Counterfactual Sensitivity and Robustness*

Timothy Christensen† Benjamin Connault‡

April 14, 2019

Abstract

Researchers frequently make parametric assumptions about the distribution of unobservables when formulating structural models. Such assumptions are typically motived by computational convenience rather than economic theory and are often untestable. Counterfactuals can be particularly sensitive to such assumptions, threatening the credibility of structural modeling exercises. To address this issue, we leverage insights from the literature on ambiguity and model uncertainty to propose a tractable econometric framework for characterizing the sensitivity of counterfactuals with respect to a researcher's assumptions about the distribution of unobservables in a class of structural models. In particular, we show how to construct the smallest and largest values of the counterfactual as the distribution of unobservables spans nonparametric neighborhoods of the researcher's assumed specification while other "structural" features of the model, e.g. equilibrium conditions, are maintained. Our methods are computationally simple to implement, with the nuisance distribution effectively profiled out via a low-dimensional convex program. Our procedure delivers sharp bounds for the identified set of counterfactuals (i.e. without parametric assumptions about the distribution of unobservables) as the neighborhoods become large. Over small neighborhoods, we relate our procedure to a measure of local sensitivity which is further characterized using an influence function representation. We provide a suitable sampling theory for plug-in estimators and apply our procedure to models of strategic interaction and dynamic discrete choice.

Keywords: Robustness, ambiguity, model uncertainty, misspecification, sensitivity analysis.

JEL codes: C14, C18, C54, D81

*We thank several seminar and conference audiences for comments and suggestions. We are grateful to R. Allen, S. Bonhomme, M. Masten, D. Miller (discussant), M. Mogstad, F. Molinari, J. Nesbit, J. Porter (discussant), and A. Torgovitsky for helpful comments.

†Department of Economics, New York University. timothy.christensen@nyu.edu

‡IEX Group. benjamin.connault@gmail.com

1 Introduction

Researchers frequently make strong parametric assumptions about the distribution of unobservables when formulating structural models, often for computational convenience.[1] Yet economic theory typically provides little or no guidance as to the correct specification of the distribution of unobservables and in many models, such as those we consider in this paper, the distribution of unobservables is not nonparametrically identified. Ex-ante policy evaluation exercises, or *counterfactuals*, can be particularly sensitive to potentially incorrect distributional assumptions. Sensitivity arises through two channels. The distributional assumptions are first used at the estimation stage, as they help define the mapping from structural parameters to observables. The assumptions are again used at the evaluation stage, when solving the model under the policy intervention at the estimated structural parameters. Counterfactual choice probabilities may be particularly sensitive to distributional assumptions, as they are essentially probabilities of tail events that are not observed in the data. The potential sensitivity of counterfactuals to such assumptions threatens the credibility of structural modeling exercises, a point made even by proponents of structural modeling (see, e.g., Section 5 of Keane, Todd, and Wolpin (2011)).

In this paper, we introduce a tractable econometric framework to characterize the sensitivity of counterfactuals with respect to distributional assumptions in a class of structural models. We show how to construct sets of counterfactuals as the distribution of interest spans neighborhoods of the researcher's assumed specification while other "structural" features of the model, e.g. equilibrium conditions, are maintained. This approach is in the spirit of global sensitivity analysis advocated by Leamer (1985). Global, rather than local, approaches to characterizing sensitivity in structural models are important, as the nonlinearity of structural models and/or policy interventions means that policies can have different effects at different points in the parameter space. Global sensitivity analyses of nonlinear models can be computationally and theoretically challenging, however.[2] Local sensitivity analyses—based on local linearization around an assumed true specification—are often more tractable. However, local approaches may fail to correctly characterize the counterfactuals predicted by the model when the researcher's assumed distribution of unobservables is misspecified by a degree that is not vanishingly small.

[1]The empirical trade literature typically specifies the distributions of latent idiosyncratic efficiencies or costs as Fréchet or Pareto for computational convenience (see, e.g., Eaton and Kortum (2002) and Allen and Arkolakis (2014)). Workhorse dynamic discrete choice models following Rust (1987) are typically implemented assuming that latent payoff shocks are extreme-value distributed for computational convenience. Prominent works analyzing strategic interaction using models of static and dynamic discrete games make strong parametric assumptions about the distribution of latent payoff shocks (see, e.g., Ericson and Pakes (1995), Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2007), Pesendorfer and Schmidt-Dengler (2008), Ciliberto and Tamer (2009), and Bajari, Hong, and Ryan (2010)). Moreover, strong parametric assumptions about the distribution of utility shocks and random coefficients are frequently made, partly for computational considerations, when evaluating policies using differentiated products demand models following Berry, Levinsohn, and Pakes (1995).

[2]See, for example, work on partially identified semiparametric models by Chen, Tamer, and Torgovitsky (2011) which is motivated partly by questions of sensitivity with respect to various modeling assumptions.

The key innovation of our approach is to leverage insights from the robustness literature in economics pioneered by Hansen and Sargent (see, e.g., Hansen and Sargent (2001, 2008)) to simplify computation using convex programming techniques. To make the analysis as tractable as possible, we restrict attention to a class of structural models whose equilibrium restrictions may be written as a set of moment in/equalities, where the expectation is taken with respect to the distribution of unobservables. This class is sufficiently broad that it accommodates many models of static and dynamic discrete choice and some models of static or dynamic discrete games. Following the robustness literature, we define nonparametric neighborhoods in terms of statistical divergence from the researcher's assumed specification, with the option to add location/scale normalizations or smoothness constraints as appropriate. Consider the problem of minimizing or maximizing the counterfactual at a particular value of structural parameters by varying the distribution over this neighborhood, subject to the equilibrium conditions summarized in the moment in/equalities. This infinite-dimensional optimization problem can be recast as convex program of fixed (low) dimension for the class of problems we consider. Similar insights also underlie other latent variable methods in econometrics (see, e.g., Schennach (2014)) and generalized empirical likelihood, though their redeployment here to target counterfactuals appears novel. The low-dimensional convex programs, when embedded in an outer optimization over structural parameters, deliver the smallest and largest counterfactuals consistent with the model as the distribution varies over the neighborhood. Moreover, our approach is robust to partial identification and irregular estimability of structural parameters, both of which may be important in applications.

We propose plug-in estimators of the smallest and largest counterfactual obtained as the distribution varies over nonparametric neighborhoods of the researcher's assumed specification and develop a suitable sampling theory. In particular, we show that the estimators are consistent and establish their joint asymptotic distribution. Although the distribution will typically be nonstandard, inference is still feasible via subsampling or modified bootstrap methods.

In addition, we characterize the properties of the sets of counterfactuals over very large or very small neighborhoods of the researcher's assumed specification. We show that our procedure delivers sharp bounds on the identified set of counterfactuals (i.e. without any parametric assumption about the distribution of unobservables) as the neighborhood size expands, provided the researcher's assumed specification satisfies a type of support condition. In this sense, our use of neighborhoods constrained by statistical divergence can be viewed as an (infinite-dimensional) sieve: although the neighborhoods exclude many distributions, as they become larger they eventually span the set of distributions relevant for characterizing the identified set of counterfactuals. Unlike finite-dimensional sieve methods, however, here the dimensionality of the optimization problem remains fixed as we consider increasingly rich classes of distributions. Our methods therefore provide a tractable way for characterizing identified sets of counterfactuals in nonlinear structural models without specifying distributions of unobservables.

Small neighborhoods are relevant for comparing our procedure with local sensitivity analyses. To this end, we describe a measure of local sensitivity of counterfactuals with respect to the researcher's assumed specification and formally relate it to our procedure when the neighborhood size is small. Our sensitivity measure is conceptually different from the local sensitivity measures proposed by Andrews, Gentzkow, and Shapiro (2017, 2018), which treat the model specification as given and quantify sensitivity of counterfactuals with respect to local misspecification of the moments used at the estimation stage. In point-identified, sufficiently regular models typically studied in the local sensitivity literature, we show that our sensitivity measure may be characterized by a particular influence function representation. Using this representation, we provide a simple and consistent plug-in estimator of local sensitivity, which researchers may easily report alongside their estimated counterfactuals in structural modeling exercises.

We illustrate the usefulness of our procedure with application to two workhorse models, namely a canonical entry game and an infinite-horizon model of dynamic discrete choice. The game provides a suitable laboratory to illustrate our procedure in a transparent way, as all calculations can be performed in closed form. The dynamic discrete choice example illustrates how our procedure can reveal important asymmetries that may be overlooked in a local sensitivity analysis.

Related literature Our approach has some similarities with the literature on global prior sensitivity in Bayesian analysis. Broadly speaking, this literature studies variation in the posterior as the prior ranges over a class of priors. Early notable references include Chamberlain and Leamer (1976), Leamer (1982), Berger (1984), and Berger and Berliner (1986). Of particular relevance are recent works by Giacomini, Kitagawa, and Uhlig (2016) and Ho (2018) who consider nonparametric classes of priors that are constrained by Kullback–Leibler (KL) divergence relative to a default prior, also partly motivated by the robustness literature in economics. The contexts and objectives of these two works are very different from ours.[3] Nevertheless, they also inherit tractability in complex, partially identified settings by specifying neighborhoods in terms of statistical divergence from a researcher's assumed (prior) distribution.

In principle, one could attack our problem by framing it as a subvector estimation/inference problem in a partially identified semiparametric model. Chen et al. (2011), Tao (2014), and Chernozhukov, Newey, and Santos (2015) study inference in general partially-identified semiparametric models using sieve approximations for the infinite-dimensional parameter (i.e., the distribution of unobservables in our setting). We consider a nonparametric class constrained by statistical divergence from a researcher's assumed distribution, rather than a ball of "smooth" functions typically assumed so as to justify a sieve approximation. In principle, one could adapt various inference methods from

[3]Both works study sensitivity with respect to priors for Bayesian inference whereas we study sensitivity with respect to a particular modeling assumption, namely a distributional over unobservables in structural models, and our inference methods are not Bayesian. Giacomini et al. (2016) emphasize application to structural VARs whereas Ho (2018) emphasizes applicability to large-scale DSGE models.

this literature to construct confidence sets for counterfactuals. This approach would require an inner optimization over the sieve coefficients—whose number must increase to infinity in order to span the full set of distributions—that would generally be non-convex. For the more restrictive class of problems we consider, our approach eliminates the infinite-dimensional nuisance parameter via a convex program of fixed (low) dimension.

The use of convex programming to eliminate distributions of latent variables has been previously noted in the important works of Ekeland, Galichon, and Henry (2010) and Schennach (2014). Both of these earlier works consider somewhat different classes of models from those that we consider. These works are also concerned primarily with characterizing the identified set of structural parameters whereas we are concerned with characterizing sensitivity of counterfactuals with respect to a researcher's modeling assumptions, so the resulting convex programs are different. In recent work that is concurrent with ours, Li (2018) relaxes some restrictions on the moment functions and the support of unobservables in Ekeland et al. (2010) and suggests performing inference on counterfactuals via subvector methods for moment inequality models. By targeting counterfactuals directly and leveraging some additional structure, our approach sidesteps this difficult subvector inference problem.

A number of other recent works construct identified sets of counterfactuals in specific models without making parametric assumptions about the distributions of unobservables. Examples of works that bound counterfactuals in discrete choice models include Manski (2007, 2014), Allen and Rehbeck (2017) and Chiong, Hsieh, and Shum (2017). Norets and Tang (2014) construct identified sets of counterfactuals in infinite-horizon dynamic binary choice models via a reparameterization which is specific to binary choice settings. Torgovitsky (2016) bounds counterfactuals in dynamic potential outcomes models. Of these works, Manski (2007), Norets and Tang (2014), and Torgovitsky (2016) study inference. Manski (2007) proposes finite-sample confidence sets that are specific to that class of models. Torgovitsky (2016) uses subsampling and inversion of a profiled test statistic. Norets and Tang (2014) and perform inference using Bayesian methods.

There is also an active literature on local sensitivity. Local sensitivity analyses typically consider deviations from a "true" limiting model that are shrinking to zero at an appropriate rate with the sample size (i.e. contiguously). The idea is that, as one observes more data, larger departures may be detected using various statistical tests. This approach does not therefore seem appropriate when we are concerned with sensitivity with respect to an (untestable) assumption regarding the distribution of unobservables. Indeed, much of the recent literature on local sensitivity is concerned with formulating estimators and inference procedures that are robust to local misspecification of moment conditions and the like, which is a different problem from that which we consider. Nevertheless, Kitamura, Otsu, and Evdokimov (2013) and Bonhomme and Weidner (2018) are notable for their use of local neighborhoods characterized by statistical divergence. Both of these papers are concerned with formulating estimators that are optimal under local misspecification. Like us, Bonhomme and Weidner (2018) target a specific aspect of model specification. Andrews et al. (2017, 2018) proposed reporting measures to characterize the sensitivity of estimators of counterfactuals or structural parameters with respect to moments used in estimation. In their framework, the distribution of *observables* is modeled as possibly (locally) misspecified. Armstrong and Kolesár (2018) discuss optimal inference in this context and Mukhin (2018) draws connections with semiparametric efficiency theory. Although we assume the model is correctly specified (i.e., there exists a distribution of unobservables and parameter vector that can rationalize the model), it would be interesting to extend our methods along the lines of these works to allow for local misspecification of the moment in/equalities also.

Finally, several recent works in statistical decision theory advocate decision rules that minimize maximum expected loss over KL neighborhoods of a reference model, motivated in part by Hansen and Sargent's work; see Watson and Holmes (2016) and references therein. Hansen and Marinacci (2016) draw connections between this approach and decision theory in economics.

The remainder of the paper is as follows. Section 2 describes the setting and outlines our approach. Section 3 illustrates how to implement our procedure in a simple but transparent entry game setting and a dynamic discrete choice model. Section 4.1 presents results from convex analysis that justify our procedure and Section 4.2 establishes the large-sample properties of plug-in estimators. Section 5 discusses the sharp bounds on the identified set of counterfactuals and Section 6 discusses local sensitivity. Section 7 concludes.

2 Procedure

2.1 Setup

Consider the following description of a structural modeling exercise. The researcher observes a sample of data, say $X_1, \ldots, X_n$. The researcher computes (i) a $d_P \times 1$ vector of targeted moments $\hat{P}$ (e.g. a vector of choice probabilities or market shares), and possibly (ii) an estimator $\hat{\gamma}$ of an auxiliary parameter (e.g. a law of motion of a Markov state). We assume that $\hat{P}$ and $\hat{\gamma}$ are consistent estimators and let P_0 and γ_0 denote their probability limits.

The researcher's structural model links $P_0 \in \mathcal{P} \subseteq \mathbb{R}^{d_P}$ and $\gamma_0 \in \Gamma$ (a metric space) to a $d_\theta \times 1$

vector of structural parameters $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ through the moment in/equality restrictions

$$\mathbb{E}^F[g_1(U,\theta,\gamma_0)] \le P_{10}\,, \tag{1}$$

$$\mathbb{E}^F[g_2(U,\theta,\gamma_0)] = P_{20}\,, \tag{2}$$

$$\mathbb{E}^F[g_3(U,\theta,\gamma_0)] \le 0\,, \tag{3}$$

$$\mathbb{E}^F[g_4(U,\theta,\gamma_0)] = 0\,, \tag{4}$$

where $g_1,\dots,g_4$ are vectors of moment conditions of dimension $d_1,\dots,d_4$ describing model-implied moments (in g_1 and g_2) and additional equilibrium conditions (in g_3 and g_4), U is a vector of unobservables with distribution F, and P_0 is partitioned conformably as $P_0 = (P_{10}', P_{20}')'$.

The distribution F is typically not nonparametrically identified. Therefore, in common practice, a seemingly reasonable and/or computationally convenient distribution, say F_*, is assumed by the researcher and maintained throughout the analysis. We refer to F_* as the researcher's *default* or *benchmark* specification.

Given F_* and first-stage estimators $\hat\gamma$ and $\hat P = (\hat P_1', \hat P_2')'$, the researcher would invert the sample moment conditions

$$\mathbb{E}^{F_*}[g_1(U,\theta,\hat\gamma)] \le \hat P_1\,, \qquad \mathbb{E}^{F_*}[g_2(U,\theta,\hat\gamma)] = \hat P_2\,,$$
$$\mathbb{E}^{F_*}[g_3(U,\theta,\hat\gamma)] \le 0\,, \qquad \mathbb{E}^{F_*}[g_4(U,\theta,\hat\gamma)] = 0\,,$$

using, e.g., a minimum distance, GMM, or moment inequality criterion, to obtain an estimator $\hat\theta$ of θ. The researcher would then estimate a (scalar) counterfactual of interest

$$\kappa = \mathbb{E}^F[k(U,\theta,\gamma_0)]$$

(e.g. a counterfactual choice probability or counterfactual measure of expected profits) using the plug-in estimator

$$\hat\kappa = \mathbb{E}^{F_*}[k(U,\hat\theta,\hat\gamma)]\,.$$

We refer to this as the *explicit-dependence* case in what follows, as the function k depends *explicitly* on the latent variables. Alternatively, the researcher may have a counterfactual of the form

$$\kappa = k(\theta,\gamma_0)$$

for which the researcher might use the plug-in estimator

$$\hat\kappa = k(\hat\theta,\hat\gamma)\,.$$

We refer to this as the *implicit-dependence* case in what follows, as the assumed distribution F_*

will still have an effect on the estimator $\hat{\theta}$. Although the preceding discussion has assumed point identification of θ and κ for sake of exposition, our methods are also robust to partial identification of structural parameters and counterfactuals.

To fix ideas, consider an infinite-horizon dynamic discrete choice (DDC) model with a discrete statespace, as in Rust (1987). In that model, U is the vector of payoff shocks in the period utility function and F_* is taken to be type-I extreme value for computational convenience. The vector of moment conditions g_2 specifies the model-implied conditional choice probabilities and the vector g_4 specifies the fixed-point equation for the ex-ante value function. There are no inequalities, so the remaining moment conditions involving g_1 and g_3 would be vacuous. The vector θ would consist of structural parameters and pre- and post-intervention ex-ante value functions, similar to MPEC implementations. Finally, γ would collect all of the transition matrices for the observable components of the state. The counterfactual κ could be a counterfactual choice probability, a welfare measure, or a measure of expected payoffs. Section 3 shows how to implement our procedure in DDC models and presents a numerical example.

In the preceding description of a structural modeling exercise, the distribution F_* is being used for estimation of the structural parameter θ and for computation of the counterfactual κ. A natural question that arises is: to what extent are the counterfactuals driven by the researcher's choice F_*, and to what extent do they rely on the underlying structure of the model? The challenge is to address this question in a way that remains computationally tractable for empirically relevant structural models.

We address this question as follows. Let $\mathcal{U}$ denote the support of U (equipped with its Borel σ-algebra), let $\mathcal{F}$ denote the set of all probability distributions on $\mathcal{U}$, and let $\mathcal{N} \subset \mathcal{F}$ denote a neighborhood of F_*. The neighborhood $\mathcal{N}$ will be nonparametric in our analysis: it will consist of all probability distributions F that are within some well-defined "distance" of F_*. If this class seems too large, the researcher may further disciple the class of distributions by incorporating shape constraints, smoothness restrictions, or location/scale normalizations within the moment conditions (1)–(4); see Section 2.4.

The object of interest is the interval

$$\left[\underline{\kappa}(\mathcal{N})\,,\,\bar{\kappa}(\mathcal{N})\right],$$

where, in the explicit-dependence case

$$\begin{aligned}
\underline{\kappa}(\mathcal{N}) &= \inf_{\theta\in\Theta, F\in\mathcal{N}} \mathbb{E}^F[k(U,\theta,\gamma_0)] \quad \text{subject to (1)–(4) holding at } (\theta, F)\,, \text{ and} \\
\bar{\kappa}(\mathcal{N}) &= \sup_{\theta\in\Theta, F\in\mathcal{N}} \mathbb{E}^F[k(U,\theta,\gamma_0)] \quad \text{subject to (1)–(4) holding at } (\theta, F)
\end{aligned}$$

are respectively the smallest and largest values of the counterfactual obtained by varying θ over its parameter space and F over the neighborhood $\mathcal{N}$ while respecting the model structure (1)–(4). Extreme counterfactuals are defined similarly in the implicit-dependence case, replacing $\mathbb{E}^F[k(U,\theta,\gamma_0)]$ in the above display with $k(\theta,\gamma_0)$.

Focusing on the extreme counterfactuals $\underline{\kappa}(\mathcal{N})$ and $\bar{\kappa}(\mathcal{N})$ has two advantages. First, it does not require point-identification of θ under F_* or any other candidate distribution F. Thus, it naturally accommodates models with point- or partially-identified structural parameters and counterfactuals, and we sidestep having to compute the full identified set of structural parameters. Moreover, like Reguant (2016), we avoid having to compute all equilibria in models with multiple equilibria, only those supporting the smallest and largest values of the counterfactual.

The main contribution of this paper is to allow researchers to conduct a *global sensitivity analysis* as the parametric assumption F_* is relaxed and F is allowed to vary over neighborhoods of various size, while other structural features of the model are maintained. Key to computational tractability of our procedure is how neighborhoods are defined, as outlined in the next subsection. Later, in Section 6, we will also describe a measure of the *local sensitivity* of the counterfactual with respect to F_*, which characterizes how counterfactuals vary over small neighborhoods of F_*.

2.2 Characterization via convex programming

This subsection gives an heuristic overview to fix ideas and notation. A formal justification is presented in Section 4.1.

Consider computing $\underline{\kappa}(\mathcal{N})$ and $\bar{\kappa}(\mathcal{N})$ using an *inner loop*, where the counterfactual is minimized or maximized over $F \in \mathcal{N}$ subject to the restrictions (1)–(4), and an *outer loop* optimizing over $\theta \in \Theta$. The inner loops are infinite-dimensional optimization problems. Their computational tractability hinges on how the neighborhood $\mathcal{N}$ is constructed. We follow the robustness literature in economics pioneered by Hansen and Sargent and define nonparametric neighborhoods in terms of a type of statistical divergence from F_*, thereby allowing the inner loops to be recast as low-dimensional convex programs.

We consider neighborhoods that are constrained by ϕ-divergence from F_*:

$$\mathcal{N}_\delta = \{F \in \mathcal{F} : D_\phi(F \| F_*) \le \delta\},$$

with

$$D_\phi(F\|F_*) = \begin{cases} \displaystyle\int \phi\left(\frac{\mathrm{d}F}{\mathrm{d}F_*}\right) \mathrm{d}F_* & \text{if } F \ll F_*\,, \\ +\infty & \text{otherwise,} \end{cases}$$

where $F \ll F_*$ denotes absolute continuity of F relative to F_*. Many default F_*, such as normal, type-I extreme value, logistic, Pareto, and Fréchet distributions have strictly positive (Lebesgue) density over $\mathcal{U}$, so the absolute continuity condition $F \ll F_*$ merely rules out F with mass points. The function $\phi : [0,\infty) \to \mathbb{R} \cup \{+\infty\}$ is a convex function representing a cost of departing from F_*. There are some easily verifiable technical conditions that the function ϕ must satisfy so as to be compatible with the model, which we describe formally in Section 4.1.

To give some examples of neighborhoods, the function $\phi(x) = x\log x - x + 1$ corresponds to Kullback–Leibler (KL) divergence:

$$\mathcal{N}_\delta = \left\{ F \in \mathcal{F} : \int \log\left(\frac{\mathrm{d}F}{\mathrm{d}F_*}\right) \mathrm{d}F \le \delta \right\} .$$

KL divergence is used extensively in the robustness literature. However, it does require that the moment functions $g_1, \ldots, g_4$ and counterfactual function k have quite thin tails under F_*, as discussed further in Section 4.1. Weaker moment conditions are required for the function $\phi(x) = \frac{x^p - 1 - p(x-1)}{p(p-1)}$ with index $p > 1$, which corresponds to neighborhoods constrained by Cressie–Read divergence (equivalently, by L^p divergence, α-divergence, or Renyi divergence). Choosing $p = 2$ yields neighborhoods constrained by χ^2 (or Pearson) divergence:

$$\mathcal{N}_\delta = \left\{ F \in \mathcal{F} : \frac{1}{2}\int \left(\frac{\mathrm{d}F}{\mathrm{d}F_*} - 1\right)^2 \mathrm{d}F_* \le \delta \right\} .$$

We also found it useful to work with a *hybrid* of KL and χ^2 divergence, whose corresponding ϕ function is

$$\phi(x) = \begin{cases} x\log x - x + 1 & x \le e\,, \\ \frac{1}{2e}(x-e)^2 + (x-e) + 1 & x > e\,, \end{cases}$$

where e denotes Euler's number. Hybrid divergence retains some attractive features of KL divergence but the technical conditions underlying the duality results are satisfied for a much broader class of models and benchmark distributions F_*.

In the explicit-dependence case, the smallest and largest values of the counterfactual obtained as we vary $F \in \mathcal{N}_\delta$ (subject to the restrictions imposed by the model) at any fixed θ are

$$\underline{\kappa}_\delta(\theta;\gamma_0,P_0) = \inf_{F\in\mathcal{N}_\delta} \mathbb{E}^F[k(U,\theta,\gamma_0)] \quad \text{subject to (1)–(4) holding at } (\theta,F)\,, \text{ and}$$

$$\bar{\kappa}_\delta(\theta;\gamma_0,P_0) = \sup_{F\in\mathcal{N}_\delta} \mathbb{E}^F[k(U,\theta,\gamma_0)] \quad \text{subject to (1)–(4) holding at } (\theta,F)$$

with the understanding that $\underline{\kappa}_\delta(\theta;\gamma_0,P_0)=+\infty$ and $\bar{\kappa}_\delta(\theta;\gamma_0,P_0)=-\infty$ if there exists no distribution in $\mathcal{N}_\delta$ for which (1)–(4) hold. Although these are infinite-dimensional optimization problems, their dual representations are low-dimensional convex programs.

We state the dual programs of $\underline{\kappa}_\delta(\theta;\gamma,P)$ and $\bar{\kappa}_\delta(\theta;\gamma,P)$ for now as a result; Section 4.1 provides a formal statement. Let $g=(g_1',\ldots,g_4')'$ denote the vector of moment functions and let $d=\sum_{i=1}^4 d_i$. Partition $\lambda\in\mathbb{R}^d$ as $\lambda=(\lambda_1',\lambda_2',\lambda_3',\lambda_4')'\in\Lambda:=\mathbb{R}_+^{d_1}\times\mathbb{R}^{d_2}\times\mathbb{R}_+^{d_3}\times\mathbb{R}^{d_4}$. The vector λ consists of the Lagrange multipliers on the moment conditions (1)–(4); the remaining multipliers are $\eta\in\mathbb{R}_+$ for the constraint on the condition $D_\phi(F\|F_*)\le\delta$ and $\zeta\in\mathbb{R}$ for the constraint that F must integrate to unity. Let $(\eta\phi)^\star(x)=\sup_{t\ge0:\eta\phi(t)<+\infty}(tx-\eta\phi(t))$ denote the convex conjugate of $\eta\phi$.

Result 2.1 *The dual programs of $\underline{\kappa}_\delta(\theta;\gamma,P)$ and $\bar{\kappa}_\delta(\theta;\gamma,P)$ are*

$$\underline{\kappa}_\delta^\star(\theta;\gamma,P)=\sup_{\eta\ge0,\zeta\in\mathbb{R},\lambda\in\Lambda}-\mathbb{E}^{F_*}\Big[(\eta\phi)^\star(-k(U,\theta,\gamma)-\zeta-\lambda'g(U,\theta,\gamma))\Big]-\eta\delta-\zeta-\lambda_{12}'P\,, \quad (5)$$

$$\bar{\kappa}_\delta^\star(\theta;\gamma,P)=\inf_{\eta\ge0,\zeta\in\mathbb{R},\lambda\in\Lambda}\mathbb{E}^{F_*}\Big[(\eta\phi)^\star(k(U,\theta,\gamma)-\zeta-\lambda'g(U,\theta,\gamma))\Big]+\eta\delta+\zeta+\lambda_{12}'P\,, \quad (6)$$

where $\lambda_{12}=(\lambda_1',\lambda_2')'$. In particular, for KL neighborhoods:

$$\underline{\kappa}_\delta^\star(\theta;\gamma,P)=\sup_{\eta>0,\lambda\in\Lambda}-\eta\log\mathbb{E}^{F_*}\Big[e^{-\eta^{-1}(k(U,\theta,\gamma)+\lambda'g(U,\theta,\gamma))}\Big]-\eta\delta-\lambda_{12}'P\,, \quad (7)$$

$$\bar{\kappa}_\delta^\star(\theta;\gamma,P)=\inf_{\eta>0,\lambda\in\Lambda}\eta\log\mathbb{E}^{F_*}\Big[e^{\eta^{-1}(k(U,\theta,\gamma)-\lambda'g(U,\theta,\gamma))}\Big]+\eta\delta+\lambda_{12}'P\,, \quad (8)$$

and for hybrid neighborhoods:

$$\underline{\kappa}_\delta^\star(\theta;\gamma,P)=\sup_{\eta>0,\zeta\in\mathbb{R},\lambda\in\Lambda}-\eta\mathbb{E}^{F_*}\Big[\Psi\Big(-\eta^{-1}\big(k(U,\theta,\gamma)+\zeta+\lambda'g(U,\theta,\gamma)\big)\Big)\Big]-\eta\delta-\zeta-\lambda_{12}'P\,, \quad (9)$$

$$\bar{\kappa}_\delta^\star(\theta;\gamma,P)=\inf_{\eta>0,\zeta\in\mathbb{R},\lambda\in\Lambda}\eta\mathbb{E}^{F_*}\Big[\Psi\Big(\eta^{-1}\big(k(U,\theta,\gamma)-\zeta-\lambda'g(U,\theta,\gamma)\big)\Big)\Big]+\eta\delta+\zeta+\lambda_{12}'P\,, \quad (10)$$

where

$$\Psi(x)=\begin{bmatrix} e^x-1 & x\le1\,,\\ \frac{e}{2}(x^2+1)-1 & x>1\,. \end{bmatrix} \quad (11)$$

The expectations in the dual programs are all under F_*, so $\underline{\kappa}_\delta^\star$ and $\bar{\kappa}_\delta^\star$ can be computed in closed form, as for the game example in Section 3, or otherwise numerically. Gradients and Hessians may also be available in closed form, facilitating fast optimization. Indeed, evaluation of $\underline{\kappa}_\delta^\star$ and $\bar{\kappa}_\delta^\star$ may be only marginally more computationally costly than evaluation of a criterion function for estimating θ.

By *weak duality*, the inequalities

$$\underline{\kappa}_\delta^\star(\theta;\gamma,P) \le \underline{\kappa}_\delta(\theta;\gamma,P)\,, \qquad\qquad \bar{\kappa}_\delta(\theta;\gamma,P) \le \bar{\kappa}_\delta^\star(\theta;\gamma,P)$$

always hold. Therefore, the dual problems always provide (possibly conservative) upper and lower bounds for the extreme counterfactuals at any given θ. We say that *strong duality* holds when the primal and dual problems are equal. Under a mild condition on the moments, strong duality may be verified via the convex program

$$\delta^\star(\theta;\gamma,P) = \sup_{\zeta\in\mathbb{R},\lambda\in\Lambda} -\mathbb{E}^{F_*}\left[\phi^\star(-\zeta-\lambda' g(U,\theta,\gamma))\right] - \zeta - \lambda_{12}' P\,. \tag{12}$$

In particular, for KL divergence

$$\delta^\star(\theta;\gamma,P) = \sup_{\lambda\in\Lambda} -\log\mathbb{E}^{F_*}\left[e^{-\lambda' g(U,\theta,\gamma)}\right] - \lambda_{12}' P\,, \tag{13}$$

and for hybrid divergence

$$\delta^\star(\theta;\gamma,P) = \sup_{\zeta\in\mathbb{R},\lambda\in\Lambda} -\mathbb{E}^{F_*}\left[\Psi\left(-\zeta-\lambda' g(U,\theta,\gamma)\right)\right] - \zeta - \lambda_{12}' P\,. \tag{14}$$

The expectations in the above displays are again under F_* and may therefore be computed in closed form for certain models or otherwise numerically.

