Econometrica, Vol. 91, No. 1 (January, 2023), 263–298

COUNTERFACTUAL SENSITIVITY AND ROBUSTNESS

TIMOTHY CHRISTENSEN
Department of Economics, New York University and Department of Economics, University College London

BENJAMIN CONNAULT
Latour Trading LLC

We propose a framework for analyzing the sensitivity of counterfactuals to parametric assumptions about the distribution of latent variables in structural models. In particular, we derive bounds on counterfactuals as the distribution of latent variables spans nonparametric neighborhoods of a given parametric specification while other "structural" features of the model are maintained. Our approach recasts the infinite-dimensional problem of optimizing the counterfactual with respect to the distribution of latent variables (subject to model constraints) as a finite-dimensional convex program. We also develop an MPEC version of our method to further simplify computation in models with endogenous parameters (e.g., value functions) defined by equilibrium constraints. We propose plug-in estimators of the bounds and two methods for inference. We also show that our bounds converge to the sharp nonparametric bounds on counterfactuals as the neighborhood size becomes large. To illustrate the broad applicability of our procedure, we present empirical applications to matching models with transferable utility and dynamic discrete choice models.

KEYWORDS: Robustness, ambiguity, model uncertainty, misspecification, global sensitivity analysis.

1. INTRODUCTION

RESEARCHERS FREQUENTLY MAKE PARAMETRIC ASSUMPTIONS about the distribution of latent variables in structural models. These assumptions are typically made for computational convenience[1] or because simulation-based methods are used for estimation. In many models, such as those we consider in this paper, the distribution of latent variables is not nonparametrically identified. This raises the possibility that model parameters and the outcomes of policy experiments, or *counterfactuals*, may be only partially identified when parametric assumptions are relaxed. That is, different distributions may fit the data equally well in-sample, but may yield different values of the counterfactual. It is therefore natural to question whether counterfactuals are sensitive or robust to researchers' parametric assumptions, especially when evaluating the credibility of structural modeling exercises.

Timothy Christensen: timothy.christensen@nyu.edu
Benjamin Connault: benjamin.connault@gmail.com

We are very grateful to four anonymous referees, R. Allen, S. Bonhomme, G. Chamberlain, A. Galvao, B. Kaplowitz, J. Lazarev, M. Masten, K. Menzel, D. Miller, M. Mogstad, F. Molinari, J. Nesbit, A. Poirier, J. Porter, T. Sargent, A. Torgovitsky, and numerous seminar and conference participants for helpful comments and suggestions. Maximilian Huber provided outstanding research assistance. Support from the National Science Foundation via Grant SES-1919034 is gratefully acknowledged.

[1]Examples include the conventional Gumbel (or type-I extreme value) assumption in discrete choice models following McFadden (1974), dynamic discrete choice models following Rust (1987), and matching models with transferable utility following Dagsvik (2000) and Choo and Siow (2006). Models of static or dynamic discrete games often impose parametric assumptions about the distribution of payoff shocks—see, for example, Berry (1992), Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2007), and Ciliberto and Tamer (2009).

 https://doi.org/10.3982/ECTA17232

This paper proposes a framework for analyzing the sensitivity of counterfactuals to parametric assumptions about the distribution of latent variables in a class of structural models. In particular, we derive bounds on counterfactuals as the distribution of latent variables spans nonparametric neighborhoods of a given parametric specification while other "structural" features of the model are maintained. This approach is in the spirit of global sensitivity analysis advocated by Leamer (1985) (see also Tamer (2015)). Global sensitivity analyses are important in this context: many structural models are nonlinear so policy interventions can have different effects at different points in the parameter space. But a major difficulty with implementing global sensitivity analyses is tractability. A more tractable alternative are local sensitivity analyses, which are based on small perturbations around a chosen specification. Because local approaches rely on linearization, they may fail to correctly characterize the range of counterfactuals predicted by a nonlinear model when the distribution differs nontrivially from the researcher's chosen parametric specification.

Our main insight is to borrow from the robustness literature in economics pioneered by Hansen and Sargent (2001, 2008) to simplify computation using convex programming.[2] Following this literature, we define neighborhoods around the researcher's parametric specification using statistical divergence (e.g., Kullback–Leibler divergence), with the option to add certain shape restrictions as appropriate. For tractability, we restrict our attention to models that may be written as a finite number of moment (in)equalities, where the expectation is with respect to the distribution of latent variables. While restrictive, this class accommodates many important models of static and dynamic discrete choice, discrete games, and matching.

To describe our procedure, consider the problem of minimizing or maximizing the counterfactual at a fixed value of structural parameters by varying the distribution of latent variables over a neighborhood, subject to the model's (in)equality restrictions. We use duality to recast this infinite-dimensional optimization problem as a finite-dimensional convex program. The value of this *inner* program is treated as a criterion function, which is optimized in an *outer* optimization with respect to structural parameters. Importantly, the dimension of the inner problem is independent of the neighborhood size, making our procedure tractable over both small and large neighborhoods. To further simplify computation, we develop an MPEC version of our procedure for models featuring endogenous parameters (e.g., value functions) defined by equilibrium constraints. We show that this implementation can produce significant computational gains for dynamic discrete choice models in particular.

Our approach is conceptually different from nonparametric partial identification analyses which derive bounds on counterfactuals under minimal distributional assumptions. But as we show, bounds computed using our procedure converge to the (sharp) nonparametric bounds in the limit as the neighborhood size becomes large. Aside from sensitivity analyses, our methods may therefore be used to approximate nonparametric bounds by taking the neighborhood size to be large but finite.

For estimation and inference, we propose simple plug-in estimators of the bounds and establish their consistency. We also propose and theoretically justify two methods for inference: a computationally simple but conservative projection procedure and a relatively more efficient bootstrap procedure.

[2]Our approach is also related to the field of *distributionally robust optimization* in operations research. See, for example, Shapiro (2017), Duchi and Namkoong (2021), and references therein.

We illustrate our procedures with two empirical applications. The first revisits the "marital college premium" estimates reported in Chiappori, Salanié, and Weiss (2017), which relied on an i.i.d. Gumbel (type-I extreme value) assumption for the distribution of individuals' idiosyncratic marital preferences (see also Choo and Siow (2006)). The second empirical application performs a counterfactual welfare analysis in the canonical dynamic discrete choice model of Rust (1987).

Related Literature. Our approach has connections with global prior sensitivity in Bayesian analysis (Chamberlain and Leamer (1976), Leamer (1982), Berger (1984)), most notably Giacomini, Kitagawa, and Uhlig (2019) and Ho (2022) who considered sets of priors constrained by Kullback–Leibler divergence relative to a default prior.

Motivated by questions of sensitivity, Chen, Tamer, and Torgovitsky (2011) studied inference in semiparametric likelihood models using sieve approximations for the infinite-dimensional nuisance parameter (the distribution of latent variables in our setting). For the class of moment-based models we consider, our approach instead eliminates the infinite-dimensional nuisance parameter via a convex program of fixed dimension.

Several other works have used convex duality to characterize identified sets in models with latent variables. Most closely related are Ekeland, Galichon, and Henry (2010) and Schennach (2014).[3] The problem we study is different, both because of its focus on counterfactuals, rather than structural parameters, and because the optimization is performed over a neighborhood, rather than over all distributions. As a consequence, our estimation and inference methods are also quite different.

Torgovitsky (2019b) used linear programming to characterize sharp identified sets in latent variable models defined by quantile restrictions. Within this class, his approach is more computationally convenient than ours for characterizing identified sets. Several important moments or counterfactuals cannot be expressed as quantile restrictions, such as social surplus in discrete choice models and Bellman equations in dynamic discrete choice models. Our approach is compatible with these moments and counterfactuals, thereby allowing the user to characterize identified sets in broader classes of model as well as to perform sensitivity analyses.

There is also a literature deriving nonparametric bounds in specific latent variable models. Examples include Manski (2007, 2014), Allen and Rehbeck (2019), Tebaldi, Torgovitsky, and Yang (2022), Lafférs (2019), Torgovitsky (2019a), and Gualdani and Sinha (2022). Most closely related is Norets and Tang (2014), who constructed identified sets of counterfactual conditional choice probabilities (CCPs) in dynamic binary choice models. Their approach is specific to counterfactual CCPs and to dynamic binary choice models. Our approach allows for a wider range of counterfactual (e.g., welfare), shape restrictions, and multinomial choice, in addition to performing sensitivity analyses.[4]

Finally, our work is complementary to the recent literature on local sensitivity—see, for example, Kitamura, Otsu, and Evdokimov (2013), Andrews, Gentzkow, and Shapiro (2017, 2020), Armstrong and Kolesár (2021), Bonhomme and Weidner (2022), and Mukhin (2018). Much of this literature is concerned with local misspecification of moment conditions, which is different from the setting we consider.

[3]Works using other notions of "duality" to construct identified sets include Beresteanu, Molchanov, and Molinari (2011), Galichon and Henry (2011), Chesher and Rosen (2017), and Li (2018).

[4]Kalouptsidi, Scott, and Souza-Rodrigues (2021) and Kalouptsidi, Kitamura, Lima, and Souza-Rodrigues (2020) considered the converse problem, in which flow payoffs are nonparametric (as they can be in our setting) but the distribution of latent payoff shocks is known.

Outline. Section 2 introduces our procedure and estimators of the bounds, and shows our approach recovers nonparametric bounds as the neighborhood size becomes large. Section 3 discusses practical aspects and implementation details. Section 4 gives guidance for interpreting the neighborhood size. Empirical applications are presented in Section 5. Section 6 discusses estimation and inference. The Supplemental Material (Christensen and Connault (2023)) presents extensions of our methodology, connections with local sensitivity analyses, additional empirical results, and proofs of our main results. A secondary appendix of our working paper version (Christensen and Connault (2022)) presents background material on Orlicz classes and supplemental proofs.

2. PROCEDURE

We begin in Section 2.1 by describing the class of models to which our procedure may be applied. Section 2.2 describes our approach, Section 2.3 shows how duality is used to simplify the bounds, and Section 2.4 introduces our estimators of the bounds. Section 2.5 shows our bounds converge to the sharp nonparametric bounds as the neighborhood size becomes large.

2.1. *Setup*

We consider a class of models that link a structural parameter $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$, a vector of targeted moments $P_0 \in \mathcal{P} \subseteq \mathbb{R}^{d_P}$, and possibly an auxiliary parameter $\gamma_0 \in \Gamma$ (a metric space) via the moment restrictions

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

$$\mathbb{E}^F\big[g_2(U, \theta, \gamma_0)\big] = P_{20}, \tag{1b}$$

$$\mathbb{E}^F\big[g_3(U, \theta, \gamma_0)\big] \le 0, \tag{1c}$$

$$\mathbb{E}^F\big[g_4(U, \theta, \gamma_0)\big] = 0, \tag{1d}$$

where $g_1, \dots, g_4$ are vectors of moment functions, $P_0 = (P_{10}, P_{20})$ is partitioned conformably, and $\mathbb{E}^F$ denotes expectation with respect to a vector of latent variables $U \sim F$. We assume that the researcher has consistent estimators $(\hat{P}, \hat{\gamma})$ of (P_0, γ_0). We also assume that the researcher is interested in a (scalar) counterfactual of the form

$$\kappa = \mathbb{E}^F\big[k(U, \theta, \gamma_0)\big]. \tag{2}$$

This setup accommodates counterfactuals that do not depend explicitly on U, in which case (2) reduces to $\kappa = k(\theta, \gamma_0)$. Note that κ will still depend on the distribution of U through θ, whose values are disciplined by the moment conditions (1a)–(1d).

Several models and counterfactuals of interest fall into this framework. We review three examples before proceeding.

EXAMPLE 2.1—Discrete choice and consumer welfare: Suppose an individual derives utility $h_j(X, \theta) + U_j$ from choice $j \in \mathcal{J}_0 := \{0, 1, \dots, J\}$, where $X \in \mathcal{X}$ are observed covariates and $U = (U_j)_{j \in \mathcal{J}_0}$ is latent (to the econometrician). We assume, as typical, that U is drawn independently across individuals from a continuous distribution F. The probability that an individual with characteristics x chooses j is

$$p(j|x) = \mathbb{P}_F\Big(h_j(x, \theta) + U_j = \max_{j' \in \mathcal{J}_0}\big(h_{j'}(x, \theta) + U_{j'}\big)\Big), \tag{3}$$

where $\mathbb{P}_F$ denotes probabilities when $U \sim F$. In empirical work, θ is typically estimated using a criterion that fits the model-implied choice probabilities (3) to probabilities observed in the data. Welfare analyses are often based on the social surplus (McFadden (1978))

$$W(x) = \mathbb{E}^F\Big[\max_{j\in\mathcal{J}_0}\big(h_j(x,\theta)+U_j\big)\Big],$$

which is the average utility consumers with characteristics x derive from the choice problem. A related welfare measure is the change in surplus $\Delta W(x_a, x_b) = W(x_a) - W(x_b)$ associated with a shift from x_b to x_a. In practice, it is common to assume the U_j are i.i.d. Gumbel (type-I extreme value), as this yields closed-form expressions for choice probabilities and the welfare measures $W(x)$ and $\Delta W(x_a, x_b)$.

Our approach may be used to perform a sensitivity analysis of $W(x)$ and $\Delta W(x_a, x_b)$ to parametric assumptions about F when $\mathcal{X}$ is finite. A leading example is matching models with finitely many agent types—see Section 5.1 and references therein. Understanding the sensitivity of $W(x)$ and $\Delta W(x_a, x_b)$ to F is important in this case because $W(x)$ and $\Delta W(x_a, x_b)$ are not nonparametrically identified.[5]

In our notation, g_2 collects indicator functions representing the choice probabilities (3) across covariates $x \in \mathcal{X}$ and choices $j \in \mathcal{J} := \{1, \dots, J\}$ ($j = 0$ is redundant):

$$g_2(U,\theta) = \Big(\mathbb{1}\Big\{h_j(x,\theta)+U_j = \max_{j'\in\mathcal{J}_0}\big(h_{j'}(x,\theta)+U_{j'}\big)\Big\}\Big)_{(j,x)\in\mathcal{J}\times\mathcal{X}}$$

and $P_{20} = (\Pr(j|x))_{(j,x)\in\mathcal{J}\times\mathcal{X}}$ is the vector of true choice probabilities. There is no g_1, g_3, g_4, or γ in this model. Finally, $k(U,\theta) = \max_{j\in\mathcal{J}_0}(h_j(x,\theta)+U_j)$ for $W(x)$ and $k(U,\theta) = \max_{j\in\mathcal{J}_0}(h_j(x_a,\theta)+U_j) - \max_{j\in\mathcal{J}_0}(h_j(x_b,\theta)+U_j)$ for $\Delta W(x_a, x_b)$.

EXAMPLE 2.2—Discrete games: Following Bresnahan and Reiss (1990, 1991), Berry (1992), and Tamer (2003), consider the complete-information game in Table I.

Here, $U = (U_1, U_2)$ is the latent (to the econometrician) component of firms' profits, which is independent of covariates X. Suppose that the solution concept is restricted to equilibria in pure strategies. The econometrician may estimate the probabilities of the potential market structures (0, 0), (0, 1), (1, 0), (1, 1) (conditional on X) from data on a large number of markets. As the model is incomplete—there are values of U for which there are multiple equilibria—moment inequality methods are typically used in empirical work to avoid restricting the equilibrium selection mechanism. However, strong paramet-

TABLE I

PAYOFF MATRIX FOR (FIRM 1, FIRM 2) WHEN $X = x$.

