ for every Q .

A larger class. By (23), on its domain $r(x, \eta)$ is the pointwise maximum of the affine functions $\underline{u}x - \frac{1-\underline{u}}{2}\eta$ and x , and, when $\bar{u} < \infty$, of $\bar{u}x - \frac{\bar{u}-1}{2}\eta$ as well. Let $\tilde{r} : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ denote this maximum of two or three affine functions on all of $\mathbb{R} \times \mathbb{R}_+$, which coincides with r on the domain of the latter and is finite everywhere, and let $\tilde{\mathcal{H}}_{\mathcal{A}}^J := \{\tilde{r}(a(\cdot; \theta, \zeta, \lambda), \eta) : \theta \in \Theta, \xi \in \mathcal{A}\}$. Then $\mathcal{H}_{\mathcal{A}}^J \subseteq \tilde{\mathcal{H}}_{\mathcal{A}}^J$, because the restriction to $\xi \in \mathcal{F}(\theta)$ places the argument of r in its domain at every z , and it suffices to show that $\tilde{\mathcal{H}}_{\mathcal{A}}^J$ is P_* -Donsker, as a subclass of a Donsker class is Donsker.

VC-subgraph in the multipliers. Fix θ . As (ζ, η, λ) ranges over $\mathcal{A}$, each affine branch $ua(\cdot; \theta, \zeta, \lambda) - \frac{|u-1|}{2}\eta$, for $u \in \{\underline{u}, 1, \bar{u}\}$, varies over the finite-dimensional vector space $V_\theta := \text{span}\{g(\cdot; \theta), m_1(\cdot; \theta), \dots, m_{d_m}(\cdot; \theta), 1\}$, of dimension at most $d_m + 2$, which is VC-subgraph (van der Vaart and Wellner, 1996, Lemma 2.6.15). Since the pairwise maximum of two VC-subgraph classes is again VC-subgraph (van der Vaart and Wellner, 1996, Lemma 2.6.18), applied once or twice, the class $\{\tilde{r}(a(\cdot; \theta, \zeta, \lambda), \eta) : \xi \in \mathcal{A}\}$ is VC-subgraph with index V bounded by a constant depending only on d_m , uniformly in θ . By Theorem 2.6.7 in van der Vaart and Wellner (1996) its uniform covering numbers obey, with envelope F_0 ,

$$\sup_Q N(\varepsilon \|F_0\|_{Q,2}, \{\tilde{r}(a(\cdot; \theta, \zeta, \lambda), \eta) : \xi \in \mathcal{A}\}, L_2(Q)) \leq C \varepsilon^{-2(V-1)}, \quad 0 < \varepsilon < 1,$$

with C not depending on θ .

Lipschitz control in θ . The θ -direction is handled by Assumption SA-4(i): since $\tilde{r}(\cdot, \eta)$ is $u^\dagger$ -Lipschitz,

$$\begin{aligned} & \|\tilde{r}(a(\cdot; \theta, \zeta, \lambda), \eta) - \tilde{r}(a(\cdot; \theta', \zeta, \lambda), \eta)\|_\infty \\ & \leq u^\dagger \left(\sup_z |g(z; \theta) - g(z; \theta')| + \|\lambda\| \sup_z \|m(z; \theta) - m(z; \theta')\| \right) \leq u^\dagger (L + L_0 L) \|\theta - \theta'\| \end{aligned}$$

uniformly over $\xi \in \mathcal{A}$, where we use $\|\lambda\| \leq L_0$ and the uniform Lipschitz-in- θ moduli of SA-4(i). Covering the compact Θ by $N(\varepsilon, \Theta, \|\cdot\|) \leq C'(\text{diam } \Theta/\varepsilon)^{d_\theta}$ balls and pairing the VC bound at each center with this modulus, which is legitimate because the envelope is the single constant F_0 , yields, for the full class,

$$\sup_Q N(\varepsilon \|F_0\|_{Q,2}, \tilde{\mathcal{H}}_{\mathcal{A}}^J, L_2(Q)) \leq C'' \varepsilon^{-(d_\theta+2(V-1))}, \quad 0 < \varepsilon < 1,$$

so the uniform entropy integral $\int_0^1 \sup_Q \sqrt{\log N(\varepsilon \|F_0\|_{Q,2}, \tilde{\mathcal{H}}_{\mathcal{A}}^J, L_2(Q))} d\varepsilon$ is finite.

Conclusion. Separability of $\Theta \times \mathcal{A}$ and continuity of $(\theta, \xi) \mapsto \tilde{r}(a(z; \theta, \zeta, \lambda), \eta)$ make the class pointwise measurable. A pointwise-measurable class with finite uniform entropy integral and a P_* -square-integrable envelope is P_* -Donsker (van der Vaart and Wellner, 1996, Theorem 2.5.2), so $\tilde{\mathcal{H}}_{\mathcal{A}}^J$, and hence $\mathcal{H}_{\mathcal{A}}^J$, is P_* -Donsker. Let $\mathcal{A}' := \mathcal{A} \cup \{(\zeta, 1, \lambda) : (\zeta, \lambda) \in \Pi_{\zeta\lambda}(\mathcal{A})\}$, which is compact as the union of $\mathcal{A}$ and a continuous image of the compact set $\Pi_{\zeta\lambda}(\mathcal{A})$. The identical argument with $g \equiv 0$ and $\mathcal{A}'$ in place of $\mathcal{A}$ shows that the class $\{\tilde{r}(-\lambda^\top m(\cdot; \theta) - \zeta, \eta) : \theta \in \Theta, \xi \in \mathcal{A}'\}$ is P_* -Donsker, and both $\mathcal{H}_{\mathcal{A}}^D$ and $\mathcal{H}_{\mathcal{A}}^J$ are subclasses of it, as the restrictions on their indices place the argument of r in its domain, where $\tilde{r}$ and r coincide. Finally, a P_* -Donsker class with bounded envelope is P_* -Glivenko–Cantelli. ■

SA8.23 Details on the Construction of Ξ and $\mathcal{H}$

This section verifies the properties claimed in Section SA4.3 for the radii ε_θ and for the sets $\mathcal{N}_0$ and Ξ constructed there.

The radii ε_θ . Fix $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$, and let θ_j be a center of the subcover with $\|\theta - \theta_j\| < \varepsilon_{\theta_j}^D/2$. On the set of triples (θ', ρ', P') with $\|\theta' - \theta\| + \|(\rho', P') - (\rho_0, P_*)\|_{\mathcal{H}_0} < \varepsilon_{\theta_j}^D/2$, the constraint qualification holds with radius $r_{\theta_j}/3$ and $D(\theta'; \rho', P')$ is within C_{θ_j} times the displayed distance of $D(\theta; \rho_0, P_*) < \delta$ (Assumption SA-4(iii)), so that, after shrinking the radius to some $\varepsilon_\theta > 0$, the strict feasibility $D(\theta'; \rho', P') < \delta$ persists with a uniform margin. On such a neighborhood Lemma SA-24(iii) makes every set $\Xi_\delta(\theta'; \rho', P')$ nonempty and compact, and the proof of part (ii) of Lemma SA-27 below, which uses only the slack inequality of Lemma SA-24(i), the persisting constraint qualification, and the uniform feasibility margin, bounds every minimizer realized on the neighborhood explicitly, so that their union is bounded. This delivers the radius ε_θ of Section SA4.3, and the set Ξ defined there is bounded as a finite union of bounded sets.

The set $\mathcal{N}_0$. Since $\|(\hat{\rho}_n, \hat{P}_n) - (\rho_0, P_*)\|_{\mathcal{H}_0} \rightarrow 0$ in probability by Assumption SA-4(iv) and the Glivenko–Cantelli property of Lemma SA-26, with probability approaching one this distance is smaller than $\min_k \varepsilon_k$. On that event, the triple $(\theta', \hat{\rho}_n, \hat{P}_n)$ lies in $\mathcal{N}_0$ for every θ' with $\|\theta' - \theta_k\|$ smaller than ε_k minus this distance for some k , and every dual optimizer realized at such a θ' has its index in Ξ .

The class $\mathcal{H}$. With $\mathcal{A}$ and $\mathcal{H}$ as fixed in Section SA4.3, $\mathcal{A}_0 \subseteq \mathcal{A}$ gives $\mathcal{H}_0 \subseteq \mathcal{H}$, so that $\|\cdot\|_{\mathcal{H}_0} \leq \|\cdot\|_{\mathcal{H}}$ and hence $\|\cdot\|_{\mathcal{H}_0} \leq \|\cdot\|_{\mathcal{H}}$. The bound obtained above for $\|\cdot\|_{\mathcal{H}_0}$ -neighborhoods therefore holds a fortiori on the smaller $\|\cdot\|_{\mathcal{H}}$ -neighborhoods, and the set Ξ remains valid when the baseline is measured in $\|\cdot\|_{\mathcal{H}}$.

SA8.24 Proof of Lemma SA-27

Proof. Throughout, $\bar{G} := \sup_{(z, \theta') \in \mathcal{Z} \times \Theta} |g(z; \theta')|$ and $\bar{M} := \sup_{(z, \theta') \in \mathcal{Z} \times \Theta} \|m(z; \theta')\|$, both finite by Assumption SA-4(i), and we use repeatedly that $|P\ell - P'\ell| \leq \|P - P'\|_{\mathcal{H}}$ for every $\ell \in \mathcal{H}$ and all $P, P' \in \Delta(\mathcal{Z})$.

Part (i). Fix any $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$. By Assumption SA-4(iii),

$$D(\theta; \rho_0, P_*) < \delta. \tag{84}$$

Let θ_j be a center of the finite subcover fixed in Section SA4.3 with $\|\theta - \theta_j\| < \varepsilon_{\theta_j}^D/2$. Lemma SA-25(iii) at (θ_j, ρ_0, P_*) , where the constraint qualification holds by Assumption SA-4(ii), gives

$$|D(\theta; \rho, P) - D(\theta; \rho_0, P_*)| \leq C_{\theta_j} (\|\rho - \rho_0\| + \|P - P_*\|_{\mathcal{H}_{\mathcal{A}_0, \theta_j}^\infty}) \leq C_{\theta_j} \|(\rho, P) - (\rho_0, P_*)\|_{\mathcal{H}}$$

whenever $\|\rho - \rho_0\| + \|P - P_*\|_{\mathcal{H}_{\mathcal{A}_0, \theta_j}^\infty} < \varepsilon_{\theta_j}^D/2$, where the second inequality follows from $\mathcal{H}_{\mathcal{A}_0, \theta_j}^\infty \subseteq \mathcal{H}_0 \subseteq \mathcal{H}$. Set $\bar{C} := \max_{1 \leq j \leq J} C_{\theta_j}$ and $\bar{\varepsilon} := \min_{1 \leq j \leq J} \varepsilon_{\theta_j}^D/2$, so that the display above holds with $\bar{C}$ in place of C_{θ_j} at every $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$ whenever $\|(\rho, P) - (\rho_0, P_*)\|_{\mathcal{H}} < \bar{\varepsilon}$. The same Lipschitz estimate, read in its θ -argument together with the Lipschitz continuity of m in θ (Assumption SA-4(i)), shows that $\theta \mapsto D(\theta; \rho_0, P_*)$ is continuous on the compact set $\bar{\Theta}_\delta(\rho_0, P_*)$, so that, by (84), $\bar{D} := \sup_{\theta \in \bar{\Theta}_\delta(\rho_0, P_*)} D(\theta; \rho_0, P_*) < \delta$. With $r := \min\{\bar{\varepsilon}, (\delta - \bar{D})/(2\bar{C})\}$ we obtain $D(\theta; \rho, P) < \bar{D} + \bar{C}r \leq \frac{1}{2}(\delta + \bar{D}) < \delta$, that is, $\theta \in \Theta_\delta(\rho, P)$, for every $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$ and every (ρ, P) in the $\|\cdot\|_{\mathcal{H}}$ -ball $\mathcal{B}_r(\rho_0, P_*)$, which eventually contains any sequence $(\rho_n, P_n) \rightarrow (\rho_0, P_*)$. This proves part (i).

Part (ii). Fix any $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$ and let θ_j be as in part (i), with constraint-qualification radius $r_{\text{cq}} := r_{\theta_j}$. By Lemma SA-25(iii) at (θ_j, ρ_0, P_*) , together with (84) and the Lipschitz estimate of part (i), there is $\varepsilon > 0$ such that, on the ball $\mathcal{B}_\varepsilon(\theta, \rho_0, P_*)$ of the statement, the constraint qualification holds with radius $r_{\text{cq}}/3$ and $\text{D}(\theta'; \rho', P') \leq \delta - c$, where $c := \frac{1}{2}(\delta - \text{D}(\theta; \rho_0, P_*)) > 0$. By Lemma SA-24(iii), the set $\bar{\Xi}_\delta(\theta'; \rho', P')$ is then nonempty, compact, and convex at every $(\theta', \rho', P') \in \mathcal{B}_\varepsilon(\theta, \rho_0, P_*)$, and it remains to bound these sets uniformly, which we do through the slack inequality of Lemma SA-24(i). Fix $(\theta', \rho', P') \in \mathcal{B}_\varepsilon(\theta, \rho_0, P_*)$ and $\xi^* = (\zeta^*, \eta^*, \lambda^*) \in \bar{\Xi}_\delta(\theta'; \rho', P')$. Since the point $(\sup_z g(z; \theta'), 0, 0)$ is feasible with objective at most $\bar{G}$, $J(\xi^*, \theta'; \rho', P') = \bar{K}_\delta(\theta'; \rho', P') \leq \bar{G}$. By the attainment part of Lemma SA-25, there is $Q^\circ \in \mathcal{P}_\mathbb{U}(P')$ with $\mathbb{E}_{Q^\circ}[m(Z; \theta')] \in \mathcal{M}(\rho')$ and $\text{TV}(Q^\circ \| P') = \text{D}(\theta'; \rho', P') \leq \delta - c$, so Lemma SA-24(i) gives $\bar{G} \geq J(\xi^*, \theta'; \rho', P') \geq -\bar{G} + \eta^* c$, that is, $\eta^* \leq 2\bar{G}/c$. Next, let $w'_0 \in \mathcal{M}(\rho')$ be the point of the constraint qualification at (θ', ρ', P') , so that $\mathcal{B}_{r_{\text{cq}}/3}(w'_0) \subseteq \Gamma_\mathbb{U}(\theta'; P', \delta)$. If $\lambda^* \neq 0$, the point $w := w'_0 - \frac{r_{\text{cq}}}{6} \lambda^* / \|\lambda^*\|$ lies in that ball and is attained by some $Q_w \in \mathcal{P}_\mathbb{U}(P')$ with $\text{TV}(Q_w \| P') \leq \delta$, and Lemma SA-24(i) gives

$$\bar{G} \geq J(\xi^*, \theta'; \rho', P') \geq -\bar{G} + \eta^*(\delta - \text{TV}(Q_w \| P')) + (\lambda_{12}^{*\top} \rho' - \lambda^{*\top} w'_0) + \frac{r_{\text{cq}}}{6} \|\lambda^*\| \geq -\bar{G} + \frac{r_{\text{cq}}}{6} \|\lambda^*\|,$$

using $\lambda_{12}^{*\top} \rho' \geq \lambda^{*\top} w'_0$, whence $\|\lambda^*\| \leq 12\bar{G}/r_{\text{cq}}$. Finally, (r4) and the bounds just obtained give

$$\bar{G} \geq J(\xi^*, \theta'; \rho', P') \geq (1 - \underline{u})\zeta^* - \|\lambda^*\| \|\rho'\| - \underline{u}(\bar{G} + \|\lambda^*\| \bar{M}) - \frac{1 - \underline{u}}{2} \eta^*,$$

which bounds ζ^* from above by a constant depending only on $\bar{G}, \bar{M}, r_{\text{cq}}, c, \mathbb{U}$, and $\sup\{\|\rho'\| : \|\rho' - \rho_0\| < \varepsilon\}$. From below, ζ^* is bounded by $\sigma(\lambda^*, \theta') - \eta^*/2 \geq -\bar{G} - \|\lambda^*\| \bar{M} - \eta^*/2$ when $\bar{u} = +\infty$, by dual feasibility, and by the third inequality of (r4), which gives $\bar{G} \geq (1 - \bar{u})\zeta^* - \|\lambda^*\| \|\rho'\| - \bar{u}(\bar{G} + \|\lambda^*\| \bar{M}) - \frac{\bar{u} - 1}{2} \eta^*$, when $\bar{u} < \infty$. Hence every minimizer realized on $\mathcal{B}_\varepsilon(\theta, \rho_0, P_*)$ lies in a single bounded set, which proves part (ii).

Properties of the dual. Before turning to parts (iii)–(vi), we record four properties of the dual program that are used repeatedly below. First. The objective $J(\xi, \theta'; \rho', P')$ is jointly continuous in $(\xi, \theta', \rho', P')$ along every sequence $(\xi_n, \theta_n, \rho_n, P_n) \rightarrow (\xi, \theta', \rho', P')$ with $\xi_n \in \mathcal{F}(\theta_n)$ for every n , $\xi \in \mathcal{F}(\theta')$, and $\|P_n - P'\|_{\mathcal{H}_\mathcal{K}} \rightarrow 0$ for some compact $\mathcal{K}$ containing $\{\xi_n\}_{n \geq 1}$. Indeed, J is affine in ρ' and satisfies

$$|P_n \ell_{\xi_n, \theta_n} - P' \ell_{\xi, \theta'}| \leq \|P_n - P'\|_{\mathcal{H}_\mathcal{K}} + |P' \ell_{\xi_n, \theta_n} - P' \ell_{\xi, \theta'}| \rightarrow 0,$$

where $\mathcal{K}$ is the compact set of the hypothesis, the first term vanishes by hypothesis, and the second by the dominated convergence theorem, because the integrand is continuous in (ξ, θ) at every z on the domain of r , by continuity of $g(\cdot; \theta)$ and $m(\cdot; \theta)$ on $\mathcal{Z} \times \Theta$ (Assumption SA-4(i)), and bounded by (r3). Second. The correspondence $\theta' \mapsto \mathcal{F}(\theta')$ has closed graph and is lower hemicontinuous. It is constant when $\bar{u} < \infty$. When $\bar{u} = +\infty$, the right-hand side of the constraint $\zeta + \eta/2 \geq \sigma(\lambda, \theta')$ is jointly continuous in (θ', λ) by Berge's maximum theorem on the compact $\mathcal{Z}$, which gives the closed graph. For lower hemicontinuity at θ , take a strictly feasible point ξ° of $\mathcal{F}(\theta)$, that is, one with $\zeta^\circ + \eta^\circ/2 > \sigma(\lambda^\circ, \theta)$, which exists because $\sigma(\cdot, \theta)$ is finite. For $\xi \in \mathcal{F}(\theta)$ and $t \in (0, 1]$, convexity of the constraint function in ξ makes $(1 - t)\xi + t\xi^\circ$ strictly feasible at θ , hence feasible at every θ' near θ by joint continuity, and letting $t \rightarrow 0$ along a diagonal produces the required approximating sequence. Third. As a consequence, the dual value $(\theta', \rho', P') \mapsto \bar{K}_\delta(\theta'; \rho', P')$ is jointly upper semicontinuous along every sequence $(\theta_n, \rho_n, P_n) \rightarrow (\theta, \rho, P)$ with $\|P_n - P\|_{\mathcal{H}_\mathcal{K}} \rightarrow 0$ for every compact $\mathcal{K}$: given $\xi \in \mathcal{F}(\theta)$, lower hemicontinuity supplies $\xi_n \in \mathcal{F}(\theta_n)$ with $\xi_n \rightarrow \xi$, whence $\limsup_n \bar{K}_\delta(\theta_n; \rho_n, P_n) \leq \lim_n J(\xi_n, \theta_n; \rho_n, P_n) = J(\xi, \theta; \rho, P)$ by the first property, and taking the infimum over $\xi \in \mathcal{F}(\theta)$ gives $\limsup_n \bar{K}_\delta(\theta_n; \rho_n, P_n) \leq \bar{K}_\delta(\theta; \rho, P)$. In particular, $\bar{K}_\delta(\cdot; \rho, P)$ is upper semicontinuous on Θ for every fixed (ρ, P) . Finally, at every (θ, ρ_0, P_*) with $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$ the dual value is jointly continuous along sequences $(\theta_n, \rho_n, P_n) \rightarrow (\theta, \rho_0, P_*)$ converging in $\|\cdot\|_{\mathcal{H}}$. Indeed, writing

$$\bar{K}_\delta(\theta_n; \rho_n, P_n) - \bar{K}_\delta(\theta; \rho_0, P_*) = \{\bar{K}_\delta(\theta_n; \rho_n, P_n) - \bar{K}_\delta(\theta_n; \rho_0, P_*)\} + \{\bar{K}_\delta(\theta_n; \rho_0, P_*) - \bar{K}_\delta(\theta; \rho_0, P_*)\},$$

the first term is bounded by $M\|(\rho_n, P_n) - (\rho_0, P_*)\|_{\mathcal{H}} \rightarrow 0$ by the Lipschitz estimate of part (v) below, whose proof rests on part (ii) and (L) alone, and the second term vanishes because $\theta' \mapsto \bar{K}_\delta(\theta'; \rho_0, P_*)$ is continuous at θ . The latter follows from Berge's maximum theorem: by part (ii) the optimizer sets $\bar{\Xi}_\delta(\theta'; \rho_0, P_*)$ for θ' near θ are nonempty and contained in a compact ball $\bar{\mathcal{B}}^*$, which we enlarge to contain the strictly feasible point ξ° , so that the infimum over $\mathcal{F}(\theta')$ coincides with the minimum over the compact set $\mathcal{F}(\theta') \cap \bar{\mathcal{B}}^*$. The correspondence $\theta' \mapsto \mathcal{F}(\theta') \cap \bar{\mathcal{B}}^*$ is compact-valued and continuous. Upper hemicontinuity follows from the closed graph, and lower hemicontinuity from the argument just given, as the points $(1 - t)\xi + t\xi^\circ$ lie in the convex set $\bar{\mathcal{B}}^*$. At the fixed baseline P_* , the objective is jointly continuous in (ξ, θ') by the first property. Fourth. For every $h \in \mathbb{R}^{d_\theta}$ and every $\gamma \in UC(\mathcal{H}, e_{P_*})$, the map $(\theta', \xi) \mapsto \lambda_{12}^\top h + \gamma \ell_{\xi, \theta'}$ is continuous, and hence uniformly continuous, on the set of pairs $(\theta', \xi) \in \Theta \times \mathcal{A}$ with $\xi \in \mathcal{F}(\theta')$, on which $\ell_{\xi, \theta'} \in \mathcal{H}$. Indeed, this set is compact by the closed graph of $\mathcal{F}(\cdot)$ and the compactness of Θ and $\mathcal{A}$, the linear part of the map is continuous in ξ , and $e_{P_*}(\ell_{\xi_n, \theta_n}, \ell_{\xi, \theta'}) \rightarrow 0$ along every sequence $(\theta_n, \xi_n) \rightarrow (\theta', \xi)$ in this set, by the dominated convergence theorem applied as in the first property, while γ is uniformly continuous with respect to e_{P_*} .