When $\delta^\star(\theta;\gamma,P) < \infty$, the program $\delta^\star$ identifies a distribution $F_{\theta,\gamma,P}$ that minimizes $D_\phi(\cdot\|F_*)$ among all distributions satisfying the constraints (1)–(4) at (θ,γ,P). If $\delta^\star(\theta;\gamma,P) \le \delta$ then there exists at least one distribution satisfying (1)–(4) at (θ,γ,P) that belongs to $\mathcal{N}_\delta$. If the inequality is strict, strong duality holds under a mild constraint qualification condition and $\underline{\kappa}_\delta^\star$ and $\bar{\kappa}_\delta^\star$ give the correct values for the inner loop at θ. On the other hand, if $\delta^\star(\theta,\gamma,P) = \delta$, then $F_{\theta,\gamma,P}$ is the unique distribution in $\mathcal{N}_\delta$ satisfying the moment conditions at (θ,γ,P), in which case $\underline{\kappa}_\delta(\theta;\gamma,P) = \bar{\kappa}_\delta(\theta;\gamma,P) = \mathbb{E}^{F_{\theta,\gamma,P}}[k(U,\theta,\gamma)] = \mathbb{E}^{F_*}[m_{\theta,\gamma,P}(U)k(U,\theta,\gamma)]$ where $m_{\theta,\gamma,P}$ denotes the Radon–Nikodym derivative of $F_{\theta,\gamma,P}$ with respect to F_*. In particular, for KL divergence

$$m_{\theta,\gamma,P}(U) = \frac{e^{-\lambda_{\theta,\gamma,P}' g(U,\theta,\gamma)}}{\mathbb{E}^{F_*}[e^{-\lambda_{\theta,\gamma,P}' g(U,\theta,\gamma)}]}\,,$$

where $\lambda_{\theta,\gamma,P}$ solves (13), and for hybrid divergence

$$m_{\theta,\gamma,P}(U) = \dot{\Psi}\left(-\zeta_{\theta,\gamma,P} - \lambda_{\theta,\gamma,P}' g(U,\theta,\gamma)\right)\,,$$

where $(\zeta_{\theta,\gamma,P}, \lambda_{\theta,\gamma,P})$ solves (14) and

$$\dot{\Psi}(x) = \left[\begin{array}{ll} e^x & x \leq 1\,, \\ ex & x > 1\,. \end{array} \right.$$

2.3 Estimation

The preceding description provides roadmap for how to estimate the extreme counterfactuals, given first-stage estimates $(\hat{\gamma}, \hat{P})$ of (γ_0, P_0). In the explicit-dependence case (where k depends non-trivially on u), define the sample criterion functions

$$\underline{\hat{K}}_\delta(\theta) = \left[\begin{array}{l} \underline{\kappa}^\star_\delta(\theta; \hat{\gamma}, \hat{P}) \\ \mathbb{E}^{F_{\theta,\hat{\gamma},\hat{P}}}[k(U, \theta, \hat{\gamma})] \\ +\infty\,, \end{array} \right. \quad , \quad \hat{\overline{K}}_\delta(\theta) = \left[\begin{array}{ll} \bar{\kappa}^\star_\delta(\theta; \hat{\gamma}, \hat{P}) & \text{if } \delta^\star(\theta; \hat{\gamma}, \hat{P}) < \delta, \\ \mathbb{E}^{F_{\theta,\hat{\gamma},\hat{P}}}[k(U, \theta, \hat{\gamma})] & \text{if } \delta^\star(\theta; \hat{\gamma}, \hat{P}) = \delta, \\ -\infty & \text{if } \delta^\star(\theta; \hat{\gamma}, \hat{P}) > \delta. \end{array} \right.$$

Note that the knife-edge case $\delta^\star(\theta; \hat{\gamma}, \hat{P}) = \delta$ will almost certainly never occur if $\delta^\star$ is computed by numerical optimization. In the implicit-dependence case (where k does not depend on u), then the sample criterion functions simplify to

$$\underline{\hat{K}}_\delta(\theta) = \left[\begin{array}{l} k(\theta, \hat{\gamma}) \\ +\infty \end{array} \right. \quad , \quad \hat{\overline{K}}_\delta(\theta) = \left[\begin{array}{ll} k(\theta, \hat{\gamma}) & \text{if } \delta^\star(\theta; \hat{\gamma}, \hat{P}) \leq \delta, \\ -\infty & \text{if } \delta^\star(\theta; \hat{\gamma}, \hat{P}) > \delta. \end{array} \right.$$

The estimators of the smallest and largest counterfactuals are obtained by optimizing the criterion functions $\underline{\hat{K}}_\delta$ and $\hat{\overline{K}}_\delta$ with respect to θ:

$$\hat{\underline{\kappa}}(\mathcal{N}_\delta) = \inf_{\theta \in \Theta} \underline{\hat{K}}_\delta(\theta)\,, \quad \hat{\bar{\kappa}}(\mathcal{N}_\delta) = \sup_{\theta \in \Theta} \hat{\overline{K}}_\delta(\theta)\,.$$

If (γ_0, P_0) are known then we may to compute the smallest and largest counterfactuals $\underline{\kappa}(\mathcal{N}_\delta)$ and $\bar{\kappa}(\mathcal{N}_\delta)$ by simply replacing $(\hat{\gamma}, \hat{P})$ with (γ_0, P_0) in the above description.

In Section 4.2, we show that the plug-in estimators $\hat{\underline{\kappa}}(\mathcal{N}_\delta)$ and $\hat{\bar{\kappa}}(\mathcal{N}_\delta)$ are consistent estimators of the extreme counterfactuals $\underline{\kappa}(\mathcal{N}_\delta)$ and $\bar{\kappa}(\mathcal{N}_\delta)$ and we establish their joint asymptotic distribution. The distribution will typically be nonstandard, however, as extreme counterfactuals may occur under multiple distributions in $\mathcal{N}_\delta$ and at multiple structural parameter values. Nevertheless, one may still use subsampling or various modified bootstraps to perform inference on the extreme counterfactuals.

When δ is very small and the model is sufficiently regular, $\hat{\underline{\kappa}}(\mathcal{N}_\delta)$ and $\hat{\bar{\kappa}}(\mathcal{N}_\delta)$ may be approximated as $\hat{\kappa} - \sqrt{\hat{s}\delta}$ and $\hat{\kappa} + \sqrt{\hat{s}\delta}$, respectively, where $\hat{s}$ is an estimate of local sensitivity; see Section 6.

2.4 Shape restrictions

The moment conditions (1)–(4) may also contain further shape restrictions on F. Examples include: (i) location normalizations, e.g. $\mathbb{E}^F[U] = 0$ or $\mathbb{E}^F[1\!\!1\{U_i \leq 0\}] = \frac{1}{2}$ for each entry U_i of U for a median normalization; (ii) scale normalizations, e.g. $\mathbb{E}^F[U_i^2] = 1$ or $\mathbb{E}^F[1\!\!1\{U_i \leq a\}] - \mathbb{E}^F[1\!\!1\{U_i \leq -a\}] = b$ for a normalization of the inter-quantile range; (iii) covariance normalizations, e.g. $\mathbb{E}^F[UU'] = I$, or bounds, e.g. $\mathbb{E}^F[UU'] \leq \Sigma$; and (iv) smoothness, e.g. $\mathbb{E}^F[1\!\!1\{U_i \leq a_{k+1}\}] - \mathbb{E}^F[1\!\!1\{U_i \leq a_k\}] \leq C$ for $a_1 < a_2 < \ldots < a_K$. Researchers may add and remove shape restrictions as appropriate to investigate how such restrictions affect the sets of counterfactuals.

3 Numerical examples

3.1 Discrete game of complete information

To illustrate our procedure in a simple and transparent way, we consider a complete-information entry game similar to that studied by Bresnahan and Reiss (1990, 1991), Berry (1992), and Tamer (2003). Payoffs are described in Table 1.

		Firm 2: 0	Firm 2: 1
Firm 1	0	$(0, 0)$	$(0, \beta_2 + \beta z + U_2)$
	1	$(\beta_1 + \beta z + U_1, 0)$	$(\beta_1 + \beta z - \Delta + U_1, \beta_2 + \beta z - \Delta + U_2)$

Table 1: Payoff matrix for (Firm 1, Firm 2) when $Z = z$.

In this specification, $Z \in \{0, 1, 2\}$ denotes a market-specific regressor and $U = (U_1, U_2)'$ is a random vector representing unobserved (to the econometrician but not to the firms) profits which is distributed independently of Z. The structural parameters are $(\beta_1, \beta_2, \beta, \Delta)$ where β_1 and β_2 represent firm-specific fixed costs and Δ (assumed positive) represents a loss of profitability from competing as a duopolist. The solution concept is restricted to equilibria in pure strategies.

The econometrician observes choice probabilities of the four market structures (conditional on Z). A standard approach for estimation is to match the observed conditional choice probabilities with the model-implied conditional choice probabilities. As the model is incomplete—there are certain realizations of U for which there are two equilibria in pure strategies (Firm 1 enters and Firm 2 does not, or vice versa)—moment inequality methods are typically used so as to be robust to potential misspecification of the equilibrium selection mechanism.

Nevertheless, it is often the case in applied work that strong parametric assumptions are made about the distribution of payoff shocks to map the structural parameters θ into model-implied conditional choice probabilities. For example, Berry (1992) and Ciliberto and Tamer (2009) both assume U is distributed as bivariate normal.[4] Given the emphasis on robustness with respect to equilibrium selection, it seems natural to also question the sensitivity of counterfactuals to the researcher's assumed parametric distribution of unobservables.

Suppose the researcher wants to investigate the effect of a tax on market structure. Payoffs under the tax are presented in Table 2. We will focus on the conditional probability of observing a monopoly, though our approach could be used equally for other choice probabilities or measures of expected firm profits.

		Firm 2: 0	Firm 2: 1
Firm 1	0	$(0,0)$	$(0,\beta_2+\beta z-\tau+U_2)$
Firm 1	1	$(\beta_1+\beta z-\tau+U_1,0)$	$(\beta_1+\beta z-\Delta+U_1,\beta_2+\beta z-\Delta+U_2)$

Table 2: Counterfactual payoff matrix for (Firm 1, Firm 2) when $Z=z$.

Under the counterfactual payoffs described in Table 2, neither firm enters if $U_j \le \tau-\beta_j-\beta z$ for $j=1,2$ where $\tau>0$ is the tax, and both firms enter (duopoly) if $U_j \ge \Delta-\beta_j-\beta z$ for $j=1,2$. The moment condition defining the counterfactual of interest is therefore

$$\begin{aligned} k(U,\theta;z) = 1 - 1\!\!1\{&(U_1 \le \tau-\beta_1-\beta z;\, U_2 \le \tau-\beta_2-\beta z) \\ &\cup (U_1 \ge \Delta-\beta_1-\beta z;\, U_2 \ge \Delta-\beta_2-\beta z)\} \end{aligned}$$

and $\mathbb{E}^F[k(U,\theta;z)]$ is the probability of observing a monopoly under the tax when $Z=z$.

To describe the moment functions g_1 and g_2, let $p(0,0|z)$, $p(0,1|z)$, $p(1,0|z)$, and $p(1,1|z)$ denote the model-implied conditional choice probabilities for $z\in\{0,1,2\}$. The model predicts a unique equilibrium in pure strategies for $p(0,0|z)$ and $p(1,1|z)$, so there are six equality restrictions. We follow convention and construct (standard) inequalities for the conditional choice probabilities $p(0,1|z)$ and $p(1,0|z)$. This yields six equality restrictions and six inequality restrictions:

$$g_1(U,\theta)=\begin{bmatrix} -1\!\!1\{U_1\ge-\beta_1;U_2\le\Delta-\beta_2\} \\ -1\!\!1\{U_1\le\Delta-\beta_1;U_2\ge-\beta_2\} \\ -1\!\!1\{U_1\ge-\beta_1-\beta;U_2\le\Delta-\beta_2-\beta\} \\ -1\!\!1\{U_1\le\Delta-\beta_1-\beta;U_2\ge-\beta_2-\beta\} \\ -1\!\!1\{U_1\ge-\beta_1-2\beta;U_2\le\Delta-\beta_2-2\beta\} \\ -1\!\!1\{U_1\le\Delta-\beta_1-2\beta;U_2\ge-\beta_2-2\beta\} \end{bmatrix}, \quad g_2(U,\theta)=\begin{bmatrix} -1\!\!1\{U_1\le-\beta_1;U_2\le-\beta_2\} \\ -1\!\!1\{U_1\ge\Delta-\beta_1;U_2\ge\Delta-\beta_2\} \\ -1\!\!1\{U_1\le-\beta_1-\beta;U_2\le-\beta_2-\beta\} \\ -1\!\!1\{U_1\ge\Delta-\beta_1-\beta;U_2\ge\Delta-\beta_2-\beta\} \\ -1\!\!1\{U_1\le-\beta_1-2\beta;U_2\le-\beta_2-2\beta\} \\ -1\!\!1\{U_1\ge\Delta-\beta_1-2\beta;U_2\ge\Delta-\beta_2-2\beta\} \end{bmatrix}.$$

[4] There are some exceptions, e.g., Aradillas-Lopez (2011).

The reduced-form parameter $\hat{P}_1$ stacks the (negative of the) observed conditional probabilities of $(0,1)$ and $(1,0)$ and $\hat{P}_2$ stacks the (negative of the) observed conditional probabilities of $(0,0)$ and $(1,1)$. There is no auxiliary parameter (i.e., no γ) in this setting.

As a benchmark specification, we assume $U \sim N(0, I_2)$ under F_*. Note, however, that we will allow for arbitrary correlation between U_1 and U_2 as the neighborhoods expand. We impose a location normalization $\mathbb{E}^F[U] = 0$ using

$$g_4(U,\theta) = \left[\begin{array}{cc} U_1 & U_2 \end{array} \right]' .$$

We also impose a scale normalization by setting $\beta = 1$. Alternatively, one could impose a scale normalization on the variance or inter-quantile range of U_1 and U_2. In total, we have 6 inequality constraints and 8 equality constraints. The vector of structural parameters is $\theta = (\beta_1, \beta_2, \Delta)$.[5]

We define neighborhoods via KL divergence relative to F_*. Although the moment functions in g_4 are unbounded, this neighborhood definition is still compatible with the key regularity condition underlying our procedure (Assumption Φ in Section 4.1). With this notion of divergence, the criterion functions for the dual programs $\underline{\kappa}^\star$, $\bar{\kappa}^\star$, and $\delta^\star$ as well as the local sensitivity measure s can all be computed in closed form. Full details are deferred to Appendix A.1.

Suppose that the observed conditional probabilities of observing the various market structures (pre-intervention) are as described in Table 3.

Z	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$
0	0.619	0.003	0.226	0.152
1	0.175	0.075	0.450	0.300
2	0.013	0.427	0.335	0.225

Table 3: Observed conditional probabilities.

Matching the model-implied conditional choice probabilities under F_* to the observed conditional choice probabilities as above yields parameter estimates of $\hat{\beta}_1 = -0.70$, $\hat{\beta}_2 = -0.90$, and $\hat{\Delta} = 0.80$ (there are enough equality restrictions to point-identify these parameters under F_*). We will focus on the probability of observing a monopoly when $Z = 1$, where pre-intervention presence of monopolies is highest. Suppose that the tax is $\tau = 1.5$. The counterfactual probability of a monopoly under F_* at $\hat{\theta}$ is $\hat{\kappa} = 0.143$. Therefore, under the independent bivariate normal assumption, the model predicts the probability of observing a monopoly when $Z = 1$ will fall from 0.750 to 0.143 under the tax. How reliant is this prediction on the assumed distribution of unobserved profits?

[5] We could have allowed the duopoly cost to be Δ_1 and Δ_2 for firms 1 and 2, respectively. However, when $\Delta_1 - \tau < 0$ and $\Delta_2 - \tau > 0$ (or vice versa) there is no equilibrium in pure strategies for certain values of U for the game in Table 2. We therefore impose the restriction $\Delta_1 = \Delta_2 = \Delta$ and maintain pure-strategy equilibrium as the solution concept. One could alternatively consider equilibria in mixed strategies and modify the moment conditions accordingly.

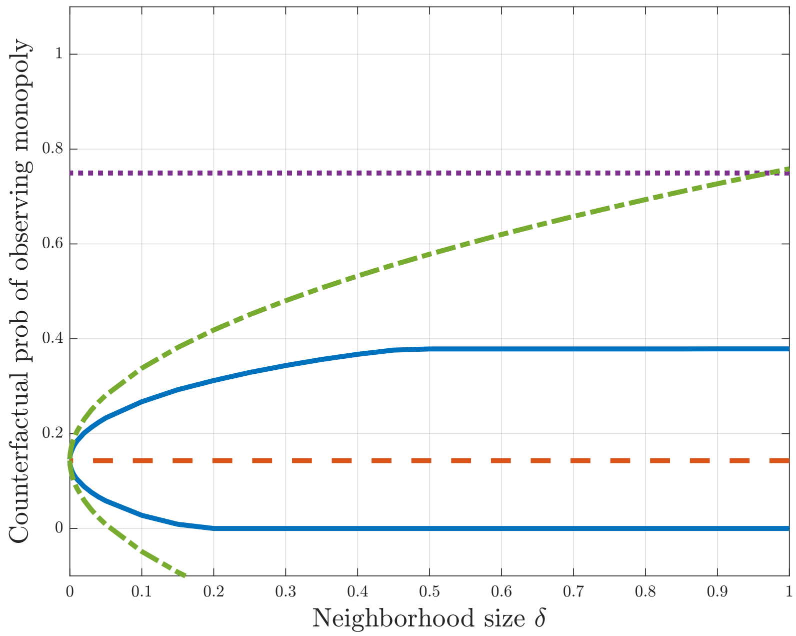

Figure 1: Solid lines: $\underline{\hat{\kappa}}(\mathcal{N}_\delta)$ and $\hat{\bar{\kappa}}(\mathcal{N}_\delta)$ for the counterfactual probability of observing a monopoly under a tax when $Z = 1$. Dashed line: estimated counterfactual under the $N(0, I_2)$ parametric benchmark. Dotted line: pre-tax probability of observing a monopoly. Dot-dashed line: bounds based on extrapolating a local sensitivity measure.

To answer this question, Figure 1 plots the upper and lower bounds of the set of counterfactuals over KL neighborhoods of F_* of various sizes. These bounds represent the smallest and largest values of the counterfactual that can be obtained under a distribution in $\mathcal{N}_\delta$ which is constrained so that (i) the model explains the pre-intervention conditional choice probabilities in Table 3, and (ii) the mean-zero restriction on U holds. As can be seen, the bounds are equal to the counterfactual under F_* when $\delta = 0$. As δ increases the bounds expand until they span the interval $[0, 0.379]$, which represents the identified set of the counterfactual probability of observing a monopoly when $Z = 1$.

To interpret the neighborhood size δ, consider the benchmark specification where F_* is $N(0, I_2)$. A shift in the fix cost parameters $(\beta_1, \beta_2)'$ could be offset by shifting the mean from 0 to μ, say. The KL divergence between the $N(0, I_2)$ distribution and $N(\mu, I_2)$ distribution is $\frac{1}{2}\|\mu\|^2$. So, a neighborhood of size $\delta = \frac{1}{2}$ would contain distributions that are as far from F_* as distributions that shift fixed costs of one of the firms by one unit of profits. Similarly, neighborhoods of size $\delta = 0.125$ contain distributions as far from F_* as distributions that shift fixed costs of one of the firms by half a unit of profits. With this in mind, we see that the lower bound of 0 is achieved by $\delta = 0.2$. Yet there are distributions equally close to F_* under which the probability of observing a monopoly would be around 0.32. The neighborhoods continue to expand upwards until the upper bound of 0.379 is achieved, slightly over half the probability of observing a monopoly without the tax.

Figure 1 also plots the bounds that are obtained by extrapolating a measure of local sensitivity of the counterfactual with respect to F_* (see Section 6 and Appendix A.1 for computational details). Extrapolation gives a good approximation when δ is very small. However, the approximation breaks down outside of very small neighborhoods. The reason is that at F_* none of the moment equality restrictions are binding. Thus, none of the inequality restrictions are relevant in characterizing local sensitivity. The moment inequality restrictions are, however, relevant outside of very small neighborhoods of F_*. These inequality restrictions further constrain the sets of distributions and, therefore, the set of counterfactuals.

The preceding exercise could be repeated under further restrictions on the parameter space Θ or by imposing additional shape constraints on the class of distributions F. One can then explore the extent to which these restrictions further sharpen the identified sets for counterfactuals.

3.2 Dynamic discrete choice

As a second numerical illustration, we consider a dynamic discrete choice (DDC) model following Rust (1987). The DDC literature is extensive; see Aguirregabiria and Mira (2010) and Arcidiacono and Ellickson (2011) for surveys. The basic setup of a discrete-time, infinite-horizon model is as follows. At each date t, agent i chooses $a_{i,t} \in \{0, 1, \ldots, n_a\}$ so as to maximize discounted expected payoffs. There is an observable state $x_{i,t} \in \{1, 2, \ldots, n_x\}$ which evolves as a controlled Markov process with transition kernel M. The agent's problem can be summarized by the ex-ante value function

$$V(x_{i,t}) = \mathbb{E}^F \left[\max_a \left(\pi(a, x_{i,t}; \theta_\pi) + U_{i,t}(a) + \beta \mathbb{E}^M [V(x_{i,t+1}) | x_{i,t}, a] \right) \right] ,$$

where π is a deterministic per-period payoff indexed by parameters θ_π, $\beta \in (0, 1)$ is a discount factor, and $U_{i,t} = (U_{i,t}(0), \ldots, U_{i,t}(n_a))'$ is a $\mathbb{R}^{n_a+1}$-valued vector of latent (to the econometrician) payoff shocks that are independently (of $x_{i,t}$) and identically distributed with distribution F. The testable implications of the model are summarized by the conditional choice probabilities (CCPs)

$$\begin{aligned} p(a|x_{it}) = \mathbb{E}^F \Bigg[\mathbb{1} \Bigg\{ & \pi(a, x_{i,t}; \theta_\pi) + U_{i,t}(a) + \beta \mathbb{E}^M [V(x_{i,t+1}) | x_{i,t}, a] \\ & \geq \max_{a'} \left(\pi(a', x_{i,t}; \theta_\pi) + U_{i,t}(a') + \beta \mathbb{E}^M [V(x_{i,t+1}) | x_{i,t}, a'] \right) \Bigg\} \Bigg] , \quad a = 0, \ldots, n_a \, . \end{aligned}$$

Researchers typically observe panel data on $(x_{i,t}, a_{i,t})$ and estimate structural parameters assuming the latent utility shocks have a particular parametric distribution, say F_*.[6] A standard parametric assumption used in the estimation procedures of Rust (1987) and Hotz and Miller (1993) is that

[6] A few papers study estimation without parametric assumptions on the distribution of payoff shocks. See Norets and Tang (2014) for the case of models with finite states, as above, and Blevins (2014), Chen (2017), and Buchholz, Shum, and Xu (2018) for models with continuous states.

the payoff shocks are i.i.d. type-I extreme value. This assumption is motivated by computational considerations, as it leads to closed-form expressions for expectations of maxima. Given an estimate of θ_π, the dynamic program can be solved again under counterfactual changes to the environment to investigate the effect on choice probabilities, welfare, or other quantities of interest. As before, we see that the researcher's parametric assumption F_* plays a role both at the estimation stage and again when solving the dynamic program to compute counterfactuals.

To map this setup into our framework, identify V with a n_x-vector v solving the moment condition

$$\mathbb{E}^{F_*}\left[\max_a \Big(\pi(a;\theta_\pi) + U_{i,t}(a) + \beta M_a v\Big) - v\right] = 0\,, \tag{15}$$

where $\pi(a;\theta_\pi) = (\pi(a,1;\theta_\pi),\ldots,\pi(a,n_x;\theta_\pi))'$, M_a is the $n_x \times n_x$ transition matrix representing $M(x_{i,t+1}|x_{i,t},a)$, and the maximum is applied row-wise. The CCPs $p(a|x)$ may be identified with a n_x-vector p_a which solves the moment condition

$$\mathbb{E}^{F_*}\left[1\!\!1\Big\{\pi(a;\theta_\pi) + U_{i,t}(a) + \beta M_a v \geq \max_{a'}\big(\pi(a';\theta_\pi) + U_{i,t}(a') + \beta M_{a'} v\big)\Big\}\right] = p_a\,, \tag{16}$$

where the maximum, inequality, and indicator function are all applied row-wise. Consider a counterfactual transforming payoffs π to $\tilde\pi$ and/or transition probabilities M_a to $\tilde M_a$. The value function $\tilde V$ under the counterfactual may be identified with a n_x-vector $\tilde v$ solving the moment condition

$$\mathbb{E}^{F_*}\left[\max_a \Big(\tilde\pi(a;\theta_\pi) + U_{i,t}(a) + \beta \tilde M_a \tilde v\Big) - \tilde v\right] = 0\,, \tag{17}$$

where $\tilde\pi(a;\theta_\pi) = (\tilde\pi(a,1;\theta_\pi),\ldots,\tilde\pi(a,n_x;\theta_\pi))'$.

In the notation of Section 2, there are no inequalities so g_1, g_3, and P_1 are vacuous. The function g_2 collects the model-implied conditional choice probabilities from equation (16) for $a = 1,\ldots,n_a$ and the vector $\hat P_2 = (\hat p_1',\ldots,\hat p_{n_a}')'$ collects the estimated CCPs. The auxiliary parameter is $\hat\gamma = (\hat M_0,\ldots,\hat M_{n_a})$. Both $\hat P_2$ and $\hat\gamma$ are computed from the panel of data on $(x_{i,t}, a_{i,t})$. The function g_4 collects the fixed-point equations (15) and (17) for v and $\tilde v$, respectively. Finally, $\theta = (\theta_\pi', v', \tilde v')'$. Our approach to augmenting the parameter space and adding equilibrium conditions for v and $\tilde v$ is somewhat similar to MPEC implementations. As our methods are robust to partial identification, one could also treat β as unknown and include it as a component of θ. In the example below we follow much of the literature and treat β as known, however. One could add further moment conditions embodying location/scale normalizations or smoothness restrictions to g_4, as described in Section 2.4. Unlike the game example, we do not impose a mean-zero normalization on U here.

We use a dynamic model of monopolist entry and exit from Kalouptsidi, Scott, and Souza-Rodrigues (2017) as a numerical example. Each period a monopolist decides whether to participate ($a = 1$) or not participate ($a = 0$) in a market. The observed state is $x_{i,t} = (s_{i,t}, a_{i,t-1})'$ where $s_{i,t}$ is a

$p(1\|x)$	$(H,0)$	$(M,0)$	$(L,0)$	$(H,1)$	$(M,1)$	$(L,1)$
Estimated	0.9361	0.8748	0.7299	0.9999	0.8091	0.0048
Counterfactual	0.9495	0.9027	0.8033	0.9999	0.6959	0.0029

Table 4: Estimated CCPs ($\hat{P}_2$) and counterfactual CCPs ($\hat{\kappa}$) under F_*.

market-level variable (high, medium or low) which evolves exogenously. Payoffs are

$$\pi(a, x_{i,t}; \theta_\pi) = \left[\begin{array}{cl} -ac_e & \text{if } a_{i,t-1} = 0\,, \\ a(\pi_1(s_{i,t};\theta_\pi)) + (1-a)c_s & \text{if } a_{i,t-1} = 1\,, \end{array} \right.$$

where c_e is an entry cost, c_s is scrap value, and variable profits $\pi_1(s;\theta_\pi)$ are

$$\pi_1(s;\theta_\pi) = \frac{(x(s) - c_m)^2}{4c_d} - c_f\,,$$

where c_f is a fixed cost, c_m is the monopolist's constant marginal cost, and $x(s)$ and c_d are the intercept the slope, respectively, of the linear demand curve faced by the monopolist in state s. As in Kalouptsidi et al. (2017), we take $x(s) = 20$, 17, and 12 in the high, medium, and low states, respectively. We also normalize $\beta = 0.95$ and $\phi_s = 10$, so $\theta_\pi = (c_d, c_e, c_f, c_m)$.

Suppose we observe a panel of data on $(x_{i,t}, a_{i,t})$ from which we estimate CCPs (see Table 4) and transition law q for s (the matrices M_0 and M_1 are known up to q):

$$\hat{q}(s_{t+1}|s_t) = \begin{bmatrix} 0.40 & 0.35 & 0.25 \\ 0.30 & 0.40 & 0.30 \\ 0.20 & 0.20 & 0.60 \end{bmatrix},$$

where the first row/column correspond to the high state and the third to the low state. Fixing F_* so that the payoff shocks are type-I extreme value, we invert the estimated CCPs to obtain estimates $\hat{\theta}_\pi = (11.0, 9.0, 5.5, 1.5)$.

We wish to investigate the effect of a subsidy that reduces the cost of entry by 0.9 units, a reduction of 10%. Counterfactual CCPs displayed in Table 4 are obtained by solving the model using $(\hat{\theta}_\pi, F_*)$. How reliant are the counterfactual CCPs on the assumed distribution of payoff shocks?

To investigate sensitivity, we apply our procedure over neighborhoods of F_* of various sizes. We focus on the first two counterfactual CCPs in Table 4, which represent the conditional probabilities of an inactive monopolist entering the market when the market-level state is high and medium, respectively. As the type-I extreme value distribution has slightly fatter tails than the normal distribution used in the game example, here we constrain neighborhoods by hybrid divergence, rather than KL divergence, so as to satisfy the key regularity condition underlying our procedure

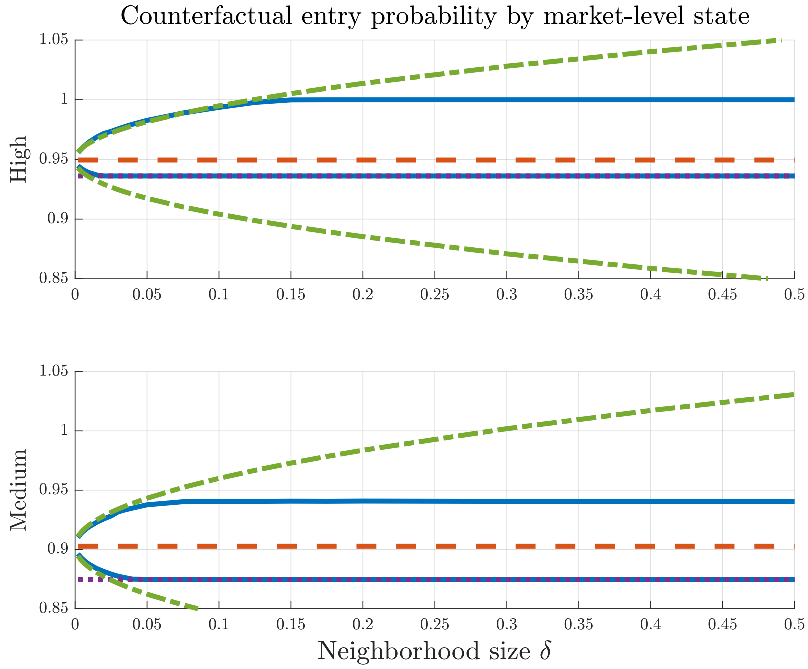

Figure 2: Solid lines: $\hat{\underline{\kappa}}(\mathcal{N}_\delta)$ and $\hat{\bar{\kappa}}(\mathcal{N}_\delta)$ for the entry CCP (i.e., $P(1|s, a = 0)$) by market-level state. Dashed line: estimated counterfactual under the type-I extreme value parametric benchmark. Dotted line: observed CCP without the subsidy. Dot-dashed line: bounds based on extrapolating a local sensitivity measure.

(Assumption Φ in Section 4.1).[7] Figure 2 plots the smallest and largest counterfactuals that could be obtained under distributions in $\mathcal{N}_\delta$ that explain the CCPs in the first line of Table 4. Both figures show that the counterfactual CCPs are bounded below by the CCPs without the subsidy. The counterfactual CCP in the medium state is constrained above, however, showing that the model has some structure that disciplines the set of counterfactual predictions.

To interpret the neighborhood size δ, the hybrid divergence between a type-I extreme value distribution with mean zero and type-I extreme value distribution with location shifted by μ units behaves like $\frac{1}{2}\mu^2$ for small values of μ. So, a neighborhood of size $\delta = 0.10$ contains distributions that are as far from F_* as distributions that have a location shift of around 0.45 profit units, a magnitude equivalent to half the size of the subsidy. Similarly, neighborhoods of size $\delta = 0.025$ contain distributions that are as far from F_* as distributions that have a location shift around one quarter the size of the subsidy. With this in mind, we see that the lower bound for the counterfactual CCP in the high state is achieved by around $\delta = 0.025$, the lower and upper bounds for the CCPs in the medium state are attained by around $\delta = 0.05$, and the upper bound of 1 for the CCP in the high

[7] The moment conditions g_4 grow like $\|U\|$ for large values of $\|U\|$. Compatibility with KL divergence requires that $\mathbb{E}^{F_*}[e^{c\|U\|}] < \infty$ for each $c > 0$; see Section 4.1. The type-I extreme value distribution does not satisfy this condition as it tails are too thick. It does, however, satisfy the weaker condition $\mathbb{E}^{F_*}[\|U\|^2] < \infty$ required by hybrid divergence.

state is achieved by around $\delta = 0.15$. The bounds for the CCP in the high state expand more slowly than for the medium state, indicating that this counterfactual CCP is relatively less sensitive to specification of F_*. After $\delta = 0.15$, the bounds for both counterfactual CCPs are stable and equal to the limits of the endpoints of their respective identified sets.

Figure 2 also plots the bounds that are obtained by extrapolating a local sensitivity measure (see Section 6 and Appendix A.2 for computational details). The measure is $\hat{s} = 0.020$ for the counterfactual CCP in the high state and $\hat{s} = 0.035$ for the medium state, showing again that the counterfactual CCP in the high state is relatively less sensitive to specification of F_*. As with the game example, extrapolation gives a good approximation when δ is very small. Outside of small neighborhoods, however, these sets fail to capture the true nature of the identified sets of counterfactual CCPs. One reason for this is that extrapolation produces bounds that are symmetric around the counterfactual at F_*, whereas the upper panel of Figure 2 reveals asymmetries in the true set of counterfactuals obtained using our procedure.

We close this section by comparing our approach with Norets and Tang (2014), who study identified sets of CCPs in dynamic binary choice models without parametric assumptions on F. They eliminate the nuisance distribution via a reparameterization and a linear program, which is solved for each θ_π-counterfactual CCP pair. If one's aim is to recover the identified set of counterfactual CCPs, then their approach is computationally lighter than ours as it uses linear programming and involves no numerical integration. However, their analysis is more restrictive than ours, as it is specific to counterfactual CCPs in dynamic binary choice models whereas we allow for general counterfactuals (e.g. welfare measures, expected profits, and so on) in dynamic multinomial choice models. We also accommodate a broader range of location/scale normalizations and shape restrictions.

4 Theory

In this section we first outline some theoretical results which underlie the dual representation of the extreme counterfactuals and justify our choice of criterion function. We then go on to discuss estimation and inference results.

4.1 Duality

To ensure the dual representations are well defined, we impose two easily verifiable conditions on the cost function ϕ and its compatibility with the thickness of the tails of the distribution of $g_1, \ldots, g_4$ and k under F_*.

To introduce the condition, let $\psi(x) = \phi^\star(x) - x$ where $\phi^\star(x) = \sup_{t \geq 0 : \phi(t) < +\infty}(tx - \phi(t))$ is the convex conjugate of ϕ. Define

$$\mathcal{E} = \{f : \mathcal{U} \to \mathbb{R} \text{ such that } \mathbb{E}^{F_*}[\psi(c|f(U)|)] < \infty \text{ for all } c > 0\}.$$

The space $\mathcal{E}$ is an Orlicz class of functions. Further details are deferred to Appendix C. For now, we just compare $\mathcal{E}$ with L^p spaces, namely $L^p(F_*) = \{f : \mathcal{U} \to \mathbb{R} \text{ such that } \mathbb{E}^{F_*}[|f(U)|^p] < \infty\}$ for $1 \leq p < \infty$ and the space of (essentially) bounded functions when $p = \infty$. For KL neighborhoods, we have $\mathcal{E} = \{f : \mathcal{U} \to \mathbb{R} : \mathbb{E}^{F_*}[e^{c|f(U)|}] < \infty \text{ for all } c > 0\}$, so $L^\infty(F_*) \subset \mathcal{E} \subset L^p(F_*)$ for each $p < \infty$. For ϕ corresponding to Cressie–Read divergence with exponent $p > 1$, we have $\mathcal{E} = L^q(F_*)$ where $\frac{1}{p} + \frac{1}{q} = 1$. Moreover, $\mathcal{E} = L^2(F_*)$ for hybrid and χ^2 divergence.