		Firm 2: 0	Firm 2: 1
Firm 1	0	$(0,0)$	$(0, \beta_2'x + U_2)$
	1	$(\beta_1'x + U_1, 0)$	$(\beta_1'x - \Delta_1 + U_1, \beta_2'x - \Delta_2 + U_2)$

[5]See, for example, Berry and Haile (2010, 2014) and Allen and Rehbeck (2019) for nonparametric identification of utilities and welfare measures in discrete choice models when characteristics have continuous support.

ric assumptions are often made about the distribution of U (typically bivariate Normal) to derive the model-implied probabilities for different market structures; see, for example, Berry (1992), Ciliberto and Tamer (2009), Beresteanu, Molchanov, and Molinari (2011), and Kline and Tamer (2016). It therefore seems natural to also question the sensitivity of counterfactuals to parametric assumptions for U.

This model falls into our setup when the regressors X have finite support $\mathcal{X}$.[6] In our notation, g_1 collects the moment inequalities that bound the probabilities of $(0, 1)$ and $(1, 0)$ across $x \in \mathcal{X}$, with P_{10} denoting the corresponding true probabilities. The inequalities are typically expressed as upper bounds on the probabilities of $(0, 1)$ and $(1, 0)$; we flip the sign to be compatible with (1a):

$$g_1(U,\theta) = \begin{bmatrix} \big(-\mathbb{1}\{U_1 \ge -\beta_1' x;\, U_2 \le \Delta_2 - \beta_2' x\}\big)_{x\in\mathcal{X}} \\ \big(-\mathbb{1}\{U_1 \le \Delta_1 - \beta_1' x;\, U_2 \ge -\beta_2' x\}\big)_{x\in\mathcal{X}} \end{bmatrix},$$

$$P_{10} = \begin{bmatrix} \big(-\Pr\big((1,0)|X = x\big)\big)_{x\in\mathcal{X}} \\ \big(-\Pr\big((0,1)|X = x\big)\big)_{x\in\mathcal{X}} \end{bmatrix},$$

where $\theta = (\Delta_1, \Delta_2, \beta_1, \beta_2)$. Similarly, g_2 and P_{20} collect the moment conditions and probabilities for outcomes $(0, 0)$ and $(1, 1)$, which are always realized as the result of unique equilibria:

$$g_2(U,\theta) = \begin{bmatrix} \big(\mathbb{1}\{U_1 \le -\beta_1' x;\, U_2 \le -\beta_2' x\}\big)_{x\in\mathcal{X}} \\ \big(\mathbb{1}\{U_1 \ge \Delta_1 - \beta_1' x;\, U_2 \ge \Delta_2 - \beta_2' x\}\big)_{x\in\mathcal{X}} \end{bmatrix},$$

$$P_{20} = \begin{bmatrix} \big(\Pr\big((0,0)|X = x\big)\big)_{x\in\mathcal{X}} \\ \big(\Pr\big((1,1)|X = x\big)\big)_{x\in\mathcal{X}} \end{bmatrix}.$$

There is no g_3, g_4, or γ in this model. Ciliberto and Tamer (2009) computed upper bounds on the probability of entrants under a counterfactual payoff shift, say $\tau(\theta)$. The function $k(U, \theta) = \mathbb{1}\{U_1 \ge \tau(\theta) - \beta_1' x\}$ corresponds to the upper bound on the probability of firm 1 entering when $X = x$ under this counterfactual.

EXAMPLE 2.3—Dynamic discrete choice: Consider a canonical dynamic discrete choice (DDC) model following Rust (1987). The decision maker solves

$$V(s) = \mathbb{E}^F\Big[\max_{d\in\mathcal{D}_0}\big(\pi_{d,s}(\theta_\pi) + U_d + \beta E\big[V(s')|d, s\big]\big)\Big], \tag{4}$$

where $s \in \mathcal{S}$ is a Markov state variable, $\mathcal{D}_0 = \{0, 1, \dots, D\}$ is the set of actions, $\pi_{d,s}$ is the flow payoff for action d in state s which is parameterized by θ_π, U_d is a latent payoff shock, $\beta \in (0, 1)$ is a discount parameter, and $E[\cdot|d, s]$ denotes expectation with respect to the future state s'. The distribution F of $U = (U_d)_{d\in\mathcal{D}_0}$ is typically assumed to be continuous

[6]Continuous regressors are often discretized in empirical applications; see, for example, Ciliberto and Tamer (2009), Grieco (2014), Kline and Tamer (2016), and Chen, Christensen, and Tamer (2018).

and independent of s. The CCP of action d in state s is

$$
\begin{aligned}
&p(d|s) \\
&\quad = \mathbb{P}_F\Big(\pi_{d,s}(\theta_\pi) + U_d + \beta E\big[V(s')|d,s\big] = \max_{d'\in\mathcal{D}_0}\big(\pi_{d',s}(\theta_\pi) + U_{d'} + \beta E\big[V(s')|d',s\big]\big)\Big), \quad (5)
\end{aligned}
$$

where $\mathbb{P}_F$ denotes probabilities when $U \sim F$.

It is standard to assume the U_d are i.i.d. Gumbel, as this yields closed-form expressions for the expectation in (4) and multinomial-logit expressions for the CCPs (5). Parameters θ_π or (θ_π, β) are typically estimated using a criterion function that fits the model-implied CCPs (5) to probabilities observed in the data. Counterfactuals are then computed by solving (4) under alternative laws of motion, flow payoffs, or other interventions.

When $\mathcal{S}$ is finite, model parameters, counterfactual CCPs, and counterfactual welfare measures are typically not identified without parametric restrictions on F. Our procedure may be used to perform a sensitivity analysis of counterfactuals to parametric assumptions on F as follows. Let $\theta = (\theta_\pi, v, \tilde{v})$ or $\theta = (\theta_\pi, \beta, v, \tilde{v})$, where $v = (V(s))_{s\in\mathcal{S}}$ and $\tilde{v} = (\tilde{V}(s))_{s\in\mathcal{S}}$ collect the baseline and counterfactual value functions across $s \in \mathcal{S}$. Also let $\gamma = (M_d)_{d\in\mathcal{D}_0}$ collect the transition matrices for s, g_2 collect indicator functions for the CCPs (5) across states $s \in \mathcal{S}$ and choices $d \in \mathcal{D} := \{1, \dots, D\}$ ($d = 0$ is redundant):

$$
g_2(U,\theta,\gamma) = \Big(\mathbb{1}\Big\{\pi_{d,s}(\theta_\pi) + U_d + \beta M_{d,s} v = \max_{d'\in\mathcal{D}_0}\big(\pi_{d',s}(\theta_\pi) + U_{d'} + \beta M_{d',s} v\big)\Big\}\Big)_{(d,s)\in\mathcal{D}\times\mathcal{S}}
$$

with $M_{d,s}$ denoting the sth row of M_d, and $P_{20} = (\Pr(d|s))_{(d,s)\in\mathcal{D}\times\mathcal{S}}$ collect the corresponding true CCPs. Finally, g_4 collects moment functions representing (4) in the baseline model and under the counterfactual:

$$
g_4(U,\theta,\gamma) = \begin{bmatrix} \Big(\max\limits_{d\in\mathcal{D}_0}\big\{\pi_{d,s}(\theta_\pi) + U_d + \beta M_{d,s} v\big\} - v_s\Big)_{s\in\mathcal{S}} \\ \Big(\max\limits_{d\in\mathcal{D}_0}\big\{\tilde{\pi}_{d,s}(\theta_\pi) + U_d + \tilde{\beta}\tilde{M}_{d,s}\tilde{v}\big\} - \tilde{v}_s\Big)_{s\in\mathcal{S}} \end{bmatrix}, \quad (6)
$$

where $v_s = V(s)$, $\tilde{v}_s = \tilde{V}(s)$, and $\tilde{\pi}$, $\tilde{\beta}$, $\tilde{M}_d$ denote counterfactual flow payoffs, discount factor, and law of motion.[7] We recommend including the location normalizations $\mathbb{E}^F[U_d] = 0$ for $d \in \mathcal{D}_0$ in g_4 for interpretability. We also recommend including scale normalizations in g_4 so that $\mathbb{E}^F[\max_{d\in\mathcal{D}_0} U_d]$ is finite. For instance, in Section 5.2 we normalize $\mathbb{E}^F[U_d^2]$ for all $d \in \mathcal{D}_0$.

Counterfactual CCPs can be computed using

$$
k(U,\theta,\gamma) = \mathbb{1}\Big\{\tilde{\pi}_{d,s}(\theta_\pi) + U_d + \tilde{\beta}\tilde{M}_{d,s}\tilde{v} = \max_{d'\in\mathcal{D}_0}\big(\tilde{\pi}_{d',s}(\theta_\pi) + U_{d'} + \tilde{\beta}\tilde{M}_{d',s}\tilde{v}\big)\Big\}.
$$

Change in average welfare corresponds to $k(\theta,\gamma) = w'(\tilde{v} - v)$ for a weight vector w.

[7]If $\mathbb{E}^F[\max_{d\in\mathcal{D}_0} U_d]$ is finite, then $v \mapsto (\mathbb{E}^F[\max_{d\in\mathcal{D}_0}\{\pi_{d,s}(\theta_\pi) + U_d + \beta M_{d,s} v\}])_{s\in\mathcal{S}}$ is a ℓ^∞-contraction of modulus β on $\mathbb{R}^{|\mathcal{S}|}$. Hence, there is a unique $(v,\tilde{v})$ solving $\mathbb{E}^F[g_4(U,\theta,\gamma)] = 0$ at any fixed $(\theta_\pi, \beta, \tilde{\beta}, F)$. The solution $(v,\tilde{v})$ must collect the solutions to (4) in the baseline model and counterfactual across states: $v = (V(s))_{s\in\mathcal{S}}$ and $\tilde{v} = (\tilde{V}(s))_{s\in\mathcal{S}}$. It follows that F satisfies $\mathbb{E}^F[g_4(U,\theta,\gamma)] = 0$ at $\theta = (\theta_\pi, \beta, v, \tilde{v})$ if and only if $(v,\tilde{v})$ corresponds to the value functions V and $\tilde{V}$ under F.

REMARK 2.1: We allow for conditional moment models with $\mathbb{E}[g_1(U, X, \theta, \gamma)|X = x] \leq P_{10}(x)$ (and similarly for (1b)–(1d)) if U is independent of X and X takes values in a finite set $\mathcal{X}$. Moment functions are then stacked across $x \in \mathcal{X}$ to form g_1, g_2, g_3, and g_4 (see Examples 2.1–2.3). Appendix A discusses extensions to conditional moment models where the distribution of U may vary with the value of (discrete) covariates, and to non-separable models with discrete covariates. Models with continuous covariates fall outside the scope of our procedure.

REMARK 2.2: Our setup relies on the counterfactual being expressible as (2). If k is vector-valued, our procedure can be applied to compute the support function[8] of the identified set of counterfactuals: set $k^\tau(U, \theta, \gamma) = \tau' k(U, \theta, \gamma)$ for a conformable unit vector τ and replace (2) with $\kappa^\tau = \mathbb{E}^F[k^\tau(U, \theta, \gamma_0)]$. Our setup excludes counterfactuals that are infinite-dimensional, such as the distribution of the number of firms in a market.

REMARK 2.3: The distribution F is not nonparametrically identified in any of the above examples or, more generally, in the class of models (1a)–(1d) when the support of U contains many more points than there are moment conditions (e.g., when U is continuously distributed).

In common practice, a seemingly reasonable or computationally convenient distribution, say F_*, is assumed by the researcher and maintained throughout the analysis (e.g., bivariate Normal in Example 2.2 and i.i.d. Gumbel in Examples 2.1 and 2.3). Given F_* and estimates $\hat{P} = (\hat{P}_1, \hat{P}_2)$ of P_0 and $\hat{\gamma}$ of γ_0, the researcher computes an estimate $\hat{\theta}$ of θ using a criterion function based on the moment conditions

$$\begin{aligned} &\mathbb{E}^{F_*}\big[g_1(U, \theta, \hat{\gamma})\big] \leq \hat{P}_1, \qquad \mathbb{E}^{F_*}\big[g_2(U, \theta, \hat{\gamma})\big] = \hat{P}_2, \\ &\mathbb{E}^{F_*}\big[g_3(U, \theta, \hat{\gamma})\big] \leq 0, \qquad \mathbb{E}^{F_*}\big[g_4(U, \theta, \hat{\gamma})\big] = 0. \end{aligned} \tag{7}$$

Finally, the researcher estimates the counterfactual using $\hat{\kappa} = \mathbb{E}^{F_*}[k(U, \hat{\theta}, \hat{\gamma})]$. If k does not depend on U, then the estimated counterfactual is simply $\hat{\kappa} = k(\hat{\theta}, \hat{\gamma})$. In this case, $\hat{\kappa}$ will still depend implicitly on F_* through $\hat{\theta}$.[9]

The researcher's chosen specification F_* is used both for estimation of θ and again when computing the counterfactual. A natural question is: to what extent does the counterfactual depend on the choice of distribution? The main contribution of this paper is to provide a tractable econometric framework for answering this question.

2.2. *Our Approach*

As a sensitivity analysis, we shall relax the researcher's parametric assumption and allow F to vary over nonparametric neighborhoods $\mathcal{N}_\delta$ of F_*, where δ is a measure of neighborhood "size." When we do so, there may be multiple pairs $(\theta, F) \in \Theta \times \mathcal{N}_\delta$ that satisfy (1a)–(1d) but which yield different values of the counterfactual. Our objects of interest

[8]A closed convex set is determined by its support function—see Rockafellar (1970, Section 13).

[9]While this discussion has assumed point identification of θ and κ for sake of exposition, our methods allow structural parameters and counterfactuals to be partially identified.

are the smallest and largest values of the counterfactual over all such (θ, F) pairs:

$$\underline{\kappa}_\delta = \inf_{\theta\in\Theta, F\in\mathcal{N}_\delta} \mathbb{E}^F\big[k(U,\theta,\gamma_0)\big] \quad \text{subject to (1a)–(1d)}, \tag{8}$$

$$\overline{\kappa}_\delta = \sup_{\theta\in\Theta, F\in\mathcal{N}_\delta} \mathbb{E}^F\big[k(U,\theta,\gamma_0)\big] \quad \text{subject to (1a)–(1d)}. \tag{9}$$

By focusing on $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$, our approach naturally accommodates models with partially-identified structural parameters and counterfactuals. Our approach also sidesteps having to compute the identified set of structural parameters.

The optimization problems (8) and (9) are made tractable by a convenient choice of $\mathcal{N}_\delta$. Following Hansen and Sargent (2001) and Maccheroni, Marinacci, and Rustichini (2006), we consider neighborhoods constrained by ϕ-divergence (Csiszár (1967)):

$$\mathcal{N}_\delta = \big\{F\in\mathcal{F} : D_\phi(F\|F_*) \le \delta\big\},$$
$$D_\phi(F\|F_*) = \begin{cases} \displaystyle\int \phi\left(\frac{\mathrm{d}F}{\mathrm{d}F_*}\right)\mathrm{d}F_* & \text{if } F \ll F_*, \\ +\infty & \text{otherwise,}\end{cases} \tag{10}$$

where $\mathcal{F}$ denotes all probability measures on the support[10] $\mathcal{U}$ of U and $F\ll F_*$ denotes absolute continuity of F with respect to F_*. The convex function $\phi:[0,\infty)\to\mathbb{R}_+\cup\{+\infty\}$ penalizes deviations of F from F_*. For example, $\phi(x) = x\log x - x + 1$ corresponds to Kullback–Leibler (KL) divergence, $\phi(x) = \frac{1}{2}(x-1)^2$ corresponds to Pearson χ^2 divergence, and

$$\phi(x) = \frac{x^p - 1 - p(x-1)}{p(p-1)} \quad (p>1),$$

corresponds to L^p divergence. If F_* has positive (Lebesgue) density, then the absolute continuity condition merely rules out F with mass points.