Part (iii). Let $(\theta_n, \rho_n, P_n) \rightarrow (\theta, \rho_0, P_*)$ with $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$. By part (ii) and the construction of Section SA4.3, for n large the sets $\bar{\Xi}_\delta(\theta_n; \rho_n, P_n)$ are nonempty and contained in $\Xi \subseteq \mathcal{A}$, so any selection $\xi_n \in \bar{\Xi}_\delta(\theta_n; \rho_n, P_n)$ admits a subsequence $\xi_{n_l} \rightarrow \xi^\infty$, with $\xi^\infty \in \mathcal{F}(\theta)$ by the closed graph of $\mathcal{F}(\cdot)$. Since $\xi_{n_l} \in \mathcal{A}$, the first property above gives $J(\xi_{n_l}, \theta_{n_l}; \rho_{n_l}, P_{n_l}) \rightarrow J(\xi^\infty, \theta; \rho_0, P_*)$, while $J(\xi_{n_l}, \theta_{n_l}; \rho_{n_l}, P_{n_l}) = \bar{K}_\delta(\theta_{n_l}; \rho_{n_l}, P_{n_l}) \rightarrow \bar{K}_\delta(\theta; \rho_0, P_*)$ by the joint continuity of the dual value at (θ, ρ_0, P_*) . Hence $J(\xi^\infty, \theta; \rho_0, P_*)$ equals $\bar{K}_\delta(\theta; \rho_0, P_*)$, that is, $\xi^\infty \in \bar{\Xi}_\delta(\theta; \rho_0, P_*)$, and thus $\lambda_{12, n_l} \rightarrow \lambda_{12}^\infty \in \bar{\Lambda}_\delta(\theta; \rho_0, P_*)$. This proves part (iii).

Part (iv). We apply Berge's maximum theorem to $v(\theta; h, \gamma) = \min_{\xi \in \Xi_\delta(\theta; \rho_0, P_*)} \{\lambda_{12}^\top h + \gamma \ell_{\xi, \theta}\}$ at any $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$. By part (v), $\mathcal{N}$ is a neighborhood of θ and, for every $\theta' \in \mathcal{N} \cap \Theta$, the set $\Xi_\delta(\theta'; \rho_0, P_*)$ is contained in $\Xi \subseteq \mathcal{A}$, so that $\ell_{\xi, \theta'} \in \mathcal{H}$ and $\gamma \ell_{\xi, \theta'}$ is defined for every ξ in it. The inner objective is continuous in (θ, ξ) by the fourth property above. The correspondence $\theta' \mapsto \Xi_\delta(\theta'; \rho_0, P_*)$ is nonempty- and compact-valued for θ' near θ by part (ii), lower hemicontinuous at θ by Assumption SA-4(v), and upper hemicontinuous at θ by part (iii) applied with $(\rho_n, P_n) \equiv (\rho_0, P_*)$. Berge's theorem therefore gives continuity of $v(\cdot; h, \gamma)$ at θ , which proves part (iv).

Part (v). We first construct a compact neighborhood of $\bar{\Theta}_\delta(\rho_0, P_*)$ on which the optimizer sets are uniformly controlled. For each $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$, let $\varepsilon_\theta > 0$ be the radius supplied by part (ii). The balls $\mathcal{B}_{\varepsilon_\theta/2}(\theta)$ cover the compact set $\bar{\Theta}_\delta(\rho_0, P_*)$, so we can extract a finite subcover with centers $\theta_1, \dots, \theta_K$ and set

$$\mathcal{N} := \bigcup_{k=1}^K \bar{\mathcal{B}}_{\varepsilon_{\theta_k}/2}(\theta_k), \quad r := \min_{1 \leq k \leq K} \frac{\varepsilon_{\theta_k}}{2} > 0.$$

Then $\mathcal{N}$ is compact and contains $\bar{\Theta}_\delta(\rho_0, P_*)$ in its interior. Any $\theta \in \mathcal{N} \cap \Theta$ lies within $\varepsilon_{\theta_k}/2$ of some center θ_k , so that, for any (ρ, P) with $\|(\rho, P) - (\rho_0, P_*)\|_{\mathcal{H}} < r$,

$$\|\theta - \theta_k\| + \|(\rho, P) - (\rho_0, P_*)\|_{\mathcal{H}} < \frac{\varepsilon_{\theta_k}}{2} + r \leq \varepsilon_{\theta_k},$$

and part (ii) applies at the center θ_k . Hence $\Xi_\delta(\theta; \rho, P)$ is nonempty and compact, and these sets are bounded uniformly over all such (θ, ρ, P) . Since $\|\cdot\|_{\mathcal{H}_0} \leq \|\cdot\|_{\mathcal{H}}$, we may shrink $\mathcal{N}$ and r so that every such triple lies in the cover $\mathcal{N}_0$ from which the set Ξ was formed, whence $\Xi_\delta(\theta; \rho, P) \subseteq \Xi \subseteq \mathcal{A}$ throughout. Let $M \in [1, \infty)$ bound $\|\lambda_{12}\|$ over $\Pi_{12}(\mathcal{A})$. This establishes every claim of the second assertion of part (v) except the Lipschitz estimate, which we prove next.

Lipschitz condition. Fix any $\theta \in \mathcal{N} \cap \Theta$ and two pairs $(\rho, P), (\rho', P')$ in the $\|\cdot\|_{\mathcal{H}}$ -ball $\mathcal{B}_r(\rho_0, P_*)$. By (L), $(\rho, P) \mapsto \bar{K}_\delta(\theta; \rho, P)$ is concave, as a pointwise infimum of the jointly affine maps $(\rho, P) \mapsto J(\xi, \theta; \rho, P) = \zeta + \eta\delta + \lambda_{12}^\top \rho + P\ell_{\xi, \theta}$, and at any optimizer $\xi \in \Xi_\delta(\theta; \rho, P) \subseteq \mathcal{A}$ it admits the supergradient $(\lambda_{12}, \ell_{\xi, \theta})$. Concavity then gives

$$\lambda_{12}^\top(\rho' - \rho) + (P' - P)\ell_{\xi, \theta} \geq \bar{K}_\delta(\theta; \rho', P') - \bar{K}_\delta(\theta; \rho, P) \geq \lambda_{12}^\top(\rho' - \rho) + (P' - P)\ell_{\xi', \theta},$$

with $\xi' \in \Xi_\delta(\theta; \rho', P') \subseteq \mathcal{A}$. Since $\|\lambda_{12}\|, \|\lambda_{12}'\| \leq M$ and $\ell_{\xi, \theta}, \ell_{\xi', \theta} \in \mathcal{H}$, each side is bounded in absolute value by $M\|\rho - \rho'\| + \|P - P'\|_{\mathcal{H}} \leq M\|(\rho, P) - (\rho', P')\|_{\mathcal{H}}$, using $M \geq 1$. Hence $|\bar{K}_\delta(\theta; \rho, P) - \bar{K}_\delta(\theta; \rho', P')| \leq M\|(\rho, P) - (\rho', P')\|_{\mathcal{H}}$, uniformly over $\theta \in \mathcal{N} \cap \Theta$.

Continuity of the bound. We prove lower and upper semicontinuity separately. For lower semicontinuity, fix $\varepsilon' > 0$ and any $\theta^* \in \bar{\Theta}_\delta(\rho_0, P_*)$, which attains $\bar{b}(\delta, \rho_0, P_*)$. By part (i), $\theta^* \in \Theta_\delta(\rho, P)$ on a neighborhood of (ρ_0, P_*) , and by the Lipschitz condition just established $\bar{K}_\delta(\theta^*; \rho, P) > \bar{K}_\delta(\theta^*; \rho_0, P_*) - \varepsilon'$ for (ρ, P) close enough to (ρ_0, P_*) . Therefore, on that neighborhood,

$$\bar{b}(\delta, \rho, P) \geq \bar{K}_\delta(\theta^*; \rho, P) > \bar{b}(\delta, \rho_0, P_*) - \varepsilon',$$

which proves lower semicontinuity. For upper semicontinuity, suppose, towards contradiction, that there is a sequence $(\rho_n, P_n) \rightarrow (\rho_0, P_*)$, with $\|P_n - P_*\|_{\mathcal{H}_K^1} \rightarrow 0$ for every compact K , along which $\bar{b}(\delta, \rho_n, P_n) \geq \bar{b}(\delta, \rho_0, P_*) + 2\varepsilon'$ for some $\varepsilon' > 0$. Pick $\theta_n \in \Theta_\delta(\rho_n, P_n)$ with $\bar{K}_\delta(\theta_n; \rho_n, P_n) > \bar{b}(\delta, \rho_n, P_n) - \varepsilon'$, which exists because $\bar{b}(\delta, \rho_n, P_n)$ is the supremum of $\bar{K}_\delta(\cdot; \rho_n, P_n)$ over $\Theta_\delta(\rho_n, P_n)$, nonempty for n large by part (i). Compactness of Θ (Assumption SA-4(i)) gives a subsequence with $\theta_{n_l} \rightarrow \bar{\theta} \in \Theta$. By the joint upper semicontinuity of the dual value along such sequences, recorded above,

$$\bar{K}_\delta(\bar{\theta}; \rho_0, P_*) \geq \limsup_l \bar{K}_\delta(\theta_{n_l}; \rho_{n_l}, P_{n_l}) \geq \limsup_l (\bar{b}(\delta, \rho_{n_l}, P_{n_l}) - \varepsilon') \geq \bar{b}(\delta, \rho_0, P_*) + \varepsilon'.$$

On the other hand, by Assumption SA-4(iii) the argmax of $\bar{K}_\delta(\cdot; \rho_0, P_*)$ over all of Θ lies inside $\Theta_\delta(\rho_0, P_*)$, so that $\bar{b}(\delta, \rho_0, P_*) = \sup_{\theta \in \Theta} \bar{K}_\delta(\theta; \rho_0, P_*) \geq \bar{K}_\delta(\bar{\theta}; \rho_0, P_*)$, a contradiction. This proves part (v).

Part (vi). Fix an open set $\mathcal{U} \supseteq \bar{\Theta}_\delta(\rho_0, P_*)$ and set $\mathcal{K} := \Theta \setminus \mathcal{U}$, a compact set disjoint from $\bar{\Theta}_\delta(\rho_0, P_*)$. The map $\bar{K}_\delta(\cdot; \rho_0, P_*)$ is upper semicontinuous on $\mathcal{K}$, by the third property above, and bounded above by $\bar{G}$, so $\beta_{\mathcal{K}} := \sup_{\theta \in \mathcal{K}} \bar{K}_\delta(\theta; \rho_0, P_*) \in [-\infty, \infty)$ is attained on $\mathcal{K}$ whenever finite. Since $\bar{b}(\delta, \rho_0, P_*)$ is the supremum of $\bar{K}_\delta(\cdot; \rho_0, P_*)$ over all of Θ and is attained only on $\bar{\Theta}_\delta(\rho_0, P_*)$ (Assumption SA-4(iii)), we have $\beta_{\mathcal{K}} < \bar{b}(\delta, \rho_0, P_*)$. Let $\varepsilon_{\mathcal{K}} := \min\{1, (\bar{b}(\delta, \rho_0, P_*) - \beta_{\mathcal{K}})/2\} > 0$, with $\varepsilon_{\mathcal{K}} := 1$ when $\beta_{\mathcal{K}} = -\infty$. We claim that $\sup_{\theta \in \mathcal{K}} \bar{K}_\delta(\theta; \rho_n, P_n) \leq \bar{b}(\delta, \rho_0, P_*) - \varepsilon_{\mathcal{K}}$ for all n large. Otherwise there are infinitely many n and points $\theta_n \in \mathcal{K}$ with $\bar{K}_\delta(\theta_n; \rho_n, P_n) > \bar{b}(\delta, \rho_0, P_*) - \varepsilon_{\mathcal{K}}$, and by compactness a subsequence has $\theta_n \rightarrow \theta_* \in \mathcal{K}$, so that the joint upper semicontinuity of the dual value along (θ_n, ρ_n, P_n) yields

$$\bar{b}(\delta, \rho_0, P_*) - \varepsilon_{\mathcal{K}} \leq \limsup_n \bar{K}_\delta(\theta_n; \rho_n, P_n) \leq \bar{K}_\delta(\theta_*; \rho_0, P_*) \leq \beta_{\mathcal{K}} \leq \bar{b}(\delta, \rho_0, P_*) - 2\varepsilon_{\mathcal{K}},$$

a contradiction in view of $\varepsilon_{\mathcal{K}} > 0$, the last inequality holding by the definition of $\varepsilon_{\mathcal{K}}$ and trivially when $\beta_{\mathcal{K}} = -\infty$. On the other hand, $\bar{b}(\delta, \rho_n, P_n) \rightarrow \bar{b}(\delta, \rho_0, P_*)$ by the first assertion of part (v) and $\nu_n \rightarrow 0$, so that $\bar{b}(\delta, \rho_n, P_n) - \nu_n > \bar{b}(\delta, \rho_0, P_*) - \varepsilon_{\mathcal{K}}$ for all n large. For such n no $\theta \in \mathcal{K}$ satisfies $\bar{K}_\delta(\theta; \rho_n, P_n) \geq \bar{b}(\delta, \rho_n, P_n) - \nu_n$, which proves part (vi). $\blacksquare$

SA8.25 Proof of Theorem SA-28

Proof. For $j \geq 1$, let $K_j := \{\xi \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m} : \|\xi\| \leq j\}$. By Assumption SA-4(iv), $\hat{\rho}_n \xrightarrow{\mathbb{P}} \rho_0$, whereas Lemma SA-26 gives $\|\hat{P}_n - P_*\|_{\mathcal{H}} \xrightarrow{\mathbb{P}} 0$ and $\|\hat{P}_n - P_*\|_{\mathcal{H}_{K_j}^J} \xrightarrow{\mathbb{P}} 0$ for every j . By Borel-Cantelli, from any subsequence we may therefore extract a further subsequence along which all these convergences hold almost surely. We do so to be able to apply Lemma SA-27(v), which holds for deterministic sequences. Every compact $K \subseteq \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^{d_m}$ is contained in some K_j , so Lemma SA-27(v) gives $\bar{\mathbf{b}}(\delta, \hat{\rho}_n, \hat{P}_n) \rightarrow \bar{\mathbf{b}}(\delta, \rho_0, P_*)$ almost surely along that further subsequence. The subsequence characterization of convergence in probability then yields $\hat{\mathbf{b}}_n(\delta) \xrightarrow{\mathbb{P}} \bar{\mathbf{b}}(\delta, \rho_0, P_*)$. The argument for $\hat{\mathbf{b}}_n(\delta)$ is identical with $g \mapsto -g$. ■

SA8.26 Proof of Theorem SA-29

Proof. We prove the result for $\bar{\mathbf{b}}$. That for $\underline{\mathbf{b}}$ is symmetric, and the joint statement then follows because both coordinates are functionals of the same pair $(\hat{\rho}_n, \hat{P}_n)$. Recall that $\mathbf{b}_0 = (P_*\ell)_{\ell \in \mathcal{H}}$ is the element of $\ell^\infty(\mathcal{H})$ induced by P_* , as defined in Section SA4.4, and fix sequences $t_n \downarrow 0$, $h_n \rightarrow h$ in $\mathbb{R}^{d_e}$, and $\gamma_n \rightarrow \gamma$ in $\ell^\infty(\mathcal{H})$ with $\gamma \in UC(\mathcal{H}, e_{P_*})$. Write $\rho_n := \rho_0 + t_n h_n$ and $\mathbf{b}_n := \mathbf{b}_0 + t_n \gamma_n$. By (I) the feasible set $\mathcal{F}(\theta)$ does not depend on $(\rho, \mathbf{b})$, so a dual point feasible at one $(\rho, \mathbf{b})$ is feasible at every other. By (L) the objective $J(\xi, \theta; \cdot)$ is jointly affine in $(\rho, \mathbf{b})$, so that, for every $\theta \in \Theta$ and every $\xi \in \mathcal{F}(\theta) \cap \mathcal{A}$,

$$J(\xi, \theta; \rho_n, \mathbf{b}_n) - J(\xi, \theta; \rho_0, \mathbf{b}_0) = t_n \{ \lambda_{12}^\top h_n + \gamma_n \ell_{\xi, \theta} \}.$$

Step 1: directional derivative. The argument consists of a lower bound obtained by holding a population optimizer fixed and an upper bound obtained along near-optimizers of the perturbed program.