Assumption Φ

(i) $\phi : [0, \infty) \to \mathbb{R} \cup \{+\infty\}$ is twice continuously differentiable on $(0, \infty)$ and strictly convex, $\phi(1) = \phi'(1) = 0$, $\phi(0) < +\infty$, $\lim_{x \to \infty} x^{-1}\phi(x) = +\infty$, $\lim_{x \downarrow 0} \phi'(x) < 0$, $\lim_{x \to \infty} \phi'(x) > 0$, and $\lim_{x \to \infty} x\phi'(x)/\phi(x) < +\infty$.

(ii) $k(\cdot; \theta, \gamma)$ and each entry of $g(\cdot; \theta, \gamma)$ belong to $\mathcal{E}$ for each $\theta \in \Theta$ and $\gamma \in \Gamma$.

Assumption Φ(i) is satisfied by functions inducing KL, Cressie–Read, χ^2, and hybrid divergence, among many others. However, it rules out $\phi(x) = -\log x + x - 1$ used in empirical likelihood. In view of Assumption Φ(i), we may extend the domain of ϕ so that $\phi(x) = +\infty$ if $x < 0$.

Assumption Φ(ii) describes a trade-off between the tail-thickness of $F \in \mathcal{N}_\delta$ and the growth of k and $\|g\|$ which ensures the expectations in (1)–(4) are well defined.[8] As k, g, and F_* are all specified by the researcher, this condition is easily verified. In particular, Assumption Φ(ii) holds for KL, Cressie–Read, χ^2, and hybrid divergence under a strong Cramér-type condition, namely that $\mathbb{E}^{F_*}[e^{c|k(U,\theta,\gamma)|}]$ and $\mathbb{E}^{F_*}[e^{c\|g(U,\theta,\gamma)\|}]$ are finite for all $c > 0$ and all (θ, γ). This condition is satisfied, e.g., when (a) k and g are bounded, or (b) k and g are additively separable in U and the researcher's choice F_* is a sufficiently thin-tailed distribution, e.g. normal. If this Cramér-type condition fails, then a stronger divergence than KL must be used. For instance, if k and each entry of g all have finite second moments under F_* then we may use χ^2 or hybrid divergence.

Remark 4.1 *There is a tradeoff between the class of distributions over unobservables and the thickness of the tails of moment functions. A key innovation of Schennach (2014) is to allow for moment functions that are unbounded. Schennach (2014) uses ϕ corresponding to KL divergence but adjusts the reference measure (F_* in our notation) appropriately. In contrast, our objective is to*

[8] Assumption Φ is similar to conditions justifying duality results in generalized empirical likelihood estimation (see, e.g., Komunjer and Ragusa (2016)) where F_* would be replaced by the data-generating probability measure.

examine sensitivity to departures from the researcher's default choice F_*. We therefore keep F_* as the reference measure but modify ϕ appropriately. Li (2018) uses truncation to extend the framework of Ekeland et al. (2010) to accommodate unbounded moments, which introduces a tuning parameter. Rather than modify the model through truncation, our approach restricts attention to sufficiently thin-tailed classes of distributions. As we show in Section 5, these classes are sufficiently rich that our procedure delivers sharp bounds on the identified set of counterfactuals over large neighborhoods.

A sufficient condition for strong duality is the following Slater condition. Let 0_{d_i} denote a $d_i \times 1$ vector of zeros and $\mathcal{C} = \mathbb{R}^{d_1}_+ \times \{0_{d_2}\} \times \mathbb{R}^{d_3}_+ \times \{0_{d_4}\}$. Also let $\mathcal{N}_\infty = \{F \in \mathcal{F} : D_\phi(F\|F_*) < \infty\}$.

Definition 4.1 Condition S *holds at* (θ, γ, P) *if* $(P', 0'_{d_3+d_4})' \in \mathrm{int}(\{\mathbb{E}^F[g(U,\theta,\gamma)] : F \in \mathcal{N}_\infty\} + \mathcal{C})$.

Condition S requires that there exist F "in the interior" of $\mathcal{N}_\infty$ under which conditions (1)–(4) hold at (θ, γ, P). If there are no inequalities, then Condition S reduces to $(P_2', 0'_{d_4})' \in \mathrm{int}(\{\mathbb{E}^F[g(U,\theta,\gamma)] : F \in \mathcal{N}_\infty\})$. If there are only inequalities, then Condition S holds if there exists a distribution $F \in \mathcal{N}_\infty$ such that $\mathbb{E}^F[g(U,\theta,\gamma)] < (P_1', 0'_{d_3})'$ (where the strict inequality holds element-wise). It is straightforward to relax condition S by replacing "interior" with "relative interior" so as to accommodate moment functions that become linearly dependent at certain parameter values. Indeed, the dual representations all remain valid under this weaker condition. However, weakening the condition somewhat complicates the derivation of the estimation and inference results.

Lemma 4.1 *Let Assumption Φ hold. Then: the duals of $\underline{\kappa}_\delta(\theta;\gamma,P)$ and $\bar{\kappa}_\delta(\theta;\gamma,P)$ in the explicit-dependence case are the programs $\underline{\kappa}^\star_\delta(\theta;\gamma,P)$ and $\bar{\kappa}^\star_\delta(\theta;\gamma,P)$ defined in equations (5) and (6). If Condition S also holds at (θ,γ,P) and $\delta^\star(\theta;\gamma,P) < \delta$, then: the supremum can be taken over $(\eta,\zeta,\lambda) \in (0,\infty) \times \mathbb{R} \times \Lambda$, and the dual programs reduce to the programs in equations (7) and (8) for KL neighborhoods and (9) and (10) for hybrid neighborhoods.*

Recall that $F_{\theta,\gamma,P}$ minimizes $D_\phi(F\|F_*)$ subject to the restrictions (1)–(4) at (θ,γ,P). In the explicit-dependence case, define

$$\underline{K}_\delta(\theta;\gamma,P) = \left[\begin{array}{l} \underline{\kappa}^\star_\delta(\theta;\gamma,P) \\ \mathbb{E}^{F_{\theta,\gamma,P}}[k(U,\theta,\gamma)] \\ +\infty\,, \end{array}\right. , \quad \overline{K}_\delta(\theta;\gamma,P) = \left[\begin{array}{ll} \bar{\kappa}^\star_\delta(\theta;\gamma,P) & \text{if } \delta^\star(\theta;\gamma,P) < \delta, \\ \mathbb{E}^{F_{\theta,\gamma,P}}[k(U,\theta,\gamma)] & \text{if } \delta^\star(\theta;\gamma,P) = \delta, \\ -\infty & \text{if } \delta^\star(\theta;\gamma,P) > \delta, \end{array}\right.$$

where $\delta^\star$ is defined in equation (12). In the implicit-dependence case, define

$$\underline{K}_\delta(\theta;\gamma,P) = \left[\begin{array}{l} k(\theta,\gamma) \\ +\infty\,, \end{array}\right. , \quad \overline{K}_\delta(\theta;\gamma,P) = \left[\begin{array}{ll} k(\theta,\gamma) & \text{if } \delta^\star(\theta;\gamma,P) \le \delta, \\ -\infty & \text{if } \delta^\star(\theta;\gamma,P) > \delta. \end{array}\right.$$

With this notation, the sample criterion function for the smallest and largest counterfactuals are $\hat{\underline{K}}_\delta(\theta) = \underline{K}_\delta(\theta; \hat\gamma, \hat P)$ and $\hat{\overline{K}}_\delta(\theta) = \overline{K}_\delta(\theta; \hat\gamma, \hat P)$. We sometimes abbreviate the population criterion functions to $\underline{K}_\delta(\theta) := \underline{K}_\delta(\theta; \gamma_0, P_0)$ and $\overline{K}_\delta(\theta) := \overline{K}_\delta(\theta; \gamma_0, P_0)$.

Lemma 4.2 *Let Assumption* Φ *hold. Then:*

(i) If $\delta^\star(\theta; \gamma, P) > \delta$*, then: there does not exist* $F \in \mathcal{N}_\delta$ *satisfying conditions (1)–(4) at* (θ, γ, P)*.*

(ii) If Condition S holds at (θ, γ, P)*, then:* $\underline{\kappa}_\delta(\theta; \gamma, P) = \underline{K}_\delta(\theta; \gamma, P)$ *and* $\bar\kappa_\delta(\theta; \gamma, P) = \overline{K}_\delta(\theta; \gamma, P)$*.*

Lemma 4.2 justifies the description of the inner loop in Section 2. In particular, if Condition S holds at (θ, γ_0, P_0) for every $\theta \in \Theta$ with $\delta^\star(\theta; \gamma_0, P_0) \le \delta$, then

$$\underline{\kappa}(\mathcal{N}_\delta) = \inf_{\theta \in \Theta} \underline{K}_\delta(\theta)\,, \quad \bar\kappa(\mathcal{N}_\delta) = \sup_{\theta \in \Theta} \overline{K}_\delta(\theta)\,.$$

When introducing the estimators $\hat{\underline{\kappa}}(\mathcal{N}_\delta)$ and $\hat{\bar\kappa}(\mathcal{N}_\delta)$ earlier in Section 2, it was argued that the case $\delta^\star(\theta; \hat\gamma, \hat P) = \delta$ could effectively be ignored because of numerical optimization error. This knife-edge case can also be ignored at a population level under mild conditions. Equip $\mathcal{E}$ with the norm

$$\|f\|_\psi = \inf_{c>0} \frac{1}{c}(1 + \mathbb{E}^{F_*}[\psi(c|f(Z)|)])\,.$$

For χ^2 and hybrid neighborhoods, the norm $\|\cdot\|_\psi$ is equivalent to the $L^2(F_*)$ norm. A class of functions $\{g_\alpha : \alpha \in \mathcal{A}\} \subset \mathcal{E}$ is $\mathcal{E}$-continuous in α if $\|g_{\alpha_1} - g_{\alpha_2}\|_\psi \to 0$ as $\alpha_1 \to \alpha_2$ ($\mathcal{A}$ is a metric space). For χ^2 and hybrid neighborhoods this notion of continuity is equivalent to $L^2(F_*)$ continuity, i.e., $\mathbb{E}^{F_*}[(g_{\alpha_1}(U) - g_{\alpha_2}(U))^2] \to 0$ as $\alpha_1 \to \alpha_2$. Let $\Theta_\delta = \{\theta \in \Theta : \delta^\star(\theta, \gamma_0, P_0) < \delta\}$.

Assumption M

(i) $k(\cdot; \theta, \gamma)$ and each entry of $g(\cdot; \theta, \gamma)$ are $\mathcal{E}$-continuous in (θ, γ)

(ii) Θ_δ is nonempty and Condition S holds at (θ, γ_0, P_0) for each $\theta \in \Theta_\delta$

(iii) $\mathrm{cl}(\Theta_\delta) \supseteq \{\theta : \delta^\star(\theta; \gamma_0, P_0) \le \delta\}$.

Assumption M(i) may be verified under continuity conditions on k and g. If k and g each consist of indicator functions of events, Assumption M(i) holds provided the probabilities of the events under F_* are continuous in (θ, γ). In the implicit-dependence case, Assumption M(i) requires that k is continuous in (θ, γ). We only require $\mathcal{E}$-continuity in θ at γ_0 for the results in this subsection; the results in the next subsection require continuity in (θ, γ). Assumption M(i) reduces to $\mathcal{E}$-continuity in θ for models with no auxiliary parameter (i.e., no γ). Nonemptyness of Θ_δ is always satisfied when the model is correctly specified under F_*, i.e., there exists a $\theta \in \Theta$ solving (1)–(4) under F_*.

Assumption M(iii) is made for convenience and can be relaxed.[9] This condition simply ensures that there do not exist values of θ at which $\delta^\star(\theta;\gamma_0,P_0)=\delta$ but that are separated from Θ_δ.

Lemma 4.3 *Let Assumptions* Φ *and M hold. Then:*

$$\underline{\kappa}(\mathcal{N}_\delta)=\inf_{\theta\in\Theta_\delta}\underline{\kappa}^\star_\delta(\theta;\gamma_0,P_0)\,,\quad \bar{\kappa}(\mathcal{N}_\delta)=\sup_{\theta\in\Theta_\delta}\bar{\kappa}^\star_\delta(\theta;\gamma_0,P_0)$$

in the explicit-dependence case, and

$$\underline{\kappa}(\mathcal{N}_\delta)=\inf_{\theta\in\Theta_\delta}k(\theta;\gamma_0)\,,\quad \bar{\kappa}(\mathcal{N}_\delta)=\sup_{\theta\in\Theta_\delta}k(\theta;\gamma_0)$$

in the implicit-dependence case.

Lemma 4.3 is a continuity result. It formally justifies ignoring the knife-edge case $\delta^\star(\theta;\gamma_0,P_0)=\delta$ when characterizing the extreme counterfactuals at a population level.

4.2 Large-sample properties of plug-in estimators

We now show that the plug-in estimators are consistent and derive their asymptotic distribution. To do so, we first impose two more mild regularity conditions.

Assumption M (continued)

(iv) Θ is a compact subset of $\mathbb{R}^{d_\theta}$

(v) $\mathbb{E}^{F_*}[\phi^\star(c_1+c_2k(U,\theta,\gamma)+c_3'g(U,\theta,\gamma))]$ is continuous in (θ,γ) for each $c\in\mathbb{R}^{d+2}$.

Assumption M(iv) can be relaxed but simplifies some of the proofs. If k and each entry of g consist of indicator functions of events, then Assumption M(v) merely requires that the probability of the events under F_* are continuous in (θ,γ). For models without auxiliary parameters (i.e. no γ), Assumption M(v) just requires continuity in θ.

Theorem 4.1 *Let Assumptions* Φ *and M hold and let* $(\hat{\gamma},\hat{P})\to_p(\gamma_0,P_0)$ *or, if there is no auxiliary parameter,* $\hat{P}\to_p P_0$*. Then:* $(\hat{\underline{\kappa}}(\mathcal{N}_\delta),\hat{\bar{\kappa}}(\mathcal{N}_\delta))'\to_p(\underline{\kappa}(\mathcal{N}_\delta),\bar{\kappa}(\mathcal{N}_\delta))'$.

[9]This condition can be relaxed by working with sets of the form $\mathcal{N}_{\delta'}$ with $\delta'>\delta$ and taking limits of $\underline{\kappa}_{\delta'}$ and $\bar{\kappa}_{\delta'}$ as $\delta'\downarrow\delta$, since any point with $\delta^\star(\theta;\gamma_0,P_0)=\delta$ belongs to $\Theta_{\delta'}$ for all $\delta'>\delta$.

We now derive the asymptotic distribution of the lower and upper bounds. To simplify presentation, we assume γ is vacuous in the remainder of this subsection. This entails no loss of generality for the entry game example. For the dynamic discrete choice models it assumes that the Markov transition matrix is known by the econometrician. We also drop dependence of all quantities on γ for the remainder of this subsection. We may view the lower and upper bounds as functions of the reduced-form parameter, so we write $\underline{\kappa}(\mathcal{N}_\delta) = \underline{\kappa}(\mathcal{N}_\delta; P_0)$ and $\bar{\kappa}(\mathcal{N}_\delta) = \bar{\kappa}(\mathcal{N}_\delta; P_0)$, and $\hat{\underline{\kappa}}(\mathcal{N}_\delta) = \underline{\kappa}(\mathcal{N}_\delta; \hat{P})$ and $\hat{\bar{\kappa}}(\mathcal{N}_\delta) = \bar{\kappa}(\mathcal{N}_\delta; \hat{P})$. Inference results are derived by establishing differentiability properties of $\underline{\kappa}(\mathcal{N}_\delta; P)$ and $\bar{\kappa}(\mathcal{N}_\delta; P)$. In some cases, the lower and upper bounds are (fully) differentiable functions of the reduced-form parameter P and standard delta-method arguments can be applied. Lack of full differentiability can arise if there are multiple Lagrange multipliers at a particular θ, and/or if the value of θ at which the lower and/or upper counterfactual is obtained is not unique. In these cases, the lower and upper bounds satisfy a weaker notion of differentiability and inference can be performed using subsampling or various modified bootstraps.

Define

$$\underline{\Theta}_\delta = \{\theta \in \Theta : \underline{\kappa}_\delta(\theta; P) = \underline{\kappa}(\mathcal{N}_\delta)\} \,, \quad \overline{\Theta}_\delta = \{\theta \in \Theta : \bar{\kappa}_\delta(\theta; P) = \bar{\kappa}(\mathcal{N}_\delta)\} \,.$$

Also define

$$\begin{aligned}
\underline{\Xi}_\delta(\theta) &= \mathrm{argsup}_{\eta \geq 0, \zeta \in \mathbb{R}, \lambda \in \Lambda} - \mathbb{E}^{F_*}\Big[(\eta\phi)^\star(-k(U,\theta) - \zeta - \lambda' g(U,\theta))\Big] - \eta\delta - \zeta - (\lambda_1', \lambda_2')P \,, \\
\overline{\Xi}_\delta(\theta) &= \mathrm{arginf}_{\eta \geq 0, \zeta \in \mathbb{R}, \lambda \in \Lambda} \mathbb{E}^{F_*}\Big[(\eta\phi)^\star(k(U,\theta) - \zeta - \lambda' g(U,\theta))\Big] + \eta\delta + \zeta + (\lambda_1', \lambda_2')P
\end{aligned}$$

in the explicit-dependence case, and

$$\begin{aligned}
\underline{\Xi}_\delta(\theta) &= \mathrm{argsup}_{\eta \geq 0, \zeta \in \mathbb{R}, \lambda \in \Lambda} - \mathbb{E}^{F_*}\Big[(\eta\phi)^\star(-\zeta - \lambda' g(U,\theta))\Big] - \eta\delta - \zeta - (\lambda_1', \lambda_2')P \,, \\
\overline{\Xi}_\delta(\theta) &= \mathrm{arginf}_{\eta \geq 0, \zeta \in \mathbb{R}, \lambda \in \Lambda} \mathbb{E}^{F_*}\Big[(\eta\phi)^\star(-\zeta - \lambda' g(U,\theta))\Big] + \eta\delta + \zeta + (\lambda_1', \lambda_2')P
\end{aligned}$$

in the implicit-dependence case. These sets are nonempty, convex and compact under the conditions of the following Lemma and Theorem. Also let

$$\begin{aligned}
\underline{\Lambda}_\delta(\theta) &= \{(\underline{\lambda}_1', \underline{\lambda}_2')' : (\underline{\eta}, \underline{\zeta}, \underline{\lambda}_1', \underline{\lambda}_2', \underline{\lambda}_3', \underline{\lambda}_4')' \in \underline{\Xi}_\delta(\theta) \text{ for some } (\underline{\eta}, \underline{\zeta}, \underline{\lambda}_3', \underline{\lambda}_4')'\} \,, \\
\overline{\Lambda}_\delta(\theta) &= \{(\overline{\lambda}_1', \overline{\lambda}_2')' : (\overline{\eta}, \overline{\zeta}, \overline{\lambda}_1', \overline{\lambda}_2', \overline{\lambda}_3', \overline{\lambda}_4')' \in \overline{\Xi}_\delta(\theta) \text{ for some } (\overline{\eta}, \overline{\zeta}, \overline{\lambda}_3', \overline{\lambda}_4')'\} \,.
\end{aligned}$$

A function $f : \mathbb{R}^{d_1+d_2} \to \mathbb{R}$ is (Hadamard) directionally differentiable at P_0 if there is a continuous map $df(P_0)[\cdot] : \mathbb{R}^{d_1+d_2} \to \mathbb{R}$ such that

$$\lim_{n\to\infty} \frac{f(P_0 + t_n \pi_n) - f(P_0)}{t_n} = df(P_0)[\pi]$$

for each positive sequence $t_n \downarrow 0$ and sequence of vectors $\pi_n \to \pi \in \mathbb{R}^{d_1+d_2}$ (Shapiro, 1990, p. 480). If the map $df(P_0)[\cdot]$ is linear then f is (fully) differentiable at P_0.

Lemma 4.4 *Let Assumptions Φ and M hold. If $\underline{\Theta}_\delta \subseteq \Theta_\delta$ and either (i) $\underline{\Lambda}_\delta$ is lower hemicontinuous at each $\theta \in \underline{\Theta}_\delta$ or (ii) $\underline{\Lambda}_\delta(\theta)$ is a singleton for each $\theta \in \underline{\Theta}_\delta$, then: $\underline{\kappa}(\mathcal{N}_\delta;\cdot)$ is directionally differentiable at P_0:*

$$d\underline{\kappa}(\mathcal{N}_\delta;P_0)[\pi] = \inf_{\theta\in\underline{\Theta}_\delta} \sup_{(\underline{\lambda}_1',\underline{\lambda}_2')'\in\underline{\Lambda}_\delta(\theta)} -(\underline{\lambda}_1',\underline{\lambda}_2')'\pi\,.$$

Similarly, if $\overline{\Theta}_\delta \subseteq \Theta_\delta$, and either (i) $\overline{\Lambda}_\delta$ is lower hemicontinuous at each $\theta \in \overline{\Theta}_\delta$ or (ii) $\underline{\Lambda}_\delta(\theta)$ is a singleton for each $\theta \in \underline{\Theta}_\delta$ then: $\bar{\kappa}(\mathcal{N}_\delta;\cdot)$ is directionally differentiable at P_0:

$$d\bar{\kappa}(\mathcal{N}_\delta;P_0)[\pi] = \sup_{\theta\in\overline{\Theta}_\delta} \inf_{(\overline{\lambda}_1',\overline{\lambda}_2')'\in\overline{\Lambda}_\delta(\theta)} (\overline{\lambda}_1',\overline{\lambda}_2')\pi\,.$$

We are now in a position to derive the joint asymptotic distribution of lower and upper bounds. Provided the first-stage estimator $\hat{P}$ is asymptotically normally distributed, the result follows by a delta method for directionally differentiable functions (Shapiro, 1991) and Lemma 4.4.

Theorem 4.2 *Let $\sqrt{n}(\hat{P} - P_0) \to_d N(0,\Sigma)$ and let the conditions of Lemma 4.4 hold. Then:*

$$\sqrt{n}\begin{bmatrix} \hat{\bar{\kappa}}(\mathcal{N}_\delta) - \bar{\kappa}(\mathcal{N}_\delta) \\ \hat{\underline{\kappa}}(\mathcal{N}_\delta) - \underline{\kappa}(\mathcal{N}_\delta) \end{bmatrix} \to_d \begin{bmatrix} \sup_{\theta\in\overline{\Theta}_\delta} \inf_{(\overline{\lambda}_1',\overline{\lambda}_2')'\in\overline{\Lambda}_\delta(\theta)} \quad (\overline{\lambda}_1',\overline{\lambda}_2')Z \\ \inf_{\theta\in\underline{\Theta}_\delta} \sup_{(\underline{\lambda}_1',\underline{\lambda}_2')'\in\underline{\Lambda}_\delta(\theta)} -(\underline{\lambda}_1',\underline{\lambda}_2')Z \end{bmatrix},$$

where $Z \sim N(0,\Sigma)$.

In particular, if $\cup_{\theta\in\overline{\Theta}_\delta}\overline{\Lambda}_\delta(\theta) = \{(\overline{\lambda}_1',\overline{\lambda}_2')'\}$ and $\cup_{\theta\in\underline{\Theta}_\delta}\underline{\Lambda}_\delta(\theta) = \{(\underline{\lambda}_1',\underline{\lambda}_2')'\}$, then:

$$\sqrt{n}\begin{bmatrix} \hat{\bar{\kappa}}(\mathcal{N}_\delta) - \bar{\kappa}(\mathcal{N}_\delta) \\ \hat{\underline{\kappa}}(\mathcal{N}_\delta) - \underline{\kappa}(\mathcal{N}_\delta) \end{bmatrix} \to_d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \overline{\lambda}_1' & \overline{\lambda}_2' \\ -\underline{\lambda}_1' & -\underline{\lambda}_2' \end{bmatrix} \Sigma \begin{bmatrix} \overline{\lambda}_1 & -\underline{\lambda}_1 \\ \overline{\lambda}_2 & -\underline{\lambda}_2 \end{bmatrix}\right).$$

Theorem 4.2 derives the joint asymptotic distribution of plug-in estimators of the extreme counterfactuals. The asymptotic distribution is typically nonstandard, however, due to directional differentiability of the lower and upper bounds in P. Nevertheless, one may use subsampling or various modified bootstraps to perform asymptotically valid inference; see Fang and Santos (2019) and Hong and Li (2018) for related theoretical developments. In particular, note that the directional derivative $d\underline{\kappa}(\mathcal{N}_\delta;P_0)$ is convex when $\underline{\Theta}_\delta$ is a singleton and $d\bar{\kappa}(\mathcal{N}_\delta;P_0)$ is concave when $\overline{\Theta}_\delta$ is a singleton. As emphasized in Fang and Santos (2019) and Hong and Li (2018), these convexity and concavity properties are helpful for guaranteeing uniform asymptotic coverage of one-sided confidence intervals for $\underline{\kappa}(\mathcal{N}_\delta)$ of the form $[\hat{\underline{\kappa}}(\mathcal{N}_\delta) - \hat{\underline{c}}_{1-\alpha}/\sqrt{n}, \infty)$ and one-sided confidence intervals for $\bar{\kappa}(\mathcal{N}_\delta)$ of the form $(-\infty, \hat{\bar{\kappa}}(\mathcal{N}_\delta) - \hat{\bar{c}}_\alpha/\sqrt{n}]$, where $\hat{\underline{c}}_{1-\alpha}$ and $\hat{\bar{c}}_\alpha$ are critical values obtained using subsampling or various modified bootstraps.

5 Sharp bounds on the identified set of counterfactuals

In this section, we show that the extreme counterfactuals $\underline{\kappa}(\mathcal{N}_\delta)$ and $\bar{\kappa}(\mathcal{N}_\delta)$ deliver sharp bounds on the identified set of counterfactuals (i.e., the set of counterfactuals consistent with (1)–(4) where no parametric distributional assumptions are placed on F) as the neighborhood size δ becomes large. In this sense, the local neighborhoods $\mathcal{N}_\delta$ act like an infinite-dimensional sieve: although they exclude many distributions, the neighborhoods are in some sense "dense" in the set of distributions relevant for characterizing the identified set of counterfactuals. We first present results for the explicit-dependence case, before turning to the implicit-dependence case. Further related theoretical results are deferred to Appendix B.

5.1 Counterfactuals depending explicitly on latent variables

In the specification of structural models such as dynamic discrete choice models and static and dynamic games, it is common to assume that the latent variables have a density relative to some σ-finite dominating measure, say μ (usually Lebesgue measure). In defining the identified set of counterfactuals, we therefore consider only distributions that are absolutely continuous with respect to μ. We do so because absolute continuity is often itself a "structural" assumption that is used, among other things, to help establish existence of equilibria.

To introduce the identified set, let $\mathcal{F}_\theta = \{F \in \mathcal{F} : \mathbb{E}^F[g(U,\theta,\gamma_0)] \text{ is finite and } F \ll \mu\}$. The set $\mathcal{F}_\theta$ is the largest set of probability measures that are absolutely continuous with respect to μ and for which the moments (1)–(4) are defined at θ. The identified set of counterfactuals is

$$\mathcal{K}_\# = \left\{\mathbb{E}^F[k(U,\theta,\gamma_0)] \text{ such that (1)–(4) hold for some } \theta \in \Theta \text{ and } F \in \mathcal{F}_\theta\right\} .$$

The set $\mathcal{F}_\theta$ contains many fatter-tailed distributions not in $\mathcal{N}_\infty$. It therefore seems reasonable to ask: in confining ourselves to $\mathcal{N}_\infty$, do we throw away other distributions that can yield smaller or larger values of the counterfactual? As we shall see, the answer is "no" provided μ and F_* are mutually absolutely continuous, which we may interpret as a type of full support condition for F_*. In that case, the neighborhoods $\mathcal{N}_\delta$ eventually span the full set of distributions relevant for characterizing the smallest and largest elements of $\mathcal{K}_\#$.

We say that k is μ-essentially bounded if $|k(\cdot,\theta,\gamma_0)|$ has finite μ-essential supremum[10] for each $\theta \in \Theta$. This is trivially true if the function k is bounded, i.e.: $\sup_u |k(u,\theta,\gamma_0)| < \infty$ for each θ. Conditional choice probabilities always satisfy this boundedness condition because the k function is

[10] The μ-essential supremum of $f : \mathcal{U} \to \mathbb{R}$ is $\mu\text{-ess}\sup f = \inf\{c : \mu(\{u : f(u) > c\} = 0)\}$. Similarly, the μ-essential infimum is $\mu\text{-ess}\inf f = \sup\{c : \mu(\{u : f(u) < c\} = 0)\}$. Note that $\inf f \le \mu\text{-ess}\inf f \le \mu\text{-ess}\sup f \le \sup f$.

an indicator function of an event. This boundedness condition ensures the extreme counterfactuals are finite. Note, however, that we do not require any of the moment functions $g_1, \ldots, g_4$ to be bounded.

Theorem 5.1 *Let Assumption Φ hold, let Condition S hold at (θ, γ_0, P_0) for all $\theta \in \Theta$, let μ and F_* be mutually absolutely continuous, and let k be μ-essentially bounded. Then:*

$$\underline{\kappa}(\mathcal{N}_\delta) \to \inf \mathcal{K}_\# \,, \quad \bar{\kappa}(\mathcal{N}_\delta) \to \sup \mathcal{K}_\# \qquad \text{as } \delta \to \infty.$$

Remark 5.1 *One-sided versions also hold when k is not μ-essentially bounded. Suppose that k is only μ-essentially bounded from below (i.e. $\mu\text{-ess}\inf k(\cdot, \theta, \gamma_0) > -\infty$ for each $\theta \in \Theta$) and the remaining conditions of Theorem 5.1 hold. Then: $\underline{\kappa}(\mathcal{N}_\delta) \to \inf \mathcal{K}_\#$ as $\delta \to \infty$. Similarly, if k is only μ-essentially bounded from above (i.e. $\mu\text{-ess}\sup k(\cdot, \theta, \gamma_0) < \infty$ for each $\theta \in \Theta$), then: $\bar{\kappa}(\mathcal{N}_\delta) \to \sup \mathcal{K}_\#$ as $\delta \to \infty$.*

Mutual absolute continuity of μ and F_* may be interpreted as a type of full support condition on F_*. In models where μ is Lebesgue measure, it follows from Theorem 5.1 that choosing F_* with strictly positive density over $\mathcal{U}$ —which is indeed the case for all conventional benchmark choices such as normal, extreme value (Gumbel), Fréchet, Pareto, etc.— ensures that the extreme counterfactuals $\underline{\kappa}(\mathcal{N}_\delta)$ and $\bar{\kappa}(\mathcal{N}_\delta)$ will approach the bounds of the identified set of counterfactuals as δ gets large.

5.2 Counterfactuals depending implicitly on latent variables

In the implicit-dependence case, the identified set of counterfactuals is

$$\mathcal{K}_\# = \{k(\theta, \gamma_0) \text{ such that (1)–(4) hold for some } \theta \in \Theta \text{ and } F \in \mathcal{F}_\theta\} \,.$$

As k does not depend on U, an analogous version of Theorem 5.1 holds without the essential-boundedness condition.

Theorem 5.2 *Let Assumption Φ hold, let Condition S hold at (θ, γ_0, P_0) for all $\theta \in \Theta$, let μ and F_* be mutually absolutely continuous. Then:*

$$\underline{\kappa}(\mathcal{N}_\delta) \to \inf \mathcal{K}_\# \,, \quad \bar{\kappa}(\mathcal{N}_\delta) \to \sup \mathcal{K}_\# \qquad \text{as } \delta \to \infty.$$

6 Local sensitivity

Local sensitivity analyses characterize the behavior of a targeted quantity (e.g. a counterfactual or structural parameter) as a model input (e.g. a distribution or vector of moments) varies over a vanishingly small neighborhood of an assumed true specification. Here we describe a measure of local sensitivity of counterfactuals with respect to the distribution of unobservables and connect the measure to our procedure over small neighborhoods. Our measure deals with an assumption made at the modeling stage: it holds the observable implications of the model fixed and focuses on what happens "under the hood" of the model. This is conceptually distinct from the measures proposed by Andrews, Gentzkow, and Shapiro (2017, 2018; AGS hereafter), which characterize the sensitivity of estimates of counterfactuals or structural parameters with respect to possible local misspecification of the moments used in estimation. In their framework, the distribution of *observables* is possibly (locally) misspecified. In contrast, our approach holds the observables fixed and varies specification of the distribution of *unobservables*. The first part of this section describes our local sensitivity measure, derives its influence function representation, and presents a consistent and easily computable estimator. We then show how the influence function we obtain and our local sensitivity measure is complementary to though conceptually very different from the (statistical) influence function of the counterfactual and related local sensitivity measures.

Our *measure of local sensitivity* of counterfactuals with respect to F_* is

$$s = \lim_{\delta \downarrow 0} \frac{(\bar{\kappa}(\mathcal{N}_\delta) - \underline{\kappa}(\mathcal{N}_\delta))^2}{4\delta}\,.$$

The quantity s measures the curvature of the functions $\delta \mapsto \underline{\kappa}(\mathcal{N}_\delta)$ and $\delta \mapsto \bar{\kappa}(\mathcal{N}_\delta)$ at $\delta = 0$. If s is finite, then (under some regularity conditions):

$$\underline{\kappa}(\mathcal{N}_\delta) = \kappa(F_*) - \sqrt{\delta s} + o(\sqrt{\delta})\,, \quad \bar{\kappa}(\mathcal{N}_\delta) = \kappa(F_*) + \sqrt{\delta s} + o(\sqrt{\delta}) \quad \text{as } \delta \downarrow 0\,,$$

where $\theta(F_*)$ solves (1)–(4) under F_* and $\kappa(F_*) = \mathbb{E}^{F_*}[k(U, \theta(F_*), \gamma_0)]$ in the explicit-dependence case or $\kappa(F_*) = k(\theta(F_*), \gamma_0)$ in the implicit-dependence case.[11] When specialized further to point-identified, regular models studied in the local sensitivity literature, the measure s is particularly simple to characterize. We present an easily computable estimator $\hat{s}$ of s in this case, which researchers may report alongside estimated counterfactuals. Approximate bounds on counterfactuals as F varies over small neighborhoods of F_* can then be estimated using $\hat{\kappa} \pm \sqrt{\delta \hat{s}}$, though the examples presented in Section 3 indicate that these should be interpreted with some caution outside of very small neighborhoods.