REMARK 2.4: Normalizations and other shape restrictions may be added by augmenting the moment functions $g_1,\dots,g_4$. Examples include: (i) location normalizations, for example, $\mathbb{E}^F[U]=0$ or $\mathbb{E}^F[\mathbb{1}\{U_i\le 0\}-0.5]=0$ for each element U_i of U; (ii) scale normalizations, for example, $\mathbb{E}^F[U_i^2]=1$; (iii) covariance normalizations, for example, $\mathbb{E}^F[UU']=I$; and (iv) smoothness restrictions, for example, $\mathbb{E}^F[\mathbb{1}\{U_i\le a_{k+1}\}-\mathbb{1}\{U_i\le a_k\}]\le C$ for $a_1<\cdots<a_K$ and a positive constant C.

REMARK 2.5: Appendix A.1 in the Supplemental Material shows that shape restrictions including symmetry, exchangeability, and, more generally, invariance under a finite group of transformations, are also easy to impose.

2.3. *Dual Formulation*

We use convex duality to simplify computation of $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$. We start by noting $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ may be written as the solution to two profiled optimization problems:

$$\underline{\kappa}_\delta = \inf_{\theta\in\Theta}\underline{K}_\delta(\theta;\gamma_0,P_0), \qquad \overline{\kappa}_\delta = \sup_{\theta\in\Theta}\overline{K}_\delta(\theta;\gamma_0,P_0),$$

[10]That is, $\mathcal{U}$ is the set of all values that U could conceivably take according to the model, which is possibly larger than the support of the measure F_*.

where the criterion functions $\underline{K}_\delta(\theta; \gamma_0, P_0)$ and $\overline{K}_\delta(\theta; \gamma_0, P_0)$ are, respectively, the infimum and supremum of $\mathbb{E}^F[k(U, \theta, \gamma_0)]$ with respect to $F \in \mathcal{N}_\delta$ subject to the moment conditions (1a)–(1d). In what follows, it is helpful to define the criterion functions at a generic (γ, P). To do so, we say that the moment conditions (1a)–(1d) hold "at (θ, γ, P)" if they hold when γ_0 is replaced by γ and P_0 is replaced by P. Then

$$\underline{K}_\delta(\theta; \gamma, P) = \inf_{F \in \mathcal{N}_\delta} \mathbb{E}^F\big[k(U, \theta, \gamma)\big] \quad \text{subject to (1a)–(1d) holding at } (\theta, \gamma, P), \tag{11}$$

$$\overline{K}_\delta(\theta; \gamma, P) = \sup_{F \in \mathcal{N}_\delta} \mathbb{E}^F\big[k(U, \theta, \gamma)\big] \quad \text{subject to (1a)–(1d) holding at } (\theta, \gamma, P), \tag{12}$$

with the understanding that $\underline{K}_\delta(\theta; \gamma, P) = +\infty$ and $\overline{K}_\delta(\theta; \gamma, P) = -\infty$ if there does not exist a distribution in $\mathcal{N}_\delta$ for which the moment conditions (1a)–(1d) hold at (θ, γ, P).

We first impose some mild regularity conditions on F_*, ϕ, and the moment functions to justify the dual formulation. Similar conditions are used in generalized empirical likelihood estimation (see, e.g., Komunjer and Ragusa (2016)). Let Φ_0 denote the set of all $\phi : [0, \infty) \to \mathbb{R} \cup \{+\infty\}$ such that ϕ is continuously differentiable on $(0, +\infty)$ and strictly convex, with $\phi(1) = \phi'(1) = 0$, $\phi(0) < +\infty$, $\lim_{x \downarrow 0} \phi'(x) < 0$, $\lim_{x \to +\infty} \phi(x)/x = +\infty$, $\lim_{x \to +\infty} \phi'(x) > 0$, and $\lim_{x \to +\infty} x\phi'(x)/\phi(x) < +\infty$. The functions inducing KL, χ^2, and L^p divergence all belong to Φ_0.

Let $\phi^\star(x) = \sup_{t \geq 0 : \phi(t) < +\infty}(tx - \phi(t))$ denote the convex conjugate of $\phi \in \Phi_0$ and let $\psi(x) = \phi^\star(x) - x$. Define $\mathcal{E} = \{f : \mathcal{U} \to \mathbb{R}$ for which $\mathbb{E}^{F_*}[\psi(c|f(U)|)] < \infty$ for all $c > 0\}$. The class $\mathcal{E}$ is an Orlicz class of functions (see Appendix F of Christensen and Connault (2022) for details). For example,

$$\mathcal{E} = \big\{f : \mathcal{U} \to \mathbb{R} : \mathbb{E}^{F_*}\big[e^{c|f(U)|}\big] < \infty \text{ for all } c > 0\big\} \quad \text{for KL divergence,}$$

$$\mathcal{E} = \big\{f : \mathcal{U} \to \mathbb{R} : \mathbb{E}^{F_*}\big[f(U)^2\big] < \infty\big\} \quad \text{for } \chi^2 \text{ divergence, and}$$

$$\mathcal{E} = \big\{f : \mathcal{U} \to \mathbb{R} : \mathbb{E}^{F_*}\big[|f(U)|^q\big] < \infty\big\} \quad \text{for } L^p \text{ divergence } \big(p^{-1} + q^{-1} = 1\big).$$

Let $g = (g_1, g_2, g_3, g_4)$ denote the vector formed by stacking each of the moment functions from (1a)–(1d). Our key regularity condition is the following:

ASSUMPTION Φ: (i) $\phi \in \Phi_0$.
(ii) *$k(\cdot, \theta, \gamma)$ and each entry of $g(\cdot, \theta, \gamma)$ belong to $\mathcal{E}$ for each $\theta \in \Theta$ and $\gamma \in \Gamma$.*

For KL divergence, the class $\mathcal{E}$ contains bounded functions (e.g., indicator functions) and functions that are additively separable in U provided F_* has tails that decay faster than exponentially (e.g., Gaussian but not Gumbel). Assumption Φ therefore fails for KL divergence in Examples 2.1 and 2.3, but holds for χ^2 or L^p divergence as these only require finite second or qth moments, respectively.

Let $d = \sum_{i=1}^4 d_i$ where d_i is the dimension of g_i, let $\Lambda = \mathbb{R}_+^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}_+^{d_3} \times \mathbb{R}^{d_4}$, and let λ_{12} denote the first $d_1 + d_2$ elements of $\lambda \in \Lambda$. A derivation of the following criterion functions is presented in Appendix G.2 of Christensen and Connault (2022).

PROPOSITION 2.1: *Suppose that Assumption* Φ *holds. Then the criterion functions* (11) *and* (12) *may be restated as*

$$\underline{K}_\delta(\theta;\gamma,P) = \sup_{\eta>0,\zeta\in\mathbb{R},\lambda\in\Lambda} -\eta\mathbb{E}^{F_*}\left[\phi^\star\left(\frac{k(U,\theta,\gamma)+\zeta+\lambda' g(U,\theta,\gamma)}{-\eta}\right)\right] - \eta\delta - \zeta - \lambda_{12}' P, \tag{13}$$

$$\overline{K}_\delta(\theta;\gamma,P) = \inf_{\eta>0,\zeta\in\mathbb{R},\lambda\in\Lambda} \eta\mathbb{E}^{F_*}\left[\phi^\star\left(\frac{k(U,\theta,\gamma)-\zeta-\lambda' g(U,\theta,\gamma)}{\eta}\right)\right] + \eta\delta + \zeta + \lambda_{12}' P. \tag{14}$$

Moreover, the value of (13) *is* $+\infty$ (*equivalently, the value of* (14) *is* $-\infty$) *if and only if there is no distribution in* $\mathcal{N}_\delta$ *under which* (1a)–(1d) *holds at* (θ,γ,P).

REMARK 2.6: Problems (13) and (14) are convex in (η,ζ,λ). The parameter η is the Lagrange multiplier for the constraint $D_\phi(F\|F_*)\le\delta$. Similarly, λ collects the Lagrange multipliers for the moment (in)equalities (1a)–(1d). These multipliers are non-negative if they correspond to inequality restrictions and unconstrained otherwise. Finally, ζ is the Lagrange multiplier for the constraint $\int \mathrm{d}F = 1$, which ensures that the optimization is over probability measures.

Problems (13) and (14) simplify in some special cases. For KL neighborhoods, $\phi^\star(x) = e^x - 1$ and the multiplier ζ has a closed-form solution, leading to

$$\underline{K}_\delta(\theta;\gamma,P) = \sup_{\eta>0,\lambda\in\Lambda} -\eta\log\mathbb{E}^{F_*}\left[e^{-(k(U,\theta,\gamma)+\lambda' g(U,\theta,\gamma))/\eta}\right] - \eta\delta - \lambda_{12}' P,$$

$$\overline{K}_\delta(\theta;\gamma,P) = \inf_{\eta>0,\lambda\in\Lambda} \eta\log\mathbb{E}^{F_*}\left[e^{(k(U,\theta,\gamma)-\lambda' g(U,\theta,\gamma))/\eta}\right] + \eta\delta + \lambda_{12}' P.$$

Another special case is when $k(u,\theta,\gamma)$ does not depend on u. To analyze this case, consider

$$\Delta(\theta;\gamma,P) := \inf_F D_\phi(F\|F_*) \quad \text{subject to (1a)–(1d) holding at } (\theta,\gamma,P). \tag{15}$$

The value $\Delta(\theta;\gamma,P)$ is the minimum ϕ-divergence between F_* and a distribution F for which the moment conditions hold at (θ,γ,P). Proposition G.2 of Christensen and Connault (2022) shows that $\Delta(\theta;\gamma,P)$ has an equivalent dual formulation:

$$\Delta(\theta;\gamma,P) = \sup_{\zeta\in\mathbb{R},\lambda\in\Lambda} -\mathbb{E}^{F_*}\left[\phi^\star\left(-\zeta - \lambda' g(U,\theta,\gamma)\right)\right] - \zeta - \lambda_{12}' P. \tag{16}$$

For KL divergence, ζ may be solved for in closed form and problem (16) simplifies to

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

If k does not depend on u, then by a change of variables[11] we may restate problems (13) and (14) as

$$\underline{K}_\delta(\theta;\gamma,P)=\begin{cases}k(\theta,\gamma),\\+\infty,\end{cases}\qquad \overline{K}_\delta(\theta;\gamma,P)=\begin{cases}k(\theta,\gamma) & \text{if } \Delta(\theta;\gamma,P)\le\delta,\\-\infty & \text{if } \Delta(\theta;\gamma,P)>\delta.\end{cases}\tag{17}$$

An important feature of our approach is that the optimization problems (13), (14), and (16) are convex and their dimension does not increase with δ. This feature is not shared by other seemingly natural approaches to flexibly model F, such as mixtures or other finite-dimensional sieves. As we show in Section 2.5, our procedure may be used to approximate sharp nonparametric bounds on counterfactuals by taking δ to be large but finite.

2.4. *Estimation*

We now propose simple estimators of the bounds $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ based on "plugging in" consistent estimators $(\hat{P},\hat{\gamma})$ of (P_0,γ_0). Estimators $\hat{\underline{\kappa}}_\delta$ and $\hat{\overline{\kappa}}_\delta$ are computed by optimizing criterion functions with respect to θ:

$$\hat{\underline{\kappa}}_\delta=\inf_{\theta\in\Theta}\hat{\underline{K}}_\delta(\theta),\qquad \hat{\overline{\kappa}}_\delta=\sup_{\theta\in\Theta}\hat{\overline{K}}_\delta(\theta),$$

where

$$\hat{\underline{K}}_\delta(\theta)=\begin{cases}\underline{K}_\delta(\theta;\hat{\gamma},\hat{P}),\\+\infty,\end{cases}\qquad \hat{\overline{K}}_\delta(\theta)=\begin{cases}\overline{K}_\delta(\theta;\hat{\gamma},\hat{P}) & \text{if } \Delta(\theta;\hat{\gamma},\hat{P})<\delta,\\-\infty & \text{if } \Delta(\theta;\hat{\gamma},\hat{P})\ge\delta,\end{cases}$$

and $\underline{K}_\delta(\theta;\hat{\gamma},\hat{P})$, $\overline{K}_\delta(\theta;\hat{\gamma},\hat{P})$, and $\Delta(\theta;\hat{\gamma},\hat{P})$ are the criterion functions (13), (14), and (16) evaluated at $(\hat{\gamma},\hat{P})$. If $k(u,\theta,\gamma)=k(\theta,\gamma)$, then we simply have

$$\hat{\underline{K}}_\delta(\theta)=\begin{cases}k(\theta,\hat{\gamma}),\\+\infty,\end{cases}\qquad \hat{\overline{K}}_\delta(\theta)=\begin{cases}k(\theta,\hat{\gamma}) & \text{if } \Delta(\theta;\hat{\gamma},\hat{P})<\delta,\\-\infty & \text{if } \Delta(\theta;\hat{\gamma},\hat{P})\ge\delta.\end{cases}$$

In Section 6.1, we establish consistency of $\hat{\underline{\kappa}}_\delta$ and $\hat{\overline{\kappa}}_\delta$ and derive their asymptotic distribution.

2.5. *Nonparametric Bounds on Counterfactuals*

We define the (nonparametric) identified set of counterfactuals as

$$\mathcal{K}=\left\{\mathbb{E}^F\left[k(U,\theta,\gamma_0)\right]:\text{(1a)–(1d) holds for some } \theta\in\Theta \text{ and } F\in\mathcal{F}_\theta\right\},$$

where $\mathcal{F}_\theta=\{F\in\mathcal{F}:\mathbb{E}^F[g(U,\theta,\gamma_0)]\text{ is finite and } F\ll\mu\}$ denotes all distributions on $\mathcal{U}$ that are absolutely continuous with respect to a σ-finite dominating measure μ and for which the moments in (1a)–(1d) are finite at θ. We impose existence of a density with respect to μ as it is often a structural assumption used, for example, to avoid ties in CCPs

[11]Substitute $\eta\zeta-k(\theta,\gamma)$ in place of ζ in (13) and $\eta\zeta+k(\theta,\gamma)$ in place of ζ in (14), then substitute $\eta\lambda$ in place of λ in both (13) and (14).

or to establish existence of equilibria. The main result of this section shows that $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ approach the sharp nonparametric bounds $\inf \mathcal{K}$ and $\sup \mathcal{K}$ as δ becomes large.

We first introduce some additional regularity conditions. Say k is "μ-essentially bounded" if $|k(\cdot, \theta, \gamma_0)|$ has finite μ-essential supremum[12] for each $\theta \in \Theta$. This holds trivially if k is bounded (e.g., counterfactual CCPs in Examples 2.2 and 2.3 and change in average welfare in Example 2.3). Models with unbounded k may be reparameterized (as a proof device) by setting $\tilde{\theta} = (\theta, \kappa)$, appending $k(U, \theta, \gamma_0) - \kappa$ as an element of g_4, and setting $k(U, \tilde{\theta}, \gamma_0) = \kappa$.

We also require a constraint qualification condition. This is a sufficient condition for establishing equivalence of "nonparametric" primal and dual problems in Appendix B in the Supplemental Material, which is an intermediate step in the proof of the following result. Let 0_{d_i} denote a $d_i \times 1$ vector of zeros, $\mathcal{C} = \mathbb{R}_+^{d_1} \times \{0_{d_2}\} \times \mathbb{R}_+^{d_3} \times \{0_{d_4}\}$, $\mathcal{G}(\theta, \gamma) = \{\mathbb{E}^F[g(U, \theta, \gamma)] : F \in \mathcal{N}_\infty\}$ where $\mathcal{N}_\infty = \{F : D_\phi(F \| F_*) < \infty\}$, and $\vec{P} = (P, 0_{d_3+d_4})$. For $A, B \subseteq \mathbb{R}^d$, we let $\mathrm{ri}(A)$ denote the relative interior of A and $A + B = \{a + b : a \in A, b \in B\}$.

DEFINITION 2.1: *Condition* S holds at (θ, γ, P) if $\vec{P} \in \mathrm{ri}(\mathcal{G}(\theta, \gamma) + \mathcal{C})$.