(a) Lower bound. Fix $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$. Then $\theta \in \mathcal{N}_\Theta$, and $D(\theta; \rho_n, \mathbf{b}_n) < \delta$ for all n large. Indeed, $D(\theta; \rho_0, \mathbf{b}_0) = D(\theta; \rho_0, P_*) < \delta$ by Assumption SA-4(iii), and, by its dual definition restricted to $\mathcal{A}$, the map $\mathbf{b} \mapsto D(\theta; \rho, \mathbf{b})$ is 1-Lipschitz for the $\ell^\infty(\mathcal{H})$ norm and the map $\rho \mapsto D(\theta; \rho, \mathbf{b})$ is Lipschitz, in both cases uniformly in the remaining argument. This is the counterpart of Lemma SA-27(i) with P replaced by $\mathbf{b}$. Hence θ belongs to the constraint set of the supremum defining $\bar{\mathbf{b}}(\delta, \rho_n, \mathbf{b}_n)$ for all n large. Note that $\bar{K}_\delta(\theta; \rho_n, \mathbf{b}_n)$ and $\bar{K}_\delta(\theta; \rho_n, P_*)$ are infima of objectives that differ by $(\mathbf{b}_n - \mathbf{b}_0)(\ell_{\xi, \theta}) = t_n \gamma_n \ell_{\xi, \theta}$, where $\ell_{\xi, \theta} \in \mathcal{H}_{\mathcal{A}}^J \subseteq \mathcal{H}$ for every $\xi \in \mathcal{F}(\theta) \cap \mathcal{A}$, so that the two values differ by at most $t_n \|\gamma_n\|_{\ell^\infty(\mathcal{H})}$. Moreover, $\bar{K}_\delta(\theta; \rho_n, P_*)$ is finite for all n large. We may thus choose $\xi_n^* = (\zeta_n^*, \eta_n^*, \lambda_n^*) \in \mathcal{F}(\theta) \cap \mathcal{A}$ with $J(\xi_n^*, \theta; \rho_n, \mathbf{b}_n) \leq \bar{K}_\delta(\theta; \rho_n, \mathbf{b}_n) + t_n^2$. Since ξ_n^* is t_n^2 -optimal at $(\rho_n, \mathbf{b}_n)$ and feasible at $(\rho_0, \mathbf{b}_0)$, so that $\bar{K}_\delta(\theta; \rho_0, \mathbf{b}_0) \leq J(\xi_n^*, \theta; \rho_0, \mathbf{b}_0)$, subtracting the two inequalities and using (L) yields

$$\bar{K}_\delta(\theta; \rho_n, \mathbf{b}_n) - \bar{K}_\delta(\theta; \rho_0, \mathbf{b}_0) \geq t_n \{ (\lambda_{12, n}^*)^\top h_n + \gamma_n \ell_{\xi_n^*, \theta} \} - t_n^2.$$

The sequence $\{\xi_n^*\}$ lies in the compact set $\mathcal{F}(\theta) \cap \mathcal{A}$. Let $\xi_{n_i}^* \rightarrow \xi^\infty \in \mathcal{F}(\theta) \cap \mathcal{A}$ along a subsequence. By (L) and boundedness of $\mathcal{A}$, $|J(\xi, \theta; \rho_{n_i}, \mathbf{b}_{n_i}) - J(\xi, \theta; \rho_0, \mathbf{b}_0)| \rightarrow 0$ uniformly over $\xi \in \mathcal{F}(\theta) \cap \mathcal{A}$, so that $\bar{K}_\delta(\theta; \rho_{n_i}, \mathbf{b}_{n_i}) \rightarrow \bar{K}_\delta(\theta; \rho_0, \mathbf{b}_0)$ and, by the near-optimality of $\xi_{n_i}^*$, $J(\xi_{n_i}^*, \theta; \rho_0, \mathbf{b}_0) \rightarrow \bar{K}_\delta(\theta; \rho_0, \mathbf{b}_0)$, whereas $J(\xi_{n_i}^*, \theta; \rho_0, \mathbf{b}_0) \rightarrow J(\xi^\infty, \theta; \rho_0, \mathbf{b}_0)$ by the dominated convergence theorem, as $\mathbf{b}_0$ is induced by P_* . Hence $J(\xi^\infty, \theta; \rho_0, \mathbf{b}_0) = \bar{K}_\delta(\theta; \rho_0, \mathbf{b}_0)$, that is, $\xi^\infty \in \bar{\Xi}_\delta(\theta; \rho_0, \mathbf{b}_0)$. Since, moreover, $|\gamma_n \ell_{\xi, \theta} - \gamma \ell_{\xi, \theta}| \leq \|\gamma_n - \gamma\|_{\ell^\infty(\mathcal{H})} \rightarrow 0$ uniformly in ξ and $\xi \mapsto \gamma \ell_{\xi, \theta}$ is continuous on $\mathcal{F}(\theta) \cap \mathcal{A}$ by the fourth property of the dual in the proof of Lemma SA-27, passing to a subsequence along which the lim inf is attained yields

$$\liminf_n \{ (\lambda_{12, n}^*)^\top h_n + \gamma_n \ell_{\xi_n^*, \theta} \} \geq \min_{\xi^* \in \bar{\Xi}_\delta(\theta; \rho_0, \mathbf{b}_0)} \{ (\lambda_{12}^*)^\top h + \gamma \ell_{\xi^*, \theta} \},$$

where the minimum on the right-hand side is attained because the map $\xi \mapsto \lambda_{12}^\top h + \gamma \ell_{\xi, \theta}$ is continuous on the nonempty compact set $\bar{\Xi}_\delta(\theta; \rho_0, \mathbf{b}_0) = \bar{\Xi}_\delta(\theta; \rho_0, P_*)$ and where the equality follows from $\bar{\Xi}_\delta(\theta; \rho_0, P_*) \subseteq \bar{\Xi} \subseteq \mathcal{A}$. Next, $\bar{\mathbf{b}}(\delta, \rho_n, \mathbf{b}_n) \geq \bar{K}_\delta(\theta; \rho_n, \mathbf{b}_n)$ for all n large, because θ belongs to the constraint set of the supremum, and $\bar{\mathbf{b}}(\delta, \rho_0, \mathbf{b}_0) = \bar{K}_\delta(\theta; \rho_0, \mathbf{b}_0)$, by the identity $\bar{\mathbf{b}}(\delta, \rho_0, \mathbf{b}_0) = \bar{\mathbf{b}}(\delta, \rho_0, P_*)$ and $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$. Dividing by t_n and taking the supremum over $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$, we obtain

$$\liminf_n \frac{\bar{\mathbf{b}}(\delta, \rho_n, \mathbf{b}_n) - \bar{\mathbf{b}}(\delta, \rho_0, \mathbf{b}_0)}{t_n} \geq \max_{\theta^*} \min_{\xi^*} \{ (\lambda_{12}^*)^\top h + \gamma \ell_{\xi^*, \theta^*} \}. \quad (85)$$

(b) Upper bound. Fix an arbitrary subsequence of $\{(t_n, h_n, \gamma_n)\}_{n \geq 1}$, which we continue to index by n . Along it, choose $\theta_n \in \mathcal{N}_\Theta$ with $D(\theta_n; \rho_n, \mathbf{b}_n) < \delta$ and $\bar{K}_\delta(\theta_n; \rho_n, \mathbf{b}_n) \geq \bar{\mathbf{b}}(\delta, \rho_n, \mathbf{b}_n) - t_n^2$, and, by compactness of $\mathcal{N}_\Theta$, extract a further subsequence along which $\theta_n \rightarrow \bar{\theta} \in \mathcal{N}_\Theta$. We first show that $\bar{\theta} \in \bar{\Theta}_\delta(\rho_0, P_*)$. Since the difference quotient in (85) has a finite limit inferior and $t_n \rightarrow 0$, part (a) gives $\liminf_n \bar{\mathbf{b}}(\delta, \rho_n, \mathbf{b}_n) \geq \bar{\mathbf{b}}(\delta, \rho_0, \mathbf{b}_0)$. Conversely, the extended criterion is jointly upper semicontinuous along $(\theta_n, \rho_n, \mathbf{b}_n) \rightarrow (\bar{\theta}, \rho_0, \mathbf{b}_0)$, as it is 1-Lipschitz in $\mathbf{b}$ uniformly in the remaining arguments and equal, at $\mathbf{b}_0$ and on $\mathcal{N}_\Theta$, to the jointly upper semicontinuous map $(\theta, \rho) \mapsto \bar{K}_\delta(\theta; \rho, P_*)$ for ρ in a neighborhood of ρ_0 . Hence

$$\bar{\mathbf{b}}(\delta, \rho_0, \mathbf{b}_0) \geq \bar{K}_\delta(\bar{\theta}; \rho_0, \mathbf{b}_0) \geq \limsup_n \bar{K}_\delta(\theta_n; \rho_n, \mathbf{b}_n) \geq \limsup_n \bar{\mathbf{b}}(\delta, \rho_n, \mathbf{b}_n),$$

where the first inequality holds because $\bar{\mathbf{b}}(\delta, \rho_0, P_*)$ is the supremum of $\bar{K}_\delta(\cdot; \rho_0, P_*)$ over all of Θ (Assumption SA-4(iii)), and the last inequality follows from the choice of θ_n . Together with the limit inferior bound, this shows that

$\bar{\mathbf{b}}(\delta, \rho_n, \mathbf{b}_n) \rightarrow \bar{\mathbf{b}}(\delta, \rho_0, \mathbf{b}_0)$ and $\bar{K}_\delta(\bar{\theta}; \rho_0, P_*) = \bar{\mathbf{b}}(\delta, \rho_0, P_*)$, that is, $\bar{\theta} \in \bar{\Theta}_\delta(\rho_0, P_*)$.

Next, we argue as in part (a) with the roles of $(\rho_0, \mathbf{b}_0)$ and $(\rho_n, \mathbf{b}_n)$ interchanged. Let ξ_n^* minimize $\xi \mapsto \lambda_{12}^\top h_n + \gamma \ell_{\xi, \theta_n}$ over the compact set $\Xi_\delta(\theta_n; \rho_0, \mathbf{b}_0)$, where the minimum is attained by the continuity of $\xi \mapsto \gamma \ell_{\xi, \theta_n}$ noted in part (a). Since ξ_n^* is optimal at $(\rho_0, \mathbf{b}_0)$ and feasible at $(\rho_n, \mathbf{b}_n)$, (L) bounds $\bar{K}_\delta(\theta_n; \rho_n, \mathbf{b}_n) - \bar{K}_\delta(\theta_n; \rho_0, \mathbf{b}_0)$ from above by $t_n \{(\lambda_{12, n}^*)^\top h_n + \gamma_n \ell_{\xi_n^*, \theta_n}\}$. Combining this bound with the near-optimality of θ_n , with $\bar{K}_\delta(\theta_n; \rho_0, \mathbf{b}_0) \leq \bar{\mathbf{b}}(\delta, \rho_0, \mathbf{b}_0)$, which holds as in the first inequality of the preceding display, and with the replacement of (h_n, γ_n) by (h, γ) at a cost that vanishes uniformly in ξ , exactly as in part (a), we obtain, along the subsequence,

$$\limsup_n \frac{\bar{\mathbf{b}}(\delta, \rho_n, \mathbf{b}_n) - \bar{\mathbf{b}}(\delta, \rho_0, \mathbf{b}_0)}{t_n} \leq \limsup_n \min_{\xi \in \Xi_\delta(\theta_n; \rho_0, \mathbf{b}_0)} \{\lambda_{12}^\top h + \gamma \ell_{\xi, \theta_n}\}.$$

It remains to bound the right-hand side. Fix $\xi^* \in \Xi_\delta(\bar{\theta}; \rho_0, \mathbf{b}_0)$. By the lower hemicontinuity of $\theta \mapsto \Xi_\delta(\theta; \rho_0, P_*)$ at $\bar{\theta} \in \bar{\Theta}_\delta(\rho_0, P_*)$ (Assumption SA-4(v)) and the equality $\Xi_\delta(\theta_n; \rho_0, \mathbf{b}_0) = \Xi_\delta(\theta_n; \rho_0, P_*)$ on $\mathcal{N}_\Theta$, there exist $\xi_n \in \Xi_\delta(\theta_n; \rho_0, \mathbf{b}_0)$ with $\xi_n \rightarrow \xi^*$. The map $(\theta, \xi) \mapsto \lambda_{12}^\top h + \gamma \ell_{\xi, \theta}$ is continuous on the pairs (θ, ξ) with $\xi \in \mathcal{F}(\theta) \cap \mathcal{A}$, by the fourth property of the dual in the proof of Lemma SA-27, so that the right-hand side of the preceding display is at most $(\lambda_{12}^*)^\top h + \gamma \ell_{\xi^*, \bar{\theta}}$. Minimizing over ξ^* and using $\bar{\theta} \in \bar{\Theta}_\delta(\rho_0, P_*)$, we obtain, along the subsequence,

$$\limsup_n \frac{\bar{\mathbf{b}}(\delta, \rho_n, \mathbf{b}_n) - \bar{\mathbf{b}}(\delta, \rho_0, \mathbf{b}_0)}{t_n} \leq \min_{\xi^* \in \Xi_\delta(\bar{\theta}; \rho_0, \mathbf{b}_0)} \{(\lambda_{12}^*)^\top h + \gamma \ell_{\xi^*, \bar{\theta}}\} \leq \max_{\theta^*} \min_{\xi^*} \{(\lambda_{12}^*)^\top h + \gamma \ell_{\xi^*, \theta^*}\}.$$

Since the initial subsequence was arbitrary and the bound does not depend on it, the same bound holds for the limit superior along the full sequence. Together with (85), this proves the expression for $d\bar{\mathbf{b}}_{(\rho_0, P_*)}[h, \gamma]$ in (33). Finally, the map $(h, \gamma) \mapsto d\bar{\mathbf{b}}_{(\rho_0, P_*)}[h, \gamma]$ is continuous on $\mathbb{R}^{d_e} \times UC(\mathcal{H}, e_{P_*})$. Indeed, it is Lipschitz,

$$|d\bar{\mathbf{b}}_{(\rho_0, P_*)}[h, \gamma] - d\bar{\mathbf{b}}_{(\rho_0, P_*)}[h', \gamma']| \leq M_{\mathcal{A}} \|h - h'\| + \|\gamma - \gamma'\|_{\ell^\infty(\mathcal{H})}, \quad M_{\mathcal{A}} := \sup\{\|\lambda_{12}\| : \xi \in \mathcal{A}\},$$

where $M_{\mathcal{A}} < \infty$ as $\mathcal{A}$ is compact. The bound follows from $|\max_S f - \max_S f'| \leq \sup_S |f - f'|$ and the same inequality for minima, because, for every $\theta^* \in \bar{\Theta}_\delta(\rho_0, P_*)$ and $\xi^* \in \Xi_\delta(\theta^*; \rho_0, P_*) \subseteq \Xi \subseteq \mathcal{A}$, we have $\|\lambda_{12}^*\| \leq M_{\mathcal{A}}$ and $\ell_{\xi^*, \theta^*} \in \mathcal{H}$, so that $|(\gamma - \gamma') \ell_{\xi^*, \theta^*}| \leq \|\gamma - \gamma'\|_{\ell^\infty(\mathcal{H})}$.

Step 2: asymptotic distribution. By Lemma SA-26 and Assumption SA-4(iv), the class $\mathcal{H}$ augmented by the d_e coordinates of the influence function ψ is P_* -Donsker, so that

$$\sqrt{n}((\hat{\rho}_n - \rho_0), (\hat{P}_n - P_*)) \rightsquigarrow (\mathbb{Z}, \mathbb{G}) \quad \text{in } \mathbb{R}^{d_e} \times \ell^\infty(\mathcal{H}),$$

where the limit $(\mathbb{Z}, \mathbb{G})$ is a tight Gaussian process whose support is contained in $\mathbb{R}^{d_e} \times UC(\mathcal{H}, e_{P_*})$. This is the tangent set relative to which Step 1 established the Hadamard directional differentiability of $(\rho, \mathbf{b}) \mapsto (\underline{\mathbf{b}}, \bar{\mathbf{b}})(\delta, \rho, \mathbf{b})$. The hypotheses of the directional functional delta method (Fang and Santos, 2019, Theorem 2.1) are therefore met, and, in view of the two identities established below, it yields

$$\sqrt{n} \begin{pmatrix} \hat{\underline{\mathbf{b}}}_n(\delta) - \underline{\mathbf{b}}(\delta, \rho_0, P_*) \\ \hat{\bar{\mathbf{b}}}_n(\delta) - \bar{\mathbf{b}}(\delta, \rho_0, P_*) \end{pmatrix} \rightsquigarrow \begin{pmatrix} d\underline{\mathbf{b}}_{(\rho_0, P_*)}[\mathbb{Z}, \mathbb{G}] \\ d\bar{\mathbf{b}}_{(\rho_0, P_*)}[\mathbb{Z}, \mathbb{G}] \end{pmatrix}.$$

Two identities underlie the passage from the extended functional to the estimators and population bounds appearing in that display. First, the centering constants equal the population bounds, by the identity $\bar{\mathbf{b}}(\delta, \rho_0, \mathbf{b}_0) = \bar{\mathbf{b}}(\delta, \rho_0, P_*)$. Second, $\hat{\underline{\mathbf{b}}}_n(\delta) = \bar{\mathbf{b}}(\delta, \hat{\rho}_n, \hat{\mathbf{b}}_n)$, with $\hat{\mathbf{b}}_n := (\hat{P}_n \ell)_{\ell \in \mathcal{H}}$, with probability approaching one, so that the conclusion of the delta method for the extended functional transfers to $\hat{\underline{\mathbf{b}}}_n(\delta)$. We prove the second identity by showing that, with probability approaching one, the supremum over Θ defining $\hat{\underline{\mathbf{b}}}_n(\delta)$ is unaffected by restricting it to $\mathcal{N}_\Theta$. On the event $\|(\hat{\rho}_n, \hat{P}_n) - (\rho_0, P_*)\|_{\mathcal{H}_0} < \varepsilon_*$, whose probability tends to one by Assumption SA-4(iv) and the Glivenko–Cantelli property of Lemma SA-26, the extended program coincides with the sample program on $\mathcal{N}_\Theta$, so that $\bar{\mathbf{b}}(\delta, \hat{\rho}_n, \hat{\mathbf{b}}_n) = \sup\{\bar{K}_\delta(\theta; \hat{\rho}_n, \hat{P}_n) : \theta \in \mathcal{N}_\Theta, D(\theta; \hat{\rho}_n, \hat{P}_n) < \delta\} \leq \hat{\underline{\mathbf{b}}}_n(\delta)$. For the reverse inequality, we pass, as in the proof of Theorem SA-28, to a subsequence along which $\|(\hat{\rho}_n, \hat{P}_n) - (\rho_0, P_*)\|_{\mathcal{H}} \rightarrow 0$ and $\|\hat{P}_n - P_*\|_{\mathcal{H}_K^*} \rightarrow 0$ for every compact K almost surely, and apply Lemma SA-27(vi) with the open set $\mathcal{U} = \{\theta : d(\theta, \bar{\Theta}_\delta(\rho_0, P_*)) < \varepsilon_\Theta\}$ and $\nu_n = 1/n$. For all n large, every $\theta \in \Theta_\delta(\hat{\rho}_n, \hat{P}_n)$ with $\bar{K}_\delta(\theta; \hat{\rho}_n, \hat{P}_n) \geq \hat{\underline{\mathbf{b}}}_n(\delta) - 1/n$ then lies in $\mathcal{U} \cap \Theta \subseteq \mathcal{N}_\Theta$, so that the supremum of $\bar{K}_\delta(\cdot; \hat{\rho}_n, \hat{P}_n)$ over $\Theta_\delta(\hat{\rho}_n, \hat{P}_n) \setminus \mathcal{N}_\Theta$ is at most $\hat{\underline{\mathbf{b}}}_n(\delta) - 1/n < \hat{\underline{\mathbf{b}}}_n(\delta)$, where $\hat{\underline{\mathbf{b}}}_n(\delta)$ is finite for n large by Lemmas SA-27(i) and SA-27(v). Hence the supremum over $\Theta_\delta(\hat{\rho}_n, \hat{P}_n)$ defining $\hat{\underline{\mathbf{b}}}_n(\delta)$ equals the supremum over $\Theta_\delta(\hat{\rho}_n, \hat{P}_n) \cap \mathcal{N}_\Theta$, and, as the subsequence is arbitrary, this holds with probability approaching one, that is, $\hat{\underline{\mathbf{b}}}_n(\delta) = \bar{\mathbf{b}}(\delta, \hat{\rho}_n, \hat{\mathbf{b}}_n)$. The same argument applies to $\hat{\bar{\mathbf{b}}}_n(\delta)$ with $g \mapsto -g$, which completes the proof. ■

SA8.27 Proof of Theorem SA-30

Proof. Throughout, we let $\mathcal{N}$, r , and M be the neighborhood, radius, and constant of Lemma SA-27(v), and we use the Hausdorff distance d_H . To prove the main result, by Theorem 3.2 in Fang and Santos (2019), it suffices to verify the four conditions (FS1)–(FS4) below.

(FS1). Hadamard directional differentiability of $(\rho, P) \mapsto (\underline{\mathbf{b}}, \bar{\mathbf{b}})(\delta, \rho, P)$ at (ρ_0, P_*) tangentially to $\mathbb{R}^{d_e} \times UC(\mathcal{H}, e_{P_*})$, with continuous derivative, is Theorem SA-29.

(FS2). $\sqrt{n}((\hat{\rho}_n, \hat{P}_n) - (\rho_0, P_*)) \rightsquigarrow (\mathbb{Z}, \mathbb{G})$ is established in Step 2 of the proof of Theorem SA-29.

(FS3). The conditional bootstrap law of $\sqrt{n}((\hat{\rho}_n^*, \hat{P}_n^*) - (\hat{\rho}_n, \hat{P}_n))$ converges to the same $(\mathbb{Z}, \mathbb{G})$ in probability. This is the consistency of the nonparametric bootstrap for the P_* -Donsker class $\mathcal{H}$ adjoined with ψ (van der Vaart and Wellner, 1996, Theorems 3.6.1–3.6.2), which gives the claim for the pair $(\mathbb{G}_n^* \psi, \mathbb{G}_n^*)$, in the bounded-Lipschitz metric and in outer probability. The bootstrap statistic differs from this pair only in its first coordinate, by $R_n^* := \sqrt{n}(\hat{\rho}_n^* - \hat{\rho}_n) - \mathbb{G}_n^* \psi$, which is $o_{\mathbb{P}}(1)$ by the linearization assumed in Theorem SA-30. As every function f bounded by one and 1-Lipschitz satisfies $|f(a) - f(a')| \leq \min\{\|a - a'\|, 2\}$, the bounded-Lipschitz distances of the two conditional laws from the law of $(\mathbb{Z}, \mathbb{G})$ differ by at most $\mathbb{E}[\min\{\|R_n^*\|, 2\} \mid Z_1, \dots, Z_n]$. This term converges to zero in outer probability by Markov's inequality, as its expectation $\mathbb{E}[\min\{\|R_n^*\|, 2\}]$ vanishes by bounded convergence.