[11] Finiteness of s implies that the counterfactual $\kappa(F_*)$ is point identified. Note that this may be true even if $\theta(F_*)$ is not point identified by the moment conditions (1)–(4); see, e.g., Aguirregabiria (2005), Norets and Tang (2014), and Kalouptsidi et al. (2017) for the case of dynamic discrete choice models.

6.1 Counterfactuals depending explicitly on latent variables

To draw further comparison with the local sensitivity literature, we restrict attention to models with equality restrictions only, i.e., $d_1 = d_3 = 0$, and impose some further (standard) GMM-type regularity conditions. First, assume that the moment conditions (2) and (4) point identify a structural parameter $\theta(F_*) \in \text{int}(\Theta)$ when evaluated under F_* at (γ_0, P_{20}). We write $\theta(F_*)$ to make explicit the dependence of this structural parameter on the assumed specification of F_*. Let

$$h(u,\theta,\gamma,P_2) := \begin{bmatrix} g_2(u,\theta,\gamma) - P_2 \\ g_4(u,\theta,\gamma) \end{bmatrix}$$

and define $h_0(u) = h(u,\theta(F_*),\gamma_0,P_{20})$ and $k_0(u) = k(u,\theta(F_*),\gamma_0)$. We assume $\mathbb{E}^{F_*}[h(U,\theta,\gamma_0,P_{20})]$ is continuously differentiable with respect to θ at $\theta(F_*)$,

$$H := \frac{\partial}{\partial\theta'}\mathbb{E}^{F_*}[h(U,\theta,\gamma_0,P_{20})]\bigg|_{\theta=\theta(F_*)}$$

has full rank, $V := \mathbb{E}^{F_*}[h_0(U)h_0(U)']$ is finite and positive definite, $\mathbb{E}^{F_*}[k(U,\theta,\gamma_0)^2]$ is finite, $k(\cdot,\theta,\gamma_0)$ and each entry of $h(\cdot,\theta,\gamma_0,P_{20})$ are $L^2(F_*)$ continuous in θ at $\theta(F_*)$, and $\mathbb{E}^{F_*}[k(U,\theta,\gamma_0)]$ is continuously differentiable with respect to θ at $\theta(F_*)$. Let

$$J = \frac{\partial}{\partial\theta'}\mathbb{E}^{F_*}[k(U,\theta,\gamma_0)]\bigg|_{\theta=\theta(F_*)}.$$

Define

$$\iota(u) = \Pi k_0(u) - J(H'V^{-1}H)^{-1}H'V^{-1}h_0(u)\,, \tag{18}$$

where

$$\Pi k_0(u) = k_0(u) - \kappa(F_*) - \mathbb{E}^{F_*}[k_0(U)h_0(U)'](V^{-1} - V^{-1}H(H'V^{-1}H)^{-1}H'V^{-1})h_0(u)\,.$$

In just-identified models (i.e. $d_2 + d_4 = d_\theta$), the expression for ι simplifies:

$$\iota(u) = k_0(u) - \kappa(F_*) - JH^{-1}h_0(u)\,.$$

The function ι is the *influence function* of the counterfactual with respect to F at F_*. This is a different notion of influence function from that which is usually encountered when analyzing semiparametric estimators, as ι measures sensitivity of an estimand to a modeling assumption rather than sensitivity of an estimator to the data.[12] Nevertheless, ι is derived by similar arguments to those used to in semiparametric efficiency bound calculations for GMM-type models.

[12] We use the term influence function because the expansion $\kappa(F) - \kappa(F_*) = \int \iota\,\mathrm{d}(F - F_*) + \text{remainder}$ is valid for distributions F suitably close to F_*, where $\kappa(F) = \mathbb{E}^F[k(U,\theta(F),\gamma_0)]$. This mimics the usual asymptotic linear expansion for estimators, where F and F_* are replaced by the empirical and true probability measures, respectively.

The following theorem relates local sensitivity to the variance of ι in point-identified, regular models. We will restrict attention to neighborhoods characterized by χ^2 divergence. Other ϕ-divergences are locally equivalent to χ^2 divergence, so this restriction entails no great loss of generality.[13] Neighborhoods constrained by χ^2 divergence are also compatible with the above regularity conditions, which assume k and the entries of g have finite second moments under F_* (cf. Assumption Φ(ii)).

Theorem 6.1 *Let Assumptions Φ(ii) and M(i)(iv) hold for χ^2 divergence, and let the above GMM-type regularity conditions hold. Then: $s = 2\mathbb{E}^{F_*}[\iota(U)^2]$ where ι is defined in (18).*

In this setting, the researcher will have consistent estimates $(\hat{\gamma}, \hat{P}_2)$ of (γ_0, P_{20}), which can be used to consistently estimate $\theta(F_*)$. Let $\hat{\theta}$ denote such an estimator. The researcher would estimate the counterfactual (under F_*) using

$$\hat{\kappa} = \mathbb{E}^{F_*}[k(U, \hat{\theta}, \hat{\gamma})] \,.$$

In addition to the estimated counterfactual $\hat{\kappa}$, the researcher could also report an estimate of the local sensitivity of the counterfactual with respect to F_*:

$$\hat{s} = 2\mathbb{E}^{F_*}[(\hat{k}(U) - \hat{\kappa})^2] + 2\hat{Q}'\hat{V}\hat{Q} - 4\mathbb{E}^{F_*}[\hat{h}(U)(\hat{k}(U) - \hat{\kappa})]'\hat{Q} \,,$$

where $\hat{k}(u) = k(u, \hat{\theta}, \hat{\gamma})$, $\hat{h}(u) = h(u, \hat{\theta}, \hat{\gamma}, \hat{P}_2)$, $\hat{V} = \mathbb{E}^{F_*}[\hat{h}(U)\hat{h}(U)']$, and

$$\hat{Q} = \mathbb{E}^{F_*}[\hat{k}(U)\hat{h}(U)'](\hat{V}^{-1} - \hat{V}^{-1}\hat{H}(\hat{H}'\hat{V}^{-1}\hat{H})^{-1}\hat{H}'\hat{V}^{-1}) + \hat{J}(\hat{H}'\hat{V}^{-1}\hat{H})^{-1}\hat{H}'\hat{V}^{-1} \,,$$

with

$$\hat{H} = \frac{\partial}{\partial \theta'}\mathbb{E}^{F_*}[h(U, \theta, \hat{\gamma}, \hat{P}_2)]\bigg|_{\theta = \hat{\theta}} \,, \qquad \hat{J} = \frac{\partial}{\partial \theta'}\mathbb{E}^{F_*}[k(U, \theta, \hat{\gamma})]\bigg|_{\theta = \hat{\theta}} \,.$$

If the model is just identified, then the expression for $\hat{Q}$ simplifies to $\hat{Q} = \hat{J}\hat{H}^{-1}$. In either case, the plug-in estimator $\hat{s}$ is consistent under very mild smoothness conditions.

Lemma 6.1 *Let the conditions of Theorem 6.1 hold. Also let $(\hat{\theta}, \hat{\gamma}, \hat{P}_2) \to_p (\theta(F_*), \gamma_0, P_{20})$, and let $\frac{\partial}{\partial \theta'}\mathbb{E}^{F_*}[h(U, \theta, \gamma, P_2)]$, $\frac{\partial}{\partial \theta'}\mathbb{E}^{F_*}[k(U, \theta, \gamma)]$, $\mathbb{E}^{F_*}[h(U, \theta, \gamma, P_2)h(U, \theta, \gamma, P_2)']$, $\mathbb{E}^{F_*}[h(U, \theta, \gamma, P_2)k(U, \theta, \gamma)]$, $\mathbb{E}^{F_*}[h(U, \theta, \gamma, P_2)]$, $\mathbb{E}^{F_*}[k(U, \theta, \gamma)]$, and $\mathbb{E}^{F_*}[k(U, \theta, \gamma)^2]$ be continuous in (θ, γ, P_2) at $(\theta(F_*), \gamma_0, P_{20})$. Then: $\hat{s} \to_p s$.*

Some of the terms used to construct $\hat{s}$ will already be computed when estimating the model using minimum-distance or GMM methods. Therefore, $\hat{s}$ should be easy to report alongside $\hat{\kappa}$ in practice.

[13] See Theorem 4.1 in Csiszár and Shields (2004). If ϕ-divergence different from χ^2 divergence is used, the quantity $2\mathbb{E}^{F_*}[\iota(U)^2]$ may be rescaled by a factor of $\phi''(1)$ to obtain s. No such rescaling is required for KL or hybrid divergence as $\phi''(1) = 1$ in both cases.

6.2 Counterfactuals depending implicitly on latent variables

Turning to the implicit-dependence case, here we make the same GMM-type assumptions on h_0, H, and V, but instead assume that $k(\theta, \gamma_0)$ is continuously differentiable with respect to θ at $\theta(F_*)$. The measure of local sensitivity again takes the form $s = 2\mathbb{E}^{F_*}[\iota(U)^2]$ under the above conditions, where the influence function is

$$\iota(u) = -J(H'V^{-1}H)^{-1}H'V^{-1}h_0(u) \tag{19}$$

with H, V, and h_0 as described in the previous subsection, and $J = \frac{\partial}{\partial \theta'} k(\theta, \gamma_0)\big|_{\theta=\theta(F_*)}$. A result identical to Theorem 6.1 holds in this setting.

Theorem 6.2 *Let Assumptions Φ(ii) and M(i)(iv) hold for ϕ corresponding to χ^2 divergence, and let the above GMM-type regularity conditions hold. Then: $s = 2\mathbb{E}^{F_*}[\iota(U)^2]$ where ι is defined in (19).*

In this setting, the researcher will have consistent estimates $(\hat{\theta}, \hat{\gamma}, \hat{P}_2)$ of $(\theta(F_*), \gamma_0, P_{20})$. The researcher's estimator of $\kappa(F_*)$ would be

$$\hat{\kappa} = k(\hat{\theta}, \hat{\gamma})\,.$$

The local sensitivity of $\kappa(F_*)$ with respect to specification of F_* can be estimated using

$$\hat{s} = 2\hat{Q}'\hat{V}\hat{Q}$$

where $\hat{Q} = \hat{J}(\hat{H}'\hat{V}^{-1}\hat{H})^{-1}\hat{H}'\hat{V}^{-1}$ with $\hat{H}$ and $\hat{V}$ as described in the previous subsection and $\hat{J} = \frac{\partial}{\partial \theta'} k(\theta, \hat{\gamma})\big|_{\theta=\hat{\theta}}$. As in the previous subsection, the plug-in estimator $\hat{s}$ is consistent under very mild smoothness conditions.

Lemma 6.2 *Let the conditions of Theorem 6.2 hold. Also let $(\hat{\theta}, \hat{\gamma}, \hat{P}_2) \to_p (\theta(F_*), \gamma_0, P_{20})$, and let $\frac{\partial}{\partial \theta'}\mathbb{E}^{F_*}[h(U, \theta, \gamma, P_2)]$, $\frac{\partial}{\partial \theta'} k(\theta, \gamma)$, and $\mathbb{E}^{F_*}[h(U, \theta, \gamma, P_2)h(U, \theta, \gamma, P_2)']$, be continuous in (θ, γ, P_2) at $(\theta(F_*), \gamma_0, P_{20})$. Then: $\hat{s} \to_p s$.*

6.3 Comparison with other notions of sensitivity

We now briefly compare our local sensitivity measure and influence function representation with the (statistical) influence function of the counterfactual and with AGS's measures of sensitivity and informativeness. Consider a class of models whose only moments are of the form (2) with $d_2 \geq d_\theta$

and for which the auxiliary parameter γ is vacuous. Given a first-stage estimator $\hat{P}_2$ and assumed distribution F_*, the researcher could estimate θ by minimizing the criterion function

$$(\mathbb{E}^{F_*}[g_2(U,\theta)] - \hat{P}_2)'\hat{W}(\mathbb{E}^{F_*}[g_2(U,\theta)] - \hat{P}_2)$$

given some positive-definite and symmetric matrix $\hat{W}$ whose probability limit W is also positive-definite and symmetric. Assume $\hat{P}_2$ is a regular estimator with (statistical) influence function ι_{P_2}, i.e.:

$$\sqrt{n}(\hat{P}_2 - P_{20}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \iota_{P_2}(X_i) + o_p(1)$$

where P_{20} is the probability limit of $\hat{P}_2$. The pseudo-true parameter $\theta(F_*)$ solves $\mathbb{E}^{F_*}[g_2(U,\theta)] = P_{20}$. Given an estimator $\hat{\theta}$ of $\theta(F_*)$ obtained using this procedure, the researcher would then estimate the counterfactual as $\hat{\kappa} = \mathbb{E}^{F_*}[k(U,\hat{\theta})]$.

By standard delta-method arguments, the (statistical) influence function of $\hat{\kappa}$ is seen to be

$$\iota_\kappa(x) = J'(H'WH)^{-1}H'W\iota_{P_2}(x)\,, \tag{20}$$

where $H = \frac{\partial}{\partial\theta'}\mathbb{E}^{F_*}[g_2(U,\theta)]\big|_{\theta=\theta(F_*)}$. The function $\iota_\kappa(x)$ characterizes sensitivity of the estimator $\hat{\kappa}$ with respect to perturbations of the data $X_1,\ldots,X_n$. This is conceptually very different from the function $\iota(u)$ obtained in the previous subsections, which characterizes sensitivity of the estimand $\kappa(F_*)$ with respect to perturbations of F_*.

One may verify that AGS's measure of *sensitivity* of $\hat{\kappa}$ to $\hat{P}_2$ is $J'(H'WH)^{-1}H'W$, the adjustment required to obtain $\iota_\kappa(x)$ from $\iota_{P_2}(x)$. AGS's measure of *informativeness* of $\hat{P}_2$ for $\hat{\kappa}$ is 1, meaning that all (statistical) variation in $\hat{\kappa}$ is explained by variation in $\hat{P}_2$. In contrast, our measure of sensitivity characterizes "specification variation" in κ as the researcher varies F_*. AGS's sensitivity and informativeness measures and our measure of sensitivity therefore represent distinct but complementary quantities.

7 Conclusion

This paper introduces a framework to study the sensitivity of counterfactuals with respect to strong parametric assumptions about the distribution of unobservables that are often made in structural modeling exercises. Using insights from the model uncertainty literature, we show how to construct the smallest and largest counterfactuals obtained as the distribution of unobservables varies over fully nonparametric neighborhoods of the researcher's assumed specification while other structural features of the model, such as equilibrium conditions, are maintained. We provide a suitable sampling theory for plug-in estimators of the extreme counterfactuals and illustrate our procedure with applications to two workhorse models. Further, we show that our procedure delivers sharp bounds on the identified set of counterfactuals as the neighborhoods expand and we explore connections with a measure of local sensitivity as the neighborhoods shrink. Going forward, we plan to further extend our methods to accommodate local misspecification in the reduced form and to consider optimal estimation and inference in this setting.

References

Aguirregabiria, V. (2005). Nonparametric identification of behavioral responses to counterfactual policy interventions in dynamic discrete decision processes. *Economics Letters 87*(3), 393–398.

Aguirregabiria, V. and P. Mira (2007). Sequential estimation of dynamic discrete games. *Econometrica 75*(1), 1–53.

Aguirregabiria, V. and P. Mira (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics 156*(1), 38–67. Structural Models of Optimization Behavior in Labor, Aging, and Health.

Allen, R. and J. Rehbeck (2017). Identification with additively separable heterogeneity. SSRN working paper, `https://ssrn.com/abstract=3073945`.

Allen, T. and C. Arkolakis (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics 129*(3), 1085–1140.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017). Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics 132*(4), 1553–1592.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2018). On the informativeness of descriptive statistics for structural estimates. NBER working paper no. 25217, `https://www.nber.org/papers/w25217`.

Aradillas-Lopez, A. (2011). Nonparametric probability bounds for nash equilibrium actions in a simultaneous discrete game. *Quantitative Economics 2*(2), 135–171.

Arcidiacono, P. and P. B. Ellickson (2011). Practical methods for estimation of dynamic discrete choice models. *Annual Review of Economics 3*(1), 363–394.

Armstrong, T. B. and M. Kolesár (2018). Sensitivity analysis using approximate moment condition models. Cowles Foundation Discussion Paper no. 2158, `https://ssrn.com/abstract=3309781`.

Bajari, P., C. L. Benkard, and J. Levin (2007). Estimating dynamic models of imperfect competition. *Econometrica 75*(5), 1331–1370.

Bajari, P., H. Hong, and S. P. Ryan (2010). Identification and estimation of a discrete game of complete information. *Econometrica 78*(5), 1529–1568.

Berger, J. O. (1984). The robust Bayesian viewpoint (with discussion). In J. B. Kadane (Ed.), *Robustness of Bayesian analyses*. North-Holland.

Berger, J. O. and L. M. Berliner (1986). Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors. *The Annals of Statistics 14*(2), 461–486.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica 60*(4), 889–917.

Blevins, J. R. (2014). Nonparametric identification of dynamic decision processes with discrete and continuous choices. *Quantitative Economics 5*(3), 531–554.

Bonhomme, S. and M. Weidner (2018). Minimizing sensitivity to model misspecification. cemmap Working Paper, cwp59/18, `https://www.cemmap.ac.uk/publication/id/13486`.

Bresnahan, T. F. and P. C. Reiss (1990). Entry in monopoly market. *The Review of Economic Studies 57*(4), 531–553.

Bresnahan, T. F. and P. C. Reiss (1991). Empirical models of discrete games. *Journal of Econometrics 48*(1), 57–81.

Buchholz, N., M. Shum, and H. Xu (2018). Semiparametric estimation of dynamic discrete choice models. arxiv:1605.08369 [stat.ap].

Chamberlain, G. and E. E. Leamer (1976). Matrix weighted averages and posterior bounds. *Journal of the Royal Statistical Society. Series B (Methodological) 38*(1), 73–84.

Chen, L.-Y. (2017). Identification of discrete choice dynamic programming models with nonparametric distribution of unobservables. *Econometric Theory 33*(3), 551–577.

Chen, X., E. Tamer, and A. Torgovitsky (2011). Sensitivity analysis in semiparametric likelihood models. Cowles Foundation Discussion Paper no. 1836.

Chernozhukov, V., W. K. Newey, and A. Santos (2015). Constrained conditional moment restriction models. cemmap Working Paper, cwp59/15, `https://www.cemmap.ac.uk/publication/id/7992`.

Chiong, K., Y.-W. Hsieh, and M. Shum (2017). Counterfactual evaluation in semiparametric multinomial choice models. USC-INET research paper no. 17-20.

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. *Econometrica 77*(6), 1791–1828.

Csiszár, I. and P. C. Shields (2004). Information theory and statistics: A tutorial. *Communications and Information Theory 1*(4), 417–528.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

Ekeland, I., A. Galichon, and M. Henry (2010). Optimal transportation and the falsifiability of incompletely specified economic models. *Economic Theory 42*(2), 355–374.

Ericson, R. and A. Pakes (1995). Markov-perfect industry dynamics: A framework for empirical work. *The Review of Economic Studies 62*(1), 53–82.

Fang, Z. and A. Santos (2019). Inference on directionally differentiable functions. *Review of Economic Studies 86*(1), 377–412.

Giacomini, R., T. Kitagawa, and H. Uhlig (2016). Estimation under ambiguity. Technical report, UCL and Chicago.

Hansen, L. P. and M. Marinacci (2016). Ambiguity aversion and model misspecification: An economic perspective. *Statistical Science 31*(4), 511–515.

Hansen, L. P. and T. J. Sargent (2001). Robust control and model uncertainty. *The American Economic Review 91*(2), 60–66.

Hansen, L. P. and T. J. Sargent (2008). *Robustness*. Princeton.

Ho, P. (2018). Global robust bayesian analysis in large models. Technical report, Princeton.

Hong, H. and J. Li (2018). The numerical delta method. *Journal of Econometrics 206*(2), 379–394. Special issue on Advances in Econometric Theory: Essays in honor of Takeshi Amemiya.

Hotz, V. J. and R. A. Miller (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies 60*(3), 497–529.

Kalouptsidi, M., P. T. Scott, and E. Souza-Rodrigues (2017). Identification of counterfactuals in dynamic discrete choice models. *NBER Working Paper No. 21527*.

Keane, M. P., P. E. Todd, and K. I. Wolpin (2011, May). *The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications*, Volume 4 of *Handbook of Labor Economics*, Chapter 4, pp. 331–461. Elsevier.

Kitamura, Y., T. Otsu, and K. Evdokimov (2013). Robustness, infinitesimal neighborhoods, and moment restrictions. *Econometrica 81*(3), 1185–1201.

Komunjer, I. and G. Ragusa (2016). Existence and characterization of conditional density projections. *Econometric Theory 32*(4), 947–987.

Leamer, E. E. (1982). Sets of posterior means with bounded variance priors. *Econometrica 50*(3), 725–736.

Leamer, E. E. (1985). Sensitivity analyses would help. *The American Economic Review 75*(3), 308–313.

Li, L. (2018). Identification of structural and counterfactual parameters in a large class of structural econometric models. Technical report, Pennsylvania State University.

Manski, C. F. (2007). Partial identification of counterfactual choice probabilities. *International Economic Review 48*(4), 1393–1410.

Manski, C. F. (2014). Identification of income–leisure preferences and evaluation of income tax policy. *Quantitative Economics 5*(1), 145–174.

Mukhin, Y. (2018). Sensitivity of regular estimators. arxiv:1805.08883 [econ.em].

Norets, A. and X. Tang (2014). Semiparametric inference in dynamic binary choice models. *The Review of Economic Studies 81*(3), 1229–1262.

Pesendorfer, M. and P. Schmidt-Dengler (2008). Asymptotic least squares estimators for dynamic games. *The Review of Economic Studies 75*(3), 901–928.

Reguant, M. (2016). Bounding outcomes in counterfactual analysis. Working paper, Northwestern University and NBER.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica 55*(5), 999–1033.

Schennach, S. M. (2014). Entropic latent variable integration via simulation. *Econometrica 82*(1), 345–385.

Shapiro, A. (1990). On concepts of directional differentiability. *Journal of Optimization Theory and Applications 66*(3), 477–487.

Shapiro, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research 30*(1), 169–186.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies 70*(1), 147–165.

Tao, J. (2014). Inference for point and partially identified semi-nonparametric conditional moment models. Technical report, University of Washington.

Torgovitsky, A. (2016). Nonparametric inference on state dependence with applications to employment dynamics. SSRN working paper, `https://ssrn.com/abstract=2564305`.

Watson, J. and C. Holmes (2016). Approximate models and robust decisions. *Statistical Science 31*(4), 465–489.

A Details for numerical examples

A.1 Discrete game of complete information

This appendix presents closed-form expressions for the programs $\delta^\star$, $\underline{\kappa}^\star$ and $\bar{\kappa}^\star$ and local sensitivity measure for the game example studied in Section 3.

A.1.1 Objective functions

To develop intuition, we first discuss the case without regressors (i.e. we only use the moment conditions corresponding to $z = 0$). Consider the program $\delta^\star$ in display (13). There are nine cell probabilities associated with different realizations of U_j in the intervals $(-\infty, -\beta_j]$, $(\beta_j, \Delta - \beta_j]$ and $(\Delta - \beta_j, \infty)$ for $j = 1, 2$. We may split the expectation $\mathbb{E}^{F_*}[e^{\lambda' g(U,\theta)}]$ into the probability weighted sum of conditional expectations over each of the nine cells. Using the moment generating function (mgf) for the truncated normal distribution, we may deduce

$$\log \mathbb{E}^{F_*}\left[e^{\lambda' g(U,\theta)}\right] = \frac{\|\lambda_4\|^2}{2} + \log\left(\mathbf{1}'\left((q_0^2(\theta,\lambda_4) q_0^1(\theta,\lambda_4)') \circ R_0(\lambda)\right)\mathbf{1}\right)$$

where "$\circ$" denotes element-wise (Hadamard) product, $\mathbf{1}$ is a conformable vector of ones,

$$q_0^1(\theta,\lambda_4) = \begin{bmatrix} G(-\beta_1-\lambda_{4,1}) \\ G(\Delta-\beta_1-\lambda_{4,1})-G(-\beta_1-\lambda_{4,1}) \\ G(\beta_1+\lambda_{4,1}-\Delta) \end{bmatrix}, \qquad q_0^2(\theta,\lambda_4) = \begin{bmatrix} G(\beta_2+\lambda_{4,2}-\Delta) \\ G(\Delta-\beta_2-\lambda_{4,2})-G(-\beta_2-\lambda_{4,2}) \\ G(-\beta_2-\lambda_{4,2}) \end{bmatrix},$$

where G denotes standard normal cumulative distribution function, and

$$R_0(\lambda) = \begin{bmatrix} e^{-\lambda_{1,2}} & e^{-\lambda_{1,2}} & e^{-\lambda_{2,2}} \\ e^{-\lambda_{1,2}} & e^{-\lambda_{1,1}-\lambda_{1,2}} & e^{-\lambda_{1,1}} \\ e^{-\lambda_{2,1}} & e^{-\lambda_{1,1}} & e^{-\lambda_{1,1}} \end{bmatrix}$$

with $\lambda = (\lambda_1', \lambda_2', \lambda_4')'$ where $\lambda_1 = (\lambda_{1,1}, \lambda_{1,2})'$, $\lambda_2 = (\lambda_{2,1}, \lambda_{2,2})'$ and $\lambda_4 = (\lambda_{4,1}, \lambda_{4,2})'$.

Similar computations apply for $\underline{\kappa}^\star$ and $\bar{\kappa}^\star$. There are two cases to consider.

Case 1: $\tau \leq \Delta$. We partition the interval for U_1 and U_2 into four regions: $(-\infty, -\beta_j]$, $(-\beta_j, \tau-\beta_j]$, $(\tau - \beta_j, \Delta - \beta_j]$, $(\Delta - \beta_j, \infty)$ for $j = 1, 2$. The counterfactual game is identical to the original game with β_j transformed to $\beta_j - \tau$ and Δ transformed to $\Delta - \tau$ for $j = 1, 2$. Thus neither firm enters if

$U_j \le \tau - \beta_j$ for $j = 1,2$ and both firms enter if $U_j \ge \Delta - \beta_j$ for $j = 1,2$. We may then deduce

$$\eta \log \mathbb{E}^{F_*}\left[e^{\eta^{-1}(k(U,\theta)+\lambda' g(U,\theta))}\right] = \frac{\|\lambda_4\|^2}{2\eta} + \eta \log \left(\mathbf{1}' \left((q^2(\theta,\lambda_4,\eta) q^1(\theta,\lambda_4,\eta)') \circ R(\lambda,\eta)\right) \mathbf{1}\right) \tag{21}$$

where

$$q^1(\theta,\lambda_4,\eta) = \begin{bmatrix} G(-\beta_1 - \frac{\lambda_{4,1}}{\eta}) \\ G(\tau-\beta_1-\frac{\lambda_{4,1}}{\eta}) - G(-\beta_1-\frac{\lambda_{4,1}}{\eta}) \\ G(\Delta-\beta_1-\frac{\lambda_{4,1}}{\eta}) - G(\tau-\beta_1-\frac{\lambda_{4,1}}{\eta}) \\ G(\beta_1+\frac{\lambda_{4,1}}{\eta}-\Delta) \end{bmatrix}, \quad q^2(\theta,\lambda_4,\eta) = \begin{bmatrix} G(\beta_2+\frac{\lambda_{4,2}}{\eta}-\Delta) \\ G(\Delta-\beta_2-\frac{\lambda_{4,2}}{\eta}) - G(\tau-\beta_2-\frac{\lambda_{4,2}}{\eta}) \\ G(\tau-\beta_2-\frac{\lambda_{4,2}}{\eta}) - G(-\beta_2-\frac{\lambda_{4,2}}{\eta}) \\ G(-\beta_2-\frac{\lambda_{4,2}}{\eta}) \end{bmatrix},$$

and

$$R(\lambda,\eta) = \begin{bmatrix} e^{(1-\lambda_{1,2})/\eta} & e^{(1-\lambda_{1,2})/\eta} & e^{(1-\lambda_{1,2})/\eta} & e^{-\lambda_{2,2}/\eta} \\ e^{(1-\lambda_{1,2})/\eta} & e^{(1-\lambda_{1,1}-\lambda_{1,2})/\eta} & e^{(1-\lambda_{1,1}-\lambda_{1,2})/\eta} & e^{(1-\lambda_{1,1})/\eta} \\ e^{-\lambda_{1,2}/\eta} & e^{(-\lambda_{1,1}-\lambda_{1,2})/\eta} & e^{(1-\lambda_{1,1}-\lambda_{1,2})/\eta} & e^{(1-\lambda_{1,1})/\eta} \\ e^{-\lambda_{2,1}/\eta} & e^{-\lambda_{1,1}/\eta} & e^{(1-\lambda_{1,1})/\eta} & e^{(1-\lambda_{1,1})/\eta} \end{bmatrix}.$$

Case 2: $\tau > \Delta$**.** We partition the interval for U_1 and U_2 into four regions: $(-\infty, -\beta_j]$, $(-\beta_j, \Delta-\beta_j]$, $(\Delta-\beta_j, \tau-\beta_j]$, $(\tau-\beta_j, \infty)$ for $j = 1,2$. When $\Delta - \beta_j < U_j \le \tau - \beta_j$ for $j = 1,2$ the game has two equilibria, namely $(0,0)$ and $(1,1)$. We do not need to deal with the problem of equilibrium selection here for the purposes of the counterfactual, however, as neither equilibrium is a monopoly. The log-mgf term is now computed as in (21), with

$$q^1(\theta,\lambda_4,\eta) = \begin{bmatrix} G(-\beta_1 - \frac{\lambda_{4,1}}{\eta}) \\ G(\Delta-\beta_1-\frac{\lambda_{4,1}}{\eta}) - G(-\beta_1-\frac{\lambda_{4,1}}{\eta}) \\ G(\tau-\beta_1-\frac{\lambda_{4,1}}{\eta}) - G(\Delta-\beta_1-\frac{\lambda_{4,1}}{\eta}) \\ G(\beta_1+\frac{\lambda_{4,1}}{\eta}-\tau) \end{bmatrix}, \quad q^2(\theta,\lambda_4,\eta) = \begin{bmatrix} G(\beta_2+\frac{\lambda_{4,2}}{\eta}-\tau) \\ G(\tau-\beta_2-\frac{\lambda_{4,2}}{\eta}) - G(\Delta-\beta_2-\frac{\lambda_{4,2}}{\eta}) \\ G(\Delta-\beta_2-\frac{\lambda_{4,2}}{\eta}) - G(-\beta_2-\frac{\lambda_{4,2}}{\eta}) \\ G(-\beta_2-\frac{\lambda_{4,2}}{\eta}) \end{bmatrix},$$

and

$$R(\lambda,\eta) = \begin{bmatrix} e^{(1-\lambda_{1,2})/\eta} & e^{(1-\lambda_{1,2})/\eta} & e^{-\lambda_{2,2}/\eta} & e^{-\lambda_{2,2}/\eta} \\ e^{-\lambda_{1,2}/\eta} & e^{-\lambda_{1,2}/\eta} & e^{-\lambda_{2,2}/\eta} & e^{-\lambda_{2,2}/\eta} \\ e^{-\lambda_{1,2}/\eta} & e^{(-\lambda_{1,1}-\lambda_{1,2})/\eta} & e^{-\lambda_{1,1}/\eta} & e^{(1-\lambda_{1,1})/\eta} \\ e^{-\lambda_{2,1}/\eta} & e^{-\lambda_{1,1}/\eta} & e^{-\lambda_{1,1}/\eta} & e^{(1-\lambda_{1,1})/\eta} \end{bmatrix}.$$

Closed-form expressions for the full case with regressors follow similarly. For each player we first construct a grid by sorting $-\beta_j$, $\Delta - \beta_j$, $-\beta_j - \beta$, etc, in ascending order. The vectors $q^1(\theta,\lambda,\eta)$ and $q^2(\theta,\lambda,\eta)$ are then formed similarly to the above case without regressors using the mgf for truncated normal random variables. The matrix $R(\lambda,\eta)$ is also formed similarly, with multipliers $\lambda_{1,1},\dots,\lambda_{1,6}$

and $\lambda_{2,1},\ldots,\lambda_{2,6}$ placed in the exponents in cells corresponding to the relevant events in the equality and inequality restrictions (1) and (2), and with 1 placed in the exponents corresponding to events in which a monopoly is observed under the policy intervention.