Using relative interior instead of interior allows for moment functions that are collinear at some θ (i.e., some moments are redundant). To give some intuition, consider moment equality models. Condition S requires that (1a)–(1d) holds at (θ, γ, P) under some $F \in \mathcal{N}_\delta$ that is "interior" to $\mathcal{N}_\infty$, in the sense that one can perturb the (non-redundant) moments in any direction by perturbing F. For moment inequality models, Condition S also requires that there is $F \in \mathcal{N}_\infty$ under which all moment inequalities hold strictly at (θ, γ, P).

Let $\Theta_I = \{\theta \in \Theta : \text{(1a)–(1d) holds for some } F \in \mathcal{F}_\theta\}$ denote the (nonparametric) identified set for θ. Define the "nonparametric" objective function

$$\underline{K}_{\mathrm{np}}(\theta; \gamma, P) = \inf_{F \in \mathcal{F}_\theta} \mathbb{E}^F\big[k(U, \theta, \gamma)\big] \quad \text{subject to (1a)–(1d) holding at } (\theta, \gamma, P), \tag{18}$$

with the understanding that $\underline{K}_{\mathrm{np}}(\theta; \gamma, P) = +\infty$ if the infimum runs over an empty set. Let $\overline{K}_{\mathrm{np}}(\theta; \gamma, P)$ denote the analogous supremum. Evidently,

$$\inf \mathcal{K} = \inf_{\theta \in \Theta} \underline{K}_{\mathrm{np}}(\theta; \gamma_0, P_0) \quad \text{and} \quad \sup \mathcal{K} = \sup_{\theta \in \Theta} \overline{K}_{\mathrm{np}}(\theta; \gamma_0, P_0).$$

DEFINITION 2.2: Θ_I is *S-regular* if, for all $\epsilon > 0$, there exist $\underline{\theta}, \overline{\theta} \in \Theta_I$ such that Condition S holds at $(\underline{\theta}, \gamma_0, P_0)$ and $(\overline{\theta}, \gamma_0, P_0)$, $\underline{K}_{\mathrm{np}}(\underline{\theta}; \gamma_0, P_0) < \inf \mathcal{K} + \epsilon$, and $\overline{K}_{\mathrm{np}}(\overline{\theta}; \gamma_0, P_0) > \sup \mathcal{K} - \epsilon$.

Intuitively, S-regularity requires that the values the counterfactual takes at "boundary" points of Θ_I (i.e., at which Condition S fails) are not materially more extreme than values it can take at points "inside" Θ_I (i.e., at which Condition S holds). This condition can be verified under more primitive continuity conditions on k and g. A sufficient (but not necessary) condition for S-regularity is that Condition S holds at (θ, γ_0, P_0) for all $\theta \in \Theta_I$.

[12] The μ-essential supremum of a function f is denoted $\mu\text{-ess}\sup f$ and is the smallest value c for which $\mu(\{u : f(u) > c\}) = 0$. The μ-essential infimum, denoted $\mu\text{-ess}\inf$, is defined analogously.

THEOREM 2.1: *Suppose that Assumption* Φ *holds*, k *is* μ*-essentially bounded*, Θ_I *is S-regular, and* μ *and* F_* *are mutually absolutely continuous*. *Then*

$$\lim_{\delta\to\infty} \underline{\kappa}_\delta = \inf \mathcal{K}, \qquad \lim_{\delta\to\infty} \overline{\kappa}_\delta = \sup \mathcal{K}.$$

Theorem 2.1 shows that our procedure can be used to approximate the sharp nonparametric bounds $\inf \mathcal{K}$ and $\sup \mathcal{K}$ by setting δ to be large but finite. If μ is Lebesgue measure—which it often is in applications—then the mutual absolute continuity condition in Theorem 2.1 is satisfied whenever F_* has strictly positive density over $\mathcal{U}$.

REMARK 2.7: Appendix B presents the dual forms of $\underline{K}_{\text{np}}$ and $\overline{K}_{\text{np}}$. Unlike $\underline{K}_\delta$ and $\overline{K}_\delta$, the duals of $\underline{K}_{\text{np}}$ and $\overline{K}_{\text{np}}$ are min-max and max-min problems which involve an inner optimization over u. These problems may be computationally challenging, especially when u is multivariate. Comparing Proposition 2.1 with the duals in Appendix B, we see that setting $\delta < \infty$ replaces a "hard-max" (an optimization over u) with a "soft-max" (a convex expectation). In this respect, adding the constraint $F \in \mathcal{N}_\delta$ may be viewed as a regularization of the nonparametric objective functions, similar to the use of entropic penalization to regularize objective functions in optimal transport problems—see, for example, Cuturi (2013). Smaller values of δ impose a stronger regularization.

Theorem 2.1 is silent on the issue of how large δ needs to be so that $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ are close to the nonparametric bounds. While this is model- and counterfactual-specific, the following toy example suggests that relatively small values of δ may suffice in some problems where the counterfactual is a choice probability.

EXAMPLE 2.4: Consider the problem

$$\overline{\kappa}_\delta = \sup_{\theta\in\mathbb{R}, F\in\mathcal{N}_\delta} \mathbb{E}^F\big[\mathbb{1}\{U \le \theta\}\big] \quad \text{subject to } \mathbb{E}^F[U-\theta] = 0,$$

where $\mathcal{N}_\delta$ is defined by KL divergence and F_* is the $N(0,1)$ distribution. When $F = F_*$, the only solution to $\mathbb{E}^F[U-\theta]=0$ is $\theta=0$. Therefore, the value of the counterfactual under F_* is $\mathbb{E}^{F_*}[\mathbb{1}\{U\le 0\}] = \frac{1}{2}$ whereas $\sup \mathcal{K} = 1$. The large-δ approximation $\overline{\kappa}_\delta = 1 - 2\pi e^{-2\delta-1}(1+o(1))$ is derived in Appendix H of Christensen and Connault (2022). By symmetry, $\underline{\kappa}_\delta = 2\pi e^{-2\delta-1}(1+o(1))$ and $\inf \mathcal{K} = 0$. Therefore, in this example, $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ converge rapidly to $\inf \mathcal{K}$ and $\sup \mathcal{K}$ as δ increases.

More generally, suppose the dual problems (13) and (14) have unique solutions $\underline{\eta}$ and $\overline{\eta}$ for η, where the optimization is performed over $\eta \ge 0$.[13] Under appropriate regularity conditions (see, e.g., Milgrom and Segal (2002)), it follows that

$$\frac{\partial \underline{K}_\delta(\theta;\gamma,P)}{\partial\delta} = -\underline{\eta}, \qquad \frac{\partial \overline{K}_\delta(\theta;\gamma,P)}{\partial\delta} = \overline{\eta}.$$

One can therefore infer from $\underline{\eta}$ and $\overline{\eta}$ the extent to which, if at all, the bounds at any fixed θ would widen further if δ was increased.

[13]Optimizing over $\eta \ge 0$ rather than $\eta > 0$ does not affect the optimal value—see Proposition G.1 of Christensen and Connault (2022).

3. PRACTICAL CONSIDERATIONS

We now discuss practical details for implementing our procedure. Section 3.1 discusses computational methods, Section 3.2 presents our MPEC approach, and Section 3.3 discusses methods for dealing with over-identified models.

3.1. *Computation*

There are three aspects to computation: (i) computing the expectations with respect to F_* in the objective functions, (ii) solving the inner optimization problems over Lagrange multipliers, and (iii) solving the outer optimization problems over θ.

The expectations in the objective functions (13), (14), and (16) are available in closed form for certain settings,[14] in which case the dimension of u does not play a role in the computational complexity of our procedure. Otherwise, the expectations will need to be computed numerically. If so, the dimension of u will play a role in terms of determining how many quadrature points or Monte Carlo draws are needed to control numerical approximation error. In the empirical applications, we used a randomized quasi-Monte Carlo approach based on scrambled Halton sequences as in Owen (2017).

The inner optimization with respect to Lagrange multipliers can be solved rapidly: it is convex and gradients and Hessians are available in closed form. The envelope theorem can be used to derive gradients for the outer optimization when k and g are differentiable in θ.[15] Our procedures were all implemented in Julia with the inner and outer optimizations solved using Knitro. A general-purpose implementation of our methods in Julia is provided in the Supplemental Material.

As with parameter estimation in nonlinear structural models, the outer optimization with respect to θ is typically non-convex. In applications, we used an iterative multi-start procedure in an attempt to converge to global optima. Computation times are reported in the applications below.

3.2. *MPEC Approach*

We now describe and formally justify an MPEC version of our procedure in the spirit of Su and Judd (2012). This approach simplifies computation in models with endogenous parameters defined by equilibrium conditions (e.g., value functions defined by Bellman equations), resulting in significant computational gains for DDC models in particular.

Suppose $\theta = (\theta_s, \theta_e)$ and $g_4 = (g_{4s}, g_{4e})$ where θ_s are "deep" structural parameters and θ_e are "endogenous" parameters that are defined implicitly by g_{4e}. That is, for any (θ_s, γ, F), the parameter $\theta_e = \theta_e(\theta_s, \gamma, F)$ solves

$$\mathbb{E}^F\big[g_{4e}\big(U, (\theta_s, \theta_e), \gamma\big)\big] = 0.$$

For instance, in Example 2.3 we have $\theta_s = \theta_\pi$ or (θ_π, β), while $\theta_e = (v, \tilde{v})$ collects the value functions in the baseline model and counterfactual, and g_{4e} collects the functions representing the corresponding Bellman equations, as in display (6). Although our procedure can be implemented as described in Section 2, that implementation does not make use of the fact that θ_e is defined implicitly by g_{4e}.

[14] An earlier draft derived closed-form expressions for a discrete game of complete information with Gaussian payoff shocks and KL neighborhoods—see https://arxiv.org/abs/1904.00989v2.

[15] In practice, we smoothed any non-smooth moments and used automatic differentiation to compute derivatives with respect to θ if these were not easily available analytically.

To leverage this structure, consider the subset of moments conditions excluding g_{4e}:

$$\begin{aligned} &\mathbb{E}^F\big[g_1(U,\theta,\gamma_0)\big] \le P_{10}, \qquad && \mathbb{E}^F\big[g_2(U,\theta,\gamma_0)\big] = P_{20}, \\ &\mathbb{E}^F\big[g_3(U,\theta,\gamma_0)\big] \le 0, \qquad && \mathbb{E}^F\big[g_{4s}(U,\theta,\gamma_0)\big] = 0, \end{aligned} \tag{19}$$

and define criterion functions using these only:

$$\underline{K}^s_\delta(\theta;\gamma,P) = \inf_{F\in\mathcal{N}_\delta} \mathbb{E}^F\big[k(U,\theta,\gamma)\big] \quad \text{subject to (19) holding at } (\theta,\gamma,P), \tag{20}$$

$$\overline{K}^s_\delta(\theta;\gamma,P) = \sup_{F\in\mathcal{N}_\delta} \mathbb{E}^F\big[k(U,\theta,\gamma)\big] \quad \text{subject to (19) holding at } (\theta,\gamma,P). \tag{21}$$

Under the conditions of Proposition 2.1, these criterion functions may be restated as

$$\begin{aligned} \underline{K}^s_\delta(\theta;\gamma,P) = \sup_{\eta>0,\zeta\in\mathbb{R},\lambda\in\Lambda_s} & -\eta\mathbb{E}^{F_*}\left[\phi^\star\left(\frac{k(U,\theta,\gamma)+\zeta+\lambda' g_s(U,\theta,\gamma)}{-\eta}\right)\right] \\ & -\eta\delta - \zeta - \lambda_{12}' P, \end{aligned} \tag{22}$$

$$\begin{aligned} \overline{K}^s_\delta(\theta;\gamma,P) = \inf_{\eta>0,\zeta\in\mathbb{R},\lambda\in\Lambda_s} & \eta\mathbb{E}^{F_*}\left[\phi^\star\left(\frac{k(U,\theta,\gamma)-\zeta-\lambda' g_s(U,\theta,\gamma)}{\eta}\right)\right] \\ & +\eta\delta + \zeta + \lambda_{12}' P, \end{aligned} \tag{23}$$

with $g_s = (g_1, g_2, g_3, g_{4s})$ and $\Lambda_s = \mathbb{R}^{d_1}_+ \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3}_+ \times \mathbb{R}^{d_{4s}}$ with $d_{4s} = \dim(g_{4s})$. Problems (22) and (23) simplify analogously to (17) when k does not depend on u, with the minimum divergence problem Δ defined using g_s in place of g.

In our MPEC approach, the criterion functions (22) and (23) are optimized with respect to θ, with the remaining moment conditions involving g_{4e} appended as constraints. Importantly, these constraints are evaluated under the "least favorable" distributions $\underline{F}_{\delta,\theta}$ and $\overline{F}_{\delta,\theta}$ that solve problems (20) and (21), respectively. The following proposition formally justifies this approach.

PROPOSITION 3.1: *Suppose that Assumption Φ holds. Then the problems*

$$\inf_{\theta\in\Theta} \underline{K}_\delta(\theta;\gamma,P)$$

and

$$\inf_{\theta\in\Theta} \underline{K}^s_\delta(\theta;\gamma,P) \quad \textit{subject to } \mathbb{E}^{\underline{F}_{\delta,\theta}}\big[g_{4e}(U,\theta,\gamma)\big] = 0$$

have the same value. An analogous result holds for the upper bound.

To implement our MPEC approach, note that the expectations in the constraints may be expressed in terms of changes of measure. Let $\underline{m}_{\delta,\theta} = \mathrm{d}\underline{F}_{\delta,\theta}/\mathrm{d}F_*$ and $\overline{m}_{\delta,\theta} = \mathrm{d}\overline{F}_{\delta,\theta}/\mathrm{d}F_*$ so that

$$\mathbb{E}^{\underline{F}_{\delta,\theta}}[\cdot] = \mathbb{E}^{F_*}\big[\underline{m}_{\delta,\theta}(U)\cdot\big], \qquad \mathbb{E}^{\overline{F}_{\delta,\theta}}[\cdot] = \mathbb{E}^{F_*}\big[\overline{m}_{\delta,\theta}(U)\cdot\big].$$

If k depends on u, then we construct $\underline{m}_{\delta,\theta}$ and $\overline{m}_{\delta,\theta}$ from solutions to (22) and (23), say $(\underline{\eta}, \underline{\zeta}, \underline{\lambda})$ and $(\overline{\eta}, \overline{\zeta}, \overline{\lambda})$ (these solutions exist under the regularity conditions below). If

$\underline{\eta} > 0$, then the distribution solving (20) is unique and is induced by the change of measure

$$\underline{m}_{\delta,\theta}(u) = \dot{\phi}^\star\left(\frac{k(u,\theta,\gamma) + \underline{\zeta} + \underline{\lambda}' g_s(u,\theta,\gamma)}{-\underline{\eta}}\right), \tag{24}$$

where $\dot{\phi}^\star(x) = \frac{d\phi^\star(x)}{dx}$. The function $\overline{m}_{\delta,\theta}(u)$ is constructed similarly, replacing $(\underline{\eta}, \underline{\zeta}, \underline{\lambda})$ in (24) by $(-\overline{\eta}, -\overline{\zeta}, -\overline{\lambda})$. For KL divergence, the change of measure simplifies to

$$\underline{m}_{\delta,\theta}(u) = \frac{e^{(k(u,\theta,\gamma)+\underline{\lambda}' g_s(u,\theta,\gamma))/-\underline{\eta}}}{\mathbb{E}^{F_*}\left[e^{(k(u,\theta,\gamma)+\underline{\lambda}' g_s(u,\theta,\gamma))/-\underline{\eta}}\right]},$$

and similarly for $\underline{m}_{\delta,\theta}(u)$.