(FS4). As in the extension of Section SA4.4, we consider the localized derivative, defined as $\widehat{d\mathbf{b}}_n$ but with the outer supremum restricted to $\widehat{\Theta}_{\delta,n} \cap \mathcal{N}_{\Theta}$ and the inner infimum to $\Xi_{\delta}(\theta; \hat{\rho}_n, \hat{P}_n) \cap \mathcal{A}$, which is what places $\ell_{\xi,\theta}$ in $\mathcal{H}$. Neither restriction binds with probability approaching one. Indeed, by Assumption SA-4(iv) and Lemma SA-26, $\|(\hat{\rho}_n, \hat{P}_n) - (\rho_0, P_*)\|_{\mathcal{H}} \xrightarrow{\mathbb{P}} 0$ and $\|\hat{P}_n - P_*\|_{\mathcal{H}_K^1} \xrightarrow{\mathbb{P}} 0$ for every compact K , while $\hat{\nu}_n \sqrt{\log n/n} \xrightarrow{\mathbb{P}} 0$, as $\hat{\nu}_n = O_{\mathbb{P}}(1)$, so that, exactly as in the proof of Theorem SA-28, from any subsequence we may extract a further subsequence along which these convergences hold almost surely. Along it, Lemma SA-27(vi), applied with $\mathcal{U}$ the intersection of $\text{int}(\mathcal{N})$ and the open set $\{\theta : d(\theta, \bar{\Theta}_{\delta}(\rho_0, P_*)) < \varepsilon_{\Theta}\}$, which contains $\bar{\Theta}_{\delta}(\rho_0, P_*)$, shows that $\widehat{\Theta}_{\delta,n} \subseteq \mathcal{U}$, and hence $\widehat{\Theta}_{\delta,n} \subseteq \mathcal{U} \cap \Theta \subseteq \mathcal{N}_{\Theta} \cap \mathcal{N}$, for all n large. Since the subsequence was arbitrary, the probability of the event

$$\widehat{\Theta}_{\delta,n} \subseteq \mathcal{N}_{\Theta} \cap \mathcal{N} \text{ and } \|(\hat{\rho}_n, \hat{P}_n) - (\rho_0, P_*)\|_{\mathcal{H}} < r \quad (86)$$

approaches 1. On the event in (86), Lemma SA-27(v) makes the sets $\Xi_{\delta}(\theta; \hat{\rho}_n, \hat{P}_n)$, $\theta \in \widehat{\Theta}_{\delta,n}$, nonempty, compact, and contained in $\Xi \subseteq \mathcal{A}$. Both restrictions are then vacuous: $\widehat{\Theta}_{\delta,n} \subseteq \mathcal{N}_{\Theta}$ gives $\widehat{\Theta}_{\delta,n} \cap \mathcal{N}_{\Theta} = \widehat{\Theta}_{\delta,n}$, and $\Xi_{\delta}(\theta; \hat{\rho}_n, \hat{P}_n) \subseteq \mathcal{A}$ gives $\Xi_{\delta}(\theta; \hat{\rho}_n, \hat{P}_n) \cap \mathcal{A} = \Xi_{\delta}(\theta; \hat{\rho}_n, \hat{P}_n)$ for every $\theta \in \widehat{\Theta}_{\delta,n}$, so that on this event the localized derivative coincides with the expression of Section SA4.5,

$$\sup_{\theta \in \widehat{\Theta}_{\delta,n} \cap \mathcal{N}_{\Theta}} \inf_{\xi \in \Xi_{\delta}(\theta; \hat{\rho}_n, \hat{P}_n) \cap \mathcal{A}} \{\lambda_{12}^{\top} h + \gamma \ell_{\xi,\theta}\} = \sup_{\theta \in \widehat{\Theta}_{\delta,n}} \inf_{\xi \in \Xi_{\delta}(\theta; \hat{\rho}_n, \hat{P}_n)} \{\lambda_{12}^{\top} h + \gamma \ell_{\xi,\theta}\} = \widehat{d\mathbf{b}}_n[h, \gamma].$$

Let $\widetilde{d\mathbf{b}}_n$ be equal to the localized derivative, and hence to $\widehat{d\mathbf{b}}_n$, on the intersection of the event in (86) with $\{\widehat{\Theta}_{\delta,n} \neq \emptyset\}$, an event whose probability also tends to one by part (i) of the first fact below, and set $\widetilde{d\mathbf{b}}_n := 0$ off it. On that event the outer supremum is over a nonempty set and each inner infimum over a nonempty subset of the compact set $\mathcal{A}$, so that $\widetilde{d\mathbf{b}}_n$ is Lipschitz on all of $\mathbb{R}^{d_e} \times \ell^{\infty}(\mathcal{H})$,

$$|\widetilde{d\mathbf{b}}_n[h, \gamma] - \widetilde{d\mathbf{b}}_n[h', \gamma']| \leq M_{\mathcal{A}} \|h - h'\| + \|\gamma - \gamma'\|_{\ell^{\infty}(\mathcal{H})},$$

with $M_{\mathcal{A}}$ as in the proof of Theorem SA-29, by $|\sup_S f - \sup_S f'| \leq \sup_S |f - f'|$, the same inequality for infima, and $|(\gamma - \gamma')\ell_{\xi,\theta}| \leq \|\gamma - \gamma'\|_{\ell^{\infty}(\mathcal{H})}$ for $\ell_{\xi,\theta} \in \mathcal{H}$. By Fang and Santos (2019, Remark 3.4 and Lemma S.3.6), their Assumption 4 therefore holds for $\widetilde{d\mathbf{b}}_n$ once we establish pointwise convergence on the tangent set, which, as $\widetilde{d\mathbf{b}}_n$ and $\widehat{d\mathbf{b}}_n$ coincide with probability approaching one, amounts to

$$\widehat{d\mathbf{b}}_n[h, \gamma] \xrightarrow{\mathbb{P}} d\bar{\mathbf{b}}_{(\rho_0, P_*)}[h, \gamma] \quad \text{for every } (h, \gamma) \in \mathbb{R}^{d_e} \times UC(\mathcal{H}, e_{P_*}). \quad (87)$$

Fix $(h, \gamma) \in \mathbb{R}^{d_e} \times UC(\mathcal{H}, e_{P_*})$ and write

$$\Phi(\theta, \xi) := \lambda_{12}^{\top} h + \gamma \ell_{\xi,\theta}, \quad v(\theta) := v(\theta; h, \gamma), \quad \widehat{v}_n(\theta) := \min_{\xi \in \Xi_{\delta}(\theta; \hat{\rho}_n, \hat{P}_n)} \Phi(\theta, \xi),$$

where $v(\theta; h, \gamma) = \min_{\xi \in \Xi_{\delta}(\theta; \rho_0, P_*)} \Phi(\theta, \xi)$ is the inner value of Lemma SA-27(iv), so that $d\bar{\mathbf{b}}_{(\rho_0, P_*)}[h, \gamma] = \max_{\theta \in \bar{\Theta}_{\delta}(\rho_0, P_*)} v(\theta)$ and, on the event in (86), $\widehat{d\mathbf{b}}_n[h, \gamma] = \sup_{\theta \in \widehat{\Theta}_{\delta,n}} \widehat{v}_n(\theta)$, with v well defined on $\mathcal{N} \cap \Theta$ by Lemma SA-27(v). By the fourth property of the dual in the proof of Lemma SA-27, Φ is uniformly continuous on the compact set of pairs $(\theta, \xi) \in \Theta \times \mathcal{A}$ with $\xi \in \mathcal{F}(\theta)$, with a nondecreasing modulus ω satisfying $\omega(t) \rightarrow 0$ as $t \downarrow 0$. The minima above, taken over nonempty compact sets, are therefore attained, and so is the maximum of v over the compact set $\bar{\Theta}_{\delta}(\rho_0, P_*)$, on which v is continuous by Lemma SA-27(iv). The set $\widehat{\Theta}_{\delta,n}$ need not be closed, as $\Theta_{\delta}(\hat{\rho}_n, \hat{P}_n)$ is defined by a strict inequality, so we keep the supremum over it. For any $\theta \in \Theta$ and nonempty compact $B, B' \subseteq \mathcal{F}(\theta) \cap \mathcal{A}$, choosing for each $\xi \in B$ a point $\xi' \in B'$ with $\|\xi - \xi'\| \leq d_H(B, B')$ and conversely gives

$$\left| \min_{\xi \in B} \Phi(\theta, \xi) - \min_{\xi \in B'} \Phi(\theta, \xi) \right| \leq \omega(d_H(B, B')). \quad (88)$$

The convergence (87) rests on two facts.

First fact. We show that $d_H(\widehat{\Theta}_{\delta,n}, \bar{\Theta}_{\delta}(\rho_0, P_*)) \xrightarrow{\mathbb{P}} 0$. We show the two inclusions $\bar{\Theta}_{\delta}(\rho_0, P_*) \subseteq \widehat{\Theta}_{\delta,n}$ and $\widehat{\Theta}_{\delta,n} \subseteq \mathcal{U}_{\varepsilon} := \{\theta : d(\theta, \bar{\Theta}_{\delta}(\rho_0, P_*)) < \varepsilon\}$, for arbitrary $\varepsilon > 0$, each with probability approaching one, which bound the two directed distances by 0 and by ε , respectively.

(i) $\bar{\Theta}_\delta(\rho_0, P_*) \subseteq \widehat{\bar{\Theta}}_{\delta,n}$. By Lemma SA-27(i), $\bar{\Theta}_\delta(\rho_0, P_*) \subseteq \Theta_\delta(\widehat{\rho}_n, \widehat{P}_n)$ on the event $\{\|\widehat{\rho}_n, \widehat{P}_n - (\rho_0, P_*)\|_{\mathcal{H}} < r\}$, whose probability tends to one, and on that event, for every $\theta^* \in \bar{\Theta}_\delta(\rho_0, P_*) \subseteq \mathcal{N}$,

$$\bar{K}_\delta(\theta^*; \widehat{\rho}_n, \widehat{P}_n) - \widehat{\mathbf{b}}_n(\delta) = \{\bar{K}_\delta(\theta^*; \widehat{\rho}_n, \widehat{P}_n) - \bar{K}_\delta(\theta^*; \rho_0, P_*)\} + \{\bar{\mathbf{b}}(\delta, \rho_0, P_*) - \widehat{\mathbf{b}}_n(\delta)\} \geq -n^{-1/2}R_n,$$

where we used $\bar{K}_\delta(\theta^*; \rho_0, P_*) = \bar{\mathbf{b}}(\delta, \rho_0, P_*)$ and set $R_n := \sqrt{n}(M\|\widehat{\rho}_n, \widehat{P}_n - (\rho_0, P_*)\|_{\mathcal{H}} + |\widehat{\mathbf{b}}_n(\delta) - \bar{\mathbf{b}}(\delta, \rho_0, P_*)|)$. The first brace is bounded by the Lipschitz estimate of Lemma SA-27(v). Now $R_n = O_{\mathbb{P}}(1)$, by the root- n rates of Assumption SA-4(iv) and Lemma SA-26 for the first term and by the weak limit of Theorem SA-29 for the second. Since $\widehat{\nu}_n^{-1} = O_{\mathbb{P}}(1)$ and $\sqrt{\log n} \rightarrow \infty$, we have $\widehat{\nu}_n \sqrt{\log n} \rightarrow \infty$ in probability, so that $\mathbb{P}(\widehat{\nu}_n \sqrt{\log n} \geq R_n) \rightarrow 1$ by the tightness of R_n . On the intersection of these events every $\theta^* \in \bar{\Theta}_\delta(\rho_0, P_*)$ satisfies the inequality $\bar{K}_\delta(\theta^*; \widehat{\rho}_n, \widehat{P}_n) \geq \widehat{\mathbf{b}}_n(\delta) - \widehat{\nu}_n \sqrt{\log n/n}$ defining $\widehat{\bar{\Theta}}_{\delta,n}$, which proves (i).

(ii) $\widehat{\bar{\Theta}}_{\delta,n} \subseteq \mathcal{U}_\varepsilon$. This is the argument that gave (86), with $\mathcal{U}_\varepsilon$ in place of $\mathcal{U}$: along any almost-sure subsequence, Lemma SA-27(vi) places $\widehat{\bar{\Theta}}_{\delta,n}$ inside $\mathcal{U}_\varepsilon$ for all n large, and hence $\mathbb{P}(\widehat{\bar{\Theta}}_{\delta,n} \subseteq \mathcal{U}_\varepsilon) \rightarrow 1$. In particular, along any almost-sure subsequence, $\sup_{\theta \in \widehat{\bar{\Theta}}_{\delta,n}} d(\theta, \bar{\Theta}_\delta(\rho_0, P_*)) \rightarrow 0$ almost surely, as ε is arbitrary.

Second fact. For every $\theta \in \bar{\Theta}_\delta(\rho_0, P_*)$ and every sequence $\theta_n \rightarrow \theta$, along any almost-sure subsequence,

$$d_{\mathcal{H}}(\bar{\Xi}_\delta(\theta_n; \widehat{\rho}_n, \widehat{P}_n), \bar{\Xi}_\delta(\theta; \rho_0, P_*)) \rightarrow 0.$$

Both sets are nonempty and compact for n large, by Lemma SA-27(ii). The directed distance from $\bar{\Xi}_\delta(\theta_n; \widehat{\rho}_n, \widehat{P}_n)$ to $\bar{\Xi}_\delta(\theta; \rho_0, P_*)$ vanishes by the joint upper hemicontinuity of Lemma SA-27(iii). To see this, were it bounded below by some $\varepsilon > 0$ along a subsequence, selections ξ_n with $d(\xi_n, \bar{\Xi}_\delta(\theta; \rho_0, P_*)) \geq \varepsilon$ would admit a further subsequence converging to a point of $\bar{\Xi}_\delta(\theta; \rho_0, P_*)$, a contradiction. The directed distance from $\bar{\Xi}_\delta(\theta; \rho_0, P_*)$ to $\bar{\Xi}_\delta(\theta_n; \widehat{\rho}_n, \widehat{P}_n)$ vanishes by Assumption SA-5, which gives $d(\xi^*, \bar{\Xi}_\delta(\theta_n; \widehat{\rho}_n, \widehat{P}_n)) \rightarrow 0$ for each $\xi^* \in \bar{\Xi}_\delta(\theta; \rho_0, P_*)$. As these maps are 1-Lipschitz in ξ^* , the pointwise convergence is uniform on the compact set $\bar{\Xi}_\delta(\theta; \rho_0, P_*)$. Applied with $(\widehat{\rho}_n, \widehat{P}_n)$ replaced by the constant sequence (ρ_0, P_*) , the same argument shows that $\theta' \mapsto \bar{\Xi}_\delta(\theta'; \rho_0, P_*)$ is Hausdorff continuous at θ .

Conclusion. On the event in (86), using $|\sup_S f - \sup_S f'| \leq \sup_S |f - f'|$ for bounded f, f' ,

$$|\widehat{d\mathbf{b}}_n[h, \gamma] - d\bar{\mathbf{b}}_{(\rho_0, P_*)}[h, \gamma]| \leq \underbrace{\sup_{\theta \in \widehat{\bar{\Theta}}_{\delta,n}} |\widehat{v}_n(\theta) - v(\theta)|}_{(a)} + \underbrace{\sup_{\theta \in \widehat{\bar{\Theta}}_{\delta,n}} v(\theta) - \max_{\theta \in \bar{\Theta}_\delta(\rho_0, P_*)} v(\theta)}_{(b)},$$

and it suffices to show that (a) $\xrightarrow{\mathbb{P}} 0$ and (b) $\xrightarrow{\mathbb{P}} 0$. For (a), (88) gives (a) $\leq \omega(S_n)$, where

$$S_n := \sup_{\theta \in \widehat{\bar{\Theta}}_{\delta,n}} d_{\mathcal{H}}(\bar{\Xi}_\delta(\theta; \widehat{\rho}_n, \widehat{P}_n), \bar{\Xi}_\delta(\theta; \rho_0, P_*)),$$

so it suffices that $S_n \xrightarrow{\mathbb{P}} 0$, which we verify along an arbitrary almost-sure subsequence. Were $S_n \geq \varepsilon$ for infinitely many n on an event of positive probability, we could select $\theta_n \in \widehat{\bar{\Theta}}_{\delta,n}$ at which the Hausdorff distance in the definition of S_n is at least $\varepsilon/2$ and, by part (ii) of the first fact and the compactness of Θ , a further subsequence with $\theta_n \rightarrow \theta \in \bar{\Theta}_\delta(\rho_0, P_*)$. The triangle inequality

$$d_{\mathcal{H}}(\bar{\Xi}_\delta(\theta_n; \widehat{\rho}_n, \widehat{P}_n), \bar{\Xi}_\delta(\theta_n; \rho_0, P_*)) \leq d_{\mathcal{H}}(\bar{\Xi}_\delta(\theta_n; \widehat{\rho}_n, \widehat{P}_n), \bar{\Xi}_\delta(\theta; \rho_0, P_*)) + d_{\mathcal{H}}(\bar{\Xi}_\delta(\theta; \rho_0, P_*), \bar{\Xi}_\delta(\theta_n; \rho_0, P_*))$$

and the second fact, applied to each term on the right, then give a contradiction. Hence (a) $\xrightarrow{\mathbb{P}} 0$. For (b), fix $\varepsilon > 0$. By Lemma SA-27(iv), v is continuous at every $\theta^* \in \bar{\Theta}_\delta(\rho_0, P_*)$, so each θ^* has a radius $s_{\theta^*} > 0$ with $|v(\theta) - v(\theta^*)| < \varepsilon$ whenever $\|\theta - \theta^*\| < s_{\theta^*}$. Extracting a finite subcover of $\bar{\Theta}_\delta(\rho_0, P_*)$ by the balls of half these radii and letting $\varsigma > 0$ be the smallest half radius, every θ with $d(\theta, \bar{\Theta}_\delta(\rho_0, P_*)) < \varsigma$ lies within s_{θ_k} of some center θ_k and so satisfies $v(\theta) < v(\theta_k) + \varepsilon \leq \max_{\bar{\Theta}_\delta(\rho_0, P_*)} v + \varepsilon$. On the event $\{\bar{\Theta}_\delta(\rho_0, P_*) \subseteq \widehat{\bar{\Theta}}_{\delta,n} \subseteq \mathcal{U}_\varsigma\}$, whose probability tends to one by the first fact, we therefore have $0 \leq (b) \leq \varepsilon$. As ε is arbitrary, (b) $\xrightarrow{\mathbb{P}} 0$, which proves (87) and hence (FS4).

With (FS1)–(FS4) verified, Fang and Santos (2019, Theorem 3.2) delivers the conditional weak convergence of $\widetilde{d\mathbf{b}}_n$, evaluated at the bootstrap directions, to $d\bar{\mathbf{b}}_{(\rho_0, P_*)}[\mathbb{Z}, \mathbb{G}]$. The event on which $\widetilde{d\mathbf{b}}_n = \widehat{d\mathbf{b}}_n$ depends on the data alone and has probability approaching one. The same conditional weak convergence therefore holds for the bootstrap statistic of Algorithm SA-1.