A.1.2 Local sensitivity

Under the parameterization in Section 3, each of the six moment inequalities are slack at (θ_0, P_0). Therefore, only the eight moment equalities (six for the model-implied conditional choice probabilities of no entry and duopoly, plus the two mean-zero restrictions) are relevant for characterizing local sensitivity. In the notation of Section 6, we have

$$h(u,\theta,P_2)=\begin{bmatrix} -1\!\!1\{U_1\leq-\beta_1;U_2\leq-\beta_2\}+P_{00,0}\\ -1\!\!1\{U_1\geq\Delta-\beta_1;U_2\geq\Delta-\beta_2\}+P_{11,0}\\ -1\!\!1\{U_1\leq-\beta_1-\beta;U_2\leq-\beta_2-\beta\}+P_{00,1}\\ -1\!\!1\{U_1\geq\Delta-\beta_1-\beta;U_2\geq\Delta-\beta_2-\beta\}+P_{11,1}\\ -1\!\!1\{U_1\leq-\beta_1-2\beta;U_2\leq-\beta_2-2\beta\}+P_{00,2}\\ -1\!\!1\{U_1\geq\Delta-\beta_1-2\beta;U_2\geq\Delta-\beta_2-2\beta\}+P_{11,2}\\ U_1\\ U_2 \end{bmatrix},$$

where $P_{00,z}$ and $P_{11,z}$ denote the conditional probabilities of no entry and duopoly, respectively, when $Z=z$. Recall that we normalize $\beta\equiv 1$ so $\theta=(\beta_1,\beta_2,\Delta)'$. Therefore

$$H=\begin{bmatrix} G'(-\beta_1)G(-\beta_2) & G(-\beta_1)G'(-\beta_2) & 0\\ -G'(\beta_1-\Delta)G(\beta_2-\Delta) & -G(\beta_1-\Delta)G'(\beta_2-\Delta) & \begin{pmatrix} G'(\beta_1-\Delta)G(\beta_2-\Delta)\\ +G(\beta_1-\Delta)G'(\beta_2-\Delta)\end{pmatrix}\\ G'(-\beta_1-\beta)G(-\beta_2-\beta) & G(-\beta_1-\beta)G'(-\beta_2-\beta) & 0\\ -G'(\beta_1+\beta-\Delta)G(\beta_2+\beta-\Delta) & -G(\beta_1+\beta-\Delta)G'(\beta_2+\beta-\Delta) & \begin{pmatrix} G'(\beta_1+\beta-\Delta)G(\beta_2+\beta-\Delta)\\ +G(\beta_1+\beta-\Delta)G'(\beta_2+\beta-\Delta)\end{pmatrix}\\ G'(-\beta_1-2\beta)G(-\beta_2-2\beta) & G(-\beta_1-2\beta)G'(-\beta_2-2\beta) & 0\\ -G'(\beta_1+2\beta-\Delta)G(\beta_2+2\beta-\Delta) & -G(\beta_1+2\beta-\Delta)G'(\beta_2+2\beta-\Delta) & \begin{pmatrix} G'(\beta_1+2\beta-\Delta)G(\beta_2+2\beta-\Delta)\\ +G(\beta_1+2\beta-\Delta)G'(\beta_2+2\beta-\Delta)\end{pmatrix}\\ 0 & 0 & 0\\ 0 & 0 & 0 \end{bmatrix},$$

where G' denotes standard normal probability density function. As $\Delta-\beta_j-\beta\leq\tau-\beta_j-\beta$ for $j=1,2$, here we also have

$$\mathbb{E}^{F_*}[k(U,\theta)]=G(\Delta-\beta_1-\beta)G(\beta_2+\beta-\tau)+G(\beta_1+\beta-\tau)G(\Delta-\beta_2-\beta)$$

so

$$J = \begin{bmatrix} G'(\beta_1+\beta-\tau)G(\Delta-\beta_2-\beta)-G'(\Delta-\beta_1-\beta)G(\beta_2+\beta-\tau) \\ G(\Delta-\beta_1-\beta)G'(\beta_2+\beta-\tau)-G(\beta_1+\beta-\tau)G'(\Delta-\beta_2-\beta) \\ G'(\Delta-\beta_1-\beta)G(\beta_2+\beta-\tau)+G(\beta_1+\beta-\tau)G'(\Delta-\beta_2-\beta) \end{bmatrix}'$$

and:

$$\mathbb{E}^{F_*}[h_0(U)k_0(U)] = \begin{bmatrix} \kappa P_{00,0} \\ \kappa P_{11,0} \\ \kappa P_{00,1} \\ \kappa P_{11,1} \\ \kappa P_{00,2} \\ \begin{pmatrix} G(\beta_1+\beta-\tau)(G(\Delta-\beta_2-\beta)-G(\Delta-\beta_2-2\beta)) \\ +G(\beta_2+\beta-\tau)(G(\Delta-\beta_1-\beta)-G(\Delta-\beta_1-2\beta)) \end{pmatrix} + \kappa P_{11,2} \\ G'(\tau-\beta_1-\beta)G(\Delta-\beta_2-\beta)-G'(\Delta-\beta_1-\beta)G(\beta_2+\beta-\tau) \\ G'(\tau-\beta_2-\beta)G(\Delta-\beta_1-\beta)-G'(\Delta-\beta_2-\beta)G(\beta_1+\beta-\tau) \end{bmatrix}.$$

Finally, we partition V conformably:

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix},$$

where $V_{21} = V_{12}'$, V_{22} is a 2×2 identity matrix,

$$V_{12} = \begin{bmatrix} G'(-\beta_1)G(-\beta_2) & G(-\beta_1)G'(-\beta_2) \\ -G'(\Delta-\beta_1)G(\beta_2-\Delta) & -G(\beta_1-\Delta)G'(\Delta-\beta_2) \\ G'(-\beta_1-\beta)G(-\beta_2-\beta) & G(-\beta_1-\beta)G'(-\beta_2-\beta) \\ -G'(\Delta-\beta_1-\beta)G(\beta_2+\beta-\Delta) & -G(\beta_1+\beta-\Delta)G'(\Delta-\beta_2-\beta) \\ G'(-\beta_1-2\beta)G(-\beta_2-2\beta) & G(-\beta_1-2\beta)G'(-\beta_2-2\beta) \\ -G'(\Delta-\beta_1-2\beta)G(\beta_2+2\beta-\Delta) & -G(\beta_1+2\beta-\Delta)G'(\Delta-\beta_2-2\beta) \end{bmatrix},$$

and

$$V_{11} = \begin{bmatrix} P_{00,0} & 0 & P_{00,1} & \begin{pmatrix} (G(-\beta_1)-G(\Delta-\beta_1-\beta)) \\ \times(G(-\beta_2)-G(\Delta-\beta_2-\beta)) \end{pmatrix} & P_{00,2} & \begin{pmatrix} (G(-\beta_1)-G(\Delta-\beta_1-2\beta)) \\ \times(G(-\beta_2)-G(\Delta-\beta_2-2\beta)) \end{pmatrix} \\ \bullet & P_{11,0} & 0 & P_{11,0} & 0 & P_{11,0} \\ \bullet & \bullet & P_{00,1} & 0 & P_{00,2} & \begin{pmatrix} (G(-\beta_1-\beta)-G(\Delta-\beta_1-2\beta)) \\ \times(G(-\beta_2-\beta)-G(\Delta-\beta_2-2\beta)) \end{pmatrix} \\ \bullet & \bullet & \bullet & P_{11,1} & 0 & P_{11,1} \\ \bullet & \bullet & \bullet & \bullet & P_{00,2} & 0 \\ \bullet & \bullet & \bullet & \bullet & \bullet & P_{11,2} \end{bmatrix} - P_2P_2',$$

where "$\bullet$" denotes corresponding upper-triangular element.

A.2 Dynamic discrete choice

This appendix contains additional calculations for the local sensitivity measure for the DDC example studied in Section 3. In the notation of Section 6, we have

$$h(u,\theta,\gamma,P_2) = \begin{bmatrix} \mathbb{1}\Big\{\pi(1;\theta_\pi) + u(1) + \beta M_1 v \geq \pi(0;\theta_\pi) + u(0) + \beta M_0(v)\Big\} - P_2 \\ \max\Big(\pi(1;\theta_\pi) + u(1) + \beta M_1 v\,,\, \pi(0;\theta_\pi) + u(0) + \beta M_0 v\Big) - v \\ \max\Big(\tilde{\pi}(1;\theta_\pi) + u(1) + \beta \tilde{M}_1 \tilde{v}\,,\, \tilde{\pi}(0;\theta_\pi) + u(0) + \beta \tilde{M}_0 \tilde{v}\Big) - \tilde{v} \end{bmatrix},$$

where the indicator function, inequality, and maximum are applied row-wise. We also have

$$k(u,\theta,\gamma) = \mathbb{1}\Big\{\tilde{\pi}(1;j,\theta_\pi) + u(1) + \beta(\tilde{M}_1\tilde{v})_j \geq \tilde{\pi}(0;j,\theta_\pi) + u(0) + \beta(\tilde{M}_0\tilde{v})_j\Big\}$$

for the CCP of being active in state j, where $(M_a\tilde{v})_j$ denotes the corresponding entry of $M_a\tilde{v}$.

For the Jacobian terms, we may use the i.i.d. type-I extreme value specification to deduce

$$\frac{\partial}{\partial\theta'}\mathbb{E}^{F_*}[k(U,\theta,\gamma)] = \frac{e^{\Delta\tilde{\pi}(j,\theta_\pi)+\beta((\tilde{M}_1-\tilde{M}_0)\tilde{v})_j}}{(1+e^{\Delta\tilde{\pi}(j,\theta_\pi)+\beta((\tilde{M}_1-\tilde{M}_0)\tilde{v})_j})^2}\begin{bmatrix} \frac{\partial\Delta\tilde{\pi}(j,\theta_\pi)}{\partial\theta'} & 0'_{n_x} & \beta(\tilde{M}_1-\tilde{M}_0)_j \end{bmatrix},$$

where $\theta = (\theta_\pi', v', \tilde{v}')'$, $\Delta\tilde{\pi}(j,\theta_\pi) = \tilde{\pi}(1;j,\theta_\pi) - \tilde{\pi}(0;j,\theta_\pi)$, and $(\tilde{M}_1 - \tilde{M}_0)_j$ denotes the row of $\tilde{M}_1 - \tilde{M}_0$ corresponding to state j. Similarly,

$$\begin{aligned} &\frac{\partial}{\partial\theta'}\mathbb{E}^{F_*}[h(U,\theta,\gamma,P_2)] \\ &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \frac{\partial\pi(0;\theta_\pi)}{\partial\theta_\pi'} & \beta M_0 - I & \mathbf{0} \\ \frac{\partial\tilde{\pi}(0;\theta_\pi)}{\partial\theta_\pi'} & \mathbf{0} & \beta\tilde{M}_0 - I \end{bmatrix} + D\begin{bmatrix} \frac{\partial\Delta\pi(\theta_\pi)}{\partial\theta_\pi'} & \beta(M_1-M_0) & \mathbf{0} \\ \frac{\partial\Delta\pi(\theta_\pi)}{\partial\theta_\pi'} & \beta(M_1-M_0) & \mathbf{0} \\ \frac{\partial\Delta\tilde{\pi}(\theta_\pi)}{\partial\theta_\pi'} & \mathbf{0} & \beta(\tilde{M}_1-\tilde{M}_0) \end{bmatrix}, \end{aligned}$$

where $\Delta\pi(\theta_\pi) = \pi(1;\theta_\pi) - \pi(0;\theta_\pi)$, $\mathbf{0}$ denotes a conformable matrix of zeros, and D is a diagonal matrix, whose diagonal entries are (in order):

$$\begin{aligned} &\frac{e^{\Delta\pi(x,\theta_\pi)+\beta(M_1-M_0)v)_x}}{(1+e^{\Delta\tilde{\pi}(x,\theta_\pi)+\beta((M_1-M_0)v)_x})^2}\,, \; x = 1,\ldots,n_x\,, \\ &\frac{e^{\Delta\pi(x,\theta_\pi)+\beta((M_1-M_0)v)_x}}{1+e^{\Delta\pi(x,\theta_\pi)+\beta((M_1-M_0)v)_x}}\,, \; x = 1,\ldots,n_x\,, \text{ and} \\ &\frac{e^{\Delta\tilde{\pi}(x,\theta_\pi)+\beta((\tilde{M}_1-\tilde{M}_0)\tilde{v})_x}}{1+e^{\Delta\tilde{\pi}(x,\theta_\pi)+\beta((\tilde{M}_1-\tilde{M}_0)\tilde{v})_x}}\,, \; x = 1,\ldots,n_x\,. \end{aligned}$$

The remaining matrices $\hat{V}$ and $\mathbb{E}^{F_*}[\hat{k}(U)\hat{h}(U)']$ may be computed numerically.

B Supplementary results on identified sets of counterfactuals

Define

$$\mathcal{K}_\infty = \{\mathbb{E}^F[k(U,\theta,\gamma_0)] \text{ such that (1)–(4) hold for some } \theta \in \Theta \text{ and } F \in \mathcal{N}_\infty\}$$

in the explicit-dependence case, and

$$\mathcal{K}_\infty = \{k(\theta,\gamma_0) \text{ such that (1)–(4) hold for some } \theta \in \Theta \text{ and } F \in \mathcal{N}_\infty\}$$

in the implicit-dependence case. Clearly $\mathcal{K}_\infty \subseteq \mathcal{K}_\#$ because each $\mathcal{F}_\theta$ will generally contain fatter-tailed distributions not in $\mathcal{N}_\infty$. The set $\mathcal{K}_\infty$ may not be the quite the identified set of counterfactuals: there may exist $F \notin \mathcal{N}_\infty$ that are consistent with the model and which yield smaller or larger values of the counterfactuals. The set $\mathcal{K}_\infty$ is however relevant as it is eventually spanned by the counterfactuals over $\mathcal{N}_\delta$ as δ gets large.

Lemma B.1 *Let Assumption Φ hold. Then:*

$$\underline{\kappa}(\mathcal{N}_\delta) \to \inf \mathcal{K}_\infty\,, \quad \bar{\kappa}(\mathcal{N}_\delta) \to \sup \mathcal{K}_\infty \qquad \text{as } \delta \to \infty.$$

In the explicit-dependence case, the smallest and largest elements of $\mathcal{K}_\infty$ may be characterized in terms of low-dimensional convex optimization problems. Define

$$\begin{aligned}
\underline{\kappa}_\infty(\theta;\gamma,P) &= \inf_{F \in \mathcal{N}_\infty} \mathbb{E}^F[k(U,\theta,\gamma)] \quad \text{subject to (1)–(4) holding at } (\theta,F)\,,\\
\bar{\kappa}_\infty(\theta;\gamma,P) &= \sup_{F \in \mathcal{N}_\infty} \mathbb{E}^F[k(U,\theta,\gamma)] \quad \text{subject to (1)–(4) holding at } (\theta,F)\,.
\end{aligned}$$

The above programs again have a dual representation as finite-dimensional convex optimization problems. Let F_*-ess inf and F_*-ess sup denote essential infimum and supremum defined relative to the measure F_*.

Lemma B.2 *Let Assumption Φ hold. Then: the duals of $\underline{\kappa}_\infty(\theta;\gamma,P)$ and $\bar{\kappa}_\infty(\theta;\gamma,P)$ are*

$$\begin{aligned}
\underline{\kappa}^\star_\infty(\theta;\gamma,P) &= \sup_{\lambda \in \Lambda : F_*\text{-ess}\inf(k(\cdot,\theta,\gamma)+\lambda' g(\cdot,\theta,\gamma)) > -\infty} \left(F_*\text{-ess}\inf(k(\cdot,\theta,\gamma)+\lambda' g(\cdot,\theta,\gamma)) - \lambda_{12}' P\right)\,,\\
\bar{\kappa}^\star_\infty(\theta;\gamma,P) &= \inf_{\lambda \in \Lambda : F_*\text{-ess}\sup(k(\cdot,\theta,\gamma)-\lambda' g(\cdot,\theta,\gamma)) < +\infty} \left(F_*\text{-ess}\sup(k(\cdot,\theta,\gamma)-\lambda' g(\cdot,\theta,\gamma)) + \lambda_{12}' P\right)\,.
\end{aligned}$$

If condition S also holds at (θ,γ,P), then: strong duality holds: $\underline{\kappa}_\infty(\theta;\gamma,P) = \underline{\kappa}^\star_\infty(\theta;\gamma,P)$ and $\bar{\kappa}_\infty(\theta;\gamma,P) = \bar{\kappa}^\star_\infty(\theta;\gamma,P)$.

Remark B.1 *The above programs are non-smooth optimization problems which may be difficult to solve for certain models. Constraining the set to $\mathcal{N}_\delta$ with $\delta < +\infty$ exerts a sort of smoothing effect on the optimization problem. For instance, with KL or hybrid divergence the above non-smooth optimization problem is replaced with the smooth programs (7) and (8), and (9) and (10), respectively.*

We may characterize the smallest and largest elements of $\mathcal{K}_\#$ under a mild constraint qualification condition. Let $\mathrm{ri}(A)$ denote the relative interior of a set $A \subset \mathbb{R}^n$ (i.e. the interior within the affine hull of A). If A has positive volume in $\mathbb{R}^n$ then $\mathrm{ri}(A) = \mathrm{int}(A)$.

Condition $\mathbf{S}_\#$ $\quad (P', 0'_{d_3+d_4})' \in \mathrm{ri}(\{\mathbb{E}^F[g(U,\theta,\gamma_0)] : F \in \mathcal{F}_\theta\} + \mathcal{C})$.

This condition is weaker than Condition S, as the class of distributions $\mathcal{N}_\infty \subseteq \mathcal{F}_\theta$ for each θ. Hence, if Condition S holds at (θ, γ_0, P_0) then Condition $\mathrm{S}_\#$ holds there also.

The smallest and largest elements of $\mathcal{K}_\#$ may be computed by solving low-dimensional convex optimization problems in the explicit-dependence case. Define

$$\begin{aligned}
\underline{\kappa}_\#(\theta;\gamma_0,P_0) &= \inf_{F \in \mathcal{F}_\theta} \mathbb{E}^F[k(U,\theta,\gamma_0)] \quad \text{subject to (1)–(4) holding at } (\theta,F)\,, \\
\bar{\kappa}_\#(\theta;\gamma_0,P_0) &= \sup_{F \in \mathcal{F}_\theta} \mathbb{E}^F[k(U,\theta,\gamma_0)] \quad \text{subject to (1)–(4) holding at } (\theta,F)\,.
\end{aligned}$$

These programs also have a dual representation as finite-dimensional convex optimization problems.

Lemma B.3 *Let Condition $S_\#$ hold at (θ, γ_0, P_0) and let μ-$\operatorname{ess\,sup}|k(\cdot,\theta,\gamma_0)| < \infty$. Then: the duals of $\underline{\kappa}_\#(\theta;\gamma_0,P_0)$ and $\bar{\kappa}_\#(\theta;\gamma_0,P_0)$ are*

$$\begin{aligned}
\underline{\kappa}^\star_\#(\theta;\gamma_0,P_0) &= \sup_{\lambda \in \Lambda : \mu\text{-}\operatorname{ess\,inf}(k(\cdot,\theta,\gamma_0)+\lambda' g(\cdot,\theta,\gamma_0)) > -\infty} \left(\mu\text{-}\operatorname{ess\,inf}(k(\cdot,\theta,\gamma_0)+\lambda' g(\cdot,\theta,\gamma_0)) - \lambda'_{12}P_0\right)\,, \\
\bar{\kappa}^\star_\#(\theta;\gamma_0,P_0) &= \inf_{\lambda \in \Lambda : \mu\text{-}\operatorname{ess\,sup}(k(\cdot,\theta,\gamma_0)-\lambda' g(\cdot,\theta,\gamma_0)) < +\infty} \left(\mu\text{-}\operatorname{ess\,sup}(k(\cdot,\theta,\gamma_0)-\lambda' g(\cdot,\theta,\gamma_0)) + \lambda'_{12}P_0\right)\,,
\end{aligned}$$

and strong duality holds: $\underline{\kappa}_\#(\theta;\gamma_0,P_0) = \underline{\kappa}^\star_\#(\theta;\gamma_0,P_0)$ and $\bar{\kappa}_\#(\theta;\gamma_0,P_0) = \bar{\kappa}^\star_\#(\theta;\gamma_0,P_0)$.

C Background material on Orlicz spaces

Our results rely on the theory of paired Orlicz spaces. We refer the reader to Krasnosel'skii and Rutickii (1961) for a textbook treatment. Komunjer and Ragusa (2016) apply similar results to characterize and study existence of information projections in conditional moment models. Define

$$\begin{aligned} \mathcal{L} &= \{f : \mathcal{U} \to \mathbb{R} \text{ such that } \mathbb{E}^{F_*}[\phi(1 + c|f(U)|)] < \infty \text{ for some } c > 0\} \\ \mathcal{E} &= \{f : \mathcal{U} \to \mathbb{R} \text{ such that } \mathbb{E}^{F_*}[\psi(c|f(U)|)] < \infty \text{ for all } c > 0\}\,. \end{aligned}$$

The class $\mathcal{L}$ is an Orlicz class of functions corresponding to the function $x \mapsto \phi(1+|x|)$ whereas the class $\mathcal{E}$ is the Orlicz heart corresponding to the conjugate function ψ. To summarize the results that we use, note that the condition $\lim_{x\to\infty} x\phi'(x)/\phi(x) < \infty$ in Assumption Φ(i) verifies the so-called Δ_2-condition in Krasnosel'skii and Rutickii (1961). The space $\mathcal{L}$ is a separable Banach space when equipped with the norm

$$\|f\|_\phi = \inf_{c>0} \frac{1}{c}(1 + \mathbb{E}^{F_*}[\phi(1 + c|f(U)|)])\,,$$

and the space $\mathcal{E}$ is a separable Banach space when equipped with the norm

$$\|f\|_\psi = \inf_{c>0} \frac{1}{c}(1 + \mathbb{E}^{F_*}[\psi(c|f(U)|)])$$

(Krasnosel'skii and Rutickii, 1961, Chapter II, Section 10).

Given two functions ϕ_1, ϕ_2 satisfying Assumption Φ(i), write $\phi_1 \prec \phi_2$ if there exists positive constants c and x_0 such that $\phi_1(x) \leq \phi_2(cx)$ for all $x \geq x_0$. If $\phi_1 \prec \phi_2$ and $\phi_2 \prec \phi_1$ then ϕ_1 and ϕ_2 are said to be equivalent. Equivalent ϕ functions induce the same spaces $\mathcal{L}$ and $\mathcal{E}$ and their corresponding norms $\|\cdot\|_{\phi_1}$ and $\|\cdot\|_{\phi_2}$ are equivalent (Krasnosel'skii and Rutickii, 1961, Theorems 13.1 and 13.3). For example, the ϕ functions inducing hybrid and χ^2 divergence are equivalent.

A sequence $\{f_n : n \geq 1\} \subset \mathcal{L}$ is $\mathcal{E}$-weakly convergent if $\{\mathbb{E}^{F_*}[f_n(U)g(U)] : n \geq 1\}$ converges for each $g \in \mathcal{E}$. The space $\mathcal{L}$ is $\mathcal{E}$-weakly complete: any $\mathcal{E}$-weakly convergent sequence of functions $\{f_n : n \geq 1\} \subset \mathcal{L}$ has a unique limit, say $f^* \in \mathcal{L}$, for which

$$\lim_{n\to\infty} \mathbb{E}^{F_*}[f_n(U)g(U)] = \mathbb{E}^{F_*}[f^*(U)g(U)]$$

for each $g \in \mathcal{E}$; it is also $\mathcal{E}$-weakly compact: every $\|\cdot\|_\phi$-norm bounded sequence in $\mathcal{L}$ has an $\mathcal{E}$-weakly convergent subsequence (Krasnosel'skii and Rutickii, 1961, Theorem 14.4). A version of Hölder's inequality also holds:

$$|\mathbb{E}^{F_*}[f(U)g(U)]| \leq \|f\|_\phi \|g\|_\psi$$

for each $f \in \mathcal{L}$ and $g \in \mathcal{E}$ (Krasnosel'skii and Rutickii, 1961, Theorem 9.3).

D Proof of main results

We start with some preliminary Lemmas. Let $\mathcal{L}_+ := \{m \in \mathcal{L} : m \geq 0 \text{ (F_*-almost everywhere)}\}$.

Lemma D.1 *Let Assumption Φ hold. Then: $\mathbb{E}^{F_*}[\phi(m(U))] < \infty$ if and only if $m \in \mathcal{L}_+$.*

Proof of Lemma D.1. To prove $\mathbb{E}^{F_*}[\phi(m(U))] < \infty$ implies $m \in \mathcal{L}_+$, first note that we must have $m \geq 0$ (F_*-almost everywhere) because $\phi(x) = +\infty$ if $x < 0$. Taking $c = \frac{1}{2}$ in the definition of $\|\cdot\|_\phi$, we obtain $\|m\|_\phi \leq 2 + \phi(2) + \mathbb{E}^{F_*}[\phi(m(U))] < \infty$.

To prove $m \in \mathcal{L}_+$ implies $\mathbb{E}^{F_*}[\phi(m(U))] < \infty$, first note that the condition $\lim_{x\to\infty} x\phi'(x)/\phi(x) < \infty$ in Assumption Φ(i) verifies the so-called Δ_2-condition. Thus, $m \in \mathcal{L}$ implies $\mathbb{E}^{F_*}[\phi(1 + c|m(U)|)] < \infty$ for all $c > 0$. As F_* is a finite measure, $\mathcal{L}$ also contains constant functions. As $\mathcal{L}$ is closed under addition, we therefore have

$$\infty > \mathbb{E}^{F_*}[\phi(1 + |m(U) - 1|)] = \mathbb{E}^{F_*}[\phi(m(U))1\!\!1\{m(U) \geq 1\}] + \mathbb{E}^{F_*}[\phi(2 - m(U))1\!\!1\{m(U) \leq 1\}]$$

which implies $\mathbb{E}^{F_*}[\phi(m(U))1\!\!1\{m(U) \geq 1\}]$ is finite. Finiteness of $\mathbb{E}^{F_*}[\phi(m(U))1\!\!1\{m(U) \leq 1\}]$ follows because $\infty > \phi(0) \geq \phi(x) \geq \phi(1) = 0$ for $x \in [0,1]$ under Assumption Φ. ■

We identify each $F \in \mathcal{N}_\delta$ with its Radon–Nikodym derivative with respect to F_*. Let $\mathcal{M}_\delta$ denote the set of all measurable $m : \mathcal{U} \to \mathbb{R}$ for which $\mathbb{E}^{F_*}[\phi(m(U))] \leq \delta$. Note that $\mathcal{M}_\delta$ is a $\|\cdot\|_\phi$-norm bounded subset of $\mathcal{L}$ by the proof of Lemma D.1. Therefore, $|\mathbb{E}^F[k(U,\theta,\gamma)]| \leq \|k(\cdot,\theta,\gamma)\|_\psi \|m\|_\phi$ holds for any $F \in \mathcal{N}_\delta$ by a version of Hölder's inequality for Orlicz classes. Finiteness of $|\mathbb{E}^F[k(U,\theta,\gamma)]|$ follows by Assumption Φ(ii) whenever $F \in \mathcal{N}_\delta$.

In each of the following proofs, we only prove the results for the lower value $\underline{\kappa}$. The result for the upper value $\bar{\kappa}$ follows by parallel arguments.

In view of Lemma D.1, an equivalent formulation of $\underline{\kappa}_\delta(\theta;\gamma,P)$ is

$$\begin{aligned}
\underline{\kappa}_\delta(\theta;\gamma,P) = \inf_{m \in \mathcal{M}_\delta} \mathbb{E}^{F_*}[m(U)k(U,\theta,\gamma)] \quad \text{subject to} \quad & \mathbb{E}^{F_*}[m(U)] = 1\,, \\
& \mathbb{E}^{F_*}[m(U)g_1(U,\theta,\gamma)] \leq P_1\,, \\
& \mathbb{E}^{F_*}[m(U)g_2(U,\theta,\gamma)] = P_2\,, \\
& \mathbb{E}^{F_*}[m(U)g_3(U,\theta,\gamma)] \leq 0\,, \\
& \mathbb{E}^{F_*}[m(U)g_4(U,\theta,\gamma)] = 0\,,
\end{aligned} \tag{22}$$

where $\underline{\kappa}_\delta(\theta;\gamma,P) = +\infty$ if infimum runs over an empty set. Similarly, an equivalent formulation of the program

$$I(\theta;\gamma,P) := \inf_F D_\phi(F\|F_*) \quad \text{subject to (1)–(4) holding under } F \text{ at } (\theta,\gamma,P)$$

is

$$\inf_{m\in\mathcal{L}_+} \mathbb{E}^{F_*}[\phi(m(U))] \quad \text{subject to the moment conditions in (22)}\,. \tag{23}$$

We verify a Slater constraint qualification condition for the programs $I(\theta;\gamma,P)$ and $\underline{\kappa}_\delta(\theta;\gamma,P)$ and $\bar{\kappa}_\delta(\theta;\gamma,P)$. The condition corresponds to display (2.312) in Bonnans and Shapiro (2000).

Lemma D.2 *Let Assumption Φ hold and let Condition S hold at (θ,γ,P). Then:*

$$(1,P',0'_{d_3+d_4})' \in \operatorname{int}(\{\mathbb{E}^{F_*}[m(U)(1,g(U,\theta,\gamma)')']: m\in\mathcal{L}_+\} + \{0\}\times\mathcal{C})\,,$$

which is the constraint qualification for $I(\theta;\gamma,P)$. If $I(\theta;\gamma,P)<\delta$ also holds, then:

$$(\delta,1,P',0'_{d_3+d_4})' \in \operatorname{int}(\{(\mathbb{E}^{F_*}[\phi(m(U))],\mathbb{E}^{F_*}[m(U)(1,g(U,\theta,\gamma)')])': m\in\mathcal{L}_+\} + \mathbb{R}_+\times\{0\}\times\mathcal{C})\,,$$

which is the constraint qualification for $\underline{\kappa}_\delta(\theta;\gamma,P)$ and $\bar{\kappa}_\delta(\theta;\gamma,P)$.

Proof of Lemma D.2. As condition S holds at (θ,γ,P), we have (applying integration element-wise)

$$(P',0'_{d_3+d_4})' \in \operatorname{int}\left(\left\{\int g(u,\theta,\gamma)\,\mathrm{d}F(u): F\in\mathcal{N}_\infty\right\}+\mathcal{C}\right)\,.$$

For each $t>0$, let $t\mathcal{N}_\delta = \{tF : F\in\mathcal{N}_\delta\}$. We then have

$$\begin{aligned}(1,P',0'_{d_3+d_4})' &\in \operatorname{int}\left(\left\{\int (1,g(u,\theta,\gamma)')'\,\mathrm{d}G(u): G\in\cup_{t\in[\frac{1}{2},\frac{3}{2}]}t\mathcal{N}_\delta\right\}+\{0\}\times\mathcal{C}\right)\\ &\subseteq \operatorname{int}(\{\mathbb{E}^{F_*}[m(U)(1,g(U,\theta,\gamma)')']: m\in\mathcal{L}_+\}+\{0\}\times\mathcal{C})\,,\end{aligned} \tag{24}$$

where the second inclusion is by Lemma D.1 and the fact that $\mathcal{L}$ is a linear space (so $m\in\mathcal{L}_+$ implies $tm\in\mathcal{L}_+$ for all $t\geq 0$). This verifies the constraint qualification for $I(\theta;\gamma,P)$.

Now, if $\delta^\star(\theta;\gamma,P)<\delta$ then $m_{\theta,\gamma,P}$ is feasible for (22) and $\mathbb{E}^{F_*}[\phi(m_{\theta,\gamma,P}(U))]<\delta$. By the inclusion (24), we have

$$\begin{aligned}(\delta,1,P',0'_{d_3+d_4})' \in \operatorname{int}(\{&(\mathbb{E}^{F_*}[\phi(\tilde{m}(U))],\mathbb{E}^{F_*}[\tilde{m}(U)(1,g(U,\theta,\gamma)')])':\\ &\tilde{m}=tm_{\theta,\gamma,P}+(1-t)m, m\in\mathcal{L}_+, t\in[0,1]\}+\mathbb{R}_+\times\{0\}\times\mathcal{C})\\ \subseteq \operatorname{int}(\{&(\mathbb{E}^{F_*}[\phi(m(U))],\mathbb{E}^{F_*}[m(U)(1,g(U,\theta,\gamma)')])': m\in\mathcal{L}_+\}+\mathbb{R}_+\times 0\times\mathcal{C})\,.\end{aligned}$$

This verifies the constraint qualification for $\underline{\kappa}_\delta(\theta;\gamma,P)$ and $\bar{\kappa}_\delta(\theta;\gamma,P)$. ∎

Proof of Lemma 4.1. The proof extends arguments from Section 3.2 of Shapiro (2017) to deal with (i) the equalities/inequalities in (1)–(4) representing the equilibrium conditions of the model, (ii) the function class $\mathcal{L}$, and (iii) additional issues that arise at $\eta = 0$.

Using the equivalent formulation in display (22), the Lagrangian for $\underline{\kappa}_\delta(\theta;\gamma,P)$ is

$$L(m,\eta,\zeta,\lambda) = \mathbb{E}^{F_*}[m(U)(k(U,\theta,\gamma)+\zeta+\lambda' g(U,\theta,\gamma)) + \eta\phi(m(U))] - \eta\delta - \zeta - \lambda_{12}' P\,,$$

where $m \in \mathcal{L}_+$, $\eta \in \mathbb{R}_+$, $\zeta \in \mathbb{R}$, and $\lambda \in \Lambda$. The Lagrangian dual problem is therefore $\underline{\kappa}_\delta^\star(\theta;\gamma,P) = \sup_{\eta\geq 0,\zeta\in\mathbb{R},\lambda\in\Lambda} \inf_{m\in\mathcal{L}_+} L(m,\eta,\zeta,\lambda)$. As $\mathcal{L}$ is decomposable (Rockafellar and Wets, 1998, Definition 14.59 and Theorem 14.60), we may bring the infimum inside the expectation to obtain

$$\begin{aligned}
\inf_{m\in\mathcal{L}_+} L(m,\eta,\zeta,\lambda) &= \mathbb{E}^{F_*}\Big[\inf_{x\geq 0} x(k(U,\theta,\gamma)+\zeta+\lambda' g(U,\theta,\gamma)) + \eta\phi(x)\Big] - \eta\delta - \zeta - \lambda_{12}' P \\
&= -\mathbb{E}^{F_*}\Big[\sup_{x\geq 0} x(-k(U,\theta,\gamma)-\zeta-\lambda' g(U,\theta,\gamma)) - \eta\phi(x)\Big] - \eta\delta - \zeta - \lambda_{12}' P \\
&= -\mathbb{E}^{F_*}\Big[(\eta\phi)^\star(-k(U,\theta,\gamma)-\zeta-\lambda' g(U,\theta,\gamma))\Big] - \eta\delta - \zeta - \lambda_{12}' P\,.
\end{aligned}$$

The dual formulation in display (5) now follows.