If $\underline{\eta} = 0$, then there may be multiple minimizing distributions. As shown in the proof of Proposition 3.2, each such distribution must be supported on

$$\underline{A}_{\delta,\theta} := \left\{u : k(u,\theta,\gamma) + \underline{\lambda}' g_s(u,\theta,\gamma) = F_*\text{-ess}\inf\left(k(\cdot,\theta,\gamma) + \underline{\lambda}' g_s(\cdot,\theta,\gamma)\right)\right\}.$$

Note $F_*(\underline{A}_{\delta,\theta}) > 0$ is required for $\underline{\eta} = 0$ to be a solution. Otherwise, any distribution supported on $\underline{A}_{\delta,\theta}$ is not absolutely continuous with respect to F_* and is therefore not in $\mathcal{N}_\delta$. If $\underline{\eta} = 0$ and $F_*(\underline{A}_{\delta,\theta}) > 0$, then we construct $\underline{m}_{\delta,\theta}$ by restricting F_* to $\underline{A}_{\delta,\theta}$ and rescaling:

$$\underline{m}_{\delta,\theta}(u) = \mathbb{1}\{u \in \underline{A}_{\delta,\theta}\}/F_*(\underline{A}_{\delta,\theta}).$$

The function $\overline{m}_{\delta,\theta}(u)$ is constructed analogously, replacing $\underline{\lambda}$ with $-\overline{\lambda}$ and the set $\underline{A}_{\delta,\theta}$ with $\overline{A}_{\delta,\theta} = \{u : k(u,\theta,\gamma) - \overline{\lambda}' g_s(u,\theta,\gamma) = F_*\text{-ess}\sup(k(\cdot,\theta,\gamma) - \overline{\lambda}' g_s(\cdot,\theta,\gamma))\}$.

If k does not depend on u, then $\underline{m}_{\delta,\theta}$ and $\overline{m}_{\delta,\theta}$ are constructed from solutions to a version of problem (16) with g_s in place of g. Under the regularity conditions below, this program has a solution, say $(\underline{\zeta}, \underline{\lambda})$. In this case, we define

$$\underline{m}_{\delta,\theta}(u) = \overline{m}_{\delta,\theta}(u) = \dot{\phi}^\star\left(-\underline{\zeta} - \underline{\lambda}' g_s(u,\theta,\gamma)\right). \tag{25}$$

For KL divergence, the change of measure simplifies to

$$\underline{m}_{\delta,\theta}(u) = \overline{m}_{\delta,\theta}(u) = \frac{e^{-\underline{\lambda}' g_s(u,\theta,\gamma)}}{\mathbb{E}^{F_*}\left[e^{-\underline{\lambda}' g_s(u,\theta,\gamma)}\right]}.$$

PROPOSITION 3.2: *Suppose that Assumption Φ holds, Condition S holds at (θ,γ,P), and there exists a distribution F with $D(F\|F_*) < \delta$ under which (19) holds at (θ,γ,P). Then the distributions $\underline{F}_{\delta,\theta}$ and $\overline{F}_{\delta,\theta}$ induced by $\underline{m}_{\delta,\theta}$ and $\overline{m}_{\delta,\theta}$ solve (20) and (21), respectively.*

Example. We consider a numerical example for the DDC model of Rust (1987) based on the parameterization in Section 5.4 of Norets and Tang (2014). The counterfactual they considered is a hypothetical change in the law of motion of the state. We follow these papers and use state-space of dimension 90. As $|\mathcal{S}| = 90$ and $\mathcal{D}_0 = \{0, 1\}$, there are 90 functions in g_2 representing the observed CCPs. There are another 180 functions in g_{4e} representing the Bellman equations in the baseline model and counterfactual across states. We also impose the normalization $\mathbb{E}^F[U_d] = 0$ for $d = 0, 1$. Hence, $g_{4s}(U,\theta,\gamma) = (U_0, U_1)$. Our MPEC approach has 92 moments in the inner optimization (90 for CCPs and two

TABLE II

COMPUTATION TIMES (IN SECONDS) FOR THE INNER PROBLEMS.

Implementation	Objective			
	$\overline{K}_{0.01}$	$\overline{K}_{0.10}$	$\overline{K}_{1.00}$	Δ
MPEC (92 moments)	0.207	0.232	0.256	0.108
Full (272 moments)	4.317	12.978	43.699	3.365

Note: Expectations are computed using 50,000 scrambled Halton draws. Computations are performed in Julia v1.6.4 and Knitro v12.4.0 on a 2.7 GHz MacBook Pro with 16 GB memory.

mean-zero normalizations on the shocks) with the remaining 180 moments representing the Bellman equations appended as constraints. The full approach uses all 272 moments in the inner optimization.

Table II reports computation times for the inner optimization problems (14) and (23) (denoted $\overline{K}_\delta$) for maximizing the counterfactual CCP in the highest mileage state.[16] We also report times for solving the minimum divergence problem (16) (denoted Δ) using the full set of moment functions g and its MPEC analogue using g_s. Neighborhoods are constrained by a hybrid of KL and χ^2 divergence as in the empirical applications—see Section 5. As can be seen, the inner optimization problems are solved at least 20 times faster for the MPEC implementation, with the relative efficiency increasing in δ.

3.3. *Over-Identification*

In over-identified models (i.e., where the number of moment conditions d exceeds the dimension d_θ of θ), there might not exist $\theta \in \Theta$ for which the sample moment conditions (7) hold under F_*. We propose two methods for handling over-identified models.

First, one may compute the smallest value of δ for which there exists $F \in \mathcal{N}_\delta$ consistent with the sample moment conditions (7) by solving the optimization problem

$$\hat{\delta} = \inf_{\theta \in \Theta} \Delta(\theta; \hat{\gamma}, \hat{P}).$$

The interval $[\hat{\underline{\kappa}}_\delta, \hat{\overline{\kappa}}_\delta]$ will be nonempty for $\delta > \hat{\delta}$. If the model is correctly specified under F_*,[17] then $\hat{\delta}$ will converge in probability to zero under the conditions of Theorem 6.1. In this case, the interval $[\hat{\underline{\kappa}}_\delta, \hat{\overline{\kappa}}_\delta]$ will be nonempty with probability approaching 1 for each fixed $\delta > 0$.

It is also possible that $\hat{\delta} = +\infty$ in correctly specified but over-identified models when $\hat{P}$ is incompatible with certain model restrictions. For instance, CCPs are often estimated nonparametrically using empirical choice frequencies. If some choices are not observed in the data, then the estimated CCPs will be zero even though model-implied CCPs are strictly positive.

This issue can be circumvented in models defined by equality restrictions only (hence $P_0 \equiv P_{20}$) using the following two-step approach. First, compute a preliminary estima-

[16]The times in Table II are based on initializing the solver at $\eta = 1$, $\zeta = 0$, and $\lambda = 0$. When embedded in the outer optimization over θ, computation times for the inner problem are reduced significantly by using a warm start that initializes at the solution to the inner problem at the previous value of θ.

[17]Neither our theoretical results developed in Section 2 nor the estimation and inference results in Section 6 require correct specification of the model under F_*.

tor $\tilde{\theta}$ of θ based on (7). Then, set $\hat{P} = \mathbb{E}^{F_*}[g_2(U, \tilde{\theta}, \hat{\gamma})]$. This second-step estimator $\hat{P}$ is compatible with the model by construction, thereby ensuring that the interval $[\hat{\underline{\kappa}}_\delta, \hat{\overline{\kappa}}_\delta]$ is nonempty for each $\delta > 0$. The estimator $\hat{P}$ will be consistent and asymptotically normal under mild regularity conditions provided the model is correctly specified under F_*, so the consistency and inference results developed in Section 6 will also apply.

4. INTERPRETING THE NEIGHBORHOOD SIZE

This section presents some theoretical results and practical methods to help interpret the neighborhood size δ. Sections 4.1 and 4.4 discuss properties of ϕ-divergences and their implications for interpreting δ. Section 4.2 shows how to construct the "least favorable" distributions that minimize or maximize the counterfactual. Section 4.3 gives a practical, model-based metric for interpreting δ.

4.1. *Invariance*

A defining property of ϕ-divergences is their invariance to invertible transformations. That is, if T is an invertible transformation and G and G_* denote the distributions of $T(U)$ when $U \sim F$ and $U \sim F_*$, respectively, then $D_\phi(F \| F_*) = D_\phi(G \| G_*)$.[18] An important consequence of invariance is that δ has the same interpretation under a change in units. For instance, if one researcher writes a model in terms of dollars with $U \sim F_*$ and another researcher uses thousands of dollars with $U \sim G_*$ for $G_*(u) = F_*(10^{-3}u)$, then F is in $\mathcal{N}_\delta$ if and only if its rescaled counterpart G is in a δ-neighborhood of G_*. A second consequence is that neighborhood size is invariant under invertible location and scale transformations of F_* (e.g., $N(\mu, \Sigma)$ versus $N(0, I)$).

4.2. *Least Favorable Distributions*

A useful feature of our approach is that the "least favorable" distributions (LFDs) that attain the smallest or largest values of the counterfactual may easily be recovered. To help interpret δ, one may plot the LFDs and compute other quantities of interest (e.g., correlations or welfare measures) under them.

Section 3.2 describes how to construct LFDs when our MPEC approach is used. LFDs for our full (i.e., non-MPEC) approach are a special case with $g_4 = g_{4s}$. To briefly summarize, consider the LFD $\underline{F}_{\delta,\theta}$ solving the minimization problem (11). First suppose that k depends on u. Let $(\underline{\eta}, \underline{\zeta}, \underline{\lambda})$ solve problem (13). If $\underline{\eta} > 0$, then $\underline{F}_{\delta,\theta}$ is unique and its change of measure $\underline{m}_{\delta,\theta} = \mathrm{d}\underline{F}_{\delta,\theta}/\mathrm{d}F_*$ is given by

$$\underline{m}_{\delta,\theta}(u) = \dot{\phi}^\star\left(\frac{k(u,\theta,\gamma) + \underline{\zeta} + \underline{\lambda}' g(u,\theta,\gamma)}{-\underline{\eta}}\right). \tag{26}$$

The LFD $\overline{F}_{\delta,\theta}$ solving the maximization problem (12) is constructed similarly, replacing $(\underline{\eta}, \underline{\zeta}, \underline{\lambda})$ in (26) with $(-\overline{\eta}, -\overline{\zeta}, -\overline{\lambda})$, where $(\overline{\eta}, \overline{\zeta}, \overline{\lambda})$ solves (14). If $\underline{\eta} = 0$ or $\overline{\eta} = 0$, then there may exist multiple distributions solving (11) and (12) at θ. LFDs in this case are constructed analogously to the method described in Section 3.2. Note that $\underline{\eta} = 0$ or $\overline{\eta} = 0$

[18]See, for example, Liese and Vajda (1987). A more direct statement is in Qiao and Minematsu (2010).

is unlikely if k and/or elements of g are unbounded in u—see the discussion in Section 3.2. If k does not depend on u, then we set

$$\underline{m}_{\delta,\theta}(u) = \overline{m}_{\delta,\theta}(u) = \dot{\phi}^{\star}\big(-\underline{\zeta} - \underline{\lambda}' g(u, \theta, \gamma)\big), \tag{27}$$

where $(\underline{\zeta}, \underline{\lambda})$ solves (16). While there may exist multiple distributions solving (11) and (12) in this case, the distribution induced by (27) has smallest ϕ-divergence relative to F_* among all such distributions.

4.3. *Viewing Neighborhood Size Through the Lens of the Model*

Another method for interpreting δ is based on measuring the variation in the moments at the distributions solving (8) and (9) relative to their values under F_*.

Consider the sets of minimizing and maximizing values of θ at which $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ are attained, say $\underline{\Theta}_\delta$ and $\overline{\Theta}_\delta$. These are nonempty under the regularity conditions in Section 6. While the moment conditions (1a)–(1d) hold at any $\theta \in \underline{\Theta}_\delta \cup \overline{\Theta}_\delta$ under the corresponding LFD, they will typically not hold at θ under F_*. We therefore define

$$\begin{aligned} \mathrm{size}(\delta) = \sup_{\theta \in \underline{\Theta}_\delta \cup \overline{\Theta}_\delta} \max\Big\{ & \big\| \big(\mathbb{E}^{F_*}\big[g_1(U, \theta, \gamma_0)\big] - P_{10}\big)_+ \big\|_\infty, \big\| \mathbb{E}^{F_*}\big[g_1(U, \theta, \gamma_0)\big] - P_{20} \big\|_\infty, \\ & \big\| \big(\mathbb{E}^{F_*}\big[g_3(U, \theta, \gamma_0)\big]\big)_+ \big\|_\infty, \big\| \mathbb{E}^{F_*}\big[g_4(U, \theta, \gamma_0)\big] \big\|_\infty \Big\}, \end{aligned}$$

where $(v)_+ = (\max\{v_i, 0\})_{i=1}^d$ for a vector $v \in \mathbb{R}^d$. The quantity $\mathrm{size}(\delta)$ is the maximum degree to which the moments at $\theta \in \underline{\Theta}_\delta \cup \overline{\Theta}_\delta$ violate (1a)–(1d) under F_*.

This measure is informative about the extent to which the distortions to F_* required to attain the smallest and largest values of the counterfactual over $\mathcal{N}_\delta$ are reflected in (1a)–(1d). Small values of $\mathrm{size}(\delta)$ indicate that the LFDs supporting $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ distort F_* in a way that moves the counterfactual but barely moves the moments. Conversely, large values of $\mathrm{size}(\delta)$ indicate that distortions required to increase or decrease the counterfactual also have a material impact on the moments. In practice, this measure can be computed by replacing (P_0, γ_0) by estimators $(\hat{P}, \hat{\gamma})$ and $\underline{\Theta}_\delta$ and $\overline{\Theta}_\delta$ by the minimizers and maximizers of the sample criterion functions from Section 2.4 or by the estimators of $\underline{\Theta}_\delta$ and $\overline{\Theta}_\delta$ introduced in Section 6.2.

4.4. *Relating Different Divergences*

It is well known that ϕ-divergences are equivalent over local neighborhoods (see, e.g., Theorem 4.1 of Csiszár and Shields (2004)). However, $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ may depend on the choice of ϕ when δ is not arbitrarily small. Bounds induced by different ϕ functions may be related as follows. Let $\mathcal{N}_{\delta,1}$ and $\mathcal{N}_{\delta,2}$ denote δ-neighborhoods induced by ϕ_1 and ϕ_2, respectively. The quantity

$$\bar{a} = \sup_{x \geq 0, x \neq 1} \frac{\phi_1(x)}{\phi_2(x)}$$

is a measure of relative neighborhood size: if $\bar{a} < \infty$, then $\mathcal{N}_{\delta,2} \subseteq \mathcal{N}_{\bar{a}\delta,1}$ for each $\delta > 0$, as shown formally in the proof of Proposition 4.1 below. For instance, when comparing KL divergence ($\phi_1(x) = x \log x - x + 1$) and χ^2 divergence ($\phi_2(x) = \frac{1}{2}(x-1)^2$), we obtain $\bar{a} = 2$. Therefore, δ-neighborhoods under χ^2 divergence are contained in 2δ-neighborhoods under KL divergence. Interchanging ϕ_1 and ϕ_2 produces $\bar{a} = +\infty$, which reflects the fact that KL divergence is weaker than χ^2 divergence.

Let $\underline{\kappa}_{\delta,1}$ and $\underline{\kappa}_{\delta,2}$ denote the smallest counterfactual from display (8) over $\mathcal{N}_{\delta,1}$ and $\mathcal{N}_{\delta,2}$, respectively. Define $\overline{\kappa}_{\delta,1}$ and $\overline{\kappa}_{\delta,2}$ analogously.

PROPOSITION 4.1: *Suppose that Assumption* Φ *holds for both* ϕ_1 *and* ϕ_2 *and* $\bar{a}$ *is finite. Then* $[\underline{\kappa}_{\delta,2}, \overline{\kappa}_{\delta,2}] \subseteq [\underline{\kappa}_{\bar{a}\delta,1}, \overline{\kappa}_{\bar{a}\delta,1}]$ *for each* $\delta > 0$.