Coverage. Coverage follows by inverting the conditional convergence in distribution just established. Write

$$F^{\text{up}}(t) := \mathbb{P}(d\bar{\mathbf{b}}_{(\rho_0, P_*)}[\mathbb{Z}, \mathbb{G}] \leq t), \quad F^{\text{low}}(t) := \mathbb{P}(d\mathbf{b}_{(\rho_0, P_*)}[\mathbb{Z}, \mathbb{G}] \leq t)$$

for the limiting distribution functions and let $q^{\text{up}}(\tau) := \inf\{t \in \mathbb{R} : F^{\text{up}}(t) \geq \tau\}$ and $q^{\text{low}}(\tau) := \inf\{t \in \mathbb{R} : F^{\text{low}}(t) \geq \tau\}$ denote the associated quantile functions. Let $\widehat{\tau}_\tau^{\text{up}}$ and $\widehat{\tau}_\tau^{\text{low}}$ denote the τ -quantiles of the conditional laws of the two bootstrap statistics, which the empirical quantiles over the B bootstrap draws of Algorithm SA-1 approximate. Fix $\tau \in \{\alpha/2, \alpha\}$ and $\tau' \in \{1 - \alpha, 1 - \alpha/2\}$. By hypothesis, F^{up} is continuous and strictly increasing at $q^{\text{up}}(\tau)$,

so that q^{up} is continuous at τ , and likewise q^{low} is continuous at τ' . The conditional weak convergence holds in the bounded-Lipschitz metric in outer probability, so that every subsequence admits a further subsequence along which the conditional laws converge weakly almost surely. Along such a subsequence Lemma 21.2 in van der Vaart (1998), which gives the convergence of quantiles at every continuity point of the limiting quantile function, yields $\hat{c}_\tau^{\text{up}} \rightarrow q^{\text{up}}(\tau)$ and $\hat{c}_{\tau'}^{\text{low}} \rightarrow q^{\text{low}}(\tau')$ almost surely. As the initial subsequence was arbitrary, this gives

$$\hat{c}_\tau^{\text{up}} \xrightarrow{\mathbb{P}} q^{\text{up}}(\tau), \quad \hat{c}_{\tau'}^{\text{low}} \xrightarrow{\mathbb{P}} q^{\text{low}}(\tau'). \quad (89)$$

Finally, combining (89) with the marginal limit $\sqrt{n}(\hat{\mathbf{b}}_n(\delta) - \bar{\mathbf{b}}(\delta, \rho_0, P_*)) \rightsquigarrow d\bar{\mathbf{b}}_{(\rho_0, P_*)}[\mathbb{Z}, \mathbb{G}]$ of Theorem SA-29, the continuity of F^{up} at $q^{\text{up}}(\alpha)$, and Slutsky's lemma, we obtain

$$\liminf_{n \rightarrow \infty} \mathbb{P}\left(\bar{\mathbf{b}}(\delta, \rho_0, P_*) \in (-\infty, \hat{\mathbf{b}}_n(\delta) - \hat{c}_\alpha^{\text{up}}/\sqrt{n}]\right) = 1 - F^{\text{up}}(q^{\text{up}}(\alpha)^-) \geq 1 - \alpha.$$

The analogous identities for the lower one-sided and two-sided sets complete the proof. $\blacksquare$

SA.8.28 Proof of Proposition SA-31

Proof. We prove the results for $\bar{\tau}$. Those for $\underline{\tau}$ follow by substituting $-g$ for g , which leaves the integrability condition $\mathbb{E}_{P_*}[|g(Z)|] < \infty$ unaffected. Write $b := (1 - \kappa)\delta$ and $\bar{v}(b) := \bar{\tau}(\delta)$, so that by (34)

$$\bar{v}(b) = \sup\{\mathbb{E}_P[g(Z)] : P \in \mathcal{P}(\mathcal{Z}), \kappa P \leq P_*, \text{TV}(P \| P_*) \leq b\}, \quad b \in \mathbb{R}_+. \quad (90)$$

The reparametrization is affine with slope $1 - \kappa$, and the dual problem (36) depends on the budget only through $\eta(1 - \kappa)\delta = \eta b$, so that $\bar{\Xi}_\delta$ is its solution set at b and $\partial_\pm \bar{\tau}(\delta) = (1 - \kappa) \partial_\pm \bar{v}(b)$. It therefore suffices to prove the claims for $\bar{v}$ on $(0, \infty)$, with η in place of $(1 - \kappa)\eta$ and $[1 - \kappa, \infty)$ in place of $[1, \infty)$. Let

$$f_b(\xi) := \zeta + \eta b + \mathbb{E}_{P_*}[\text{r}_{\text{TV}}(g(Z) - \zeta, \eta)], \quad \xi = (\zeta, \eta) \in \mathcal{F} := \mathbb{R} \times \mathbb{R}_+,$$

be the dual objective at budget b , so that

$$f_b(\xi) = f_{b_0}(\xi) + \eta(b - b_0) \quad \text{for all } \xi \in \mathcal{F} \text{ and } b, b_0 \in \mathbb{R}_+. \quad (91)$$

By properties (r2) and (r3) of the integrand (23) with $\mathbb{U} = [0, 1/\kappa]$, r_{TV} is jointly convex with $|\text{r}_{\text{TV}}(x, \eta)| \leq \kappa^{-1}|x| + (2\kappa)^{-1}\eta$ and $|\text{r}_{\text{TV}}(x, \eta) - \text{r}_{\text{TV}}(x', \eta')| \leq \kappa^{-1}|x - x'| + (2\kappa)^{-1}|\eta - \eta'|$, so that, as $\mathbb{E}_{P_*}[|g(Z)|] < \infty$, f_b is finite, convex, and Lipschitz continuous on the closed convex set $\mathcal{F}$.

Step 1: regularity of $\bar{v}$. Fix $b \in \mathbb{R}_+$. The baseline is feasible in (90), so $\bar{v}(b) \geq \mathbb{E}_{P_*}[g(Z)] > -\infty$, and every feasible P satisfies $P \ll P_*$ with density $w \leq 1/\kappa$ by Lemma SA-12(i), so $\mathbb{E}_P[g(Z)] \leq \kappa^{-1} \mathbb{E}_{P_*}[|g(Z)|] < \infty$. Monotonicity holds because the feasible set grows with b , and $\bar{v}$ is constant on $[1 - \kappa, \infty)$ because $\kappa P \leq P_*$ already implies $\text{TV}(P \| P_*) \leq 1 - \kappa$ (Lemma SA-17). For concavity, let P_1 and P_2 be feasible at b_1 and b_2 , let $t \in (0, 1)$, and set $P_t := tP_1 + (1 - t)P_2$. Then $\kappa P_t \leq P_*$ and $\text{TV}(P_t \| P_*) \leq t \text{TV}(P_1 \| P_*) + (1 - t) \text{TV}(P_2 \| P_*) \leq tb_1 + (1 - t)b_2$, by the convexity of $P \mapsto \text{TV}(P \| P_*) = \sup_{A \in \mathcal{F}} |P(A) - P_*(A)|$ (Lemma SA-2), a supremum of convex functions. Hence $\bar{v}(tb_1 + (1 - t)b_2) \geq \mathbb{E}_{P_t}[g(Z)] = t \mathbb{E}_{P_1}[g(Z)] + (1 - t) \mathbb{E}_{P_2}[g(Z)]$, and taking suprema over P_1 and P_2 gives concavity on $\mathbb{R}_+$. Consequently, $\bar{v}$ is continuous on $(0, \infty)$, and its one-sided derivatives exist and are finite at every $b > 0$ and are nonincreasing in b , with $\partial_+ \bar{v}(b) \leq \partial_- \bar{v}(b)$ (Rockafellar, 1970, Theorem 10.1 and Section 24). The map $\partial_+ \bar{v}$ is nonnegative by monotonicity, and, as a concave function on an open interval, $\bar{v}$ is differentiable at all but countably many $b > 0$. Finally, for continuity at 0, a feasible P at b has $|w - 1| \leq 1/\kappa$ and $\mathbb{E}_{P_*}[|w - 1|] = 2 \text{TV}(P \| P_*) \leq 2b$ (Lemma SA-2), so that for every $M > 0$

$$|\mathbb{E}_P[g(Z)] - \mathbb{E}_{P_*}[g(Z)]| \leq \mathbb{E}_{P_*}[|w - 1| |g(Z)|] \leq 2bM + \kappa^{-1} \mathbb{E}_{P_*}[|g(Z)| \mathbf{1}\{|g(Z)| > M\}].$$

Taking the supremum over feasible P , letting $b \downarrow 0$, and then $M \rightarrow \infty$ by dominated convergence gives $\bar{v}(b) \rightarrow \bar{v}(0)$. This proves part SA-31(i).

Step 2: strong duality and dual attainment. Fix $b > 0$. We verify Assumption SA-2 for the entropy $\bar{\phi}_{\text{TV}}$ at budget b with $\mathbf{d}_m = 0$. As 0 and $1/\kappa$ lie in the domain of ϕ_{TV} , Remark SA-6 shows that $\bar{\phi}_{\text{TV}}$ has infinite recession and falls under case (b) of Assumption SA-2(i) with $\bar{x} = 1/\kappa$, and that the integrability condition in Assumption SA-2(iv), which is the one of Theorem SA-22(ii), follows from $\mathbb{E}_{P_*}[|g(Z)|] < \infty$. Since $(\bar{\phi}_{\text{TV}}^*)^\pi = \text{r}_{\text{TV}}$ by Lemma SA-23, the point $(\zeta^*, \eta^*) = (0, 1)$ satisfies Assumption SA-2(ii) by the bound on r_{TV} recorded above. Assumption SA-2(iii) holds with $P^\circ = P_*$, as $D_{\bar{\phi}_{\text{TV}}}^-(P_* \| P_*) = 0 < b$. Assumption SA-2(v) concerns case (a) only, and Assumption SA-2(vi) is vacuous because $\mathbf{d}_m = 0$. Hence Theorem SA-19 shows that $\bar{\Xi}_\delta$, the set of minimizers of f_b over $\mathcal{F}$, is nonempty, compact, and convex, where the constraint in (21) is vacuous since $\bar{\phi}_{\text{TV}}^\infty(1) = +\infty$, and Lemma SA-18(iii) shows that the recession function of f_b is strictly positive on every nonzero recession direction of $\mathcal{F}$. Moreover, Theorem SA-14(i) applies with $\phi = \phi_{\text{TV}}$ and budget b , its compatibility condition holding under the conventions $\mathcal{M}(\rho) = \Gamma_{\bar{\phi}_{\text{TV}}}^-(b) = \{0\}$ of Section SA1.2, and its robustness set is the feasible set of (90) by (15) and Lemma SA-17(i). Hence

$$\bar{v}(b) = \inf_{\xi \in \mathcal{F}} f_b(\xi) = f_b(\xi^*) \quad \text{for every } \xi^* \in \bar{\Xi}_\delta. \quad (92)$$

Step 3: optimal multipliers are supergradients. Fix $b_0 > 0$ and $\xi^* = (\zeta^*, \eta^*) \in \bar{\Xi}_{b_0}$, indexing solution sets by the budget b . For every $b > 0$, $\bar{v}(b) \leq f_b(\xi^*) = f_{b_0}(\xi^*) + \eta^*(b - b_0) = \bar{v}(b_0) + \eta^*(b - b_0)$, by (92), (91), and the

optimality of ξ^* at b_0 . Hence η^* is a supergradient of $\bar{v}$ at b_0 , and

$$\partial_+ \bar{v}(b_0) \leq \bar{\eta}_{b_0}^{\min} \leq \bar{\eta}_{b_0}^{\max} \leq \partial_- \bar{v}(b_0). \quad (93)$$

Step 4: compactness of nearby solutions. Fix $0 < b_1 \leq b_0 \leq b_2 < \infty$ and $b_a \in (0, b_1)$. For $b \in [b_1, b_2]$ and $\xi^* \in \Xi_b$ with multiplier η^* , (93) and concavity give $\eta^* \leq \partial_- \bar{v}(b) \leq \partial_+ \bar{v}(b_a) =: K < \infty$, and then (91), optimality at b , and monotonicity give $f_{b_0}(\xi^*) = \bar{v}(b) + \eta^*(b_0 - b) \leq \bar{v}(b_2) + K(b_2 - b_1) =: c$. Hence $\bigcup_{b \in [b_1, b_2]} \Xi_b \subseteq L_c := \{\xi \in \mathcal{F} : f_{b_0}(\xi) \leq c\}$, a nonempty closed convex set, since f_{b_0} is continuous and convex on the closed convex set $\mathcal{F}$. Its recession cone satisfies $0^+(L_c) \subseteq 0^+(\mathcal{F}) \cap \{d : f_{b_0}^\infty(d) \leq 0\} = \{0\}$, by Corollary 8.3.3 and Theorem 8.7 in Rockafellar (1970) and Step 2, so that L_c is compact (Rockafellar, 1970, Theorem 8.4).

Step 5: the one-sided derivatives. Fix $b_0 > 0$, let $b_k \downarrow b_0$ with $b_k \leq b_2$ for some $b_2 > b_0$, and pick $\xi_k^* = (\zeta_k, \eta_k) \in \Xi_{b_k}$, which is possible by Step 2. Step 3 applied at b_0 and evaluated at b_k , and applied at b_k and evaluated at b_0 , gives

$$\eta_k \leq \frac{\bar{v}(b_k) - \bar{v}(b_0)}{b_k - b_0} \leq \bar{\eta}_{b_0}^{\min},$$

and the difference quotients converge to $\partial_+ \bar{v}(b_0)$ by Step 1. By Step 4, along a subsequence $\xi_k^* \rightarrow \xi^\circ = (\zeta^\circ, \eta^\circ)$, and $\xi^\circ \in \mathcal{F}$ since $\mathcal{F}$ is closed. Moreover, $f_{b_0}(\xi^\circ) = \lim_k f_{b_0}(\xi_k^*) = \lim_k \{\bar{v}(b_k) + \eta_k(b_0 - b_k)\} = \bar{v}(b_0)$, by the continuity of f_{b_0} , by (91) with the optimality of ξ_k^* at b_k , and by the continuity of $\bar{v}$ together with the boundedness of (η_k) . Hence $\xi^\circ \in \Xi_{b_0}$ and $\eta^\circ \geq \bar{\eta}_{b_0}^{\min}$, so that along the subsequence $\bar{\eta}_{b_0}^{\min} \geq \partial_+ \bar{v}(b_0) \geq \lim_k \eta_k = \eta^\circ \geq \bar{\eta}_{b_0}^{\min}$, that is, $\partial_+ \bar{v}(b_0) = \bar{\eta}_{b_0}^{\min}$. The symmetric argument with $b_k \uparrow b_0$ gives $\partial_- \bar{v}(b_0) = \bar{\eta}_{b_0}^{\max}$. Translating back through $\partial_\pm \bar{\tau}(\delta) = (1 - \kappa) \partial_\pm \bar{v}(b)$ proves part SA-31(ii).

Step 6: parts SA-31(iii) and SA-31(iv). For $0 < b_1 < b_2$, Step 5 and the monotonicity of the difference quotients of the concave $\bar{v}$ give $\bar{\eta}_{b_1}^{\min} = \partial_+ \bar{v}(b_1) \geq \{\bar{v}(b_2) - \bar{v}(b_1)\}/(b_2 - b_1) \geq \partial_- \bar{v}(b_2) = \bar{\eta}_{b_2}^{\max}$, which is part SA-31(iii), as $\delta \mapsto b$ is increasing. By Step 5, $\bar{v}$ is differentiable at b if and only if $\bar{\eta}_b^{\min} = \bar{\eta}_b^{\max}$, and since $\{\eta : (\zeta, \eta) \in \Xi_b\}$ is the compact interval $[\bar{\eta}_b^{\min}, \bar{\eta}_b^{\max}]$, as the image of the compact convex set Ξ_b under a linear map, this is the uniqueness of the optimal budget multiplier. The same holds for $\bar{\tau}$ at $\delta = b/(1 - \kappa)$, with $\bar{\tau}'(\delta) = (1 - \kappa)\bar{v}'(b)$. Nonnegativity of $\bar{\eta}$ holds because $\eta \geq 0$ on $\mathcal{F}$, and its monotonicity follows from part SA-31(iii). ■

SA8.29 Proof of Proposition SA-32

Proof. Write $\hat{\delta} := \hat{\delta}_{\text{cb}}(\alpha)$, $\mathbf{c} := \mathbf{c}_{1-\alpha/2}$, $\|\cdot\|_\infty$ for the supremum norm on $[0, 1]$, $B := \bar{\tau} - \tau_*$, $G_n := \sqrt{n}(\hat{\tau}_n - \bar{\tau})$, and $c_n := \hat{\tau}_n - n^{-1/2}\mathbf{c}\hat{\sigma}_n = \bar{\tau} + n^{-1/2}(G_n - \mathbf{c}\hat{\sigma}_n)$, so that $\hat{\delta} = \inf\{\delta \in [0, 1] : c_n(\delta) \leq \tau_*\}$. By Assumption SA-6(i), B is continuous and nonincreasing, $B(0) > 0$, and $\delta_b < 1$. Moreover, $\bar{\tau}(\delta) \geq \bar{\tau}(0) \geq \bar{\tau}(0) > \tau_*$ for every $\delta \in [0, 1]$ by Proposition SA-31(i), so that the upper inequality in (37) holds throughout, $\delta_b = \inf\{\delta \in [0, 1] : B(\delta) \leq 0\}$, and $B > 0$ on $[0, \delta_b)$. Hence the closed nonempty set $\{\delta \in [0, 1] : B(\delta) \leq 0\}$ contains its infimum δ_b , whence $B(\delta_b) = 0$ by continuity from the left, and differentiability at δ_b gives

$$B(\delta) = -\eta(\delta_b)(\delta - \delta_b) + R(\delta), \quad R(\delta) = o(|\delta - \delta_b|). \quad (94)$$

Since $\eta(\delta_b) > 0$, there exists $\varepsilon_0 > 0$ with $[\delta_b - \varepsilon_0, \delta_b + \varepsilon_0] \subseteq (0, 1)$ and $B(\delta) \leq -\{\eta(\delta_b)/2\}(\delta - \delta_b) < 0$ on $(\delta_b, \delta_b + \varepsilon_0]$.

Step 1: consistency. By Assumptions SA-6(ii)–SA-6(iii) and the continuous mapping theorem, $\|G_n\|_\infty = O_{\mathbb{P}}(1)$ and $\|\hat{\sigma}_n\|_\infty = O_{\mathbb{P}}(1)$, so that $\|c_n - \bar{\tau}\|_\infty = O_{\mathbb{P}}(n^{-1/2})$. Fix $\varepsilon \in (0, \varepsilon_0]$ and set $\gamma_\varepsilon := \min[B(\delta_b - \varepsilon), \{\eta(\delta_b)/2\}\varepsilon] > 0$. On the event $\{\|c_n - \bar{\tau}\|_\infty < \gamma_\varepsilon\}$, whose probability tends to one, $c_n - \tau_* > B - \gamma_\varepsilon \geq B(\delta_b - \varepsilon) - \gamma_\varepsilon \geq 0$ on $[0, \delta_b - \varepsilon]$, since B is nonincreasing, whereas $c_n(\delta_b + \varepsilon) - \tau_* < B(\delta_b + \varepsilon) + \gamma_\varepsilon \leq 0$. Hence $|\hat{\delta} - \delta_b| \leq \varepsilon$ on that event, which proves part SA-32(i).

Step 2: the crossing identity. On the event $\{|\hat{\delta} - \delta_b| \leq \varepsilon_0/2\}$, whose probability tends to one by Step 1, $0 < \hat{\delta} < 1$. The set $\{\delta \in [0, 1] : c_n(\delta) \leq \tau_*\}$ is nonempty and closed, because c_n has continuous paths by Assumptions SA-6(ii)–SA-6(iii), so it contains $\hat{\delta}$, whereas $c_n > \tau_*$ on $[0, \hat{\delta})$ and continuity give $c_n(\hat{\delta}) \geq \tau_*$. Hence $c_n(\hat{\delta}) = \tau_*$ with probability approaching one, that is,

$$B(\hat{\delta}) = n^{-1/2}A_n, \quad A_n := \mathbf{c}\hat{\sigma}_n(\hat{\delta}) - G_n(\hat{\delta}). \quad (95)$$

Moreover, $A_n \rightsquigarrow \mathbf{c}\sigma(\delta_b) - \mathbb{G}(\delta_b)$. Indeed, $\hat{\sigma}_n(\hat{\delta}) \xrightarrow{\mathbb{P}} \sigma(\delta_b)$ by Assumption SA-6(iii), the continuity of σ , and Step 1, and $G_n(\hat{\delta}) - G_n(\delta_b) = o_{\mathbb{P}}(1)$ because $(G_n, \hat{\delta}) \rightsquigarrow (\mathbb{G}, \delta_b)$ by Slutsky's lemma, $\mathbb{G}$ has continuous paths, and the extended continuous mapping theorem (van der Vaart and Wellner, 1996, Theorem 1.11.1) applies to $(x, t) \mapsto x(t) - x(\delta_b)$, while $G_n(\delta_b) \rightsquigarrow \mathbb{G}(\delta_b) \sim \mathbf{N}(0, \sigma^2(\delta_b))$.