When Condition S holds and $\delta^\star(\theta;\gamma,P) < \delta$, it follows by Lemma D.2 and Theorem 2.165 of Bonnans and Shapiro (2000) that the set of solutions to the dual program is a nonempty, convex, compact subset of $\mathbb{R}_+ \times \mathbb{R} \times \Lambda$. Suppose that the solution is attained at $(0,\zeta^*,\lambda^*)$ for some $\zeta^* \in \mathbb{R}$ and $\lambda \in \Lambda^*$. Let $\ell(\eta,\zeta,\lambda) = \inf_{m\in\mathcal{L}_+} L(m,\eta,\zeta,\lambda)$. As ℓ is the pointwise infimum of affine functions, it is concave and upper-semicontinuous. By upper-semicontinuity, we have

$$\ell(0,\zeta^*,\lambda^*) \geq \limsup_{\eta\downarrow 0} \ell(\eta,\zeta^*,\lambda^*)\,.$$

Note the value $\ell(0,\zeta^*,\lambda^*)$ is necessarily finite: there is no duality gap (by Condition S) and the value of the primal problem is bounded (by Assumption Φ). On the other hand, for any $\tau \in (0,1)$, by concavity:

$$\ell(\tau\eta,\zeta^*,\lambda^*) \geq \tau\ell(\eta,\zeta^*,\lambda^*) + (1-\tau)\ell(0,\zeta^*,\lambda^*)\,.$$

Taking $\liminf_{\eta\downarrow 0}$ of both sides and rearranging, we obtain

$$\liminf_{\eta\downarrow 0} \ell(\eta,\zeta^*,\lambda^*) \geq \ell(0,\zeta^*,\lambda^*)\,,$$

hence $\lim_{\eta\downarrow 0} \ell(\eta,\zeta^*,\lambda^*) = \ell(0,\zeta^*,\lambda^*)$. We therefore have

$$\ell(0,\zeta^*,\lambda^*) = \sup_{\eta\geq 0,\zeta\in\mathbb{R},\lambda\in\Lambda} \ell(\eta,\zeta,\lambda) \geq \sup_{\eta> 0,\zeta\in\mathbb{R},\lambda\in\Lambda} \ell(\eta,\zeta,\lambda) \geq \lim_{\eta\downarrow 0} \ell(\eta,\zeta^*,\lambda^*) = \ell(0,\zeta^*,\lambda^*)\,.$$

Therefore, it is without loss of generality to take the supremum over $(0,\infty)\times\mathbb{R}\times\Lambda$.

For KL divergence, $\phi^\star(x) = e^x - 1$ so $(\eta\phi)^\star(x) = \eta e^{\eta^{-1}x} - \eta$ for $\eta > 0$. Therefore, for any $\eta > 0$

$$\inf_{m \in \mathcal{L}_+} L(m, \eta, \zeta, \lambda) = -\eta \mathbb{E}^{F_*}\left[e^{-\eta^{-1}(k(U,\theta,\gamma)+\zeta+\lambda' g(U,\theta,\gamma))}\right] + \eta - \eta\delta - \zeta - \lambda_{12}' P\,.$$

Optimizing with respect to ζ gives

$$\sup_{\zeta \in \mathbb{R}} \inf_{m \in \mathcal{L}_+} L(m, \eta, \zeta, \lambda) = -\eta \log \mathbb{E}^{F_*}\left[e^{-\eta^{-1}(k(U,\theta,\gamma)+\lambda' g(U,\theta,\gamma))}\right] - \eta\delta - \lambda_{12}' P\,.$$

For hybrid divergence, we have $\phi^\star(x) = \Psi(x)$ where $\Psi(x)$ is defined in equation (11). Therefore, for any $\eta > 0$

$$\inf_{m \in \mathcal{L}_+} L(m, \eta, \zeta, \lambda) = -\eta \mathbb{E}^{F_*}\left[\Psi\Big(-\eta^{-1}\big(k(U,\theta,\gamma) + \zeta + \lambda' g(U,\theta,\gamma)\big)\Big)\right] - \eta\delta - \zeta - \lambda_{12}' P$$

as claimed. ■

Proof of Lemma 4.2. Using the equivalent formulation of $I(\theta;\delta,P)$ in display (23), the Lagrangian for $I(\theta;\gamma,P)$ is $L(m,\zeta,\lambda) = \mathbb{E}^{F_*}[m(U)(-\zeta - \lambda' g(U,\theta,\gamma)) + \phi(m(U))] - \zeta - \lambda_{12}' P$. It follows by similar arguments to Lemma 4.1 that

$$\inf_{m \in \mathcal{L}_+} L(m, \zeta, \lambda) = -\mathbb{E}^{F_*}\left[\phi^\star(\zeta + \lambda' g(U,\theta,\gamma))\right] - \zeta - \lambda_{12}' P\,,$$

hence

$$\delta^\star(\theta;\gamma,P) = \sup_{\zeta \in \mathbb{R}, \lambda \in \Lambda} -\mathbb{E}^{F_*}\left[\phi^\star(-\zeta - \lambda' g(U,\theta,\gamma))\right] - \zeta - \lambda_{12}' P$$

as in display (12). For KL divergence, we have

$$\begin{aligned}
\delta^\star(\theta;\gamma,P) &= \sup_{\zeta \in \mathbb{R}, \lambda \in \Lambda} -\mathbb{E}^{F_*}\left[e^{-\zeta - \lambda' g(U,\theta,\gamma)}\right] + 1 - \zeta - \lambda_{12}' P \\
&= \sup_{\lambda \in \Lambda} -\log \mathbb{E}^{F_*}\left[e^{-\lambda' g(U,\theta,\gamma)}\right] - \lambda_{12}' P\,.
\end{aligned}$$

Similarly, for hybrid divergence, we have

$$\delta^\star(\theta;\gamma,P) = \sup_{\zeta \in \mathbb{R}, \lambda \in \Lambda} -\mathbb{E}^{F_*}\left[\Psi\left(-\zeta - \lambda' g(U,\theta,\gamma)\right)\right] - \zeta - \lambda_{12}' P\,.$$

Part (i): By weak duality, we have $I(\theta;\gamma,P) \geq \delta^\star(\theta;\gamma,P)$. Therefore, $\delta^\star(\theta;\gamma,P) > \delta$ implies there is no solution $F \in \mathcal{N}_\delta$ satisfying (1)–(4) at (θ,γ,P), so we set $\underline{\kappa}_\delta(\theta;\gamma,P) = +\infty$ and $\bar{\kappa}_\delta(\theta;\gamma,P) = -\infty$.

Part (ii): By the first part of Lemma D.2, strong duality holds for $I(\theta;\gamma,P)$ so $I(\theta;\gamma,P) = \delta^\star(\theta;\gamma,P)$. First consider the explicit-dependence case. If $\delta^\star(\theta;\gamma,P) = \delta$ then $m_{\theta,\gamma,P}$ is the unique

$m \in \mathcal{M}_\delta$ that satisfies the constraints in (22). Thus, both lower and upper values of the counterfactuals are attained under $F_{\theta,\gamma,P}$. If, in addition, $\delta^\star(\theta;\gamma,P) < \delta$ then strong duality holds for $\underline{\kappa}_\delta(\theta;\gamma,P)$ and $\bar{\kappa}_\delta(\theta;\gamma,P)$ (Bonnans and Shapiro, 2000, Theorem 2.165).

In the implicit-dependence case, if $\delta^\star(\theta;\gamma,P) \leq \delta$ then there exists a distribution $F \in \mathcal{N}_\delta$ satisfying the constraints (1)–(4) at (θ,γ,P), so the counterfactual is $k(\theta,\gamma)$. ■

Proof of Lemma 4.3. First consider the explicit-dependence case. By Lemma 4.2(ii), we have that $\inf_{\theta\in\Theta_\delta} \underline{\kappa}_\delta(\theta;\gamma_0,P_0) = \inf_{\theta\in\Theta_\delta} \underline{\kappa}^\star_\delta(\theta;\gamma_0,P_0)$ under Assumption Φ and Assumption M(ii). As $\inf_{\theta\in\Theta_\delta} \underline{\kappa}_\delta(\theta;\gamma_0,P_0) \geq \inf_{\theta\in\Theta} \underline{\kappa}_\delta(\theta;\gamma_0,P_0) =: \underline{\kappa}(\mathcal{N}_\delta)$, it therefore remains to show

$$\inf_{\theta\in\Theta_\delta} \underline{\kappa}_\delta(\theta;\gamma_0,P_0) \leq \inf_{\theta\in\Theta} \underline{\kappa}_\delta(\theta;\gamma_0,P_0)\,.$$

We prove this inequality by contradiction. Suppose there exists $\theta^* \notin \Theta_\delta$ with $\underline{\kappa}_\delta(\theta^*;\gamma_0,P_0) < \inf_{\theta\in\Theta_\delta} \underline{\kappa}_\delta(\theta;\gamma_0,P_0)$. As $\underline{\kappa}_\delta(\theta^*;\gamma_0,P_0) < +\infty$, there must exist some $m \in \mathcal{M}_\delta$ satisfying the constraints in (22) at (θ^*,γ_0,P_0). But as $\delta^\star(\theta^*;\gamma_0,P_0) = \delta$, it follows by convexity of ϕ that $m_{\theta^*,\gamma_0,P_0}$ must be the unique such $m \in \mathcal{M}_\delta$. Therefore

$$\begin{aligned}
\underline{\kappa}_\delta(\theta^*;\gamma_0,P_0) &= \mathbb{E}^{F_*}[m_{\theta^*,\gamma_0,P_0}(U)k(U,\theta^*,\gamma_0)] \\
&< \inf_{\theta\in\Theta_\delta} \underline{\kappa}_\delta(\theta;\gamma_0,P_0) \\
&\leq \inf_{\theta\in\Theta_\delta} \mathbb{E}^{F_*}[m_{\theta,\gamma_0,P_0}(U)k(U,\theta_0,\gamma_0)]\,. \qquad (25)
\end{aligned}$$

By Assumption M(iii), we must have $\theta^* \in \mathrm{cl}(\Theta_\delta)$. Take a sequence $\{\theta_n : n \geq 1\} \subset \Theta_\delta$ with $\theta_n \to \theta^*$. As $\{m_{\theta_n,\gamma_0,P_0} : n \geq 1\} \subset \mathcal{M}_\delta$, it is $\|\cdot\|_\phi$-norm bounded and hence has an $\mathcal{E}$-weakly convergent subsequence (Krasnosel'skii and Rutickii, 1961, Theorem 14.4). That is, there exists a subsequence $\{\theta_{n_i} : i \geq 1\}$ and a unique $m^* \in \mathcal{L}$ such that $m_{\theta_{n_i},\gamma_0,P_0}$ is $\mathcal{E}$-weakly convergent to m^*. By the triangle inequality and Hölder's inequality, we may deduce

$$\begin{aligned}
&\left|\mathbb{E}^{F_*}[m_{\theta_{n_i},\gamma_0,P_0}(U)k(U,\theta_{n_i},\gamma_0)] - \mathbb{E}^{F_*}[m^*(U)k(U,\theta^*,\gamma_0)]\right| \\
&\quad \leq |\mathbb{E}^{F_*}[(m_{\theta_{n_i},\gamma_0,P_0}(U) - m^*(U))k(U,\theta^*,\gamma_0)]| + \|m^*\|_\phi \|k(\,\cdot\,,\theta_{n_i},\gamma_0) - k(\,\cdot\,,\theta^*,\gamma_0)\|_\psi\,,
\end{aligned}$$

where the first term on the right-hand side vanishes by $\mathcal{E}$-weak convergence and the second term vanishes by Assumption M(i) and Hölder's inequality for Orlicz classes. By similar arguments, we may deduce

$$\begin{aligned}
\mathbb{E}^{F_*}[m^*(U)] = 1\,, \qquad &\mathbb{E}^{F_*}[m^*(U)g_1(U,\theta^*,\gamma_0)] \leq P_{10}\,, \qquad &&\mathbb{E}^{F_*}[m^*(U)g_2(U,\theta^*,\gamma_0)] = P_{20}\,, \\
&\mathbb{E}^{F_*}[m^*(U)g_3(U,\theta^*,\gamma_0)] \leq 0\,, &&\mathbb{E}^{F_*}[m^*(U)g_4(U,\theta^*,\gamma_0)] = 0\,.
\end{aligned}$$

Therefore, m^* satisfies all the constraints in (22) at $(\theta^*, \gamma_0, P_0)$. Finally, as $m \mapsto \mathbb{E}^{F_*}[\phi(m)]$ is lower semicontinuous in the $\mathcal{E}$-weak topology on $\mathcal{L}$ (Komunjer and Ragusa, 2016, p. 961), we have $\delta \geq \liminf_{i\to\infty} \mathbb{E}^{F_*}[\phi(m_{\theta_{n_i},\gamma_0,P_0}(U))] \geq \mathbb{E}^{F_*}[\phi(m^*(U))]$ hence $m^* \in \mathcal{M}_\delta$.

To complete the proof, note that Lemma 4.2 and strict convexity of ϕ implies $m^* = m_{\theta^*,\gamma_0,P_0}$, hence

$$
\begin{aligned}
\inf_{\theta\in\Theta_\delta} \mathbb{E}^{F_*}[m_{\theta,\gamma,P}(U)k(U,\theta,\gamma_0)] &\leq \lim_{i\to\infty} \mathbb{E}^{F_*}[m_{\theta_{n_i},\gamma_0,P_0}(U)k(U,\theta_{n_i},\gamma_0)] \\
&= \mathbb{E}^{F_*}[m_{\theta^*,\gamma_0,P_0}(U)k(U,\theta^*,\gamma_0)]\,,
\end{aligned}
$$

which contradicts (25).

In the implicit-dependence case, $\mathcal{E}$-continuity of $k(\theta,\gamma)$ is equivalent to continuity of the function $(\theta,\gamma) \mapsto k(\theta,\gamma)$. Then $\inf_{\theta\in\Theta_\delta} k(\theta,\gamma_0) = \inf_{\theta\in\mathrm{cl}(\Theta_\delta)} k(\theta,\gamma_0) \leq \inf_{\theta:\delta^\star(\theta;\gamma_0,P_0)\leq\delta} k(\theta,\gamma_0)$ by Assumption M(iii), and $\inf_{\theta:\delta^\star(\theta;\gamma_0,P_0)\leq\delta} k(\theta,\gamma_0) \leq \inf_{\theta\in\Theta_\delta} k(\theta,\gamma_0)$. ■

Proof of Theorem 4.1. This follows immediately from Lemmas E.4 and E.5 and Slutsky's theorem. ■

Lemma 4.4 is proved by extending some arguments from Shapiro (2008) to allow for non-compactness of the domain and possibly discontinuous objective function. These extensions are important because the multipliers take values in $\mathbb{R}_+ \times \mathbb{R} \times \Lambda$, which is not compact, and discontinuity of the objective may arise along the boundary where $\eta = 0$. Some of our extensions use techniques from Pollard (1991) on asymptotics for minimizers of convex stochastic processes.

Proof of Lemma 4.4. We first prove the result for the explicit-dependence case. Write $\underline{\Xi}_\delta = \underline{\Xi}_\delta(\theta;P_0)$ and $\underline{\Lambda}_\delta = \underline{\Lambda}_\delta(\theta;P_0)$ to denote dependence of the sets of multipliers on P. In view of Lemmas E.1, E.3 and E.6, the sets $\underline{\Xi}_\delta(\theta;P)$ and $\underline{\Lambda}_\delta(\theta;P)$ are nonempty, convex and compact for all P in a neighborhood of P_0. Let $P_n = P_0 + t_n\pi_n$. We first prove the inequality

$$
\limsup_{n\to\infty} \frac{\underline{\kappa}(\mathcal{N}_\delta;P_n) - \underline{\kappa}(\mathcal{N}_\delta;P_0)}{t_n} \leq \inf_{\theta\in\underline{\Theta}_\delta} \sup_{(\lambda_1',\lambda_2')\in\underline{\Lambda}_\delta(\theta)} -(\lambda_1',\lambda_2')\pi\,. \tag{26}
$$

Take any $\underline{\theta} \in \underline{\Theta}_\delta$. Condition S holds at $(\underline{\theta}, P_0)$ by Assumption M(ii) and hence also at $(\underline{\theta}, P_n)$ for all n sufficiently large by Lemma E.1. It follows that for n sufficiently large we have

$$
\underline{\kappa}_\delta(\underline{\theta};P_0) = \underline{\kappa}_\delta^\star(\underline{\theta};P_0)\,, \quad \underline{\kappa}_\delta(\underline{\theta};P_n) = \underline{\kappa}_\delta^\star(\underline{\theta};P_n) \tag{27}
$$

in which case, for any $(\lambda_{1n}',\lambda_{2n}') \in \underline{\Lambda}_\delta(\underline{\theta};P_n)$, the inequality

$$
\underline{\kappa}_\delta^\star(\underline{\theta};P_0) \geq \underline{\kappa}_\delta^\star(\underline{\theta};P_n) - (\lambda_{1n}',\lambda_{2n}')(P_0 - P_n) \tag{28}
$$

must hold. By (27) and (28), for n sufficiently large we must have

$$\underline{\kappa}(\mathcal{N}_\delta; P_0) = \underline{\kappa}^\star_\delta(\underline{\theta}; P_0) \geq \underline{\kappa}^\star_\delta(\underline{\theta}; P_n) - (\lambda'_{1n}, \lambda'_{2n})(P_0 - P_n) \geq \underline{\kappa}(\mathcal{N}_\delta; P_n) - (\lambda'_{1n}, \lambda'_{2n})(P_0 - P_n)$$

hence $\underline{\kappa}(\mathcal{N}_\delta; P_n) - \underline{\kappa}(\mathcal{N}_\delta; P_0) \leq -(\lambda'_{1n}, \lambda'_{2n})(P_n - P_0)$. As $d((\lambda'_{1n}, \lambda'_{2n}), \underline{\Lambda}_\delta(\underline{\theta}; P_0)) \to 0$ (cf. Lemma E.8) and $\underline{\Lambda}_\delta(\underline{\theta}; P_0)$ is bounded, we obtain the inequality

$$\limsup_{n\to\infty} \frac{\underline{\kappa}(\mathcal{N}_\delta; P_n) - \underline{\kappa}(\mathcal{N}_\delta; P_0)}{t_n} \leq \sup_{(\lambda'_1, \lambda'_2)\in\underline{\Lambda}_\delta(\underline{\theta})} -(\lambda'_1, \lambda'_2)\pi\,.$$

As $\underline{\theta} \in \underline{\Theta}_\delta(P)$ was arbitrary, this proves (26).

We now establish the corresponding lower bound

$$\liminf_{n\to\infty} \frac{\underline{\kappa}(\mathcal{N}_\delta; P_n) - \underline{\kappa}(\mathcal{N}_\delta; P_0)}{t_n} \geq \inf_{\theta\in\underline{\Theta}_\delta} \sup_{(\lambda'_1, \lambda'_2)\in\underline{\Lambda}_\delta(\theta)} -(\lambda'_1, \lambda'_2)\pi\,. \tag{29}$$

Choose a sequence $\{\theta_n : n \geq 1\} \subset \Theta$ such that $\underline{\kappa}_\delta(\theta_n; P_n) \leq \underline{\kappa}(\mathcal{N}_\delta; P_n) + o(t_n)$. By Assumption M(iv) (passing to a subsequence if necessary) we may assume that θ_n converges to some $\theta^* \in \Theta$.

We first show that $\theta^* \in \underline{\Theta}_\delta$. For each n, choose m_n solving the primal problem for $\underline{\kappa}_\delta(\theta_n; P_n)$. As $\{m_n : n \geq 1\} \subset \mathcal{M}_\delta$, it has an $\mathcal{E}$-weakly convergent subsequence $\{m_{n_i} : i \geq 1\}$ with $m_{n_i} \to m^* \in \mathcal{L}$. By similar arguments to the proof of Lemma 4.3, we may deduce that $m^* \in \mathcal{M}_\delta$, the constraints in (22) are all satisfied by m^* at (θ^*, P_0), and $\mathbb{E}^{F_*}[m_{n_i}(U)k(U, \theta_{n_i})] \to \mathbb{E}^{F_*}[m^*(U)k(U, \theta^*)]$. By definition of θ_n,

$$\underline{\kappa}(\mathcal{N}_\delta; P_{n_i}) \leq \mathbb{E}^{F_*}[m_{n_i}(U)k(U, \theta_{n_i})] \leq \underline{\kappa}(\mathcal{N}_\delta; P_{n_i}) + o(t_{n_i})\,.$$

Lemma E.4 implies $\underline{\kappa}(\mathcal{N}_\delta; P_{n_i}) \to \underline{\kappa}(\mathcal{N}_\delta; P_0)$, hence $\mathbb{E}^{F_*}[m^*(U)k(U, \theta^*)] = \underline{\kappa}(\mathcal{N}_\delta; P_0)$ and so $\theta^* \in \underline{\Theta}_\delta$.

Returning to the proof of (29), note Condition S holds at (θ^*, P_0) by Assumption M(ii). Therefore, Condition S also holds at (θ_n, P_0) and (θ_n, P_n) for all n sufficiently large (cf. Lemma E.2), and we obtain

$$\underline{\kappa}_\delta(\theta_n; P_0) = \underline{\kappa}^\star_\delta(\theta_n; P_0)\,, \quad \underline{\kappa}_\delta(\theta_n; P_n) = \underline{\kappa}^\star_\delta(\theta_n; P_n)\,.$$

By similar arguments to the proof of the upper bound, we deduce that for all n sufficiently large:

$$\begin{aligned}
\underline{\kappa}(\mathcal{N}_\delta; P_n) &\geq \underline{\kappa}^\star_\delta(\theta_n; P_n) - o(t_n) \\
&\geq \underline{\kappa}^\star_\delta(\theta_n; P_0) - (\lambda'_{1n}, \lambda'_{2n})(P_n - P_0) - o(t_n) \\
&\geq \underline{\kappa}(\mathcal{N}_\delta; P_0) - (\lambda'_{1n}, \lambda'_{2n})(P_n - P_0) - o(t_n)
\end{aligned}$$

for any $(\lambda'_{1n}, \lambda'_{2n}) \in \underline{\Lambda}_\delta(\theta_n; P_0)$. If $\underline{\Lambda}_\delta(\cdot; P_0)$ is lower hemicontinuous at θ^*, then we may choose $(\lambda'_{1n}, \lambda'_{2n}) \in \underline{\Lambda}_\delta(\theta_n; P_0)$ so that $(\lambda'_{1n}, \lambda'_{2n}) \to (\lambda'_1, \lambda'_2) \in \arg\sup_{(\lambda'_1, \lambda'_2)\in\underline{\Lambda}_\delta(\theta^*; P_0)}(\lambda'_1, \lambda'_2)\pi$. Otherwise, if $\underline{\Lambda}_\delta(\cdot; P_0)$ is a singleton then convergence follows by Lemma E.8. Therefore

$$\begin{aligned}\liminf_{n\to\infty} \frac{\underline{\kappa}(\mathcal{N}_\delta; P_n) - \underline{\kappa}(\mathcal{N}_\delta; P_0)}{t_n} &\geq -(\lambda_1', \lambda_2')'\pi \\ &= \sup_{(\lambda_1',\lambda_2')\in\underline{\Lambda}_\delta(\theta^*)} -(\lambda_1', \lambda_2')\pi \\ &\geq \inf_{\theta\in\underline{\Theta}_\delta} \sup_{(\lambda_1',\lambda_2')\in\underline{\Lambda}_\delta(\theta)} -(\lambda_1', \lambda_2')'\pi\end{aligned}$$

which completes the proof for the explicit-dependence case.

The result for the implicit-dependence case follows similarly. An equivalent primal problem in this case is described in equation (41) and its dual is in equation (42), which is why the set of multipliers $\underline{\Xi}_\delta$ is defined differently in this case. If Condition S holds at (θ, P) and $\delta^\star(\theta; P) < \delta$ then strong duality holds, i.e.: $\underline{\kappa}_\delta^\star(\theta; P) = k(\theta)$ (cf. Lemma D.2). ■

Proof of Theorem 4.2. The result follows directly from Theorem 2.1 of Shapiro (1991), using Lemma 4.4 and the condition $\sqrt{n}(\hat{P} - P) \to_d N(0, \Sigma)$. ■

Recall Condition $\mathrm{S}_\#$ and the programs $\underline{\kappa}_\#$, $\bar{\kappa}_\#$, $\underline{\kappa}_\infty$, and $\bar{\kappa}_\infty$ from Appendix B.

Proof of Theorem 5.1. As condition S holds at (θ, γ_0, P_0) for each $\theta \in \Theta$, condition $\mathrm{S}_\#$ must also hold there. So by Lemma B.3, we obtain

$$\begin{aligned}\underline{\kappa}_\#(\theta; \gamma_0, P_0) &= \sup_{\lambda\in\Lambda:\mu\text{-ess}\inf(k(\cdot,\theta,\gamma_0)+\lambda' g(\cdot,\theta,\gamma_0))>-\infty} \left(\mu\text{-ess}\inf(k(\cdot, \theta, \gamma_0) + \lambda' g(\cdot, \theta, \gamma_0)) - \lambda_{12}' P_0\right), \\ \bar{\kappa}_\#(\theta; \gamma_0, P_0) &= \inf_{\lambda\in\Lambda:\mu\text{-ess}\sup(k(\cdot,\theta,\gamma_0)-\lambda' g(\cdot,\theta,\gamma_0))<+\infty} \left(\mu\text{-ess}\sup(k(\cdot, \theta, \gamma_0) - \lambda' g(\cdot, \theta, \gamma_0)) + \lambda_{12}' P_0\right)\end{aligned}$$

for each $\theta \in \Theta$. But by Lemma B.2 and the fact that Condition S holds at (θ, γ_0, P_0) for each $\theta \in \Theta$, we also have

$$\begin{aligned}\underline{\kappa}_\infty(\theta; \gamma_0, P_0) &= \sup_{\lambda\in\Lambda:F_*\text{-ess}\inf(k(\cdot,\theta,\gamma_0)+\lambda' g(\cdot,\theta,\gamma_0))>-\infty} \left(F_*\text{-ess}\inf(k(\cdot, \theta, \gamma_0) + \lambda' g(\cdot, \theta, \gamma_0)) - \lambda_{12}' P_0\right), \\ \bar{\kappa}_\infty(\theta; \gamma_0, P_0) &= \inf_{\lambda\in\Lambda:F_*\text{-ess}\sup(k(\cdot,\theta,\gamma_0)-\lambda' g(\cdot,\theta,\gamma_0))<+\infty} \left(F_*\text{-ess}\sup(k(\cdot, \theta, \gamma_0) - \lambda' g(\cdot, \theta, \gamma_0)) + \lambda_{12}' P_0\right)\end{aligned}$$

for each $\theta \in \Theta$. As μ and F_* are mutually absolutely continuous, the μ-essential supremum and F_*-essential supremum are equal, and the same is true for essential infimum. Therefore, $\underline{\kappa}_\#(\theta; \gamma_0, P_0) = \underline{\kappa}_\infty(\theta; \gamma_0, P_0)$ and $\bar{\kappa}_\#(\theta; \gamma_0, P_0) = \bar{\kappa}_\infty(\theta; \gamma_0, P_0)$ for each $\theta \in \Theta$, so $\inf \mathcal{K}_\# = \inf \mathcal{K}_\infty$ and $\sup \mathcal{K}_\# = \sup \mathcal{K}_\infty$. The result now follows by Lemma B.1. ■

Proof of Theorem 5.2. By Lemma B.1 and the fact that $\mathcal{K}_\# \supseteq \mathcal{K}_\infty$, it's enough to show that $\inf \mathcal{K}_\# \geq \inf \mathcal{K}_\infty$ and $\sup \mathcal{K}_\# \leq \sup \mathcal{K}_\infty$.

First, suppose that $\inf \mathcal{K}_\# > -\infty$. Then for any $\varepsilon > 0$ there exists $\theta_\varepsilon \in \Theta$ and $F_\varepsilon \in \mathcal{F}_\theta$ such that conditions (1)–(4) hold at $(\theta_\varepsilon, F_\varepsilon)$ and for which $k(\theta_\varepsilon, \gamma_0) < \inf \mathcal{K}_\# + \varepsilon$. As condition S holds at $(\theta_\varepsilon, \gamma_0, P_0)$, condition $\mathrm{S}_\#$ must also hold there. As θ_ε is feasible, by Lemma E.10 we must have

$$0 = \sup_{\lambda \in \Lambda : \mu\text{-ess}\inf(\lambda' g(\cdot,\theta,\gamma_0)) > -\infty} \left(\mu\text{-ess}\inf(\lambda' g(\cdot,\theta,\gamma_0)) - \lambda_{12}' P_0 \right) .$$

As μ and F_* are mutually absolutely continuous, the μ-essential supremum and F_*-essential supremum are equal, and the same is true for essential infimum. Therefore

$$0 = \sup_{\lambda \in \Lambda : F_*\text{-ess}\inf(\lambda' g(\cdot,\theta,\gamma_0)) > -\infty} \left(F_*\text{-ess}\inf(\lambda' g(\cdot,\theta,\gamma_0)) - \lambda_{12}' P_0 \right) .$$

It follows by Lemma E.9 that there exists $F \in \mathcal{N}_\infty$ such that (1)–(4) hold at (θ_ε, F). Therefore, $\inf \mathcal{K}_\infty \leq k(\theta_\varepsilon, \gamma_0) < \inf \mathcal{K}_\# + \varepsilon$.

Now suppose that $\inf \mathcal{K}_\# = -\infty$. Then for any $n \in \mathbb{N}$ there exists $\theta_n \in \Theta$ and $F_n \in \mathcal{F}_\theta$ such that conditions (1)–(4) hold at (θ_n, F_n) and for which $k(\theta_n, \gamma_0) \leq -n$. A similar argument shows that there exists $F \in \mathcal{N}_\infty$ such that (1)–(4) hold at (θ_n, F). Therefore, $\inf \mathcal{K}_\infty \leq k(\theta_n, \gamma_0) \leq -n$. ■

Proof of Theorem 6.1. To simplify notation, we drop dependence of h on (γ, P_2) and k on γ throughout the proof.

Take any function $b \in L^2(F_*)$ with $\mathbb{E}^{F_*}[b(U)] = 0$. By analogy with standard efficiency bound calculations for GMM, define the projection $\Pi : L^2(F_*) \to L^2(F_*)$ by

$$\Pi b = b - \mathbb{E}^{F_*}[b(U) h_0(U)'](V^{-1} - V^{-1} H (H' V^{-1} H)^{-1} H' V^{-1}) h_0 .$$

If the model is just-identified, then $(V^{-1} - V^{-1} H (H' V^{-1} H)^{-1} H' V^{-1}) = 0$ and the projection reduces to the identity map.

Using a standard construction (cf. Example 3.2.1 in Bickel, Klaassen, Ritov, and Wellner (1993)), for each $t \in (-1, 1)$ we define a probability measure F_t via

$$\frac{\mathrm{d}F_t}{\mathrm{d}F_*} = \frac{\upsilon(t \Pi b)}{\mathbb{E}^{F_*}[\upsilon(t \Pi b(U))]} , \quad \text{where } \upsilon(x) = \frac{2}{1 + e^{-2x}} .$$

Thus $\{F_t : t \in (-1,1)\}$ is a smooth parametric family passing through F_* at $t = 0$. Fix any $d_\theta \times (d_2 + d_4)$ matrix A of full rank. Premultiplying h by A yields a just-identified system with moment functions $Ah(u, \theta)$. By the implicit function theorem and invertibility of AH, there exists $\varepsilon > 0$ such that the moment condition $\mathbb{E}^{F_t}[Ah(U, \theta)] = 0$ has a unique solution $\theta(F_t) \in \Theta$ for all

$t \in (-\varepsilon, \varepsilon)$, and

$$\left.\frac{d\theta(F_t)}{dt}\right|_{t=0} = -(AH)^{-1} A \mathbb{E}^{F_*}[h_0(U)\Pi b(U)]\,.$$

Writing $\kappa(F_t) = \mathbb{E}^{F_t}[k(U, \theta(F_t))]$, we therefore have

$$\begin{aligned}\left.\frac{d\kappa(F_t)}{dt}\right|_{t=0} &= \mathbb{E}^{F_*}[k_0(U)\Pi b(U)] - J(AH)^{-1} A \mathbb{E}^{F_*}[h_0(U)\Pi b(U)] \\ &= \mathbb{E}^{F_*}[\tilde{\iota}(U)\Pi b(U)]\,,\end{aligned}$$

where $\tilde{\iota}(u) = k_0(u) - \kappa(F_*) - J(AH)^{-1} A h_0(u)$. As Π is an orthogonal projection:

$$\left.\frac{d\kappa(F_t)}{dt}\right|_{t=0} = \mathbb{E}^{F_*}[\Pi\tilde{\iota}(U)\Pi b(U)]\,.$$

However, note that irrespective of the choice of A, we have

$$\begin{aligned}\Pi\tilde{\iota} &= \Pi k_0 - J(AH)^{-1} A(h_0 - \mathbb{E}^{F_*}[h_0(U)h_0(U)'](V^{-1} - V^{-1}H(H'V^{-1}H)^{-1}H'V^{-1})h_0) \\ &= \Pi k_0 - J(AH)^{-1} A(h_0 - V(V^{-1} - V^{-1}H(H'V^{-1}H)^{-1}H'V^{-1})h_0) \\ &= \Pi k_0 - J(AH)^{-1} AH(H'V^{-1}H)^{-1}H'V^{-1}h_0 \\ &= \Pi k_0 - J(H'V^{-1}H)^{-1}H'V^{-1}h_0 \\ &= \iota\end{aligned}$$

hence

$$\left.\frac{d\kappa(F_t)}{dt}\right|_{t=0} = \mathbb{E}^{F_*}[\iota(U)\Pi b(U)]$$

for all $b \in L^2(F_*)$.

As $\phi(x) = \frac{x^2-1-2(x-1)}{2}$, a Taylor series expansion of $\upsilon(x)$ around $x = 0$ yields

$$D_\phi(F_t \| F_*) = \frac{t^2}{2}\mathbb{E}^{F_*}[(\Pi b(U))^2] + o(t^2)\,.$$

Therefore, whenever $\Pi b \neq 0$ we have

$$\frac{(\kappa(F_t) - \kappa(F_{-t}))^2}{4D_\phi(F_t\|F_*)} = \frac{\mathbb{E}^{F_*}[\iota(U)\Pi b(U)]^2 + o(1)}{\frac{1}{2}\mathbb{E}^{F_*}[(\Pi b(U))^2] + o(1)}$$

hence

$$s \geq \frac{\mathbb{E}^{F_*}[\iota(U)\Pi b(U)]^2}{\frac{1}{2}\mathbb{E}^{F_*}[(\Pi b(U))^2]}\,.$$

If $\iota(u) = 0$ (F_*-almost everywhere) then the right-hand side is zero for any b and we trivially have $s \geq 2\mathbb{E}^{F_*}[\iota(U)^2]$. Otherwise, choosing $b = \iota$ yields $s \geq 2\mathbb{E}^{F_*}[\iota(U)^2]$.