It follows from Proposition 4.1 that bounds that are wide under ϕ_2 must necessarily be wide under ϕ_1. Similarly, narrow bounds under ϕ_1 must also be narrow under ϕ_2. Note also that the inclusion in Proposition 4.1 holds for any counterfactual.

5. EMPIRICAL APPLICATIONS

5.1. *Marital College Premium*

Chiappori, Salanié, and Weiss (2017), henceforth CSW, studied the evolution of marital returns to education using a frictionless matching model with transferable utility (Choo and Siow (2006)). Within this framework, the "marital college premium" is the additional expected utility that an individual would derive from the marriage market if they had a (counterfactually) higher level of education. CSW found that marital college premiums for women in the United States increased significantly across cohorts from the mid to late 20th century, particularly for the more highly educated.

As is conventional following Dagsvik (2000) and Choo and Siow (2006), CSW assumed latent variables representing individuals' idiosyncratic marital preferences are i.i.d. Gumbel. The marital college premium is only partially identified when the distribution of these latent variables is not specified. We therefore perform a sensitivity analysis of CSW's estimates to departures from this conventional parametric assumption.

Our analysis makes several findings. First, it seems difficult to draw conclusions about whether marital college premiums have increased or decreased over time under small nonparametric relaxations of the i.i.d. Gumbel assumption. Interestingly, premiums have narrow nonparametric bounds at fixed parameter values, but a slight relaxation of the i.i.d. Gumbel assumption allows for significant variation in parameters which, in turn, produces wide bounds on premiums. As parameters are just-identified under any fixed distribution of shocks (Galichon and Salanié (2022)), further restrictions on parameters or shape restrictions on the distribution are required to tighten the bounds. We show that imposing exchangeability can tighten the bounds significantly.

Model and Benchmark Estimates. Agents are male or female and one of J types (education levels). A type-a male receives utility ε_{a0} if he chooses to be unmatched and $z_{ab} + \varepsilon_{ab}$ if he matches with a type-b female. Similarly, a type-b female receives utility e_{0b} if she chooses to be unmatched and $t_{ab} + e_{ab}$ if she matches with a type-a male. The parameters $(z_{ab}, t_{ab})_{a,b=1}^{J}$ represent the common deterministic component of marital preferences. The latent shocks $(\varepsilon_{a0}, \dots, \varepsilon_{aJ})$ and $(e_{0b}, \dots, e_{Jb})$ represent individuals' idiosyncratic marital preferences. Shocks are i.i.d. across individuals and have mean zero. The type b to b' marital education premium for females is the difference in expected marital utility between types b and b':

$$\kappa = \mathbb{E}^F\Big[\max_{a=0,\dots,J}(t_{ab'} + e_{ab'})\Big] - \mathbb{E}^F\Big[\max_{a=0,\dots,J}(t_{ab} + e_{ab})\Big], \tag{28}$$

where F denotes the distribution of $(e_{0b}, \dots, e_{Jb'})$ and $t_{0b} = t_{0b'} = 0$.

CSW used data from the American Community Survey. They formed 28 cohorts indexed by female birth year from 1941 (cohort 1) to 1968 (cohort 28), each of which is treated as an independent marriage market. We focus on CSW's estimates for whites. There are $J = 5$ types: "high-school dropouts," "high-school graduates," "some college," "college graduate," and "college-plus." We center our analysis on the "some college" to "college graduate" premium, though we obtained qualitatively similar results (not reported) for the "college graduate" to "college-plus" premium. Figure 1 presents estimates and 95% confidence sets (CSs) for the premium under the i.i.d. Gumbel assumption (cf. Figure 21 in CSW) based on CSW's replication files.

Implementation. The model reduces to a standard individual-level discrete choice problem for each type (see CSW's Propositions 1 and 2). We assume that the distribution of females' preference shocks does not depend on their type, so we drop the b subscript and consider a single random vector $U = (e_0, \dots, e_J)$. We allow the distribution F of U to vary across cohorts and implement our procedures cohort-by-cohort.[19]

Under any fixed F, a cohort's parameters $(t_{ab})_{a=1}^J$ are just-identified from the marriage probabilities for that cohort's type-b women (Galichon and Salanié (2022)). We therefore impose only the moment conditions involving the parameters $\theta = (t_{ab}, t_{ab'})_{a=1}^J$ appearing in (28), as the remaining parameters can be chosen to fit the remaining marriage probabilities under the resulting least-favorable distribution. We form g_2 to explain the type b and b' marriage probabilities for women in a given cohort:

$$g_2(U, \theta) = \begin{bmatrix} \Big(\mathbb{1}\Big\{t_{ab} + e_a = \max_{a'=0,\dots,J}(t_{a'b} + e_{a'})\Big\}\Big)_{a=1}^J \\ \Big(\mathbb{1}\Big\{t_{ab'} + e_a = \max_{a'=0,\dots,J}(t_{a'b'} + e_{a'})\Big\}\Big)_{a=1}^J \end{bmatrix}$$

and form $\hat{P}_2$ using CSW's estimates of the corresponding type-b and b' marriage probabilities. We set $g_4(U, \theta) = (e_j, e_j^2 - \pi^2/6)_{j=0}^J$ so that shocks have mean zero and the same variance as the Gumbel distribution. The scale normalization also ensures that the nonparametric bounds on the premium are finite at any fixed θ. As $J = 5$, there are 22 moments (10 for marriage probabilities and 12 location/scale normalizations), and θ has dimension 10.

We consider a second implementation which imposes invariance of F under rotations and reflections of potential spouse types, so that the model-implied marriage probabilities depend on θ but not the labeling of potential spouse types (though they may depend on their ordering).[20] Formally, this shape restriction corresponds to dihedral exchangeability (see Appendix A.1 of the Supplemental Material); we refer to it simply as "exchangeability." Under this shape restriction, F must satisfy the 22 moment conditions under all 12

[19]In view of the just-identification results of Galichon and Salanié (2022), we would obtain the same bounds if F was homogeneous across cohorts. Allowing for heterogeneity in own-type would result in wider bounds.

[20]Allowing dependence on the ordering of types seems desirable here as types correspond to education levels, which are naturally ordered.

rotations and reflections of the elements of U. This implementation therefore imposes a total of 264 moment conditions. Rather than including all 264 moments separately, it suffices to form g_2 and g_4 by taking the averages of the 22 moments across the 12 permutations (see Appendix A.1). Both implementations therefore have inner optimization problems of the same dimension.

Computations are performed as described in Section 3.1. The first implementation uses 50,000 scrambled Halton draws to compute the expectations. The second uses 10,000 draws which are concatenated over the 12 permutations (see Remark A.2), for a total of 120,000 draws. Computation times are reported in Appendix D.1 of the Supplemental Material. CSs for $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ are computed using the bootstrap procedure in Section 6.2. Appendix D.1 discusses bootstrap details and presents projection CSs using the method from Section 6.3.

We define neighborhoods using a hybrid of KL and χ^2 divergence:

$$\phi(x) = \begin{cases} x \log x - x + 1 & \text{if } x \leq e, \\ \dfrac{1}{2e}(x-e)^2 + (x-e) + 1 & \text{if } x > e. \end{cases}$$

We use this divergence because Assumption Φ(ii) fails for KL divergence, whereas hybrid divergence only requires finite second moments for Assumption Φ(ii). The LFDs under hybrid divergence are also everywhere positive, which is not guaranteed under χ^2 divergence. We repeated our analysis with neighborhoods constrained by χ^2 and L^4 divergences as robustness checks. Overall, our findings are not sensitive to ϕ (see Appendix D.1 for a discussion).

Findings. Figure 1 presents a sensitivity analysis of the "some college" to "college graduate" premium. Cohort-wise estimates and CSs for $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ are presented, beginning at $\delta = 0.01$ and increasing δ by factors of 10 up to $\delta = 100$. Even with $\delta = 0.01$, estimates of $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ lie uniformly below and above zero across cohorts without exchangeability (see Figure 1(a)). Imposing exchangeability can tighten the bounds, with the bounds for $\delta = 0.01$ significantly negative in early cohorts and significantly positive in later cohorts (see Figure 1(b)). But the $\delta = 0.1$ bounds with exchangeability again contain zero across all cohorts. Bounds for larger δ presented in Figures 1(c) and 1(d) are uninformatively wide.

To understand better what is meant by "small" and "large" neighborhoods, Figure 2 plots marginal CDFs for the LFDs under which the upper bounds for cohort 1 are attained. Similar LFDs (not reported) were obtained for other cohorts and the lower bounds. Without exchangeability, the LFDs with $\delta = 0.1$ are almost identical to Gumbel (plots with $\delta = 0.01$ are indistinguishable from Gumbel). LFDs appear close to Gumbel across most potential spouse types with $\delta = 1$, while for $\delta = 10$ and $\delta = 100$ the LFDs have kinks and indicate shifts in mass from the center of the distribution to the tails.

Under exchangeability (Figure 2(b)), the marginal distribution of shocks is independent of potential spouse type. In this case, the LFDs for $\delta = 1$ or smaller are virtually indistinguishable from Gumbel. LFDs with $\delta = 10$ and $\delta = 100$ are also less kinked than Figure 2(a) because distortions are spread more evenly across potential spouse types.

We also computed the largest correlation of shocks under the LFDs at which the bounds are attained and our size measure from Section 4.3. As these quantities are stable across cohorts, we present their averages across cohorts in Table III. Shocks are independent when $\delta = 0$ and only very weakly correlated for small δ, while for large δ some shocks

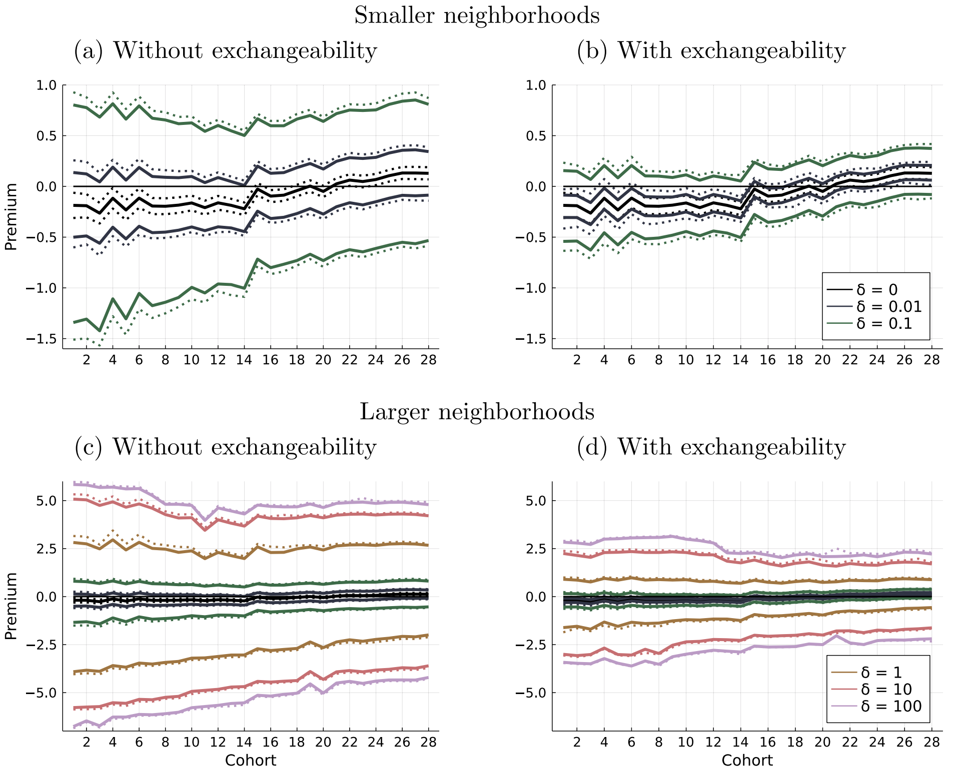

FIGURE 1.—Sensitivity analysis of the "some college" to "college graduate" premium across cohorts. *Note*: Solid lines are estimates, dotted lines are (cohort-wise) 95% CSs. CSW's estimates and CSs correspond to $\delta = 0$.

are strongly negatively correlated. The maximal correlations under exchangeability are smaller, especially for large δ. Turning to the size measure, the LFDs for $\delta = 0.01$ without exchangeability shift the model-implied marriage probabilities by 0.01 (on average, across cohorts) from their values under the i.i.d. Gumbel assumption. LFDs for $\delta = 10$ and $\delta = 100$ shift marriage probabilities around 0.25 (on average, across cohorts). Imposing exchangeability reduces the size measure by around 25% because model parameters do not vary as much under this shape restriction.

In view of the small-δ bounds in Figure 1, the LFDs in Figure 2, and the metrics in Table III, it seems difficult to draw conclusions about how the sign of the premium has changed over time under slight nonparametric relaxations of the i.i.d. Gumbel assumption. To help understand why, Figure 5 plots bounds where F is allowed to vary but θ is held fixed at CSW's estimates. These "fixed-θ" bounds for $\delta = 10$ and $\delta = 100$ are almost identical, and are roughly the same width as the $\delta = 0.01$ bounds in Figure 1. The width of the bounds in Figure 1 therefore seems largely attributable to the additional variation in θ that is permitted when parametric assumptions for F are relaxed.

Overall, our findings are complementary to Gualdani and Sinha (2022) who performed a nonparametric reanalysis of CSW using the PIES methodology of Torgovitsky (2019b). Although they did not derive nonparametric bounds on the marital education premium itself, only terms that contribute to it, they found no evidence of an increase in premiums across cohorts.

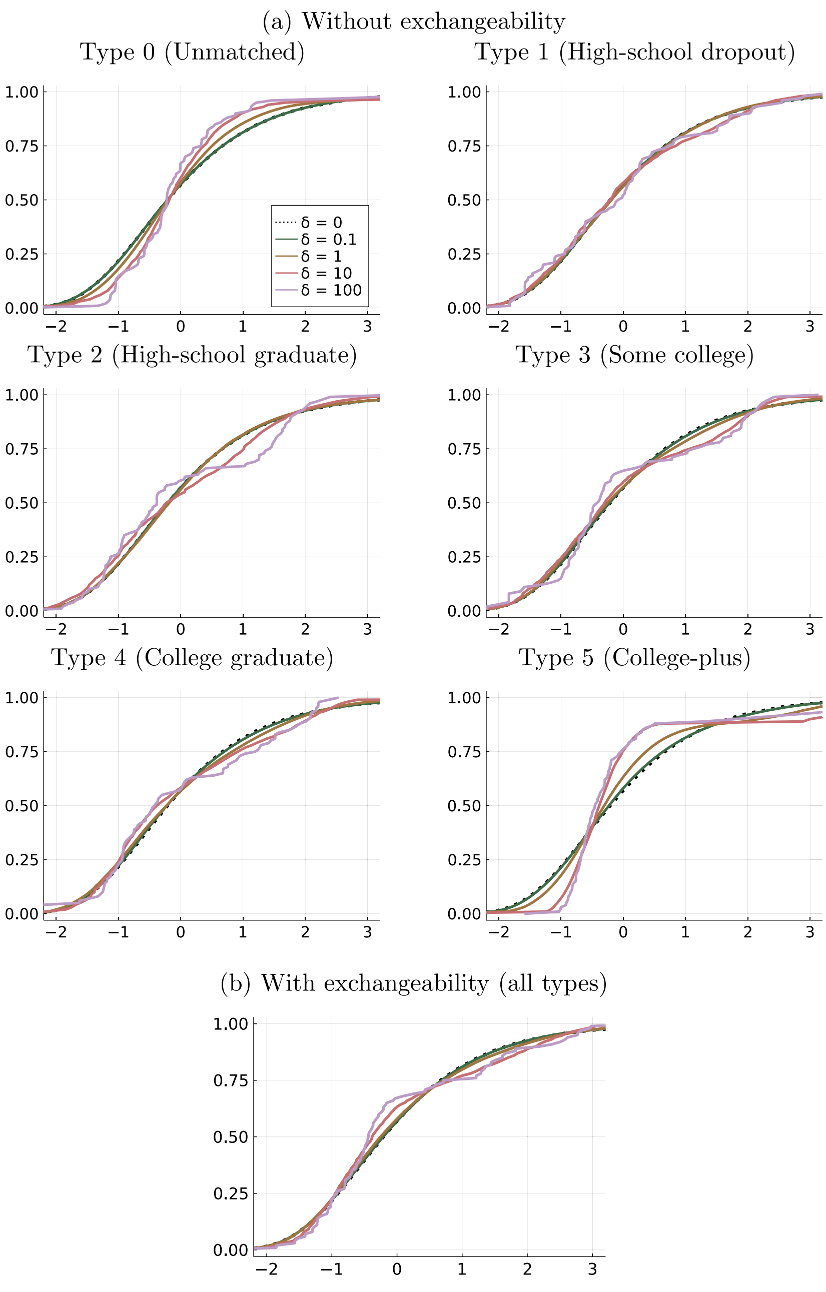

FIGURE 2.—Marginal CDFs for the LFDs maximizing the "some college" to "college graduate" premium in cohort 1 across potential spouse types.