Step 3: asymptotic distribution and coverage. Fix $\varepsilon \in (0, \eta(\delta_b))$ and $\nu \in (0, \varepsilon_0]$ with $|R(\delta)| \leq \varepsilon|\delta - \delta_b|$ whenever $|\delta - \delta_b| \leq \nu$. On the intersection of $\{|\hat{\delta} - \delta_b| \leq \nu\}$ with the event of Step 2, whose probability tends to one, (94) and (95) give $\{\eta(\delta_b) - \varepsilon\}|\hat{\delta} - \delta_b| \leq n^{-1/2}|A_n|$, so that $\sqrt{n}(\hat{\delta} - \delta_b) = O_{\mathbb{P}}(1)$ by Step 2, and hence $\sqrt{n}R(\hat{\delta}) = o_{\mathbb{P}}(1)$. Substituting (94) into (95),

$$\sqrt{n}(\hat{\delta} - \delta_b) = \frac{\sqrt{n}R(\hat{\delta}) - A_n}{\eta(\delta_b)} \rightsquigarrow \frac{\mathbb{G}(\delta_b) - \mathbf{c}\sigma(\delta_b)}{\eta(\delta_b)} \sim \mathbf{N}(-\mathbf{c}\varsigma_b, \varsigma_b^2),$$

by Step 2, Slutsky's lemma, and (40), which is part SA-32(ii). Since the limit is continuous,

$$\mathbb{P}(\widehat{\delta} \leq \delta_b) = \mathbb{P}(\sqrt{n}(\widehat{\delta} - \delta_b) \leq 0) \rightarrow \mathbb{P}\{\mathbf{N}(-\mathbf{c}_b, \mathbf{c}_b^2) \leq 0\} = \mathbb{P}\{\mathbf{N}(0, 1) \leq \mathbf{c}\} = 1 - \alpha/2,$$

which is part SA-32(iii). ■

SA8.30 Proof of Proposition SA-33

Proof. Write $B := \underline{\tau} - \tau_*$ and $\mathbf{c} := \mathbf{c}_{1-\alpha/2}$. As shown in the proof of Proposition SA-32, B is continuous, $B > 0$ on $[0, \delta_b)$, $B(\delta_b) = 0$, and B is differentiable at δ_b with slope $-\eta(\delta_b) \in (-\infty, 0)$, so that

$$B(\delta) = \underline{\tau}(\delta_b) (\delta_b - \delta) \{1 + o(1)\} \quad \text{as } \delta \uparrow \delta_b. \quad (96)$$

At every $\delta \in (0, \delta_b)$ at which $\underline{\tau}$ is differentiable, $B'(\delta) = -\eta(\delta)$ and σ is differentiable, so $\log n^*(\delta; \alpha) = 2 \log \mathbf{c} + 2 \log \sigma(\delta) - 2 \log B(\delta)$ is differentiable with derivative $2\{\sigma'(\delta)/\sigma(\delta) + \eta(\delta)/B(\delta)\}$, which is (42). For the limit, the derivative of the finite convex function $\underline{\tau}$ is nondecreasing where it exists and converges from the left to $\partial_- \underline{\tau}(\delta_b) = \underline{\tau}'(\delta_b)$ (Rockafellar, 1970, Section 24), so that $\eta(\delta) \rightarrow \eta(\delta_b)$ as $\delta \uparrow \delta_b$ and, by (96), $\eta(\delta)/B(\delta) = \{1 + o(1)\}/(\delta_b - \delta)$. Since σ is continuously differentiable and strictly positive on $[0, 1]$, σ'/σ is bounded there, so the second term in (42) is of relative order $O(\delta_b - \delta)$, and the claim follows. ■

SA8.31 Proof of Theorem SA-34

Proof. We derive in full the upper bound $\bar{\mu}^C(1; \delta)$ on the complier mean, which delivers the lower bound $\underline{\tau}(\delta)$ on τ_0 , and indicate at the end the symmetric argument for the remaining bound. Write $\phi_{\text{TV}}(u) = \frac{1}{2}|u - 1|$ for the total variation entropy, so that $D_{\phi_{\text{TV}}}(Q \| P_{\text{T}}) = \text{TV}(Q \| P_{\text{T}})$. Recall from Section SA6.1 that $\mathcal{P}_{\text{RCT}}(\delta)$ is the set $\mathcal{P}_{\text{TV}}(\delta)$ of (34) with $(\kappa, P_*) = (\pi, P_{\text{T}})$, which Proposition SA-13(i) identifies with the neighborhood generated by the induced entropy $\bar{\phi}_{\text{TV}} = \phi_{\text{TV}} + \iota_{[0, 1/\pi]}$ at budget $(1 - \pi)\delta$. Since $\pi P_{\text{T}} \leq P_{\text{T}}$ and $\text{TV}(P_{\text{T}} \| P_{\text{T}}) = 0$, we have $P_{\text{T}} \in \mathcal{P}_{\text{RCT}}(\delta)$, so the set is nonempty.

Dual reduction. We apply Theorem SA-9 to the induced entropy $\bar{\phi}_{\text{TV}}$, which is an entropy function by Proposition SA-13(i), with $g(y) = y$, no moment function, and divergence budget $(1 - \pi)\delta$. Its integrability hypothesis holds. Indeed, if $\nu \in \mathcal{M}_+(\mathcal{Y})$ satisfies $D_{\bar{\phi}_{\text{TV}}}(\nu \| P_{\text{T}}) < \infty$, then Lemma SA-1, together with $\bar{\phi}_{\text{TV}}^\infty(1) = +\infty$ and $\bar{\phi}_{\text{TV}} = +\infty$ off $[0, 1/\pi]$, gives $\nu \ll P_{\text{T}}$ with $d\nu/dP_{\text{T}} \leq 1/\pi$ P_{T} -almost surely, so that $\int |y| d\nu(y) \leq \pi^{-1} \mathbb{E}_{P_{\text{T}}} [|Y|] < \infty$. The pointwise constraint of the theorem is vacuous because $\bar{\phi}_{\text{TV}}^\infty(1) = +\infty$, and the moment multiplier is absent, so that weak duality reads

$$\bar{\mu}^C(1; \delta) \leq \inf_{(\zeta, \eta) \in \mathbb{R} \times \mathbb{R}_+} \left\{ \zeta + \eta(1 - \pi)\delta + \mathbb{E}_{P_{\text{T}}} [(\bar{\phi}_{\text{TV}}^*)^\pi(Y - \zeta, \eta)] \right\}. \quad (97)$$

The conjugate $\bar{\phi}_{\text{TV}}^*$ and its perspective $(\bar{\phi}_{\text{TV}}^*)^\pi$ are recorded in closed form in Example SA-3. In particular, for $\eta \geq 0$ the pointwise supremum defining $(\bar{\phi}_{\text{TV}}^*)^\pi(s, \eta) = \sup_{0 \leq u \leq 1/\pi} \{su - \frac{\eta}{2}|u - 1|\}$ is attained at

$$u^*(s, \eta) = \begin{cases} 0, & s \leq -\eta/2, \\ 1, & -\eta/2 \leq s \leq \eta/2, \\ 1/\pi, & s \geq \eta/2. \end{cases}$$

Since $\delta \leq 1$ we have $(1 - \pi)\delta \leq 1 - \pi\delta$, hence $t_L \leq t_U$. Set

$$\zeta^\dagger := \frac{1}{2}(t_L + t_U), \quad \eta^\dagger := t_U - t_L \geq 0,$$

so that $Y - \zeta^\dagger \leq -\eta^\dagger/2 \Leftrightarrow Y \leq t_L$ and $Y - \zeta^\dagger \geq \eta^\dagger/2 \Leftrightarrow Y \geq t_U$. The corresponding maximizer $u^*(Y - \zeta^\dagger, \eta^\dagger)$ is

$$w^\dagger(y) := \begin{cases} 0, & y \leq t_L, \\ 1, & t_L < y < t_U, \\ 1/\pi, & y \geq t_U. \end{cases}$$

Because P_{T} has no atoms at t_L or t_U ,

$$P_{\text{T}}(Y \leq t_L) = (1 - \pi)\delta, \quad P_{\text{T}}(Y \geq t_U) = \pi\delta,$$

hence

$$\mathbb{E}_{P_{\text{T}}}[w^\dagger(Y)] = (1 - (1 - \pi)\delta - \pi\delta) + \frac{1}{\pi}\pi\delta = 1, \quad \frac{1}{2}\mathbb{E}_{P_{\text{T}}} [|w^\dagger(Y) - 1|] = \frac{1}{2}(1 - \pi)\delta + \frac{1}{2}\left(\frac{1}{\pi} - 1\right)\pi\delta = (1 - \pi)\delta.$$

Thus the measure $P^\dagger$ with $dP^\dagger/dP_{\text{T}} = w^\dagger$ is a probability measure with $\text{TV}(P^\dagger \| P_{\text{T}}) = (1 - \pi)\delta$. Moreover, by Lemma SA-12(i) $0 \leq w^\dagger \leq 1/\pi$ gives $\pi P^\dagger \leq P_{\text{T}}$. Hence, $P^\dagger \in \mathcal{P}_{\text{RCT}}(\delta)$ and, by primal feasibility,

$$\bar{\mu}^C(1; \delta) \geq \mathbb{E}_{P^\dagger}[Y] = \mathbb{E}_{P_{\text{T}}}[Y w^\dagger]. \quad (98)$$

By the rescaling identity of Lemma SA-17(i), the mixing distribution $R^\dagger = (P_{\text{T}} - \pi P^\dagger)/(1 - \pi)$ satisfies $\text{TV}(P^\dagger \| R^\dagger) = \delta$: the attaining decomposition places the complier and always observed outcome distributions at total variation distance exactly δ , saturating the sensitivity restriction.

Since $w^\dagger$ is a pointwise maximizer, Example SA-3 gives, P_T -almost surely,

$$(\bar{\phi}_{TV}^*)^\pi(Y - \zeta^\dagger, \eta^\dagger) = (Y - \zeta^\dagger)w^\dagger - \frac{\eta^\dagger}{2}|w^\dagger - 1|.$$

Evaluating the objective in (97) at $(\zeta^\dagger, \eta^\dagger)$ and using $\mathbb{E}_{P_T}[w^\dagger] = 1$ together with $\frac{1}{2} \mathbb{E}_{P_T}[|w^\dagger - 1|] = (1 - \pi)\delta$,

$$\begin{aligned} & \zeta^\dagger + \eta^\dagger(1 - \pi)\delta + \mathbb{E}_{P_T}[(\bar{\phi}_{TV}^*)^\pi(Y - \zeta^\dagger, \eta^\dagger)] \\ &= \mathbb{E}_{P_T}[Y w^\dagger] + \zeta^\dagger(1 - \mathbb{E}_{P_T}[w^\dagger]) + \eta^\dagger\left((1 - \pi)\delta - \frac{1}{2} \mathbb{E}_{P_T}[|w^\dagger - 1|]\right) \\ &= \mathbb{E}_{P_T}[Y w^\dagger]. \end{aligned}$$

Therefore the infimum in (97) is at most $\mathbb{E}_{P_T}[Y w^\dagger]$. Combining this with (98) and the weak duality inequality (97) gives

$$\bar{\mu}^C(1; \delta) = \mathbb{E}_{P_T}[Y w^\dagger] = \mathbb{E}_{P_T}[Y \mathbf{1}\{t_L < Y < t_U\}] + \frac{1}{\pi} \mathbb{E}_{P_T}[Y \mathbf{1}\{Y \geq t_U\}],$$

where the last equality is the definition of $w^\dagger$. This is the sharp upper bound on the complier mean.

From the complier mean to τ_0 . By the mixture (44), the definition of the always observed bounds in (46), and $\mu_T = \mathbb{E}_{P_T}[Y]$,

$$\underline{\mu}^{AO}(1; \delta) = \frac{\mathbb{E}_{P_T}[Y] - \pi \bar{\mu}^C(1; \delta)}{1 - \pi}.$$

Splitting $\mathbb{E}_{P_T}[Y] = \mathbb{E}_{P_T}[Y \mathbf{1}\{Y \leq t_L\}] + \mathbb{E}_{P_T}[Y \mathbf{1}\{t_L < Y < t_U\}] + \mathbb{E}_{P_T}[Y \mathbf{1}\{Y \geq t_U\}]$ and inserting the closed form for $\bar{\mu}^C(1; \delta)$ just obtained, the upper-tail terms cancel and

$$\mathbb{E}_{P_T}[Y] - \pi \bar{\mu}^C(1; \delta) = \mathbb{E}_{P_T}[Y \mathbf{1}\{Y \leq t_L\}] + (1 - \pi) \mathbb{E}_{P_T}[Y \mathbf{1}\{t_L < Y < t_U\}].$$

Dividing by $1 - \pi$ gives the closed form for $\underline{\mu}^{AO}(1; \delta)$. Under **Mono** a control unit is observed only if $S(0) = 1$, in which case $S(1) \geq S(0)$ forces $S(1) = 1$, so that the observed control units are exactly the always observed ones and $\mu^{AO}(0) = \mathbb{E}_{P_0}[Y \mid D = 0, S = 1]$ does not depend on the distribution of the complier outcome. Subtracting it as in (47) yields the stated closed form for $\underline{\tau}(\delta)$. The closed form for $\bar{\tau}(\delta)$ follows from the symmetric argument, upon replacing $w^\dagger$ with the reweighting equal to $1/\pi$ on $\{Y \leq s_L\}$, to 1 on (s_L, s_U) , and to 0 on $\{Y \geq s_U\}$, which is feasible and attains $\underline{\mu}^C(1; \delta)$ by the same computation.

The bounds are attained and the value set is an interval. Both bounds are attained, the reweightings just displayed exhibiting distributions in $\mathcal{P}_{RCT}(\delta)$ at which the complier mean equals $\bar{\mu}^C(1; \delta)$ and $\underline{\mu}^C(1; \delta)$. It remains to show that every intermediate value is attained as well. The set $\mathcal{P}_{RCT}(\delta)$ is convex, as the intersection of the TV ball of radius $(1 - \pi)\delta$ around P_T with the set $\{Q \in \Delta(\mathcal{Y}) : \pi Q \leq P_T\}$, both of which are convex. Moreover, $Q \mapsto \mathbb{E}_Q[Y]$ is affine, hence so is the map (45),

$$Q \mapsto \tau(Q) = \frac{\mu_T - \pi \mathbb{E}_Q[Y]}{1 - \pi} - \mu^{AO}(0),$$

which sends a distribution of the complier outcome to the implied value of τ_0 . The image of a convex set under an affine map into $\mathbb{R}$ is an interval, and its endpoints are attained, so that image is exactly $[\underline{\tau}(\delta), \bar{\tau}(\delta)]$.

Finally, every $Q \in \mathcal{P}_{RCT}(\delta)$ is consistent with the observed distribution and the sensitivity restriction. Indeed, by Lemma SA-17(i), every such Q admits the unique mixing distribution $R = (P_T - \pi Q)/(1 - \pi)$, and $P_T = \pi Q + (1 - \pi)R$ reproduces the mixture (44) with Q in the role of the complier outcome distribution and R in that of the always observed one. By the rescaling identity of the same lemma, $\text{TV}(Q \parallel P_T) = (1 - \pi)\text{TV}(Q \parallel R)$, so the budget $\text{TV}(Q \parallel P_T) \leq (1 - \pi)\delta$ is equivalent to $\text{TV}(Q \parallel R) \leq \delta$, which is the maintained restriction. Moreover, the pair (Q, R) is rationalized by an actual data-generating process: taking Q as the complier and R as the always observed law of $Y(1)$, the identified law of $Y(0)$ on the always observed, namely that of $Y \mid D = 0, S = 1$, and any coupling of the remaining coordinates, one obtains a joint law of $(Y(0), Y(1), S(0), S(1), D)$ that satisfies random assignment and **Mono** and reproduces the observed law of (YS, S, D) . No feasible Q is therefore excluded by the data. Together with the previous paragraph, the set of values of τ_0 consistent with P_T , π , $\mu^{AO}(0)$, and the restriction is the closed interval $[\underline{\tau}(\delta), \bar{\tau}(\delta)]$. ■

SA8.32 Proof of Proposition SA-35

Proof. We work directly with the closed form bound of Theorem SA-34, and we view the plug-in estimator as a functional of the empirical measure $\hat{P}_n$ of $W_1, \dots, W_n$, where $W = (YS, S, D)$, acting on the class of functions of W

$$\mathcal{H}_{RCT} := \{f_t, g_t : t \in \mathbb{R}\} \cup \{D, 1 - D, DS, (1 - D)S, (1 - D)YS\},$$

where $f_t := DS \mathbf{1}\{YS \leq t\}$ and $g_t := DSY S \mathbf{1}\{YS > t\}$. Each of the families $\{f_t\}$ and $\{g_t\}$ multiplies a fixed function by the indicators of the half lines $\{YS \leq t\}$ or $\{YS > t\}$, so that it is VC-subgraph (van der Vaart and Wellner, 1996, Lemma 2.6.18(vi)). Hence $\mathcal{H}_{RCT}$ has a finite uniform entropy integral (van der Vaart and Wellner, 1996, Theorem 2.6.7), and its envelope $1 + |YS|$ is square integrable by Assumption SA-7. The class is also measurable, thus it is P_0 -Donsker (van der Vaart, 1998, Theorem 19.14), that is, $\sqrt{n}(\hat{P}_n - P_0) \rightsquigarrow \mathbb{G}$ in $\ell^\infty(\mathcal{H}_{RCT})$, where $\mathbb{G}$ is a tight Gaussian process whose sample paths lie in the closed subspace $\mathbb{D}_0$ of the $h \in \ell^\infty(\mathcal{H}_{RCT})$ that are uniformly continuous with respect to the $L_2(P_0)$ semimetric.

For a probability measure P on the range of W with $P(D) \in (0, 1)$, $P(DS) > 0$, and $P((1 - D)S) > 0$, let $F_P(t) := P(f_t)/P(DS)$ and $M_P(t) := P(g_t)/P(DS)$ be the distribution function and the upper-tail moment of the treated observed outcome, so that $F_{P_0} = F_{P_T}$, let $\pi_P := 1 - \{P((1 - D)S)/P(1 - D)\}/\{P(DS)/P(D)\}$ be the complier share, let $c_P := F_P^{-1}(\pi_P \delta)$ and $d_P := F_P^{-1}(1 - (1 - \pi_P)\delta)$ be the trimming points, and let

$$\Phi(P) := M_P(c_P) + \frac{\pi_P}{1 - \pi_P} M_P(d_P) - \frac{P((1 - D)YS)}{P((1 - D)S)}.$$

Write $c := s_L(\delta)$, $d := s_U(\delta)$, and $M := M_{P_0}$. Since $\vartheta_+(y; \delta) = y \mathbb{1}\{y > c\} + \frac{\pi}{1 - \pi} y \mathbb{1}\{y \geq d\}$ and P_T has no atoms, $\Phi(P_0) = \bar{\tau}(\delta)$. Moreover, $\Phi(\hat{P}_n)$ differs from $\hat{\tau}(\delta)$ only through the atom of the empirical distribution of the treated observed outcomes at its own quantile $\hat{s}_U(\delta)$, a difference of order $O_{\mathbb{P}}(n^{-1})$ because these outcomes have no ties almost surely.

The map Φ is Hadamard differentiable at P_0 tangentially to $\mathbb{D}_0$. Indeed, it is a composition of the following maps, each Hadamard differentiable at the relevant point, so that the chain rule (van der Vaart, 1998, Theorem 20.9) applies.

- The evaluations $P \mapsto P(h)$ at the five functions of the finite part of $\mathcal{H}_{RCT}$, and smooth functions of them on the stated range, which deliver π_P , $P(DS)$, equal to $p r_1$ at P_0 , and the control mean.
- The maps $P \mapsto F_P$ and $P \mapsto M_P$ into $\ell^\infty(\mathbb{R})$, whose derivatives at P_0 in a direction h are $h_F := \{h(f_\cdot) - F_{P_T} h(DS)\}/(p r_1)$ and $h_M := \{h(g_\cdot) - M h(DS)\}/(p r_1)$. For $h \in \mathbb{D}_0$, both are continuous, because for $t < t'$ the squared $L_2(P_0)$ distances between f_t and $f_{t'}$ and between g_t and $g_{t'}$ equal $p r_1 P_T((t, t'])$ and $p r_1 \mathbb{E}_{P_T}[Y^2 \mathbb{1}\{t < Y \leq t'\}]$, respectively, which vanish as $t' - t \rightarrow 0$ because P_T has no atoms.
- The quantile maps $(F, u) \mapsto F^{-1}(u)$ at $(F_{P_T}, \pi\delta)$ and at $(F_{P_T}, 1 - (1 - \pi)\delta)$. Writing $F^{-1}(u)$ as the quantile of $F + u_0 - u$ at the fixed level u_0 , Lemma 21.3 in van der Vaart (1998) and the chain rule show that each map is Hadamard differentiable tangentially to the directions $(h_F, \dot{u})$ with h_F continuous at $t_0 := F_{P_T}^{-1}(u_0)$, with derivative $(\dot{u} - h_F(t_0))/f_{P_T}(t_0)$, as f_{P_T} is continuous and positive at $t_0 \in \{c, d\}$.
- The evaluation maps $(M', t) \mapsto M'(t)$ at (M, c) and at (M, d) . For $t_0 \in \{c, d\}$, $h_s \rightarrow h_M$ uniformly with h_M continuous at t_0 , and $\dot{t}_s \rightarrow \dot{t}$,

$$\frac{(M + s h_s)(t_0 + s \dot{t}_s) - M(t_0)}{s} = \frac{M(t_0 + s \dot{t}_s) - M(t_0)}{s} + h_s(t_0 + s \dot{t}_s) \rightarrow -t_0 f_{P_T}(t_0) \dot{t} + h_M(t_0)$$

as $s \rightarrow 0$, where the limit of the first term follows from $M(t) = \mathbb{E}_{P_T}[Y \mathbb{1}\{Y > t\}]$ and the continuity of f_{P_T} at t_0 .