We prove the reverse inequality $s \leq 2\mathbb{E}^{F_*}[\iota(U)^2]$ by contradiction. Suppose that there exists a sequence $\delta_n \downarrow 0$ and $\varepsilon > 0$ such that

$$\frac{(\bar{\kappa}(\mathcal{N}_{\delta_n}) - \underline{\kappa}(\mathcal{N}_{\delta_n}))^2}{4\delta_n} \geq 2\mathbb{E}^{F_*}[\iota(U)^2] + 2\varepsilon$$

for each n. We may choose $\underline{m}_n, \overline{m}_n \in \mathcal{M}_{\delta_n}$ and $\underline{\theta}_n, \overline{\theta}_n \in \Theta$ such that

$$\mathbb{E}^{F_*}[\underline{m}_n(U)] = 1\,, \quad \mathbb{E}^{F_*}[\overline{m}_n(U)] = 1\,, \quad \mathbb{E}^{F_*}[\underline{m}_n(U)h(U,\underline{\theta}_n)] = 0\,, \quad \mathbb{E}^{F_*}[\overline{m}_n(U)h(U,\overline{\theta}_n)] = 0$$

and

$$\frac{(\mathbb{E}^{F_*}[\overline{m}_n(U)k(U,\overline{\theta}_n) - \underline{m}_n(U)k(U,\underline{\theta}_n)])^2}{4\delta_n} \geq 2\mathbb{E}^{F_*}[\iota(U)^2] + \varepsilon\,. \tag{30}$$

By Assumption M(iv) (taking a subsequence if necessary), we can assume that $\underline{\theta}_n \to \underline{\theta}^*$ and $\overline{\theta}_n \to \overline{\theta}^*$ for some $\underline{\theta}^*, \overline{\theta}^* \in \Theta$.

As neighborhoods are defined via χ^2 divergence, identify $\mathcal{L}$ and $\mathcal{E}$ with $L^2(F_*)$. Let $\|\cdot\|_2$ denote the $L^2(F_*)$ norm and observe that $\mathbb{E}^{F_*}[\phi(\underline{m}_n)] = \frac{1}{2}\|\underline{m}_n - 1\|_2^2$ and similarly for $\overline{m}_n$. We have

$$\|\underline{m}_n - 1\|_2^2, \|\overline{m}_n - 1\|_2^2 \leq 2\delta_n \downarrow 0 \quad \text{as } n \to \infty. \tag{31}$$

By similar arguments to the proof of Lemma 4.3, we may deduce $\mathbb{E}^{F_*}[h(U,\underline{\theta}^*)] = \mathbb{E}^{F_*}[h(U,\overline{\theta}^*)] = 0$. It then follows by identifiability of $\theta(F_*)$ that $\underline{\theta}^* = \overline{\theta}^* = \theta(F_*)$.

Note that $\underline{m}_n$ and $\underline{\theta}_n$ must satisfy $\mathbb{E}^{F_*}[\underline{m}_n(U)h(U,\underline{\theta}_n)] = 0$. By differentiability of $\theta \mapsto \mathbb{E}^{F_*}[h(u,\theta)]$ at $\theta(F_*)$, we therefore obtain

$$-H(\underline{\theta}_n - \theta(F_*)) + o(\|\underline{\theta}_n - \theta(F_*)\|) = \mathbb{E}^{F_*}[(\underline{m}_n(U) - 1)h(U,\underline{\theta}_n)] \quad \text{as } \underline{\theta}_n \to \theta(F_*).$$

By Cauchy-Schwartz, Assumption M(i), and the fact that H has full rank, we therefore have

$$\|\underline{\theta}_n - \theta(F_*)\| = O(\|\underline{m}_n - 1\|_2)$$

and hence, by (31), Cauchy-Schwartz, and $L^2(F_*)$ continuity of $\theta \mapsto h(\cdot,\theta,\gamma_0,P_{20})$ at $\theta(F_*)$, we obtain

$$-H(\underline{\theta}_n - \theta(F_*)) = \mathbb{E}^{F_*}[(\underline{m}_n(U) - 1)h_0(U)] + o(\delta_n^{1/2}) \tag{32}$$

and so

$$\underline{\theta}_n - \theta(F_*) = -(H'V^{-1}H)^{-1}H'V^{-1}\mathbb{E}^{F_*}[(\underline{m}_n(U) - 1)h_0(U)] + o(\delta_n^{1/2})\,.$$

Turning to the counterfactual, by similar arguments we may deduce

$$\begin{aligned}\mathbb{E}^{F_*}[\underline{m}_n(U)k(U,\underline{\theta}_n)] - \kappa(F_*) &= J(\underline{\theta}_n - \theta(F_*)) + \mathbb{E}^{F_*}[(\underline{m}_n(U)-1)(k_0(U) - \kappa(F_*))] + o(\delta_n^{1/2}) \\ &= -J(H'V^{-1}H)^{-1}H'V^{-1}\mathbb{E}^{F_*}[(\underline{m}_n(U)-1)h_0(U)] \\ &\quad + \mathbb{E}^{F_*}[(\underline{m}_n(U)-1)(k_0(U)-\kappa(F_*))] + o(\delta_n^{1/2})\,.\end{aligned}$$

However, by (32) and definition of Π we also have

$$\mathbb{E}^{F_*}[(\underline{m}_n(U)-1)(k_0(U)-\kappa(F_*) - \Pi(k_0(U)-\kappa(F_*)))] = o(\delta_n^{1/2})$$

hence

$$\mathbb{E}^{F_*}[\underline{m}_n(U)k(U,\underline{\theta}_n)] - \kappa(F_*) = \mathbb{E}^{F_*}[(\underline{m}_n(U)-1)\iota(U)] + o(\delta_n^{1/2})\,.$$

Analogous arguments apply to $\overline{m}_n$ and $\overline{\theta}_n$. We have therefore shown

$$\frac{(\mathbb{E}^{F_*}[\overline{m}_n(U)k(U,\overline{\theta}_n) - \underline{m}_n(U)k(U,\underline{\theta}_n)])^2}{4\delta_n} = \frac{(\mathbb{E}^{F_*}[(\overline{m}_n(U) - \underline{m}_n(U))\iota(U)])^2}{4\delta_n} + o(1)\,. \tag{33}$$

It remains to control the denominator. To do so, first note that we must have $\overline{m}_n \neq \underline{m}_n$ for all n sufficiently large. Otherwise, substituting (33) into (30) yields $o(1) \geq 2\mathbb{E}^{F_*}[\iota(U)^2] + \varepsilon$. As $n \to \infty$, the ε term dominates the $o(1)$ term and we obtain a contradiction.

To complete the proof, observe that

$$\|\overline{m}_n - \underline{m}_n\|_2^2 \leq 2\|\overline{m}_n - 1\|_2^2 + 2\|\underline{m}_n - 1\|_2^2 \leq 8\delta_n \tag{34}$$

by (31). Substituting (33) and (34) into (30) yields

$$\frac{2(\mathbb{E}^{F_*}[(\overline{m}_n(U) - \underline{m}_n(U))\iota(U)])^2}{\|\overline{m}_n - \underline{m}_n\|_2^2} + o(1) \geq 2\mathbb{E}^{F_*}[\iota(U)^2] + \varepsilon\,.$$

Finally, by Cauchy-Schwarz:

$$2\mathbb{E}^{F_*}[\iota(U)^2] + o(1) \geq 2\mathbb{E}^{F_*}[\iota(U)^2] + \varepsilon\,.$$

As $n \to \infty$, the ε term dominates the $o(1)$ term and we obtain a contradiction. ■

Proof of Lemma 6.1. Immediate by consistency of $(\hat{\theta}, \hat{\gamma}, \hat{P})$ and Slutsky's theorem. ■

Proof of Theorem 6.2. The proof follows similar arguments to the proof of Theorem 6.1, here we just note the necessary modifications. First note (again dropping dependence on γ to simplify notation) that here $\kappa(F_t) = k(\theta(F_t))$. In the proof of the inequality $s \geq 2\mathbb{E}^{F_*}[\iota(U)^2]$, we modify the derivative of κ to obtain

$$\left.\frac{d\kappa(F_t)}{dt}\right|_{t=0} = -J(AH)^{-1}A\mathbb{E}^{F_*}[h_0(U)\Pi b(U)]$$
$$= \mathbb{E}^{F_*}[\Pi\tilde{\iota}(U)\Pi b(U)]$$

where $\tilde{\iota}(u) = -J(AH)^{-1}Ah_0(u)$. We may again verify that irrespective of the choice of A, we have $\Pi\tilde{\iota} = -J(H'V^{-1}H)^{-1}H'V^{-1}h_0$ hence

$$\left.\frac{d\kappa(F_t)}{dt}\right|_{t=0} = \mathbb{E}^{F_*}[\iota(U)\Pi b(U)]$$

for all $b \in L^2(F_*)$.

For the proof of the reverse inequality, inequality (30) is replaced with the inequality

$$\frac{(k(\overline{\theta}_n) - k(\underline{\theta}_n))^2}{4\delta_n} \geq 2\mathbb{E}^{F_*}[\iota(U)^2] + \varepsilon\,.$$

where we may deduce similarly that

$$k(\underline{\theta}_n) - \kappa(F_*) = -J(H'V^{-1}H)^{-1}H'V^{-1}\mathbb{E}^{F_*}[(\underline{m}_n(U) - 1)h_0(U)] + o(\delta_n^{1/2})$$
$$= \mathbb{E}^{F_*}[(\underline{m}_n(U) - 1)\iota(U)] + o(\delta_n^{1/2})\,.$$

Analogous arguments apply to $\overline{m}_n$ and $\overline{\theta}_n$. In place of (33) we now have

$$\frac{(k(\overline{\theta}_n) - k(\underline{\theta}_n))^2}{4\delta_n} = \frac{(\mathbb{E}^{F_*}[(\overline{m}_n(U) - \underline{m}_n(U))\iota(U)])^2}{4\delta_n} + o(1)\,.$$

The remainder of the proof now follows similarly. ∎

Proof of Lemma 6.2. Immediate by consistency of $(\hat{\theta}, \hat{\gamma}, \hat{P})$ and Slutsky's theorem. ∎

Proof of Lemma B.1. Clearly $\underline{\kappa}(\mathcal{N}_\delta) \geq \inf \mathcal{K}_\infty$ for each $\delta > 0$. Suppose $\inf \mathcal{K}_\infty$ is finite. Fix any $\varepsilon > 0$. Then there is $F_\varepsilon \in \mathcal{N}_\infty$ and $\theta_\varepsilon \in \Theta$ such that (1)–(4) all hold at $(\theta_\varepsilon, \gamma_0, P_0)$ under F_ε and $\mathbb{E}^{F_\varepsilon}[k(U, \theta_\varepsilon, \gamma_0)] < \inf \mathcal{K}_\infty + \varepsilon$. But then for any $\delta \geq D_\phi(F_\varepsilon \| F_0)$ we necessarily have $\underline{\kappa}(\mathcal{N}_\delta) < \inf \mathcal{K}_\infty + \varepsilon$. If $\inf \mathcal{K}_\infty = -\infty$, then for each $n \in \mathbb{N}$ there exists $F_n \in \mathcal{N}_\infty$ and $\theta_n \in \Theta$ such that (1)–(4) all hold at $(\theta_n, \gamma_0, P_0)$ under F_n and $\mathbb{E}^{F_n}[k(U, \theta_n, \gamma_0)] < -n$. But then for any $\delta \geq D_\phi(F_n \| F_0)$ we necessarily have $\underline{\kappa}(\mathcal{N}_\delta) < -n$. ∎

Proof of Lemma B.2. Let $\mathcal{L}_+$ denote the cone of (F_*-almost surely) non-negative functions in

$\mathcal{L}$. We may write $\underline{\kappa}_\infty(\theta;\gamma,P)$ as a conic program:

$$\underline{\kappa}_\infty(\theta;\gamma,P) = \inf_{m\in\mathcal{L}_+} \mathbb{E}^{F_*}[m(U)k(U,\theta,\gamma)] \quad \text{subject to the constraints in (22).}$$

By standard duality results for conic programs (Bonnans and Shapiro, 2000, Section 2.5.6), the dual of $\underline{\kappa}_\infty(\theta;\gamma,P)$ is

$$\underline{\kappa}^\star_\infty(\theta;\gamma,P) = \sup_{\zeta\in\mathbb{R},\lambda\in\Lambda} -\zeta - \lambda_{12}'P \quad \text{subject to } \zeta + F_*\text{-ess}\inf(k(\cdot,\theta,\gamma)+\lambda' g(\cdot,\theta,\gamma)) \geq 0\,.$$

Whenever $F_*\text{-ess}\inf(k(u,\theta,\gamma)+\lambda' g(u,\theta,\gamma)) > -\infty$ the solution for ζ must set $\zeta = -F_*\text{-ess}\inf(k(\cdot,\theta,\gamma)+\lambda' g(\cdot,\theta,\gamma))$ in which case

$$\underline{\kappa}^\star_\infty(\theta;\gamma,P) = \sup_{\lambda\in\Lambda} \left(F_*\text{-ess}\inf(k(\cdot,\theta,\gamma)+\lambda' g(\cdot,\theta,\gamma)) - \lambda_{12}'P\right)\,.$$

Conversely, if $F_*\text{-ess}\inf(k(\cdot,\theta,\gamma)+\lambda' g(\cdot,\theta,\gamma)) = -\infty$ then λ is clearly infeasible.

To establish strong duality, the constraint qualification we require is

$$(1,P',0')' \in \mathrm{int}(\{(\mathbb{E}^{F_*}[m(U)(1,g(U,\theta,\gamma)')])' : m\in\mathcal{L}_+\} + \{0\}\times\mathcal{C})\,,$$

which holds whenever Condition S holds at (θ,γ_0,P_0) (cf. Lemma D.2). It follows by Theorem 2.187 of Bonnans and Shapiro (2000) that strong duality holds. ■

Lemma B.3 is proved by applying some results of Csiszár and Matúš (2012) that extend classical duality results relying on paired function classes to much broader classes of functions. Their results apply to optimization problems constrained by equality restrictions. Some (straightforward) modifications are required to show similar characterizations apply to classes with inequality restrictions.

Proof of Lemma B.3. Fix any $\theta\in\Theta$. With θ fixed, we drop dependence of g and k on (θ,γ_0,P_0) for the remainder of the proof. Let $L^1_+(\mu)$ denote the cone of μ-almost everywhere non-negative functions in $L^1(\mu)$. Let $\dot{\mathcal{M}} = \{m\in L^1(\mu) : \int mg\,\mathrm{d}\mu \text{ is finite}\}$ and let $\dot{\mathcal{M}}_+ = \dot{\mathcal{M}}\cap L^1_+(\mu)$. Thus, $F\in\mathcal{F}_\theta$ if and only if its derivative with respect to μ, say m, belongs to $\dot{\mathcal{M}}_+$. Similarly, any $m\in\mathcal{M}_+$ with $\int m\,\mathrm{d}\mu = 1$ corresponds to a distribution in $\mathcal{F}_\theta$. For any $c\in\mathbb{R}^{d+1}$, define the set

$$\mathcal{M}[c] = \left\{m\in\dot{\mathcal{M}}_+ : \int (1,g(u)')'m(u)\,\mathrm{d}\mu(u) = (1,P_0',0_{d_3+d_4}')' + c\right\}\,.$$

Also let the functions $q_1,q_2:\mathbb{R}^{d+1}\to(-\infty,+\infty]$ be given by

$$q_1(c) = \inf\left\{\int mk\,\mathrm{d}\mu : m\in\mathcal{M}[c]\right\}$$

and

$$q_2(c) = \left[\begin{array}{ll} 0 & \text{if } c \in \{0\} \times \mathbb{R}^{d_1}_{-} \times \{0\}^{d_2} \times \mathbb{R}^{d_3}_{-} \times \{0\}^{d_4}\,, \\ +\infty & \text{otherwise.} \end{array}\right.$$

The functions q_1 and q_2 are lower-semicontinuous, proper convex functions (Rockafellar, 1970, p. 24). Note μ-essential boundedness of k guarantees that $q_1(c) > -\infty$ for all c. The quantity $\underline{\kappa}_{\#}(\theta;\gamma_0,P_0)$ may now be written in terms of q_1 and q_2:

$$\underline{\kappa}_{\#}(\theta;\gamma_0,P_0) = \inf_{c \in \mathbb{R}^{d+1}} (q_1(c) + q_2(c))$$

If condition $\mathrm{S}_{\#}$ holds at (θ,γ_0,P_0) then $\mathrm{ri}(\{c : q_1(c) < +\infty\})$ and $\mathrm{ri}(\{c : q_2(c) < +\infty\})$ have nonempty intersection. Fenchel's duality theorem (Rockafellar, 1970, Theorem 31.3) then implies

$$\underline{\kappa}_{\#}(\theta;\gamma_0,P_0) = \sup_{\nu \in \mathbb{R}^{d+1}} (-q_1^\star(\nu) - q_2^\star(-\nu)) \tag{35}$$

where $q_1^\star$ and $q_2^\star$ are the convex conjugates of q_1 and q_2. By direct calculation, we see that

$$-\,q_2^\star(-\nu) = \left[\begin{array}{ll} 0 & \text{if } \nu \in \mathbb{R} \times -\Lambda\,, \\ -\infty & \text{otherwise.} \end{array}\right. \tag{36}$$

For $q_1^\star$, we begin by writing

$$\begin{aligned} -q_1^\star(\nu) &= \inf_{c \in \mathbb{R}^{d+1}} \big(-c'\nu + q_1(c)\big) \\ &= \inf_{c \in \mathbb{R}^{d+1}} \inf_{m \in \mathcal{M}[c]} \left(-c'\nu + \int mk \,\mathrm{d}\mu\right) \\ &= \inf_{c \in \mathbb{R}^{d+1}} \inf_{m \in \mathcal{M}[c]} \left((1, P_0', 0_{d_3+d_4}')\nu + \int \big(k(u) - (1, g(u)')\nu\big)\, m(u)\,\mathrm{d}\mu(u)\right) \\ &= \inf_{m \in \dot{\mathcal{M}}_+} \left((1, P_0', 0_{d_3+d_4}')\nu + \int \big(k(u) - (1, g(u)')\nu\big)\, m(u)\,\mathrm{d}\mu(u)\right)\,. \end{aligned}$$

Let

$$Q(u, m(u)) = \left[\begin{array}{ll} k(u)m(u) & \text{if } m(u) \geq 0\,, \\ +\infty & \text{otherwise.} \end{array}\right.$$

We therefore have

$$-q_1^\star(\nu) = \inf_{m \in \dot{\mathcal{M}}} \left((1, P_0', 0_{d_3+d_4}')\nu + \int \big(Q(u, m(u)) - (1, g(u)')\nu\big)\, m(u)\,\mathrm{d}\mu(u)\right)\,.$$

By Remark A.3 and Theorem A.4 of Csiszár and Matúš (2012), we may bring the infimum inside the expectation:

$$
\begin{aligned}
-q_1^\star(\nu) &= (1, P_0', 0_{d_3+d_4}')\nu + \int \inf_{x \in \mathbb{R}} \big(Q(u,x) - (1, g(u)')\nu\big)\, x \,\mathrm{d}\mu(u) \\
&= (1, P_0', 0_{d_3+d_4}')\nu + \int \inf_{x \geq 0} \big(k(u) - (1, g(u)')\nu\big)\, x \,\mathrm{d}\mu(u) \\
&= \left[\begin{array}{ll} -\infty & \text{if } \mu\text{-ess}\inf(k(\cdot) - (1, g(\cdot)')\nu) < 0\,, \\ (1, P_0', 0_{d_3+d_4}')\nu & \text{otherwise.} \end{array} \right.
\end{aligned}
\tag{37}
$$

Writing $\nu = (\zeta, \lambda')'$, it now follows from (35), (36), and (37) that

$$
\begin{aligned}
\underline{\kappa}_\#(\theta; \gamma_0, P_0) &= \sup_{\zeta \in \mathbb{R}, \lambda \in \Lambda : \mu\text{-ess}\inf(k(\cdot) + g(\cdot)'\lambda - \zeta) \geq 0} \zeta - \lambda_{12}' P_0 \\
&= \sup_{\lambda \in \Lambda : \mu\text{-ess}\inf(k(\cdot) + g(\cdot)'\lambda) > -\infty} \mu\text{-ess}\inf(k(\cdot) + g(\cdot)'\lambda) - \lambda_{12}' P_0
\end{aligned}
$$

as required. ■

E Supplementary results

E.1 Notation

For $x \in \mathbb{R}^n$ and $A \subset \mathbb{R}^n$ let $d(x, A) = \inf_{a \in A} \|x - a\|$ where $\|\cdot\|$ denotes Euclidean norm. Let $\vec{d}_H(A, B) = \sup_{a \in A} \inf_{b \in B} \|a - b\|$ denote the directed Hausdorff distance between sets $A, B \subset \mathbb{R}^n$. Let B_ε denote a Euclidean ball centered at the origin with radius ε, where the dimension of the ball should be obvious from the context. Let $T \subseteq \mathbb{R}^n$ be a nonempty, closed convex cone with nonempty interior and let $B \subset T$. Let ∂B denote the boundary of B. We define the exterior of B relative to T as $\text{ext}(B; T) = \text{cl}(\partial B \cap \text{int}(T))$. For example, if $n = 2$, $T = \mathbb{R} \times \mathbb{R}_+$, and $B = \{(x, y) : x^2 + y^2 \leq 1, x \geq 0\}$, then $\text{ext}(B; T) = \{(x, y) : x^2 + y^2 = 1, x \geq 0\}$.

E.2 Stability of constraint qualifications under perturbations

Lemma E.1 *Let Assumption Φ hold and let Condition S hold at (θ, γ, P). Then: there exists a neighborhood N of P such that Condition S holds at $(\theta, \gamma, \tilde{P})$ for each $\tilde{P} \in N$.*

Proof of Lemma E.1. First note $B_{2\varepsilon} \subseteq (\{\mathbb{E}^F[g(U, \theta, \gamma)] - (P', 0'_{d_3+d_4})' : F \in \mathcal{N}_\infty\} + \mathcal{C})$ must hold for some $\varepsilon > 0$. But for any $\tilde{P}$ with $\|P - \tilde{P}\| < \varepsilon$ we clearly have

$$\|(\mathbb{E}^F[g(U, \theta, \gamma)] - (P', 0'_{d_3+d_4})') - (\mathbb{E}^F[g(U, \theta, \gamma)] - (\tilde{P}', 0'_{d_3+d_4})')\| < \varepsilon$$

for all $F \in \mathcal{N}_\infty$, and so $B_\varepsilon \subseteq (\{\mathbb{E}^F[g(U, \theta, \gamma)] - (\tilde{P}', 0'_{d_3+d_4})' : F \in \mathcal{N}_\infty\} + \mathcal{C})$. ■

Lemma E.2 *Let Assumption Φ hold, let each entry of g be $\mathcal{E}$-continuous in (θ, γ), and let Condition S hold at (θ, γ, P). Then: there exists a neighborhood N of (θ, γ, P) such that Condition S holds at $(\tilde{\theta}, \tilde{\gamma}, \tilde{P})$ for each $(\tilde{\theta}, \tilde{\gamma}, \tilde{P}) \in N$.*

Proof of Lemma E.2. As Condition S holds at (θ, γ, P), there exists sufficiently large δ such that $0 \in \text{int}(\{\mathbb{E}^F[g(U, \theta, \gamma)] - (P', 0'_{d_3+d_4})' : F \in \mathcal{N}_\delta\} + \mathcal{C})$. Therefore, we may choose $\varepsilon > 0$ such that $B_{4\varepsilon} \subseteq \text{int}(\{\mathbb{E}^F[g(U, \theta, \gamma)] - (P', 0'_{d_3+d_4})' : F \in \mathcal{N}_\delta\} + \mathcal{C})$.

Identify any $F \in \mathcal{N}_\delta$ with its Radon–Nikodym derivative with respect to F_*, say $m \in \mathcal{M}_\delta$. By the proof of Lemma D.1, $\|m\|_\phi \leq 2 + \phi(2) + \delta$ for each $m \in \mathcal{M}_\delta$. By $\mathcal{E}$-continuity, there exists a neighborhood N_1 of (θ, γ) such that for any $(\tilde{\theta}, \tilde{\gamma}) \in N_1$ and with r denoting any entry of g_1, g_2,

g_3, or g_4, we have

$$\|r(\cdot,\theta,\gamma)-r(\cdot,\tilde{\theta},\tilde{\gamma})\|_\psi < \frac{\varepsilon}{\sqrt{d}(2+\phi(2)+\delta)}\,.$$

It follows by Hölder's inequality for Orlicz classes that

$$\sup_{m\in\mathcal{M}_\delta} |\mathbb{E}^{F_*}[m(U)r(U,\theta,\gamma)]-\mathbb{E}^{F_*}[m(U)r(U,\tilde{\theta},\tilde{\gamma})]| \leq \frac{\varepsilon}{\sqrt{d}}$$

for any $(\tilde{\theta},\tilde{\gamma})\in N_1$, hence $B_\varepsilon \subseteq (\{\mathbb{E}^F[g(U,\tilde{\theta},\tilde{\gamma})]-(P',0'_{d_3+d_4})' : F\in\mathcal{N}_\delta\}+\mathcal{C})$ for any $(\tilde{\theta},\tilde{\gamma})\in N_1$. Now let N_2 be an ε-neighborhood of P. For any $F\in\mathcal{N}_\delta$ and any $(\tilde{\theta},\tilde{\gamma},\tilde{P})\in N_1\times N_2$, we have

$$\|(\mathbb{E}^F[m(U)g(U,\theta,\gamma)]-(P',0')')-(\mathbb{E}^F[m(U)g(U,\tilde{\theta},\tilde{\gamma})]-(\tilde{P}',0')')\| < 2\varepsilon$$

hence $B_{2\varepsilon}\subseteq \mathrm{int}(\{\mathbb{E}^F[g(U,\tilde{\theta},\tilde{\gamma})]-(\tilde{P}',0'_{d_3+d_4})' : F\in\mathcal{N}_\delta\}+\mathcal{C})$. ■

Lemma E.3 *Let Assumption Φ hold, let $\mathbb{E}^{F_*}[\phi^\star(c_1+c_2'g(U,\theta,\gamma))]$ be continuous in (θ,γ) for every $(c_1,c_2')'\in\mathbb{R}^{d+1}$, and let Condition S hold at (θ,γ,P). Then: $\delta^\star(\theta;\gamma,P)$ is continuous at (θ,γ,P).*

Proof of Lemma E.3. Recall the definition of the program $I(\theta;\gamma,P)$ (see the discussion immediately preceding Lemma D.2). As Conditions S holds at (θ,γ,P), we must have $I(\theta;\gamma,P)<\infty$. The objective function

$$\ell(\zeta,\lambda) = -\mathbb{E}^{F_*}[\phi^\star(\zeta+\lambda'g(U,\theta,\gamma))]-\zeta-(\lambda_1',\lambda_2')'P$$

is the pointwise infimum of affine functions and is therefore concave and upper semicontinuous. By Lemma D.2 and Theorem 2.165 of Bonnans and Shapiro (2000), the set of multipliers (ζ,λ) solving the dual problem, say Ξ, is a nonempty, convex, and compact subset of $\mathbb{R}^{1+d}$. Fix $\varepsilon>0$ and let $\Xi^\varepsilon=\{(\zeta,\lambda)\in\mathbb{R}\times\Lambda : d((\zeta,\lambda),\Xi)\leq\varepsilon\}$. For any $(\tilde{\theta},\tilde{\gamma},\tilde{P})\in\Theta\times\Gamma\times\mathcal{P}$, let

$$\tilde{\ell}(\zeta,\lambda) = -\mathbb{E}^{F_*}[\phi^\star(\zeta+\lambda'g(U,\tilde{\theta},\tilde{\gamma}))]-\zeta-(\lambda_1',\lambda_2')'\tilde{P}\,.$$

By continuity of $(\theta,\gamma)\mapsto\mathbb{E}^{F_*}[\phi^\star(c_1+c_2'g(U,\theta,\gamma))]$, $\tilde{\ell}$ converges pointwise to ℓ as $(\tilde{\theta},\tilde{\gamma},\tilde{P})\to(\theta,\gamma,P)$. By concavity, convergence holds uniformly over Ξ^ε (Rockafellar, 1970, Theorem 10.8), so

$$\sup_{(\zeta,\lambda)\in\Xi^\varepsilon}\tilde{\ell}(\zeta,\lambda)\to\delta^\star(\theta;\gamma,P) \quad \text{as } (\tilde{\theta},\tilde{\gamma},\tilde{P})\to(\theta,\gamma,P). \tag{38}$$

By upper semicontinuity of ℓ and definition of Ξ:

$$\delta^\star(\theta;\gamma,P)-\sup_{(\zeta,\lambda)\in\mathrm{ext}(\partial\Xi^\varepsilon;\mathbb{R}\times\Lambda)}\ell(\zeta,\lambda) =: 2a > 0\,. \tag{39}$$

It follows by (38) and (39) that there exists a neighborhood N of (θ,γ,P) such that for any

$(\tilde{\theta}, \tilde{\gamma}, \tilde{P}) \in N$, the inequality

$$\sup_{(\zeta,\lambda)\in\Xi^\varepsilon} \tilde{\ell}(\zeta,\lambda) - \sup_{(\zeta,\lambda)\in\mathrm{ext}(\partial\Xi^\varepsilon;\mathbb{R}\times\Lambda)} \tilde{\ell}(\zeta,\lambda) > a$$

holds, hence $\arg\sup_{(\zeta,\lambda)\in\Xi^\varepsilon} \tilde{\ell}(\zeta,\lambda) \cap \mathrm{ext}(\partial\Xi^\varepsilon;\mathbb{R}\times\Lambda) = \emptyset$. Take any $(\zeta,\lambda) \in \mathbb{R}\times\Lambda$ outside Ξ^ε. Then there exists $(\zeta^\varepsilon, \lambda^\varepsilon) \in \mathrm{ext}(\partial\Xi^\varepsilon;\mathbb{R}\times\Lambda)$ that is the convex combination of (ζ,λ) and some $(\tilde{\zeta},\tilde{\lambda}) \in \arg\sup_{(\zeta,\lambda)\in\Xi^\varepsilon} \tilde{\ell}(\zeta,\lambda)$. But then by concavity of $\tilde{\ell}$:

$$\tilde{\ell}(\tilde{\zeta},\tilde{\lambda}) > \tilde{\ell}(\zeta^\varepsilon,\lambda^\varepsilon) \geq \tau\tilde{\ell}(\tilde{\zeta},\tilde{\lambda}) + (1-\tau)\tilde{\ell}(\zeta,\lambda)$$

for some $\tau \in (0,1)$. It follows that $\tilde{\ell}(\zeta,\lambda) < \sup_{(\zeta,\lambda)\in\Xi^\varepsilon} \tilde{\ell}(\zeta,\lambda)$ must hold whenever $(\tilde{\theta}, \tilde{\gamma}, \tilde{P}) \in N$. Therefore, whenever $(\tilde{\theta}, \tilde{\gamma}, \tilde{P}) \in N$ we have $\delta^\star(\tilde{\theta}; \tilde{\gamma}, \tilde{P}) = \sup_{(\zeta,\lambda)\in\Xi^\varepsilon} \tilde{\ell}(\zeta,\lambda)$. The result now follows by display (38). ■

E.3 Continuity of extreme counterfactuals

The smallest and largest counterfactuals $\underline{\kappa}(\mathcal{N}_\delta)$ and $\bar{\kappa}(\mathcal{N}_\delta)$ depend implicitly on (γ, P). In what follows, we sometimes make this dependence explicit by writing $\underline{\kappa}(\mathcal{N}_\delta; \gamma, P)$ and $\bar{\kappa}(\mathcal{N}_\delta; \gamma, P)$. The next two lemmas establish continuity of the extreme counterfactuals in (γ, P). The results are proved for $\underline{\kappa}$; the proofs for $\bar{\kappa}$ are identical. We first present results for the explicit-dependence case.