5.2. *Welfare Analysis in a Rust Model*

Our second empirical illustration is a sensitivity analysis for welfare counterfactuals in the DDC model of Rust (1987).

TABLE III

METRICS FOR INTERPRETING δ.

δ	Without exchangeability $\rho_{\max}$, $\underline{\kappa}_\delta$	$\rho_{\max}$, $\overline{\kappa}_\delta$	size	With exchangeability $\rho_{\max}$, $\underline{\kappa}_\delta$	$\rho_{\max}$, $\overline{\kappa}_\delta$	size
0.01	−0.015	−0.014	0.010	−0.022	0.013	0.006
0.10	−0.071	−0.073	0.038	−0.061	0.054	0.023
1	−0.247	−0.197	0.112	−0.139	0.115	0.099
10	−0.502	−0.496	0.242	−0.204	0.236	0.176
100	−0.620	−0.576	0.266	−0.266	0.284	0.178

Note: Averages across cohorts of the largest element of the correlation matrix for U under the LFDs at which the estimated lower bounds ($\rho_{\max}$, $\underline{\kappa}_\delta$) and upper bounds ($\rho_{\max}$, $\overline{\kappa}_\delta$) are attained, and our size measure from Section 4.3. Each is computed at the parameter values at which the estimated upper and lower bounds are attained.

Model and Benchmark Estimates. We focus on the specification in Table IX of Rust (1987) where maintenance costs are linear in the state (i.e., mileage). In the notation of Example 2.3, $|\mathcal{S}| = 90$, $\beta = 0.9999$, and $\theta_\pi = (\text{RC}, \text{MC})$ where RC is the replacement cost and MC is a maintenance cost parameter. Our counterfactual of interest is the change in average welfare arising from a 10% reduction in maintenance costs. Hence, $\pi_{1,s}(\theta_\pi) = \tilde{\pi}_{1,s}(\theta_\pi) = -\text{RC}$ and $\pi_{0,s}(\theta_\pi) = -0.001\text{MC} \times s$ (baseline) and $\tilde{\pi}_{0,s}(\theta_\pi) = 0.9\pi_{0,s}(\theta_\pi)$ (counterfactual). The counterfactual function is $k(\theta, \gamma) = w'(\tilde{v} - v)$ where w is the stationary distribution of the state in the baseline model.

Under the i.i.d. Gumbel assumption, the estimated counterfactual at the maximum likelihood estimate (MLE) of θ_π is 73.07 and its 95% CS is [48.25,101.31].[21] Note the counterfactual is point-identified under the i.i.d. Gumbel assumption because θ_π is point-identified.

Implementation. We estimate CCPs using Rust's Group 4 data. Nonparametric estimates of the 90 CCPs are zero in many states, so we proceed as in Section 3.3 and take the model-implied CCPs at the MLE of θ_π (under the i.i.d. Gumbel assumption) as our estimate $\hat{P}_2$. We drop moment conditions for CCPs in states where the replacement probability is less than 0.001 to avoid numerical instabilities induced by including near-degenerate moments. This reduces the dimension of g_2 to 71. We normalize F so that shocks have mean zero and the same variance as the Gumbel distribution by appending $\mathbb{E}^F[U_d] = 0$ and $\mathbb{E}^F[U_d^2 - \pi^2/6] = 0$, for $d = 0, 1$, to g_4. In total, there are 255 moments (71 for CCPs, 180 for Bellman equations, and 4 location/scale normalizations) and $\theta = (\theta_\pi, v, \tilde{v})$ has dimension 182.

We implement our methods as described in Section 3.2. The inner optimization uses 75 moments (71 for CCPs and 4 for normalizations), with the remaining 180 moments appended as constraints in the outer optimization. We define neighborhoods using hybrid divergence from Section 5.1 so that Assumption Φ(ii) holds. Similar results are obtained with χ^2 and L^4 neighborhoods (see Appendix D.2). Expectations are computed using 50,000 scrambled Halton draws—see Appendix D.2 for computation times. We compute 95% CSs for $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ using the bootstrap procedure from Section 6.2 and projection procedure from Section 6.3. See Appendix D.2 for details.

[21] We construct this CS by simulation. We draw $\hat{\theta}_\pi^* \sim N(\hat{\theta}_\pi, \hat{\Sigma})$ where $\hat{\theta}_\pi$ is the MLE and $\hat{\Sigma}$ is an estimate of the inverse information matrix. For each $\hat{\theta}_\pi^*$ draw, we compute the baseline and counterfactual value functions v^* and $\tilde{v}^*$, and hence the counterfactual $\hat{\kappa}^* = w'(\tilde{v}^* - v^*)$.

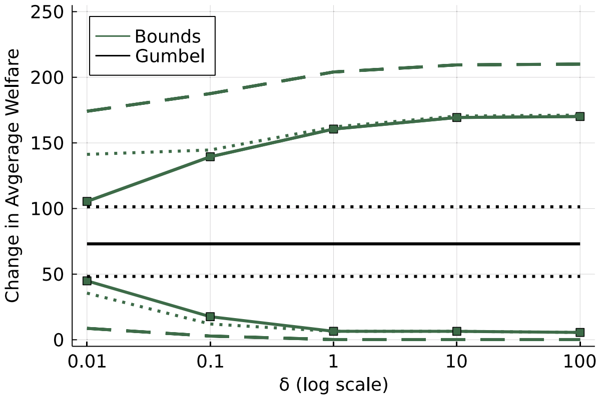

FIGURE 3.—Sensitivity analysis for change in average welfare under a 10% maintenance cost subsidy. *Note*: Solid lines are estimates, dotted lines are bootstrap CSs, dashed lines are projection CSs.

Findings. Estimates and CSs for $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ are plotted in Figure 3 for values of δ from 0.01 to 100.[22] As can be seen, the bounds expand rapidly under slight relaxations of the i.i.d. Gumbel assumption then stabilize around $\delta = 1$, where the lower bound is 6.45 and the upper bound of 160.5 represents approximately 220% of the value under the i.i.d. Gumbel assumption.

To interpret δ, in Figure 4 we plot the CDFs of $U_1 - U_0$ under the LFDs at which the estimated bounds $\hat{\underline{\kappa}}_\delta$ and $\hat{\overline{\kappa}}_\delta$ are attained. LFDs were computed as described in Section 4.2 using the construction (27). The distributions appear very close to logistic (their distribution when $\delta = 0$) for $\delta = 0.01$. Therefore, we see that large differences in welfare counterfactuals can arise under very slight departures from the i.i.d. Gumbel assumption. LFDs for the upper bound shift increasing amounts of mass to the center of the distribu-

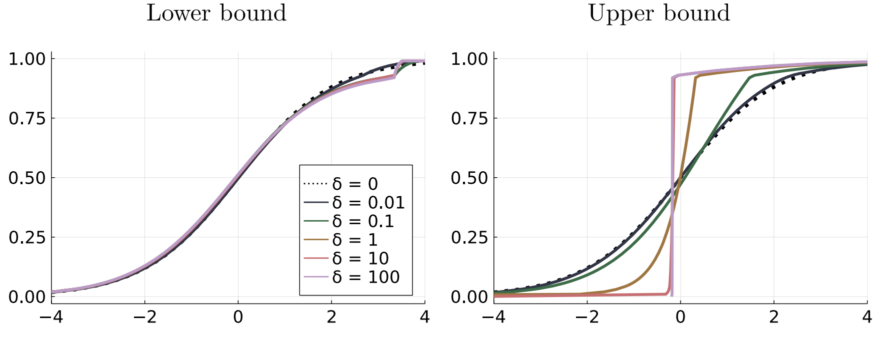

FIGURE 4.—CDFs of $U_1 - U_0$ under the LFDs at which the estimated lower and upper bounds on the welfare counterfactual are attained.

[22]The width of the bootstrap CSs relative to the bounds reduces as δ gets large. We re-estimated our bounds using several different draws of bootstrapped CCPs in place of $\hat{P}_2$ and obtained bounds that spanned a range similar to the bootstrap CSs for small δ, but which for many draws converged to values close to our estimates of the bounds for large δ. This corroborates the behavior of our bootstrap CSs. We conjecture that other features of the model are potentially more important than the numerical values of the CCPs in determining nonparametric bounds on the welfare counterfactual.

TABLE IV

METRICS FOR INTERPRETING δ.

	Lower bound				Upper bound			
δ	corr	size	RC	MC	corr	size	RC	MC
0	0.000	0.000	10.208	2.294	0.000	0.000	10.208	2.294
0.01	0.036	0.010	7.357	1.411	−0.027	0.016	13.390	3.307
0.1	−0.058	0.039	5.186	0.553	0.149	0.109	16.134	4.374
1	−0.045	0.039	4.023	0.203	0.616	0.346	17.166	5.038
10	−0.040	0.039	4.022	0.202	0.765	0.461	17.595	5.331
100	−0.063	0.039	3.931	0.176	0.764	0.469	17.626	5.365

Note: Correlation of U_0 and U_1 under the LFD at which the estimated lower and upper bounds are attained (corr), our size measure from Section 4.3, and replacement and maintenance cost parameters at which the estimated lower and upper bounds are attained.

tion of $U_1 - U_0$ as δ increases. LFDs corresponding to the lower bound are relatively less distorted, but have increasing amounts of mass shifted into the right tail. These are similar for $\delta = 0.1$ through $\delta = 100$ because the estimated lower bound stabilizes for smaller values of δ than the upper bound (cf. Figure 3).

Table IV lists other metrics to help interpret the neighborhood size. The first is the correlation of U_0 and U_1 under the LFDs at which $\hat{\underline{\kappa}}_\delta$ and $\hat{\overline{\kappa}}_\delta$ are attained. These are very small for $\delta = 0.01$ and remain small under the LFDs for $\hat{\underline{\kappa}}_\delta$ as δ increases, while U_0 and U_1 are strongly positively correlated under the LFDs for $\hat{\overline{\kappa}}_\delta$, especially for larger δ values. We compute our size measure separately for the upper and lower bounds. We measure distortions using only the moments corresponding to the CCPs as these are most directly interpretable within the context of the model. We see that the LFDs for $\delta = 0.01$ are distorting F_* in a manner that shifts the model-implied CCPs by at most 0.016. By contrast, the LFDs for $\delta = 10$ and $\delta = 100$ shift the model-implied CCPs from their values under the i.i.d. Gumbel assumption by at most 0.04 for $\hat{\underline{\kappa}}_\delta$ and 0.47 for $\hat{\overline{\kappa}}_\delta$.

The parameters at which $\hat{\underline{\kappa}}_\delta$ and $\hat{\overline{\kappa}}_\delta$ are attained are also revealing about neighborhood size. Table IV presents MLEs of MC and RC, which are similar to the values reported in Table IX of Rust (1987). We see from Table IV that $\hat{\underline{\kappa}}_\delta$ and $\hat{\overline{\kappa}}_\delta$ are attained at very different parameter values, with much smaller cost parameters for the lower bound and larger parameters for the upper bound, even for $\delta = 0.01$. Intuitively, a smaller MC means that the saving from the subsidy—which is proportional—must be small. Correspondingly, a low RC is needed to help the model to fit the observed CCPs at the smaller MC. While it is known that payoff parameters are not identified without parametric assumptions on F, it is perhaps surprising that these parameters vary by so much under slight relaxations of the i.i.d. Gumbel assumption. For instance, with $\delta = 0.01$, the lower bound is attained with cost parameters RC $= 7.357$ and MC $= 1.411$ while the upper bound is attained with cost parameters that are roughly double these values.

6. ESTIMATION AND INFERENCE

We begin in Section 6.1 by establishing consistency and the asymptotic distribution of the estimators $\hat{\underline{\kappa}}_\delta$ and $\hat{\overline{\kappa}}_\delta$ from Section 2.4. We then present a bootstrap-based inference method in Section 6.2 and a projection-based inference method in Section 6.3.

6.1. *Large-Sample Properties of Plug-in Estimators*

We first introduce some regularity conditions. Recall the space $\mathcal{E}$ from Assumption Φ. We equip $\mathcal{E}$ with the Orlicz norm (see Appendix F of Christensen and Connault (2022))

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

This norm is equivalent to the $L^2(F_*)$ norm for χ^2 and hybrid divergence and equivalent to the $L^q(F_*)$ norm for L^p divergence ($p^{-1}+q^{-1}=1$), while for KL divergence it is stronger than any $L^p(F_*)$ norm with $p<\infty$ but weaker than the sup-norm. Say that a class of functions $\{f_a : a \in \mathcal{A}\} \subset \mathcal{E}$ indexed by a metric space $\mathcal{A}$ is *$\mathcal{E}$-continuous in a* if $a' \to a$ in $\mathcal{A}$ implies $\|f_a - f_{a'}\|_\psi \to 0$. We also require a slightly stronger notion of constraint qualification than Condition S from Section 2.5.

DEFINITION 6.1: *Condition* S′ holds at (θ, γ, P) if $\vec{P} \in \operatorname{int}(\mathcal{G}(\theta, \gamma) + \mathcal{C})$.

Condition S′ replaces "relative interior" in Condition S with "interior." Finally, recall $\Delta(\theta; \gamma, P)$ from (16) and let $\Theta_\delta(\gamma, P) = \{\theta \in \Theta : \Delta(\theta; \gamma, P) < \delta\}$.

ASSUMPTION M: (i) *$k(\cdot; \theta, \gamma)$ and each entry of $g(\cdot; \theta, \gamma)$ are $\mathcal{E}$-continuous in (θ, γ);*
(ii) $(\theta, \gamma) \mapsto \mathbb{E}^{F_*}[\phi^\star(a_1 + a_2 k(U, \theta, \gamma) + a_3' g(U, \theta, \gamma))]$ *is continuous for each* $(a_1, a_2, a_3) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$;
(iii) $\Theta_\delta(\gamma_0, P_0)$ *is nonempty and Condition* S′ *holds at* (θ, γ_0, P_0) *for each* $\theta \in \Theta_\delta(\gamma_0, P_0)$;
(iv) $\operatorname{cl}(\Theta_\delta(\gamma_0, P_0)) \supseteq \{\theta \in \Theta : \Delta(\theta; \gamma_0, P_0) \le \delta\}$;
(v) Θ *is a compact subset of* $\mathbb{R}^{d_\theta}$.

Parts (i) and (ii) of Assumption M are continuity conditions. If k and g consist entirely of indicator functions of events, then these conditions hold provided the probabilities of the events under F_* are continuous in (θ, γ). In models without γ, these conditions simply require continuity in θ.