(i) Asymptotic normality. The remaining dependence of the bound on π runs through the budget $(1 - \pi)\delta$ (hence the cut-point levels), the cap $1/\pi$, and the reconstruction weight $1/(1 - \pi)$, all smooth on $\pi \in (0, 1)$. It remains to compute the derivative Φ'_{P_0} . Let $h \in \mathbb{D}_0$ induce the directions $(h_F, h_M, \dot{\pi})$, and let $\dot{c}, \dot{d}$ be the induced threshold movements. Differentiating $F_{P_T}(c) = \pi\delta$ and $F_{P_T}(d) = 1 - (1 - \pi)\delta$ gives

$$\dot{c} = \frac{\delta \dot{\pi} - h_F(c)}{f_{P_T}(c)}, \quad \dot{d} = \frac{\delta \dot{\pi} - h_F(d)}{f_{P_T}(d)}.$$

Differentiating $\bar{\mu}(\delta) = M(c) + \frac{\pi}{1 - \pi} M(d)$ with the last two maps above and substituting, the density factors $f_{P_T}(c), f_{P_T}(d)$ cancel and the derivative is linear in $(h_F, h_M, \dot{\pi})$: the part linear in (h_F, h_M) is $h_M(c) + \frac{\pi}{1 - \pi} h_M(d) + c h_F(c) + \frac{\pi}{1 - \pi} d h_F(d)$, while the coefficient on $\dot{\pi}$ is the envelope derivative $\varkappa(\delta)$ of the statement. As Φ'_{P_0} depends on h only through finitely many evaluations, it is linear and continuous on all of $\ell^\infty(\mathcal{H}_{RCT})$, and the functional delta method (van der Vaart, 1998, Theorem 20.8) gives

$$\sqrt{n}(\hat{\tau}(\delta) - \bar{\tau}(\delta)) = \Phi'_{P_0}(\sqrt{n}(\hat{P}_n - P_0)) + o_{\mathbb{P}}(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi'_{P_0}(h_{W_i}) + o_{\mathbb{P}}(1),$$

where h_W denotes the direction $f \mapsto f(W) - P_0(f)$. At $h = h_W$, the directions are $h_F = \frac{DS}{p r_1}(\mathbb{1}\{Y \leq \cdot\} - F_{P_T})$ and $h_M = \frac{DS}{p r_1}(Y \mathbb{1}\{Y > \cdot\} - M)$, the direction of π_P is $\dot{\pi} = \psi_\pi(W)$ by the delta method on the two selection frequencies, and the control mean contributes $-\psi_0(W)$, so that $\Phi'_{P_0}(h_W)$ equals the influence function $\psi_{\text{full}}(\cdot; \delta) = \psi_T(\cdot; \delta) + \varkappa(\delta) \psi_\pi - \psi_0$, whose first and third terms reproduce the fixed- π influence $\psi_2(\cdot; \delta) = \psi_T(\cdot; \delta) - \psi_0$ of the reduction. Each component is mean zero and square integrable by the second-moment condition, so $\sqrt{n}(\hat{\tau}(\delta) - \bar{\tau}(\delta)) = n^{-1/2} \sum_i \psi_{\text{full}}(W_i; \delta) + o_{\mathbb{P}}(1) \rightsquigarrow \mathbf{N}(0, \sigma_{\text{full}}^2(\delta))$.

(ii) Consistency of the variance estimators. On $\{D = 1, S = 1\}$, the support of ψ_T , and on $\{D = 0, S = 1\}$, the support of ψ_0 , the selection-frequency influence ψ_π is constant. Since ψ_T and ψ_0 are mean zero this gives $\mathbb{E}[\psi_T \psi_\pi] = \mathbb{E}[\psi_0 \psi_\pi] = 0$, hence $\mathbb{E}[\psi_2 \psi_\pi] = 0$ and $\sigma_{\text{full}}^2(\delta) = \mathbb{E}[(\psi_2(W; \delta) + \varkappa(\delta) \psi_\pi)^2] = \sigma_2^2(\delta) + \varkappa(\delta)^2 V_\pi$. The estimated influence functions $\hat{\psi}_{\text{full}}$ and $\hat{\psi}_2$ are obtained from ψ_{full} and ψ_2 by replacing the vector of population inputs $\vartheta_0 := (s_L(\delta), s_U(\delta), \pi, r_0, r_1, p, \mu^{\text{AO}}(0), \bar{\mu}(\delta), \varkappa(\delta))$ with its plug-in counterpart $\hat{\vartheta}_n$, which is consistent by the law of large numbers together with the continuous mapping theorem. Write $\psi_\bullet(\cdot; \vartheta)$, $\bullet \in \{\text{full}, 2\}$, for the influence function evaluated at a generic input vector ϑ , and consider the classes $\{w \mapsto \psi_\bullet(w; \vartheta)^2 : \vartheta \in K\}$, where K is a

compact neighborhood of ϑ_0 . Each member is a fixed linear combination, with bounded coefficients, of products of indicators of half lines in y , the functions y and $y \mathbb{1}\{y \geq t\}$, and indicators of the events $\{D = 1, S = 1\}$ and $\{D = 0, S = 1\}$, so the classes are Glivenko–Cantelli with integrable envelope proportional to $(1 + Y^2)$. The uniform law of large numbers over K , combined with the continuity of $\vartheta \mapsto \mathbb{E}[\psi_\bullet(W; \vartheta)^2]$ at ϑ_0 yields $\widehat{V}_{\text{full}}(\delta) \xrightarrow{\mathbb{P}} \sigma_{\text{full}}^2(\delta)$ and $\widehat{V}_2(\delta) \xrightarrow{\mathbb{P}} \sigma_2^2(\delta)$, that is $\widehat{\sigma}_{\text{NO}}^2(\delta) \xrightarrow{\mathbb{P}} \sigma_{\text{full}}^2(\delta)$ and $\widehat{\sigma}_{\text{naive}}^2(\delta) \xrightarrow{\mathbb{P}} \sigma_2^2(\delta)$.

Turning to the bootstrap, the class $\mathcal{H}_{\text{RCT}}$ is P_0 -Donsker with a square-integrable envelope, so that the bootstrap empirical process converges conditionally to $\mathbb{G}$ in $\ell^\infty(\mathcal{H}_{\text{RCT}})$ (van der Vaart, 1998, Theorem 23.7). As Φ is Hadamard differentiable at P_0 tangentially to $\mathbb{D}_0$, in which $\mathbb{G}$ takes its values, and the Hadamard derivative is linear, the delta method for the bootstrap (van der Vaart, 1998, Theorem 23.9) gives conditional weak convergence of $\sqrt{n}(\widehat{\tau}^*(\delta) - \widehat{\tau}(\delta))$ to a $\mathcal{N}(0, \sigma_{\text{full}}^2(\delta))$, where $\widehat{\tau}^*(\delta)$ denotes the plug-in estimator computed on a bootstrap resample. Consistency of $\widehat{\sigma}_{\text{boot}}^2(\delta)$ as $B \rightarrow \infty$ follows because the uniform integrability condition (3.21) in Section 3.1.6 of Shao and Tu (1995) is satisfied. Therefore, $\widehat{\sigma}_{\text{boot}}^2(\delta) \xrightarrow{\mathbb{P}} \sigma_{\text{full}}^2(\delta)$.

We now establish the identity $\sigma_{\text{full}}^2(1) = V_{\text{Lee}, \text{U}} + V_{\text{Lee}, \text{C}}$. At $\delta = 1$, the two trimming quantiles are identical, that is $s_L(1) = s_U(1) = F_{P_\tau}^{-1}(\pi) =: y_\pi$, so that $\vartheta_+(y; 1) = \frac{1}{1-\pi} y \mathbb{1}\{y \geq y_\pi\}$, the bound becomes $\bar{\mu}(1) = \mathbb{E}_{P_\tau}[Y \mid Y \geq y_\pi]$, and $\varkappa(1) = (\bar{\mu}(1) - y_\pi)/(1 - \pi)$ is the derivative of the population trimmed mean with respect to the trimming proportion. Define $\psi_U := \psi_\tau(\cdot; 1) + \varkappa(1)\psi_\pi$, so that $\psi_{\text{full}}(\cdot; 1) = \psi_U - \psi_0$. By the derivation of part (i) at $\delta = 1$, which applies as f_{P_τ} is continuous and strictly positive at y_π , ψ_U is the influence function of the estimated treated trimmed mean, accounting for the estimation of both the trimming threshold and the trimming proportion, so that $\mathbb{E}[\psi_U^2] = V_{\text{Lee}, \text{U}}$. Likewise ψ_0 is the influence function of the estimated control mean, so that $\mathbb{E}[\psi_0^2] = V_{\text{Lee}, \text{C}}$. Since ψ_τ and ψ_0 have disjoint supports and $\mathbb{E}[\psi_0\psi_\pi] = 0$ as established above, we get $\mathbb{E}[\psi_U\psi_0] = 0$, and therefore

$$\sigma_{\text{full}}^2(1) = \mathbb{E}[(\psi_U - \psi_0)^2] = \mathbb{E}[\psi_U^2] + \mathbb{E}[\psi_0^2] = V_{\text{Lee}, \text{U}} + V_{\text{Lee}, \text{C}}.$$

Finally, when $r_1 < 1$, $\widehat{\sigma}_{\text{Lee}}^2 = \widehat{V}_{\text{Lee}, \text{U}} + \widehat{V}_{\text{Lee}, \text{C}}$ is the GMM variance estimator of Lee (2009, Proposition 3), hence $\widehat{\sigma}_{\text{Lee}}^2 \xrightarrow{\mathbb{P}} \sigma_{\text{full}}^2(1)$, which was to be shown. $\blacksquare$

SA8.33 Proof of Lemma SA-36

Proof. We check the requirements of condition (b) of Theorem SA-15 in turn, and only the essential-supremum identity requires an argument. Because the same shock vector enters the two maxima,

$$\min_j (t_{jb} - t_{jb'}) \leq g(u; \theta) \leq \max_j (t_{jb} - t_{jb'}),$$

so g is bounded. The entropy ϕ_{TV} is continuous at 1 and boundedness of g implies $\int |g(u; \theta)| d\nu(u) < \infty$ for every finite measure ν . The singleton moment set $\{\bar{\rho}_0\}$ is nonempty, convex, and closed, and $\text{ri}(\{\bar{\rho}_0\}) = \{\bar{\rho}_0\} \subseteq \text{ri}(\Gamma_{\phi_{\text{TV}}}^{\leq}(\theta; \delta))$ by the maintained relative interior condition.

It remains to compare the supremum and essential supremum. Fix λ and write $h_\lambda := g - \lambda^\top m$. At fixed θ , the ties partition $\mathbb{R}^{J+1}$ into finitely many open sets. On each set, s_1 is constant and h_λ is continuous. At a boundary point u , consider the perturbed vector u^r , whose j -th entry is $u_j - rj$. For all sufficiently small $r > 0$, this perturbation preserves every strict comparison and selects, in both choice problems, the lowest-indexed alternative among the alternatives tied at u . It therefore approaches u through an open set on which s_1 is constant and agrees with its value at u . Since g and s_2 are continuous, we get $h_\lambda(u^r) \rightarrow h_\lambda(u)$. Continuity of h_λ on that set therefore implies that values arbitrarily close to $h_\lambda(u)$ occur on a nonempty open set. The Gumbel baseline has full support, as its coordinates are independent with strictly positive densities on $\mathbb{R}$, so every such open set has positive P_* -probability. It follows that $h_\lambda(u) \leq P_*\text{-ess sup } h_\lambda$ for every u . Taking the supremum over u gives $\sup_u h_\lambda(u) \leq P_*\text{-ess sup } h_\lambda$, whereas the reverse inequality holds because the supremum is an essential upper bound. Hence

$$\sup_{u \in \mathbb{R}^{J+1}} h_\lambda(u) = P_*\text{-ess sup}_{u \in \mathbb{R}^{J+1}} h_\lambda(u).$$

The argument uses g only through its continuity, so it applies verbatim to $-g - \lambda^\top m$. Together with the checks recorded at the start of the proof, this establishes condition (b) of Theorem SA-15 for both signs, and the equality of the bounds then follows from that theorem. $\blacksquare$

SA8.34 Proof of Proposition SA-37

Proof. Throughout the proof we use the fact that g_y and g_x are bounded by (58). By the definition of $\mathcal{P}_{\text{FI}}(\delta)$, every P in that set is of the form $P = (1 - \delta)\tilde{P} + \delta R$ for some $R \in \Delta(\mathcal{V})$, so that

$$\mathbb{E}_P[h] = (1 - \delta) \mathbb{E}_{\tilde{P}}[h] + \delta \mathbb{E}_R[h] \quad (99)$$

for every bounded h on $\mathcal{V}$. Combining (99) with the definitions of $G_y(\tilde{P})$, $G_x(\tilde{P})$, $G_y(v'; \delta)$ and $G_x(v'; \delta)$, and using that R is a probability measure, gives

$$G_y(P) = \int_{\mathcal{V}} G_y(v'; \delta) dR(v'), \quad G_x(P) = \int_{\mathcal{V}} G_x(v'; \delta) dR(v'). \quad (100)$$

Part (i). Write $\underline{G}_x := \inf_{v' \in \mathcal{V}} G_x(v'; \delta)$ and note that, by the definition of $G_x(v'; \delta)$,

$$\underline{G}_x = G_x(\tilde{P}) - \delta \sup_{v' \in \mathcal{V}} \{g_x(v') - \bar{g}_x^*\},$$

so that $\underline{G}_x > 0$ precisely when $\delta \in \mathcal{I}_{\text{FS}}$. That set is an interval containing the origin, since the displayed expression is affine and nonincreasing in δ and equals $G_x(\tilde{P}) > 0$ at $\delta = 0$, and its right endpoint is δ_{FS} , as the ratio appearing in that definition is the budget at which the displayed expression vanishes. Suppose $\underline{G}_x > 0$. Then (100) gives $G_x(P) \geq \underline{G}_x > 0$ for every $P \in \mathcal{P}_{\text{FI}}(\delta)$, so the estimator is well defined there. Moreover, $|G_y(P)| \leq \sup_{v' \in \mathcal{V}} |G_y(v'; \delta)| < \infty$ by (100) and the boundedness of g_y , so that $|\hat{\beta}(P)|$ is bounded by $\sup_{v'} |G_y(v'; \delta)| / \underline{G}_x$ and both bounds in (59) are finite. Suppose instead $\underline{G}_x < 0$. Since $\sup_{v' \in \mathcal{V}} G_x(v'; \delta) \geq G_x(\tilde{P}) > 0$, the interval of values of $G_x(P)$, which by (100) has endpoints $\underline{G}_x$ and $\sup_{v'} G_x(v'; \delta)$, contains zero in its interior, so the neighborhood contains assignment processes at which the recentered first stage vanishes.

Part (ii). Fix $\delta \in \mathcal{I}_{\text{FS}}$, so that $\underline{G}_x > 0$ by part (i), and set $\bar{c} := \sup_{v' \in \mathcal{V}} G_y(v'; \delta) / G_x(v'; \delta)$, which is finite because the numerators are bounded and the denominators are bounded below by $\underline{G}_x > 0$. For every $v' \in \mathcal{V}$ we have $G_y(v'; \delta) \leq \bar{c} G_x(v'; \delta)$, and integrating this inequality against any $R \in \Delta(\mathcal{V})$ gives $G_y(P) \leq \bar{c} G_x(P)$ by (100). Since $G_x(P) > 0$, this reads $\hat{\beta}(P) \leq \bar{c}$, so that $\bar{\beta}(\delta) \leq \bar{c}$. Conversely, taking R to be the point mass at v' gives $\hat{\beta}(P) = G_y(v'; \delta) / G_x(v'; \delta)$, so that the supremum defining $\bar{c}$ is approached within $\mathcal{P}_{\text{FI}}(\delta)$ and $\bar{\beta}(\delta) = \bar{c}$, with attainment whenever the supremum over $\mathcal{V}$ is attained. The argument for $\underline{\beta}(\delta)$ is symmetric. For the intermediate values, let $v_1, v_2 \in \mathcal{V}$ and let $R_t := (1-t)\delta_{v_1}^{\text{Dirac}} + t\delta_{v_2}^{\text{Dirac}}$ for $t \in [0, 1]$. The corresponding $\hat{\beta}$ is a continuous function of t , so by the intermediate value theorem it takes every value between the two ratios attached to v_1 and v_2 , and choosing v_1 and v_2 so that those two ratios are arbitrarily close to $\underline{\beta}(\delta)$ and to $\bar{\beta}(\delta)$ shows that every value strictly between the two bounds is taken. ■

SA8.35 Proof of Lemma SA-38

Proof. We prove the three parts in turn. The Hahn–Jordan construction behind part (i) also yields an expectation-decomposition identity, which we record along the way and then use to prove part (ii). Part (iii) follows instead from the density characterization of the decomposition neighborhood in Lemma SA-12(i), combined with the quantile change of variables already employed in the proof of part (ii).

Part (i). Define the signed measure $\nu := P - Q$. By the Hahn decomposition theorem, there exist sets $\mathcal{Y}^+, \mathcal{Y}^- \in \mathcal{F}$ such that $\mathcal{Y}^- \cup \mathcal{Y}^+ = \mathcal{Y}$, $\mathcal{Y}^- \cap \mathcal{Y}^+ = \emptyset$, $\nu(E) \geq 0$ for every measurable $E \subseteq \mathcal{Y}^+$, and $\nu(E) \leq 0$ for every measurable $E \subseteq \mathcal{Y}^-$. Define the nonnegative measures $\nu^-(E) := -\nu(E \cap \mathcal{Y}^-)$ and $\nu^+(E) := \nu(E \cap \mathcal{Y}^+)$ for each $E \in \mathcal{F}$. By construction they have disjoint supports. Since P and Q are probability measures, $\nu(\mathcal{Y}) = P(\mathcal{Y}) - Q(\mathcal{Y}) = 0$, and therefore $\nu^+(\mathcal{Y}) = \nu^-(\mathcal{Y})$ because $\nu^+(\mathcal{Y}) - \nu^-(\mathcal{Y}) = P(\mathcal{Y}) - Q(\mathcal{Y}) = 0$.

It remains to identify this common mass with $\text{TV}(P \| Q)$. For every $E \in \mathcal{F}$, $\nu(E) = \nu^+(E) - \nu^-(E) \leq \nu^+(E) \leq \nu^+(\mathcal{Y})$, so $\sup_{A \in \mathcal{F}} \nu(A) \leq \nu^+(\mathcal{Y})$. Taking $A = \mathcal{Y}^+$ gives $\nu(\mathcal{Y}^+) = \nu^+(\mathcal{Y}^+) - \nu^-(\mathcal{Y}^+) = \nu^+(\mathcal{Y}^+) - 0 = \nu^+(\mathcal{Y})$, so the supremum is achieved and $\sup_{A \in \mathcal{F}} \nu(A) = \nu^+(\mathcal{Y})$. Applying the same argument to $-\nu$, whose positive part is ν^- , gives $-\inf_{A \in \mathcal{F}} \nu(A) = \nu^-(\mathcal{Y})$. Hence

$$\text{TV}(P \| Q) = \sup_{A \in \mathcal{F}} |\nu(A)| = \max \left\{ \sup_{A \in \mathcal{F}} \nu(A), -\inf_{A \in \mathcal{F}} \nu(A) \right\} = \max\{\nu^+(\mathcal{Y}), \nu^-(\mathcal{Y})\} = \nu^+(\mathcal{Y}) = \nu^-(\mathcal{Y}),$$

which together with $\text{TV}(P \| Q) \leq \delta$ gives the asserted mass identity.

It remains to record the attainment claim. By construction $\nu^+(\mathcal{Y} \setminus \mathcal{Y}^+) = \nu((\mathcal{Y} \setminus \mathcal{Y}^+) \cap \mathcal{Y}^+) = 0$ and $\nu^-(\mathcal{Y}^+) = -\nu(\mathcal{Y}^+ \cap \mathcal{Y}^-) = 0$, so $\mathcal{Y}^+$ carries ν^+ in the stated sense. Moreover $\nu(\mathcal{Y}^+) = \nu^+(\mathcal{Y}) = \text{TV}(P \| Q) \geq 0$ by the previous display, whence

$$|P(\mathcal{Y}^+) - Q(\mathcal{Y}^+)| = |\nu(\mathcal{Y}^+)| = \nu^+(\mathcal{Y}) = \text{TV}(P \| Q),$$

so the supremum defining the total variation is attained at $\mathcal{Y}^+$. This proves fact (i).