Lemma E.4 *Let Assumptions* Φ *and M hold and let* k *depend on* u*. Then:* $\underline{\kappa}(\mathcal{N}_\delta; \cdot, \cdot)$ *and* $\bar{\kappa}(\mathcal{N}_\delta; \cdot, \cdot)$ *are continuous on a neighborhood of* (γ_0, P_0)*.*

Proof of Lemma E.4. We first show $\underline{\kappa}(\mathcal{N}_\delta; \cdot, \cdot)$ is upper semicontinuous at (γ_0, P_0). Fix $\varepsilon > 0$. By Lemma 4.3, choose $\theta_\varepsilon \in \Theta_\delta$ such that $\underline{\kappa}^\star_\delta(\theta_\varepsilon; \gamma_0, P_0) < \underline{\kappa}(\mathcal{N}_\delta; \gamma_0, P_0) + \varepsilon$. As Condition S holds at $(\theta_\varepsilon, \gamma_0, P_0)$ (cf. Assumption M(ii)), Lemma E.2 implies that Condition S also holds at $(\theta_\varepsilon, \gamma, P)$ for all (γ, P) in some neighborhood N of (γ_0, P_0). By continuity of $\delta^\star(\theta_\varepsilon; \cdot, \cdot)$ at (γ_0, P_0) (cf. Lemma E.3), the inequality $\delta^\star(\theta_\varepsilon, \cdot, \cdot) < \delta$ holds on a neighborhood N' of (γ_0, P_0). It follows by Lemma 4.2 that $\underline{\kappa}_\delta(\theta_\varepsilon; \cdot, \cdot) = \underline{\kappa}^\star_\delta(\theta_\varepsilon; \cdot, \cdot)$ holds on $N \cap N'$. Lemma E.7 implies that $\underline{\kappa}^\star_\delta(\theta_\varepsilon; \cdot, \cdot)) < \underline{\kappa}^\star_\delta(\theta_\varepsilon; \gamma_0, P_0) + \varepsilon$ holds on a neighborhood N'' of (γ_0, P_0). On $N \cap N' \cap N''$ we therefore have

$$\underline{\kappa}(\mathcal{N}_\delta; \cdot, \cdot) \leq \underline{\kappa}_\delta(\theta_\varepsilon; \cdot, \cdot) = \underline{\kappa}^\star_\delta(\theta_\varepsilon; \cdot, \cdot) < \underline{\kappa}^\star_\delta(\theta_\varepsilon; \gamma_0, P_0) + \varepsilon < \underline{\kappa}(\mathcal{N}_\delta; \gamma_0, P_0) + 2\varepsilon\,.$$

We establish lower semicontinuity by contradiction. Suppose there is $\varepsilon > 0$ and a sequence $(\gamma_n, P_n) \to (\gamma_0, P_0)$ along which

$$\underline{\kappa}(\mathcal{N}_\delta; \gamma_n, P_n) \leq \underline{\kappa}(\mathcal{N}_\delta; \gamma_0, P_0) - \varepsilon\,. \tag{40}$$

For each n, choose $\theta_n \in \Theta$ and $F_n \in \mathcal{N}_\delta$ for which $\mathbb{E}^{F_n}[k(U, \theta_n, \gamma_n)] < \underline{\kappa}_\delta(\gamma_n, P_n) + \varepsilon$. Let m_n denote the derivative of F_n with respect to F_*. By Assumption M(iv), we may assume (taking a subsequence if necessary) that $\theta_n \to \theta^*$ for some $\theta^* \in \Theta$. By $\mathcal{E}$-weak convergence, there is a subsequence $\{n_i : i \geq 1\}$ along which m_{n_i} is $\mathcal{E}$-weakly convergent to a unique $m^* \in \mathcal{L}_+$. By similar arguments to the proof of Lemma 4.3, we may use $\mathcal{E}$-weak convergence and Assumption M(i) to deduce that $m^* \in \mathcal{M}_\delta$, m^* satisfies the constraints in (22) at $(\theta^*, \gamma_0, P_0)$, and

$$\mathbb{E}^{F_*}[m_{n_i}(U)k(U, \theta_{n_i}, \gamma_{n_i})] \to \mathbb{E}^{F_*}[m^*(U)k(U, \theta^*, \gamma_0)]\,.$$

But by (40) this implies $\mathbb{E}^{F_*}[m^*(U)k(U, \theta^*, \gamma_0)] \leq \underline{\kappa}(\mathcal{N}_\delta; \gamma, P) - \varepsilon$, which contradicts the definition of $\underline{\kappa}(\mathcal{N}_\delta; \gamma_0, P_0)$. ■

Lemma E.5 *Let Assumptions Φ and M hold with k depending only on (θ, γ). Then: $\underline{\kappa}(\mathcal{N}_\delta; \cdot, \cdot)$ and $\bar{\kappa}(\mathcal{N}_\delta; \cdot, \cdot)$ are continuous on a neighborhood of (γ_0, P_0).*

Proof of Lemma E.5. The proof follows similar arguments to Lemma E.5. For upper semicontinuity, for any $\varepsilon > 0$ we may choose $\theta_\varepsilon \in \Theta_\delta$ such that $k(\theta_\varepsilon, \gamma_0) < \underline{\kappa}(\mathcal{N}_\delta; \gamma_0, P_0) + \varepsilon$. It follows from Lemmas E.2 and E.3 that there is a neighborhood N of (γ, P) such that (i) Condition S holds at $(\theta_\varepsilon, \gamma, P)$ and (ii) $\delta^\star(\theta_\varepsilon; \gamma, P) < \delta$ for all $(\gamma, P) \in N$. By continuity of $k(\theta_\varepsilon, \cdot)$ there also exists a neighborhood N' of γ_0 on which $k(\theta_\varepsilon, \cdot) < k(\theta_\varepsilon, \gamma_0) + \varepsilon$ Therefore on $N \cap (N' \times \mathbb{R}^{d_1+d_2})$ we have

$$\underline{\kappa}(\mathcal{N}_\delta; \cdot, \cdot) \leq k(\theta_\varepsilon, \cdot) < k(\theta_\varepsilon, \gamma_0) + \varepsilon < \underline{\kappa}(\mathcal{N}_\delta; \gamma_0, P_0) + 2\varepsilon\,,$$

establishing upper semicontinuity. The proof of lower semicontinuity follows similar arguments to the proof of Lemma E.4. ■

E.4 Convergence of multipliers

This section contains some ancillary results on the convergence of multipliers. As before, we prove the results only for the lower bound; corresponding results for the upper bound follow by parallel arguments.

In the explicit-dependence case, let $\Xi_\delta(\theta; \gamma, P)$ denote the set of Lagrange multipliers $(\eta, \zeta, \lambda')'$ solving $\underline{\kappa}^\star_\delta(\theta; \gamma, P)$ in equation (5). In the implicit-dependence case, we may simultaneously check feasibility of (θ, γ, P) and evaluate the counterfactual using the convex program

$$\underline{\kappa}_\delta(\theta; \gamma, P) = \inf_{F \in \mathcal{N}_\delta} k(\theta, \gamma) \quad \text{subject to (1)–(4) holding under } F \text{ at } (\theta, \gamma, P)\,. \tag{41}$$

The value of this program is $k(\theta,\gamma)$ if there is a distribution $F \in \mathcal{N}_\delta$ satisfying the constraints at (θ,γ,P), otherwise the value of the program is $+\infty$. By analogous arguments to Lemma 4.1, the program (41) has the dual representation

$$\underline{\kappa}^\star_\delta(\theta;\gamma,P) = k(\theta,\gamma) + \sup_{\eta\geq 0,\zeta\in\mathbb{R},\lambda\in\Lambda} -\mathbb{E}^{F_*}\Big[(\eta\phi)^\star(-\zeta-\lambda' g(U,\theta,\gamma))\Big] - \eta\delta - \zeta - \lambda'_{12}P\,. \tag{42}$$

Similarly, define

$$\bar{\kappa}_\delta(\theta;\gamma,P) = \sup_{F\in\mathcal{N}_\delta} k(\theta,\gamma) \quad \text{subject to (1)–(4) holding under } F \text{ at } (\theta,\gamma,P)$$

which takes the value $k(\theta,\gamma)$ if if there is a distribution $F \in \mathcal{N}_\delta$ satisfying the constraints at (θ,γ,P) and $-\infty$ otherwise. This program has the dual representation

$$\bar{\kappa}^\star_\delta(\theta;\gamma,P) = k(\theta,\gamma) + \inf_{\eta\geq 0,\zeta\in\mathbb{R},\lambda\in\Lambda} \mathbb{E}^{F_*}\Big[(\eta\phi)^\star(-\zeta-\lambda' g(U,\theta,\gamma))\Big] + \eta\delta + \zeta + \lambda'_{12}P\,. \tag{43}$$

Thus, in the implicit-dependence case, we let $\underline{\Xi}_\delta(\theta;\gamma,P)$ and $\overline{\Xi}_\delta(\theta;\gamma,P)$ denote the sets of multipliers solving (42) and (43), respectively.

Lemma E.6 *Let Assumption Φ hold, let Condition S hold at (θ,γ,P), and let $\delta^\star(\theta;\gamma,P) < \delta$. Then: $\underline{\Xi}_\delta(\theta;\gamma,P)$ is a nonempty, compact, convex subset of $\mathbb{R}_+\times\mathbb{R}\times\Lambda$.*

Proof of Lemma E.6. Follows from Theorem 2.165 of Bonnans and Shapiro (2000): the objective is the pointwise infimum of affine functions and is therefore concave and upper semicontinuous, and condition S implies a Slater constraint qualification (cf. Lemma D.2). ∎

The next lemma uses some insights from Pollard (1991). Let $T = \mathbb{R}_+\times\mathbb{R}\times\Lambda$. For each $\varepsilon > 0$ we may cover $\underline{\Xi}_\delta(\theta;\gamma,P) \subset T$ by a set $\underline{\Xi}_\delta(\theta;\gamma,P)^\varepsilon \subset T$ consisting of finitely many hypercubes with edges parallel to the coordinate axes so that $d((\eta,\zeta,\lambda),\underline{\Xi}_\delta(\theta;\gamma,P)) \leq \varepsilon$ for all $(\eta,\zeta,\lambda) \in \underline{\Xi}_\delta(\theta;\gamma,P)^\varepsilon$ and so that $\mathrm{ext}(\partial\underline{\Xi}_\delta(\theta;\gamma,P)^\varepsilon;T) \cap \underline{\Xi}_\delta(\theta;\gamma,P) = \emptyset$.

Lemma E.7 *Let Assumptions Φ and M(i)(v) hold, let Condition S hold at (θ,γ,P), and let $\delta^\star(\theta;\gamma,P) < \delta$. Then: for each $\varepsilon > 0$ there exists a neighborhood N of (θ,γ,P) such that for each $(\tilde\theta,\tilde\gamma,\tilde P) \in N$ the multipliers $\underline{\Xi}_\delta(\tilde\theta;\tilde\gamma,\tilde P)$ solving $\underline{\kappa}^\star_\delta(\tilde\theta;\tilde\gamma,\tilde P)$ are contained in $\underline{\Xi}_\delta(\theta;\gamma,P)^\varepsilon$. Moreover, $\underline{\kappa}^\star_\delta$ is continuous at (θ,γ,P).*

Proof of Lemma E.7. We prove the result for the explicit-dependence case. The result for the implicit-dependence case follows similarly.

Step 1 (preliminaries): To simplify notation, let $\Xi = \underline{\Xi}_\delta(\theta;\gamma,P)$ and $\Xi^\varepsilon = \underline{\Xi}_\delta(\theta;\gamma,P)^\varepsilon$. Lemmas E.2 and E.3 imply there is a neighborhood N' of (θ,γ,P) such that $(\tilde\theta,\tilde\gamma,\tilde P)$ is attainable and $\delta^\star(\tilde\theta,\tilde\gamma,\tilde P) < \delta$ for each $(\tilde\theta,\tilde\gamma,\tilde P) \in N'$. By Lemma E.6, for each $(\tilde\theta,\tilde\gamma,\tilde P) \in N'$ the multipliers $\tilde\Xi := \underline{\Xi}_\delta(\tilde\theta;\tilde\gamma,\tilde P)$ solving $\underline{\kappa}^\star_\delta(\tilde\theta;\tilde\gamma,\tilde P)$ are a nonempty, compact, convex subset of T. Let

$$\ell(\eta,\zeta,\lambda) = \inf_{m\in\mathcal{L}_+} L(m,\eta,\zeta,\lambda;\theta,\gamma,P)\,, \qquad \tilde\ell(\eta,\zeta,\lambda) = \inf_{m\in\mathcal{L}_+} L(m,\eta,\zeta,\lambda;\tilde\theta,\tilde\gamma,\tilde P)\,.$$

By upper semicontinuity of ℓ and definition of Ξ, we have

$$\underline{\kappa}^\star_\delta(\theta;\gamma,P) - \sup_{(\eta,\zeta,\lambda)\in \mathrm{ext}(\partial\Xi^\varepsilon;T)} \ell(\eta,\zeta,\lambda) =: 4a > 0\,. \tag{44}$$

The remaining steps of the proof depend on whether $\inf\{\eta : (\eta,\zeta,\lambda)\in\Xi\} > 0$ or not.

Step 2 (proof when $\inf\{\eta : (\eta,\zeta,\lambda')'\in\Xi\} > 0$): Without loss of generality we may choose Ξ^ε such that $\inf\{\eta : (\eta,\zeta,\lambda)\in\Xi^\varepsilon\} > 0$. As

$$\ell(\eta,\zeta,\lambda) = -\mathbb{E}^{F_*}\Big[(\eta\phi)^\star(-k(U,\theta,\gamma) - \lambda' g(U,\theta,\gamma) - \zeta)\Big] - \eta\delta - \zeta - \lambda_{12}' P$$

when $\eta > 0$, it follows by Assumption M(v) that $\tilde\ell(\eta,\zeta,\lambda) \to \ell(\eta,\zeta,\lambda)$ as $(\tilde\theta,\tilde\gamma,\tilde P) \to (\theta,\gamma,P)$. This convergence is pointwise for $(\eta,\zeta,\lambda)\in\mathbb{R}_+\times\mathbb{R}^{d+1}$ with $\eta>0$. By concavity of $\tilde\ell(\cdot,\cdot,\cdot)$ and Theorem 10.8 of Rockafellar (1970), pointwise convergence may be strengthened to uniform convergence on compact subsets of $\mathbb{R}_{++}\times\mathbb{R}^{d+1}$, hence

$$\sup_{(\eta,\zeta,\lambda)\in\Xi^\varepsilon} \tilde\ell(\eta,\zeta,\lambda) \to \underline{\kappa}^\star_\delta(\theta;\gamma,P) \text{ as } (\tilde\theta,\tilde\gamma,\tilde P)\to(\theta,\gamma,P). \tag{45}$$

By (44) and (45), there exists a neighborhood N'' of (θ,γ,P) such that for $(\tilde\theta,\tilde\gamma,\tilde P)\in N'\cap N''$, the inequality

$$\sup_{(\eta,\zeta,\lambda)\in\Xi^\varepsilon} \tilde\ell(\eta,\zeta,\lambda) - \sup_{(\eta,\zeta,\lambda)\in\mathrm{ext}(\partial\Xi^\varepsilon;T)} \tilde\ell(\eta,\zeta,\lambda) > 2a$$

holds. It follows that $\arg\sup_{(\eta,\zeta,\lambda)\in\Xi^\varepsilon} \tilde\ell(\eta,\zeta,\lambda)\cap\mathrm{ext}(\partial\Xi^\varepsilon;T) = \emptyset$ holds on $N'\cap N''$. Fix any $(\tilde\theta,\tilde\gamma,\tilde P)\in N'\cap N''$. Take any point $(\eta,\zeta,\lambda)\in T\setminus\Xi^\varepsilon$. Then there is a point $(\eta^\varepsilon,\zeta^\varepsilon,\lambda^\varepsilon)\in\mathrm{ext}(\partial\Xi^\varepsilon;T)$ that is a convex combination of (η,ζ,λ) and a point $(\tilde\eta,\tilde\zeta,\tilde\lambda)\in\arg\sup_{(\eta,\zeta,\lambda)\in\Xi^\varepsilon}\tilde\ell(\eta,\zeta,\lambda)$. But then we have

$$\tilde\ell(\tilde\eta,\tilde\zeta,\tilde\lambda) > \tilde\ell(\eta^\varepsilon,\zeta^\varepsilon,\lambda^\varepsilon) \geq \tau\tilde\ell(\tilde\eta,\tilde\zeta,\tilde\lambda) + (1-\tau)\tilde\ell(\eta,\zeta,\lambda)$$

for some $\tau\in(0,1)$, hence $\tilde\ell(\eta,\zeta,\lambda) < \sup_{(\eta,\zeta,\lambda)\in\Xi^\varepsilon}\tilde\ell(\eta,\zeta,\lambda)$. Therefore, $\tilde\Xi\subseteq\Xi^\varepsilon$ must hold on $N'\cap N''$. Continuity now follows by (45).

Step 3 (proof when $\inf\{\eta : (\eta,\zeta,\lambda')'\in\Xi\} = 0$): We break this proof into several steps.

Step 3a (proof that $\tilde{\Xi} \subseteq \Xi^\varepsilon$ on a neighborhood of (θ, γ, P)): As $\inf\{\eta : (\eta, \zeta, \lambda) \in \Xi\} = 0$ and Ξ is compact, there exists $(\bar{\zeta}, \bar{\lambda}) \in \mathbb{R} \times \Lambda$ such that $(0, \bar{\zeta}, \bar{\lambda}) \in \Xi$. By upper semicontinuity and concavity of ℓ, we may deduce that $\lim_{\eta \downarrow 0} \ell(\eta, \bar{\zeta}, \bar{\lambda}) = \ell(0, \bar{\zeta}, \bar{\lambda})$ (cf. the proof of Lemma 4.1). So for any $\varepsilon_0 \in (0, a)$ we may choose $\bar{\eta} > 0$ such that $\ell(\bar{\eta}, \bar{\zeta}, \bar{\lambda}) > \ell(0, \bar{\zeta}, \bar{\lambda}) - \varepsilon_0$ and $(\bar{\eta}, \bar{\zeta}, \bar{\lambda}) \in \mathrm{int}(\Xi^\varepsilon)$. By Assumption M(v), there exists a neighborhood N'' of (θ, γ, P) upon which the inequality

$$\tilde{\ell}(\bar{\eta}, \bar{\zeta}, \bar{\lambda}) > \underline{\kappa}^\star_\delta(\theta; \gamma, P) - 2\varepsilon_0 \tag{46}$$

holds for all $(\tilde{\theta}, \tilde{\gamma}, \tilde{P}) \in N''$.

We now show by contradiction that the inequality

$$\sup_{(\eta,\zeta,\lambda) \in \mathrm{ext}(\partial \Xi^\varepsilon; T)} \ell(\eta, \zeta, \lambda) \geq \sup_{(\eta,\zeta,\lambda) \in \mathrm{ext}(\partial \Xi^\varepsilon; T)} \tilde{\ell}(\eta, \zeta, \lambda) - 2\varepsilon_0 \tag{47}$$

holds on a neighborhood N''' of (θ, γ, P). To establish a contradiction, suppose that there is $\varepsilon_1 > 0$ and a sequence $\{(\theta_n, \gamma_n, P_n) : n \geq 1\}$ converging to (θ, γ, P) along which

$$\sup_{(\eta,\zeta,\lambda) \in \mathrm{ext}(\partial \Xi^\varepsilon; T)} \ell(\eta, \zeta, \lambda) \leq \sup_{(\eta,\zeta,\lambda) \in \mathrm{ext}(\partial \Xi^\varepsilon; T)} \ell_n(\eta, \zeta, \lambda) - \varepsilon_1\,, \tag{48}$$

where $\ell_n(\eta, \zeta, \lambda) := \inf_{m \in \mathcal{L}_+} L(m, \eta, \zeta, \lambda; \theta_n, \gamma_n, P_n)$. For each $n \geq 1$, choose

$$(\eta_n, \zeta_n, \lambda_n) \in \arg\sup_{(\eta,\zeta,\lambda) \in \mathrm{ext}(\partial \Xi^\varepsilon; T)} \ell_n(\eta, \zeta, \lambda)\,.$$

As $\mathrm{ext}(\partial \Xi^\varepsilon; T)$ is compact, we may take a subsequence $\{(\eta_{n_i}, \zeta_{n_i}, \lambda_{n_i}) : i \geq 1\}$ converging to some point $(\eta^*, \zeta^*, \lambda^*) \in \mathrm{ext}(\partial \Xi^\varepsilon; T)$. There are two cases to consider.

Case 1: if $\eta^* > 0$, then by uniform convergence of ℓ_n to ℓ on compact subsets of $\mathbb{R}_{++} \times \mathbb{R}^{d+1}$ we obtain

$$\lim_{i \to \infty} \ell_{n_i}(\eta_{n_i}, \zeta_{n_i}, \lambda_{n_i}) = \ell(\eta^*, \zeta^*, \lambda^*) \leq \sup_{(\eta,\zeta,\lambda) \in \mathrm{ext}(\partial \Xi^\varepsilon; T)} \ell(\eta, \zeta, \lambda)\,,$$

contradicting (48).

Case 2: If $\eta^* = 0$, fix any small $\varepsilon_2 > 0$ so that $(\varepsilon_2, \zeta^*, \lambda^*) \in \Xi^\varepsilon$. By upper semicontinuity and concavity of $\ell(\cdot, \zeta^*, \lambda^*)$, we may choose ε_2 sufficiently small that $\ell(\varepsilon_2, \zeta^*, \lambda^*) - \ell(2\varepsilon_2, \zeta^*, \lambda^*) < \varepsilon_1$. For all i large enough we have $\eta_{n_i} < \varepsilon_2$ and hence $\tau_{n_i} := \frac{\varepsilon_2}{2\varepsilon_2 - \eta_{n_i}} \in (0, 1)$. By concavity:

$$\ell_{n_i}(\eta_{n_i}, \zeta_{n_i}, \lambda_{n_i}) \leq \frac{1}{\tau_{n_i}} \left(\ell_{n_i}(\varepsilon_2, \zeta_{n_i}, \lambda_{n_i}) - (1 - \tau_{n_i})\ell_{n_i}(2\varepsilon_2, \zeta_{n_i}, \lambda_{n_i})\right)\,.$$

But by uniform convergence of ℓ_{n_i} on compact subsets of $\mathbb{R}_{++} \times \mathbb{R}^{d+1}$, we therefore have

$$\begin{aligned}\lim_{i\to\infty} \ell_{n_i}(\eta_{n_i}, \zeta_{n_i}, \lambda_{n_i}) &\leq 2\ell(\varepsilon_2, \zeta^*, \lambda^*) - \ell(2\varepsilon_2, \zeta^*, \lambda^*) \\ &\leq \sup_{(\eta,\zeta,\lambda)\in \mathrm{ext}(\partial\Xi^\varepsilon;T)} \ell(\eta, \zeta, \lambda) + (\ell(\varepsilon_2, \zeta^*, \lambda^*) - \ell(2\varepsilon_2, \zeta^*, \lambda^*)) \\ &< \sup_{(\eta,\zeta,\lambda)\in \mathrm{ext}(\partial\Xi^\varepsilon;T)} \ell(\eta, \zeta, \lambda) + \varepsilon_1 ,\end{aligned}$$

contradicting (48). This proves inequality (47).

It now follows from displays (44), (46), and (47) that on $N' \cap N'' \cap N'''$ we have

$$\begin{aligned}\tilde{\ell}(\bar{\eta}, \bar{\zeta}, \bar{\lambda}) > \underline{\kappa}^\star_\delta(\theta; \gamma, P) - 2\varepsilon_0 &= \sup_{(\eta,\zeta,\lambda)\in \mathrm{ext}(\partial\Xi^\varepsilon;T)} \ell(\eta, \zeta, \lambda) + 4a - 2\varepsilon_0 \\ &\geq \sup_{(\eta,\zeta,\lambda)\in \mathrm{ext}(\partial\Xi^\varepsilon;T)} \tilde{\ell}(\eta, \zeta, \lambda) + 4(a - \varepsilon_0) ,\end{aligned}$$

where $a - \varepsilon_0 > 0$. Therefore, $\sup_{(\eta,\zeta,\lambda)\in\Xi^\varepsilon} \tilde{\ell}(\eta, \zeta, \lambda) > \sup_{(\eta,\zeta,\lambda)\in \mathrm{ext}(\partial\Xi^\varepsilon;T)} \tilde{\ell}(\eta, \zeta, \lambda)$ holds on $N' \cap N'' \cap N'''$. It now follows by similar arguments to Step 2 that $\tilde{\Xi} \subseteq \Xi^\varepsilon$ on $N' \cap N'' \cap N'''$.

Step 3b (proof of continuity): By Step 3a, for any $(\tilde{\theta}, \tilde{\gamma}, \tilde{P}) \in N' \cap N'' \cap N'''$ we have

$$\underline{\kappa}^\star_\delta(\tilde{\theta}; \tilde{\gamma}, \tilde{P}) = \sup_{(\eta,\zeta,\lambda)\in\mathbb{R}_+\times\mathbb{R}\times\Lambda} \tilde{\ell}(\eta, \zeta, \lambda) = \sup_{(\eta,\zeta,\lambda)\in\Xi^\varepsilon} \tilde{\ell}(\eta, \zeta, \lambda) . \tag{49}$$

It follows by (46) that

$$\underline{\kappa}^\star_\delta(\tilde{\theta}; \tilde{\gamma}, \tilde{P}) \geq \tilde{\ell}(\bar{\eta}, \bar{\zeta}, \bar{\lambda}) > \underline{\kappa}^\star_\delta(\theta; \gamma, P) - 2\varepsilon_0 ,$$

proving lower semicontinuity. To establish upper semicontinuity, for each $\varepsilon_0 > 0$ one may deduce (by similar arguments used to establish inequality (47) in Step 3a) there is a neighborhood N'''' of (θ, γ, P) upon which

$$\sup_{(\eta,\zeta,\lambda)\in\Xi^\varepsilon} \ell(\eta, \zeta, \lambda) \geq \sup_{(\eta,\zeta,\lambda)\in\Xi^\varepsilon} \tilde{\ell}(\eta, \zeta, \lambda) - \varepsilon_0 \tag{50}$$

holds. It follows by (49) and (50) that on $N' \cap N'' \cap N''' \cap N''''$, we have

$$\underline{\kappa}^\star_\delta(\tilde{\theta}; \tilde{\gamma}, \tilde{P}) = \sup_{(\eta,\zeta,\lambda)\in\Xi^\epsilon} \tilde{\ell}(\eta, \zeta, \lambda) \leq \sup_{(\eta,\zeta,\lambda)\in\Xi^\epsilon} \ell(\eta, \zeta, \lambda) + \varepsilon_0 = \underline{\kappa}^\star_\delta(\theta; \gamma, P) + \varepsilon_0$$

as required. ■

Lemma E.8 *Let Assumptions Φ and M(i)(v) hold, let Condition S hold at (θ, γ, P), and let $\delta^\star(\theta; \gamma, P) < \delta$. Then:*

$$\vec{d}_H(\underline{\Xi}_\delta(\tilde{\theta}; \tilde{\gamma}, \tilde{P}), \underline{\Xi}_\delta(\theta; \gamma, P)) \to 0 , \quad \vec{d}_H(\overline{\Xi}_\delta(\tilde{\theta}; \tilde{\gamma}, \tilde{P}), \overline{\Xi}_\delta(\theta; \gamma, P)) \to 0 \quad \textit{as } (\tilde{\theta}, \tilde{\gamma}, \tilde{P}) \to (\theta, \gamma, P).$$

Proof of Lemma E.8. Fix $\varepsilon > 0$. By Lemma E.7, there is a neighborhood N of (θ, γ, P) such that $\Xi_\delta(\tilde{\theta}; \tilde{\gamma}, \tilde{P}) \subseteq \Xi_\delta(\theta; \gamma, P)^\varepsilon$ holds for all $(\tilde{\theta}, \tilde{\gamma}, \tilde{P}) \in N$. Therefore, the inequality

$$\vec{d}_H(\Xi_\delta(\tilde{\theta}; \tilde{\gamma}, \tilde{P}), \Xi_\delta(\theta; \gamma, P)) \leq \vec{d}_H(\Xi_\delta(\theta; \gamma, P)^\varepsilon, \Xi_\delta(\theta; \gamma, P)) \leq \varepsilon$$

holds for all $(\tilde{\theta}, \tilde{\gamma}, \tilde{P}) \in N$. ■

E.5 Feasibility

Consider characterizing the extreme counterfactuals in the implicit-dependence case. We may use conic programs to check whether there exists a distribution that satisfies the moment conditions (1)–(4) at a particular point θ. If so, we say that θ is feasible.

To check feasibility of θ for distributions belonging to $\mathcal{N}_\infty$, we may use the program

$$f_\infty(\theta; \gamma, P) = \inf_{F \in \mathcal{N}_\infty} 0 \quad \text{subject to (1)–(4) holding at } (\theta, F)$$

where $f_\infty(\theta; \gamma, P) = +\infty$ if infimum runs over an empty set.

Lemma E.9 *Let Assumption Φ hold. Then: the dual of $f_\infty(\theta; \gamma, P)$ is:*

$$f^\star_\infty(\theta; \gamma, P) = \sup_{\lambda \in \Lambda : F_*\text{-ess}\inf(\lambda' g(\cdot, \theta, \gamma)) > -\infty} \left(F_*\text{-ess}\inf(\lambda' g(\cdot, \theta, \gamma)) - \lambda'_{12} P\right) .$$

If condition S also holds at (θ, γ, P), then: strong duality holds: $f_\infty(\theta; \gamma, P) = f^\star_\infty(\theta; \gamma, P)$.

Proof of Lemma E.9. Recall $\mathcal{L}_+$ is the cone of (F_*-almost surely) non-negative functions in $\mathcal{L}$. We may write $f_\infty(\theta; \gamma, P)$ as a conic program:

$$f_\infty(\theta; \gamma, P) = \inf_{m \in \mathcal{L}_+} 0 \quad \text{subject to the constraints in (22).}$$

The result follows by similar arguments to the proof of Lemma B.2. ■

To check feasibility of θ for distributions belonging to $\mathcal{F}_\theta$, we may use the program

$$f_\#(\theta; \gamma, P) = \inf_{F \in \mathcal{F}_\theta} 0 \quad \text{subject to (1)–(4) holding at } (\theta, F)$$

where $f_\#(\theta; \gamma, P) = +\infty$ if infimum runs over an empty set.

Lemma E.10 *Let Condition $S_\#$ hold at (θ, γ_0, P_0). Then: the dual of $f_\#(\theta; \gamma_0, P_0)$ is:*

$$f^\star_\#(\theta; \gamma_0, P_0) = \sup_{\lambda \in \Lambda: \mu\text{-ess}\inf(\lambda' g(\cdot,\theta,\gamma_0)) > -\infty} \big(\mu\text{-ess}\inf(\lambda' g(\cdot, \theta, \gamma_0)) - \lambda'_{12} P_0\big)$$

and strong duality holds: $f_\#(\theta; \gamma_0, P_0) = f^\star_\#(\theta; \gamma_0, P_0)$.

Proof of Lemma E.10. We use similar arguments to the proof of Lemma B.3, stating only the necessary modifications here. Let $\dot{\mathcal{M}} = \{m \in L^1(\mu) : \int mg \,\mathrm{d}\mu \text{ is finite}\}$ and let $\dot{\mathcal{M}}_+ = \dot{\mathcal{M}} \cap L^1_+(\mu)$. For any $c \in \mathbb{R}^{d+1}$, define

$$\mathcal{M}[c] = \left\{ m \in \dot{\mathcal{M}}_+ : \int (1, g(u)')' m(u) \,\mathrm{d}\mu(u) = (1, P_0', 0'_{d_3+d_4})' + c \right\}$$

Also let the functions $q_1, q_2 : \mathbb{R}^{d+1} \to (-\infty, +\infty]$ be given by

$$q_1(c) = \left[\begin{array}{ll} 0 & \text{if } \mathcal{M}[c] \neq \emptyset\,, \\ +\infty & \text{otherwise} \end{array} \right.$$

and

$$q_2(c) = \left[\begin{array}{ll} 0 & \text{if } c \in \{0\} \times \mathbb{R}^{d_1}_- \times \{0\}^{d_2} \times \mathbb{R}^{d_3}_- \times \{0\}^{d_4}\,, \\ +\infty & \text{otherwise} \end{array} \right.$$

in which case

$$f_\#(\theta; \gamma_0, P_0) = \inf_{c \in \mathbb{R}^{d+1}} (q_1(c) + q_2(c))$$

If condition $\mathrm{S}_\#$ holds at (θ, γ_0, P_0) then $\mathrm{ri}(\{c : q_1(c) < +\infty\})$ and $\mathrm{ri}(\{c : q_2(c) < +\infty\})$ have nonempty intersection. By Fenchel's duality theorem, we therefore obtain

$$f_\#(\theta; \gamma_0, P_0) = \sup_{\nu \in \mathbb{R}^{d+1}} (-q_1^\star(\nu) - q_2^\star(-\nu)) \tag{51}$$

where $q_1^\star$ and $q_2^\star$ are the convex conjugates of q_1 and q_2; see display (36) for $q_2^\star$. For $q_1^\star$, we have

$$\begin{aligned}
-q_1^\star(\nu) &= \inf_{c \in \mathbb{R}^{d+1}} \big(-c'\nu + q_1(c)\big) \\
&= \inf_{c \in \mathbb{R}^{d+1} : \mathcal{M}[c] \neq \emptyset} \big(-c'\nu\big) \\
&= \inf_{c \in \mathbb{R}^{d+1} : \mathcal{M}[c] \neq \emptyset} \inf_{m \in \mathcal{M}[c]} \left((1, P_0', 0'_{d_3+d_4})\nu - \int (1, g(u)')\nu m(u) \,\mathrm{d}\mu(u) \right) \\
&= \inf_{m \in \dot{\mathcal{M}}_+} \left((1, P_0', 0'_{d_3+d_4})\nu - \int (1, g(u)')\nu m(u) \,\mathrm{d}\mu(u) \right) .
\end{aligned}$$

Let

$$Q(m(u)) = \left[\begin{array}{ll} 0 & \text{if } m(u) \geq 0\,, \\ +\infty & \text{otherwise.} \end{array} \right.$$

Therefore:

$$-q_1^\star(\nu) = \inf_{m\in\dot{\mathcal{M}}} \left((1, P_0', 0_{d_3+d_4}')\nu + \int \big(Q(m(u)) - (1, g(u)')\nu\big)\, m(u)\, \mathrm{d}\mu(u) \right).$$

By Remark A.3 and Theorem A.4 of Csiszár and Matúš (2012), we may bring the infimum inside the expectation:

$$\begin{aligned}
-q_1^\star(\nu) &= (1, P_0', 0_{d_3+d_4}')\nu + \int \inf_{x\in\mathbb{R}} \big(Q(x) - (1, g(u)')\nu\big)\, x\, \mathrm{d}\mu(u) \\
&= (1, P_0', 0_{d_3+d_4}')\nu + \int \inf_{x\geq 0} \big(-(1, g(u)')\nu\big)\, x\, \mathrm{d}\mu(u) \\
&= \left[\begin{array}{ll} -\infty & \text{if } \mu\text{-ess}\inf(-(1, g(\cdot)')\nu) < 0 \\ (1, P_0', 0_{d_3+d_4}')\nu & \text{otherwise.} \end{array} \right.
\end{aligned} \tag{52}$$

Writing $\nu = (\zeta, \lambda')'$, it follows from (36), (51), and (52) that

$$\begin{aligned}
q &= \sup_{\zeta\in\mathbb{R}, \lambda\in\Lambda : \mu\text{-ess}\inf(g(\cdot)'\lambda - \zeta)\geq 0} \zeta - \lambda_{12}' P_0 \\
&= \sup_{\lambda\in\Lambda : \mu\text{-ess}\inf(g(\cdot)'\lambda) > -\infty} \mu\text{-ess}\inf(g(\cdot)'\lambda) - \lambda_{12}' P_0
\end{aligned}$$

as required. ■

References

Bickel, P. J., C. A. Klaassen, Y. Ritov, and J. A. Wellner (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag, New York.

Bonnans, J. and A. Shapiro (2000). *Perturbation Analysis of Optimization Problems*. Springer.

Csiszár, I. and F. Matúš (2012). Generalized minimizers of convex integral functionals, Bregman distance, Pythagorean identities. *Kybernetika 48*(4), 637–689.

Komunjer, I. and G. Ragusa (2016). Existence and characterization of conditional density projections. *Econometric Theory 32*(4), 947–987.

Krasnosel'skii, M. A. and Ya. B. Rutickii (1961). *Convex Functions and Orlicz Spaces*. P. Noordhoff Ltd., Groningen.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory 7*(2), 186–199.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.

Rockafellar, R. T. and R. J. B. Wets (1998). *Variational Analysis*. Springer.

Shapiro, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research 30*(1), 169–186.

Shapiro, A. (2008). Asymptotics of minimax stochastic programs. *Statistics & Probability Letters 78*(2), 150–157.

Shapiro, A. (2017). Distributionally robust stochastic programming. *SIAM Journal on Optimization 27*(4), 2258–2275.