Finally, we record the decomposition identity used below. Since $\nu = P - Q$, we have $P = Q + \nu = Q + \nu^+ - \nu^-$. Moreover, $\nu^+ \leq P$ and $\nu^- \leq Q$ as measures, since for every $E \in \mathcal{F}$ we have $\nu^+(E) = P(E \cap \mathcal{Y}^+) - Q(E \cap \mathcal{Y}^+) \leq P(E)$ and $\nu^-(E) = Q(E \cap \mathcal{Y}^-) - P(E \cap \mathcal{Y}^-) \leq Q(E)$, so every $f \in L_1(P) \cap L_1(Q)$ is integrable with respect to both ν^+ and ν^- , and

$$\int f dP = \int f dQ + \int f d\nu^+ - \int f d\nu^-. \quad (101)$$

Part (ii). Let $q = F_Q^{-1}$. Since Q is the image of the uniform distribution on $(0, 1)$ under q , we have $\int h dQ = \int_0^1 h(q(u)) du$ for every $h \in L_1(Q)$, and in particular $\mathbb{E}_Q[Y] = \int_0^1 q(u) du$. By fact (i), $\nu^+(\mathcal{Y}) = \nu^-(\mathcal{Y}) = \text{TV}(P \| Q) \leq \delta$. Write $t := \nu^+(\mathcal{Y}) = \nu^-(\mathcal{Y}) = \text{TV}(P \| Q) \leq \delta$ for the common mass of the two parts. Taking $f(y) = y$ in (101),

$$\mathbb{E}_P[Y] = \mathbb{E}_Q[Y] + \int y d\nu^+(y) - \int y d\nu^-(y).$$

Because $y \in [\underline{Y}, \bar{Y}]$ on the support,

$$\int y \, d\nu^+(y) \leq \bar{Y} \nu^+(\mathcal{Y}) = t \bar{Y},$$

and so

$$\mathbb{E}_P[Y] \leq \mathbb{E}_Q[Y] + t \bar{Y} - \int y \, d\nu^-(y).$$

We now bound $\int y \, d\nu^-(y)$ from below. Since $\nu^- \leq Q$, we have $\nu^- \ll Q$, so the Radon–Nikodym theorem yields a density $r := d\nu^-/dQ$, which satisfies $0 \leq r \leq 1$ Q -a.e. in view of $\nu^- \leq Q$, together with $\int r(y) \, dQ(y) = \nu^-(\mathcal{Y}) = t$. Therefore, by the change of variables $\int h(x) \, dQ(x) = \int_0^1 h(q(u)) \, du$ applied with $h(x) = x r(x)$,

$$\int y \, d\nu^-(y) = \int y r(y) \, dQ(y) = \int_0^1 q(u) w(u) \, du,$$

where $w(u) := r(q(u))$ satisfies $0 \leq w \leq 1$ and $\int_0^1 w(u) \, du = t$. Minimizing $\int_0^1 q(u) w(u) \, du$ over all w with $0 \leq w \leq 1$ and $\int_0^1 w(u) \, du = t$ – a class at least as large as the one induced by the measures $\nu^- \leq Q$ of mass t – and using that q is weakly increasing, the minimum over this class is attained at $w^*(u) = \mathbb{1}\{u \leq t\}$, whence

$$\int y \, d\nu^-(y) \geq \int_0^t q(u) \, du \quad \text{for every } \nu^- \leq Q \text{ with } \nu^-(\mathcal{Y}) = t.$$

Substituting this lower bound, every P with $\text{TV}(P \parallel Q) = t \leq \delta$ obeys

$$\mathbb{E}_P[Y] \leq \mathbb{E}_Q[Y] + G(t), \quad G(t) := t \bar{Y} - \int_0^t q(u) \, du.$$

The map G is nondecreasing on $[0, 1]$, since $G(t) = \int_0^t (\bar{Y} - q(u)) \, du$ and the integrand is nonnegative because $q(u) \leq \bar{Y}$ for all u . Hence the bound is largest at the maximal admissible mass $t = \delta$, giving the closed form upper bound

$$\bar{\mu}_Q(\delta) = \mathbb{E}_Q[Y] + \delta \bar{Y} - \int_0^\delta q(u) \, du,$$

attained: let ν^* be the sub-measure of Q of total mass δ carried by the lowest quantiles, that is, $d\nu^*/dQ = \mathbb{1}\{y < q(\delta)\} + \gamma \mathbb{1}\{y = q(\delta)\}$ with $\gamma \in [0, 1]$ chosen so that $\nu^*(\mathcal{Y}) = \delta$ – such a γ exists since $F(q(\delta)-) \leq \delta \leq F(q(\delta))$ by definition of the quantile function – and set $P^* := Q - \nu^* + \delta \delta_Y^{\text{Dirac}}$. Then P^* is a nonnegative measure in view of $\nu^* \leq Q$, its total mass equals $1 - \delta + \delta = 1$, and $\text{TV}(P^* \parallel Q) \leq \delta$ because $P^* - Q = \delta \delta_Y^{\text{Dirac}} - \nu^*$ implies $|P^*(A) - Q(A)| \leq \max\{\delta, \nu^*(\mathcal{Y})\} = \delta$ for every $A \in \mathcal{F}$, and

$$\int y \, d\nu^* = \int_0^{F(q(\delta)-)} q(u) \, du + q(\delta)(\delta - F(q(\delta)-)) = \int_0^\delta q(u) \, du,$$

where the last equality uses that $q(u) = q(\delta)$ for $u \in (F(q(\delta)-), \delta]$. Hence both one-sided bounds hold with equality and $\mathbb{E}_{P^*}[Y] = \bar{\mu}_Q(\delta)$. A symmetric argument gives the lower bound.

Finally, because $\mathbb{E}_Q[Y] = \int_0^1 q(u) \, du$, we can also rewrite the bounds as

$$\bar{\mu}_Q(\delta) = \delta \bar{Y} + \int_\delta^1 q(u) \, du, \quad \underline{\mu}_Q(\delta) = \delta \underline{Y} + \int_0^{1-\delta} q(u) \, du.$$

Part (iii). Suppose first that $\kappa = 1$. In this case, $\mathcal{C}_1^{\ll}(Q) = \{Q\}$ and both displayed formulas reduce to the identity $\mathbb{E}_Q[Y] = \int_0^1 q(u) \, du$. Fix then $\kappa \in (0, 1)$. By Lemma SA-12(i), a probability measure $\tilde{P}$ belongs to $\mathcal{C}_\kappa^{\ll}(Q)$ if and only if $\tilde{P} \ll Q$ and its density $\tilde{r} := d\tilde{P}/dQ$ satisfies $\tilde{r} \leq \kappa^{-1}$ Q -a.e. In particular, every $\tilde{P} \in \mathcal{C}_\kappa^{\ll}(Q)$ satisfies $\mathbb{E}_{\tilde{P}}[|Y|] \leq \kappa^{-1} \mathbb{E}_Q[|Y|] < \infty$.

By the change of variables $\int h(y) \, dQ(y) = \int_0^1 h(q(u)) \, du$ applied with $h(y) = y \tilde{r}(y)$,

$$\mathbb{E}_{\tilde{P}}[Y] = \int y \tilde{r}(y) \, dQ(y) = \int_0^1 q(u) w(u) \, du,$$

where $w(u) := \tilde{r}(q(u))$ satisfies $0 \leq w \leq \kappa^{-1}$ and $\int_0^1 w(u) \, du = \int \tilde{r}(y) \, dQ(y) = 1$. We now maximize $\int_0^1 q(u) w(u) \, du$ over all w with $0 \leq w \leq \kappa^{-1}$ and $\int_0^1 w(u) \, du = 1$ and claim that the maximum over this class is attained at $w^*(u) := \kappa^{-1} \mathbb{1}\{u > 1 - \kappa\}$. Indeed, for any feasible w ,

$$\int_0^1 q(u) (w^*(u) - w(u)) \, du = \int_0^1 (q(u) - q(1 - \kappa)) (w^*(u) - w(u)) \, du \geq 0,$$

where the equality holds because $\int_0^1 (w^*(u) - w(u)) \, du = 1 - 1 = 0$, and the inequality follows because the integrand is nonnegative. To see this, note that for $u > 1 - \kappa$ we have $q(u) \geq q(1 - \kappa)$, since q is weakly increasing, and $w^*(u) - w(u) = \kappa^{-1} - w(u) \geq 0$, whereas for $u \leq 1 - \kappa$ we have $q(u) \leq q(1 - \kappa)$ and $w^*(u) - w(u) = -w(u) \leq 0$. Hence every $\tilde{P} \in \mathcal{C}_\kappa^{\ll}(Q)$ obeys

$$\mathbb{E}_{\tilde{P}}[Y] \leq \int_0^1 q(u) w^*(u) \, du = \frac{1}{\kappa} \int_{1-\kappa}^1 q(u) \, du.$$

It remains to show that the bound is attained. Let ν^* be the sub-measure of Q of total mass κ carried by the largest quantiles, that is, $d\nu^*/dQ = \mathbb{1}\{y > q(1 - \kappa)\} + \gamma \mathbb{1}\{y = q(1 - \kappa)\}$ with $\gamma \in [0, 1]$ chosen so that $\nu^*(\mathcal{Y}) = \kappa$. Note that such a γ exists because $Q(Y > q(1 - \kappa)) = 1 - F(q(1 - \kappa)) \leq \kappa$ and $Q(Y \geq q(1 - \kappa)) = 1 - F(q(1 - \kappa)-) \geq \kappa$ by the definition of the quantile function. Set $\tilde{P}^* := \kappa^{-1}\nu^*$. Then $\tilde{P}^*$ is a probability measure by construction, and $\kappa\tilde{P}^* = \nu^* \leq Q$ as measures, so $\tilde{P}^* \in \mathcal{C}_\kappa^{\leq}(Q)$ in view of Lemma SA-12(i). Moreover,

$$\mathbb{E}_{\tilde{P}^*}[Y] = \frac{1}{\kappa} \int y d\nu^* = \frac{1}{\kappa} \left(\int_{F(q(1-\kappa))}^1 q(u) du + q(1 - \kappa)(\kappa - 1 + F(q(1 - \kappa))) \right) = \frac{1}{\kappa} \int_{1-\kappa}^1 q(u) du,$$

where the second equality collects the mass that ν^* places on $\{y > q(1 - \kappa)\}$, which contributes $\int_{F(q(1-\kappa))}^1 q(u) du$ by the change of variables above and the identity $\{u \in (0, 1) : q(u) > q(1 - \kappa)\} = (F(q(1 - \kappa)), 1)$ up to null sets, and the mass $\gamma Q(\{q(1 - \kappa)\}) = \kappa - 1 + F(q(1 - \kappa))$ that ν^* places on the atom at $q(1 - \kappa)$, while the last equality uses that $q(u) = q(1 - \kappa)$ for $u \in (1 - \kappa, F(q(1 - \kappa))]$. This proves the formula for the supremum. A symmetric argument, with the mass carried by the lowest quantiles and $w^*(u) = \kappa^{-1}\mathbb{1}\{u \leq \kappa\}$, yields the formula for the infimum together with its attainment. ■

SA8.36 Proof of Theorem SA-39

Proof. Throughout, for $P \in \Delta(\mathcal{Z})$ we write $P = wP_* + P^\perp$ for its Lebesgue decomposition with respect to P_* , where $P^\perp \perp P_*$ and w is the Radon–Nikodym derivative of the absolutely continuous part of P with respect to P_* , so that $\mathbb{E}_{P_*}[w(Z)] + P^\perp(\mathcal{Z}) = P(\mathcal{Z}) = 1$. By Lemma SA-1,

$$D_\phi(P \| P_*) = \int_{\mathcal{Z}} \phi(w(z)) dP_*(z) + \phi^\infty(1) P^\perp(\mathcal{Z}). \quad (102)$$

Since ϕ is convex and finite at $\underline{x}$, and at $\bar{x}$ when the latter is finite, its effective domain is $[\underline{x}, \bar{x}]$ if $\bar{x} < \infty$ and $[\underline{x}, \infty)$ if $\bar{x} = \infty$. Moreover, $\underline{x} \leq 1 \leq \bar{x}$ because $\phi(1) = 0$.

Finite divergence and domination. We show that $\mathcal{D}_\phi(P_*)$ coincides with the set of $P \in \Delta(\mathcal{Z})$ with $D_\phi(P \| P_*) < \infty$, and that $D_\phi(P \| P_*) \leq \bar{\delta}_\phi$ on this set.

Case $\bar{x} < \infty$. Since $\phi(t) = +\infty$ for $t > \bar{x}$, Definition SA-2 with $z_0 = 1$ gives $\phi^\infty(1) = +\infty$. Hence, in view of (102) and $\phi \geq 0$, $D_\phi(P \| P_*) < \infty$ forces $P^\perp(\mathcal{Z}) = 0$ and $w \in [\underline{x}, \bar{x}]$ P_* -almost surely. By the equivalence of (2) and (3) in Lemma SA-12, which holds for every $\kappa > 0$, applied with $\kappa = 1/\bar{x}$ in part (i) and, if $\underline{x} > 0$, with $\kappa = \underline{x}$ in part (ii), these two conditions hold if and only if $\underline{x}P_* \leq P \leq \bar{x}P_*$. Conversely, let P satisfy these inequalities, so that $P = wP_*$ with $\underline{x} \leq w \leq \bar{x}$ P_* -almost surely. For every $t \in [\underline{x}, \bar{x}]$, convexity of ϕ gives

$$\phi(t) \leq \frac{(\bar{x} - t)\phi(\underline{x}) + (t - \underline{x})\phi(\bar{x})}{\bar{x} - \underline{x}},$$

and integrating this bound at $t = w(z)$ against P_* , using $\mathbb{E}_{P_*}[w(Z)] = 1$, delivers $D_\phi(P \| P_*) \leq \{(\bar{x} - 1)\phi(\underline{x}) + (1 - \underline{x})\phi(\bar{x})\}/(\bar{x} - \underline{x}) = \bar{\delta}_\phi$, which is finite by assumption.

Case $\bar{x} = \infty$. By (102) and $\phi \geq 0$, $D_\phi(P \| P_*) < \infty$ forces $w \geq \underline{x}$ P_* -almost surely, and therefore $P \geq wP_* \geq \underline{x}P_*$. Conversely, if $\underline{x}P_* \leq P$, then $w \geq \underline{x}$ P_* -almost surely, trivially if $\underline{x} = 0$ and by Lemma SA-12(ii) with $\kappa = \underline{x}$ otherwise, as the equivalence of (2) and (3) there holds for every $\kappa > 0$. For every $t \geq \underline{x}$ and every $T > \max\{t, 1\}$, convexity of ϕ gives

$$\phi(t) \leq \frac{(T - t)\phi(\underline{x}) + (t - \underline{x})\phi(T)}{T - \underline{x}},$$

and, since $\phi(1) = 0$, Definition SA-2 with $z_0 = 1$ gives $\phi^\infty(1) = \lim_{T \rightarrow \infty} \phi(T)/T$, so that letting $T \rightarrow \infty$ yields $\phi(t) \leq \phi(\underline{x}) + (t - \underline{x})\phi^\infty(1)$. Integrating this bound at $t = w(z)$ against P_* and using (102), we obtain

$$D_\phi(P \| P_*) \leq \phi(\underline{x}) + \phi^\infty(1)\{\mathbb{E}_{P_*}[w(Z)] - \underline{x} + P^\perp(\mathcal{Z})\} = \phi(\underline{x}) + (1 - \underline{x})\phi^\infty(1) = \bar{\delta}_\phi,$$

where the equality follows from $\mathbb{E}_{P_*}[w(Z)] + P^\perp(\mathcal{Z}) = 1$, and the right-hand side is finite by assumption.

Nonparametric sharp bounds. Fix $\delta \geq \bar{\delta}_\phi$ and $\theta \in \Theta$. If $P \in \mathcal{P}_\phi(\theta; \rho, P_*, \delta)$, then $D_\phi(P \| P_*) \leq \delta < \infty$, so $P \in \mathcal{D}_\phi(P_*)$ by the first part of the proof, and $P \in \mathcal{E}(\theta; \rho)$ by definition of the robustness set. Conversely, if $P \in \mathcal{D}_\phi(P_*) \cap \mathcal{E}(\theta; \rho)$, then $D_\phi(P \| P_*) \leq \bar{\delta}_\phi \leq \delta$ by the first part of the proof, so P satisfies both constraints defining $\mathcal{P}_\phi(\theta; \rho, P_*, \delta)$. The two sets therefore coincide for every $\theta \in \Theta$, and the displayed identities follow because, at every θ , the two inner optimization problems share the same feasible set and the same objective, so that the outer optimizations over $\theta \in \Theta$ coincide as well. ■

Supplement References

- Ali, Syed Mumtaz and Samuel D. Silvey, “A General Class of Coefficients of Divergence of One Distribution from Another,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 1966, 28 (1), 131–142.
- Bertsekas, Dimitri P., *Convex Optimization Theory*, Belmont, MA: Athena Scientific, 2009.
- Bonvini, Matteo and Edward H. Kennedy, “Sensitivity Analysis via the Proportion of Unmeasured Confounding,” *Journal of the American Statistical Association*, 2022, 117 (539), 1540–1550.

- Borusyak, Kirill and Peter Hull**, “Nonrandom Exposure to Exogenous Shocks,” *Econometrica*, 2023, 91 (6), 2155–2185.
- , – , and **Xavier Jaravel**, “Design-Based Identification with Formula Instruments: A Review,” *Econometrics Journal*, 2025, 28 (1), 83–108.
- Carlana, Michela and Eliana La Ferrara**, “Apart but Connected: Online Tutoring, Cognitive Outcomes, and Soft Skills,” *American Economic Review*, 2025, 115 (10), 3487–3513.
- Chiappori, Pierre-André, Bernard Salanié, and Yoram Weiss**, “Partner Choice, Investment in Children, and the Marital College Premium,” *American Economic Review*, 2017, 107 (8), 2109–2167.
- Christensen, Peter, Gustavo Nino, and Adam Osman**, “The Demand for Mobility: Evidence from an Experiment with Uber Riders,” *Journal of Political Economy*, 2026. Forthcoming.
- Christensen, Timothy and Benjamin Connault**, “Counterfactual Sensitivity and Robustness,” *Econometrica*, 2023, 91 (1), 263–298.
- Csiszár, Imre**, “Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten,” *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, 1963, 8 (1–2), 85–108.
- Currie, Janet and Jonathan Gruber**, “Health Insurance Eligibility, Utilization of Medical Care, and Child Health,” *Quarterly Journal of Economics*, 1996, 111 (2), 431–466.
- Donaldson, Dave and Richard Hornbeck**, “Railroads and American Economic Growth: A “Market Access” Approach,” *Quarterly Journal of Economics*, 2016, 131 (2), 799–858.
- Fang, Zheng and Andres Santos**, “Inference on Directionally Differentiable Functions,” *Review of Economic Studies*, 2019, 86 (1), 377–412.
- Hansen, Lars Peter and Thomas J. Sargent**, “Robust Control and Model Uncertainty,” *American Economic Review*, 2001, 91 (2), 60–66.
- Horowitz, Joel L. and Charles F. Manski**, “Identification and Robustness with Contaminated and Corrupted Data,” *Econometrica*, 1995, 63 (2), 281–302.
- Huber, Peter J.**, “Robust Estimation of a Location Parameter,” *Annals of Mathematical Statistics*, 1964, 35 (1), 73–101.
- Imai, Kosuke**, “Sharp Bounds on the Causal Effects in Randomized Experiments with “Truncation-by-Death”,” *Statistics & Probability Letters*, 2008, 78 (2), 144–149.
- Imbens, Guido W. and Charles F. Manski**, “Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 2004, 72 (6), 1845–1857.
- Kuhn, Daniel, Soroosh Shafiee, and Wolfram Wiesemann**, “Distributionally Robust Optimization,” *Acta Numerica*, 2025, 34, 579–804.
- Lee, David S.**, “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects,” *Review of Economic Studies*, 2009, 76 (3), 1071–1102.
- Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini**, “Ambiguity Aversion, Robustness, and the Variational Representation of Preferences,” *Econometrica*, 2006, 74 (6), 1447–1498.
- Manski, Charles F.**, *Partial Identification of Probability Distributions*, New York: Springer, 2003.
- Miguel, Edward and Michael Kremer**, “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities,” *Econometrica*, 2004, 72 (1), 159–217.
- Morimoto, Tetsuzo**, “Markov Processes and the H -Theorem,” *Journal of the Physical Society of Japan*, 1963, 18 (3), 328–331.
- Newey, Whitney K. and Daniel McFadden**, “Large Sample Estimation and Hypothesis Testing,” in Robert F. Engle and Daniel McFadden, eds., *Handbook of Econometrics*, Vol. 4, Amsterdam: Elsevier, 1994, pp. 2111–2245.
- Ober-Reynolds, Daniel**, “Robustness to Missing Data: Breakdown Point Analysis,” *Journal of Econometrics*, 2026, 253, 106151.
- Polyanskiy, Yury and Yihong Wu**, *Information Theory: From Coding to Learning*, Cambridge: Cambridge University Press, 2025.
- Puri, Jyotsna and Francis Rathinam**, “Challenges in Real-World Impact Evaluations: Some Learning on Costs and Timeliness,” 2019. IEU Working Paper 3, Independent Evaluation Unit, Green Climate Fund.
- Rockafellar, R. Tyrrell**, *Convex Analysis*, Princeton: Princeton University Press, 1970.
- , *Conjugate Duality and Optimization*, Philadelphia: Society for Industrial and Applied Mathematics, 1974.
- and **Roger J.-B. Wets**, *Variational Analysis*, Berlin: Springer, 2009.
- Rubin, Donald B.**, “Inference and Missing Data,” *Biometrika*, 1976, 63 (3), 581–592.
- Schennach, Susanne M. and Vincent Starck**, “Optimally-Transported Generalized Method of Moments,” *Econometrica*, 2026, 94 (2), 619–640.
- Semenova, Vira**, “Generalized Lee Bounds,” *Journal of Econometrics*, 2025, 251, 106055.
- Shao, Jun and Dongsheng Tu**, *The Jackknife and Bootstrap*, New York: Springer, 1995.
- Shao, Yulin, Liangbo Lyu, Menggang Yu, and Bingkai Wang**, “How Should Covariates Be Handled in Randomized Trials? Empirical Evidence from 50 Trials and Recommendations for Practice,” *Journal of Clinical Epidemiology*, 2026, 197, 112374.
- Shapiro, Alexander**, “On Concepts of Directional Differentiability,” *Journal of Optimization Theory and Applications*, 1990, 66 (3), 477–487.
- van der Vaart, Aad W.**, *Asymptotic Statistics*, Cambridge: Cambridge University Press, 1998.
- and **Jon A. Wellner**, *Weak Convergence and Empirical Processes: With Applications to Statistics*, 1st ed., New York: Springer, 1996.
- Zhang, Junni L. and Donald B. Rubin**, “Estimation of Causal Effects via Principal Stratification When Some Outcomes Are Truncated by “Death”,” *Journal of Educational and Behavioral Statistics*, 2003, 28 (4), 353–368.