There are two parts to Assumption M(iii). The nonemptyness condition holds when the model is correctly specified under F_* or, more generally, when there is at least one $F \in \mathcal{N}_\delta$ that satisfies (1a)–(1d) for some θ. The second part is a constraint qualification. This condition requires that for each $\theta \in \Theta_\delta(\gamma_0, P_0)$, there is a distribution F under which (1a)–(1d) holds at (θ, γ_0, P_0) that is "interior" to $\mathcal{N}_\infty$, in the sense that one can perturb the moments at (θ, γ_0, P_0) in all directions by perturbing F. Condition S′ also requires that there is $F \in \mathcal{N}_\infty$ under which any inequality restrictions at (θ, γ_0, P_0) hold strictly. Note, however, that we do not require that this F belongs to $\mathcal{N}_\delta$, only to $\mathcal{N}_\infty$. We therefore do not view this condition as overly restrictive. We also conjecture it could be relaxed using a notion similar to S-regularity from Section 2.5.

Assumption M(iv) is made for convenience and can be relaxed; this condition simply ensures that there do not exist values of θ at which $\Delta(\theta; \gamma_0, P_0) = \delta$ that are separated from $\Theta_\delta(\gamma_0, P_0)$. Assumption M(v) is standard and can be relaxed.

THEOREM 6.1: *Suppose that Assumptions Φ and M hold and $(\hat{\gamma}, \hat{P}) \to_p (\gamma_0, P_0)$ or, if there is no auxiliary parameter, $\hat{P} \to_p P_0$. Then $\hat{\underline{\kappa}}_\delta \to_p \underline{\kappa}_\delta$ and $\hat{\overline{\kappa}}_\delta \to_p \overline{\kappa}_\delta$.*

To derive the asymptotic distribution of the estimators, we assume γ_0 is known and suppress dependence of all quantities on γ for the remainder of this section. This entails no loss of generality for models without γ, such as Examples 2.1 and 2.2 and the application in Section 5.1. In DDC models, this presumes the law of motion of the state is known. The asymptotic distribution therefore reflects only sampling uncertainty from the estimated CCPs, which is the case for confidence sets reported when laws of motion are first estimated "offline." Extending our approach to accommodate sampling variation in $\hat{\gamma}$ in a tractable manner appears to require exploiting application-specific model structure, which we defer to future work.

Define

$$\underline{b}_\delta(P) = \inf_{\theta\in\Theta_\delta(P)} \underline{K}_\delta(\theta; P), \qquad \overline{b}_\delta(P) = \sup_{\theta\in\Theta_\delta(P)} \overline{K}_\delta(\theta; P). \tag{29}$$

In this notation, $\underline{\kappa}_\delta = \underline{b}_\delta(P_0)$ and $\overline{\kappa}_\delta = \overline{b}_\delta(P_0)$ (see Lemma E.3) and $\hat{\underline{\kappa}}_\delta = \underline{b}_\delta(\hat{P})$ and $\hat{\overline{\kappa}}_\delta = \overline{b}_\delta(\hat{P})$. We derive the asymptotic distribution of $\hat{\underline{\kappa}}_\delta$ and $\hat{\overline{\kappa}}_\delta$ by showing $\underline{b}_\delta$ and $\overline{b}_\delta$ are directionally differentiable and applying a suitable delta method. Say $f : \mathbb{R}^{d_1+d_2} \to \mathbb{R}$ 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

$$\lim_{n\to\infty} t_n^{-1}\big(f(P_0 + t_n h_n) - f(P_0)\big) = df_{P_0}[h]$$

for all sequences $t_n \downarrow 0$ and $h_n \to h$ (Shapiro (1990, p. 480)). If $df_{P_0}[h]$ is linear in h, then f is (fully) differentiable at P_0. We introduce some additional notation used to define the directional derivatives of $\underline{b}_\delta$ and $\overline{b}_\delta$. Let

$$\underline{\Xi}_\delta(\theta; P) = \operatorname*{argsup}_{\eta\geq 0, \zeta\in\mathbb{R}, \lambda\in\Lambda} -\mathbb{E}^{F_*}\big[(\eta\phi)^\star\big(-k(U,\theta) - \zeta - \lambda' g(U,\theta)\big)\big] - \eta\delta - \zeta - \lambda_{12}' P,$$

where $(\eta\phi)^\star$ denotes the convex conjugate of $x \mapsto \eta \cdot \phi(x)$, and let $\overline{\Xi}_\delta(\theta; P)$ denote the analogous arginf for the minimization problem corresponding to the upper bound. Recall that $\underline{\lambda}_{12} = (\underline{\lambda}_1, \underline{\lambda}_2)$ collects the first $d_1 + d_2$ elements of $\underline{\lambda}$. Let

$$\underline{\Lambda}_\delta(\theta; P) = \big\{(\lambda_1, \lambda_2) : (\eta, \zeta, \lambda_1, \lambda_2, \lambda_3, \lambda_4) \in \underline{\Xi}_\delta(\theta; P)\big\}$$

denote the projection of $\underline{\Xi}_\delta(\theta; P)$ for $\underline{\lambda}_{12}$. We let $\overline{\Lambda}_\delta(\theta; P)$ denote the analogous projection of $\overline{\Xi}_\delta(\theta; P)$. Finally, let

$$\underline{\Theta}_\delta(P_0) = \arg\min_{\theta\in\Theta} \underline{K}_\delta(\theta; P_0), \qquad \overline{\Theta}_\delta(P_0) = \arg\max_{\theta\in\Theta} \overline{K}_\delta(\theta; P_0).$$

The sets $\underline{\Theta}_\delta(P_0)$ and $\overline{\Theta}_\delta(P_0)$ are nonempty and compact under Assumptions Φ and M.

The following regularity conditions are presented for the general case in which k depends on u. It may be possible to weaken some of these conditions when k does not depend on u.

ASSUMPTION M—continued: (vi) *$\underline{\Theta}_\delta(P_0) \subseteq \Theta_\delta(P_0)$ and $\overline{\Theta}_\delta(P_0) \subseteq \Theta_\delta(P_0)$;*
(vii) *$\theta \mapsto \underline{\Lambda}_\delta(\theta; P_0)$ and $\theta \mapsto \overline{\Lambda}_\delta(\theta; P_0)$ are lower hemicontinuous at each $\theta \in \underline{\Theta}_\delta(P_0)$ and $\theta \in \overline{\Theta}_\delta(P_0)$, respectively.*

THEOREM 6.2: *Suppose that Assumptions* Φ *and* M *hold. Then* $\underline{b}_\delta$ *and* $\overline{b}_\delta$ *are directionally differentiable at* P_0, *with*

$$d\underline{b}_{\delta,P_0}[h]=\min_{\theta\in\underline{\Theta}_\delta(P_0)}\max_{\underline{\lambda}_{12}\in\underline{\Lambda}_\delta(\theta;P_0)}-\underline{\lambda}_{12}'h,\qquad d\overline{b}_{\delta,P_0}[h]=\max_{\theta\in\overline{\Theta}_\delta(P_0)}\min_{\overline{\lambda}_{12}\in\overline{\Lambda}_\delta(\theta;P_0)}\overline{\lambda}_{12}'h.$$

Moreover, if $\sqrt{n}(\hat{P}-P_0)\to_d Z\sim N(0,\Sigma)$ *with* Σ *finite, then*

$$\sqrt{n}\left(\begin{pmatrix}\hat{\overline{\kappa}}_\delta\\ \hat{\underline{\kappa}}_\delta\end{pmatrix}-\begin{pmatrix}\overline{\kappa}_\delta\\ \underline{\kappa}_\delta\end{pmatrix}\right)\to_d\begin{pmatrix}d\underline{b}_{\delta,P_0}[Z]\\ d\overline{b}_{\delta,P_0}[Z]\end{pmatrix}.$$

The asymptotic distribution presented in Theorem 6.2 may be non-Gaussian. In the special case in which $\bigcup_{\theta\in\underline{\Theta}_\delta(P_0)}\underline{\Lambda}_\delta(\theta;P_0)=\{\underline{\lambda}_{12}\}$, the asymptotic distribution of $\hat{\underline{\kappa}}_\delta$ simplifies to $N(0,\underline{\lambda}_{12}'\Sigma\underline{\lambda}_{12})$. An analogous simplification holds for $\hat{\overline{\kappa}}_\delta$ when $\bigcup_{\theta\in\overline{\Theta}_\delta(P_0)}\overline{\Lambda}_\delta(\theta;P_0)$ is a singleton.

6.2. *Inference Procedure 1: Bootstrap*

Our first inference procedure specializes the general approach of Fang and Santos (2019) for inference on directionally differentiable functions to the present setting. Define

$$\widehat{d\underline{b}}_{\delta,P_0}[h]=\inf_{\theta\in\hat{\underline{\Theta}}_{\delta,n}}\sup_{\underline{\lambda}_{12}\in\underline{\Lambda}_\delta(\theta;\hat{P})}-\underline{\lambda}_{12}'h,\qquad \widehat{d\overline{b}}_{\delta,P_0}[h]=\sup_{\theta\in\hat{\overline{\Theta}}_{\delta,n}}\inf_{\overline{\lambda}_{12}\in\overline{\Lambda}_\delta(\theta;\hat{P})}\overline{\lambda}_{12}'h,$$

where

$$\hat{\underline{\Theta}}_{\delta,n}=\{\theta\in\Theta_\delta(\hat{P}):\underline{K}_\delta(\theta;\hat{P})\le\hat{\underline{\kappa}}_\delta+\hat{\nu}\sqrt{\log n/n}\},\quad\text{and}$$
$$\hat{\overline{\Theta}}_{\delta,n}=\{\theta\in\Theta_\delta(\hat{P}):\overline{K}_\delta(\theta;\hat{P})\ge\hat{\overline{\kappa}}_\delta-\hat{\nu}\sqrt{\log n/n}\},$$

with $\hat{\nu}$ a (possibly random) positive scalar tuning parameter for which $\hat{\nu}\to_p\nu>0$. Any such $\hat{\nu}$ results in a confidence set with asymptotically correct coverage. We give some practical guidance for choosing $\hat{\nu}$ below.

Let $\hat{P}^*$ denote a bootstrapped version of $\hat{P}$. In practice, any bootstrap can be used provided it satisfies mild consistency conditions. In the empirical application in Section 5.1, we simply draw $\hat{P}^*\sim N(\hat{P},\hat{\Sigma}/n)$ where $\hat{\Sigma}$ is a consistent estimator of Σ. Let

$$\hat{\underline{c}}_\alpha=\alpha\text{-quantile of }\widehat{d\underline{b}}_{\delta,P_0}\big[\sqrt{n}(\hat{P}^*-\hat{P})\big],\qquad \hat{\overline{c}}_\alpha=\alpha\text{-quantile of }\widehat{d\overline{b}}_{\delta,P_0}\big[\sqrt{n}(\hat{P}^*-\hat{P})\big],$$

where the quantiles are computed by resampling $\hat{P}^*$ (conditional on the data). Lower, upper, and two-sided $100(1-\alpha)\%$ CSs for $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ are

$$\mathrm{CS}^{1-\alpha}_{\delta,L}=\left[\hat{\underline{\kappa}}_\delta-\frac{\hat{\underline{c}}_{1-\alpha}}{\sqrt{n}},+\infty\right),$$

$$\mathrm{CS}^{1-\alpha}_{\delta,U}=\left(-\infty,\hat{\overline{\kappa}}_\delta-\frac{\hat{\overline{c}}_\alpha}{\sqrt{n}}\right],\qquad \mathrm{CS}^{1-\alpha}_{\delta}=\left[\hat{\underline{\kappa}}_\delta-\frac{\hat{\underline{c}}_{1-\alpha/2}}{\sqrt{n}},\hat{\overline{\kappa}}_\delta-\frac{\hat{\overline{c}}_{\alpha/2}}{\sqrt{n}}\right].$$

We require a slight strengthening of Assumption M(vii) to establish validity of the procedure. As before, regularity conditions are presented for the general case where k depends on u. It may be possible to weaken these conditions when k does not depend on u.

ASSUMPTION M—continued: (vii′) $(\theta, P) \mapsto \underline{\Lambda}_\delta(\theta; P)$ *and* $(\theta, P) \mapsto \overline{\Lambda}_\delta(\theta; P)$ *are lower hemicontinuous at* (θ, P_0) *for each* $\theta \in \underline{\Theta}_\delta(P_0)$ *and* $\theta \in \overline{\Theta}_\delta(P_0)$, *respectively.*

THEOREM 6.3: *Suppose that Assumptions* Φ *and* M(i)–(vi), (vii′) *hold*, $\sqrt{n}(\hat{P} - P_0) \to_d Z \sim N(0, \Sigma)$ *with* Σ *finite, and* $\hat{P}^*$ *satisfies Assumption* 3 *of Fang and Santos* (2019). *Then the distribution of* $\widehat{d\underline{b}}_{\delta,P_0}[\sqrt{n}(\hat{P}^* - \hat{P})]$ *and* $\widehat{d\overline{b}}_{\delta,P_0}[\sqrt{n}(\hat{P}^* - \hat{P})]$ (*conditional on the data*) *is consistent for the asymptotic distribution derived in Theorem* 6.2. *Moreover, if the CDFs of* $d\underline{b}_{\delta,P_0}[Z]$ *and* $d\overline{b}_{\delta,P_0}[Z]$ *are continuous and increasing at their* $\alpha/2$, α, $1-\alpha$, *and* $1-\alpha/2$ *quantiles*, *then*

$$\lim_{n\to\infty} \Pr\big(\underline{\kappa}_\delta \in \mathrm{CS}^{1-\alpha}_{\delta,L}\big) = 1-\alpha,$$

$$\lim_{n\to\infty} \Pr\big(\overline{\kappa}_\delta \in \mathrm{CS}^{1-\alpha}_{\delta,U}\big) = 1-\alpha, \qquad \liminf_{n\to\infty} \Pr\big([\underline{\kappa}_\delta, \overline{\kappa}_\delta] \subseteq \mathrm{CS}^{1-\alpha}_{\delta}\big) \geq 1-\alpha.$$

Any $\hat{\nu}$ that satisfies $\hat{\nu} \to_p \nu > 0$ results in asymptotically valid CSs. In view of the functional forms of $\widehat{d\underline{b}}_{\delta,P_0}[\cdot]$ and $\widehat{d\underline{b}}_{\delta,P_0}[\cdot]$, smaller $\hat{\nu}$ produce (weakly) wider CSs. In the CSW application, we set $\hat{\nu}$ equal to the minimum diagonal element of the covariance matrix of the moments evaluated at $(\hat{\theta}, \hat{\gamma}, \hat{P})$ under F_*, where $\hat{\theta}$ is computed under F_*. We chose this quantity as it is related to the convexity of the inner problem for small δ. In practice, this resulted in $\hat{\nu}$ between 0.001 and 0.01. We recommend setting $\hat{\nu}$ to be of a similarly small magnitude, then performing a sensitivity analysis to check that critical values are not too dependent on $\hat{\nu}$. Setting $\hat{\nu} = 0$ and replacing $\hat{\underline{\Theta}}_{\delta,n}$ and $\hat{\overline{\Theta}}_{\delta,n}$ by $\{\hat{\underline{\theta}}_\delta\}$ and $\{\hat{\overline{\theta}}_\delta\}$ where $\hat{\underline{\theta}}_\delta$ and $\hat{\overline{\theta}}_\delta$ minimize and maximize the sample criterions is also valid, but may be conservative.

6.3. *Inference Procedure 2: Projection*

This second approach is computationally simple but possibly conservative.[23] Suppose we have random vectors $\hat{P}^{1-\alpha}_{1,U}$, $\hat{P}^{1-\alpha}_{2,U}$, and $\hat{P}^{1-\alpha}_{2,L}$ that form a $100(1-\alpha)\%$ rectangular CS for P_0:

$$\liminf_{n\to\infty} \Pr\big(P_{10} \leq \hat{P}^{1-\alpha}_{1,U}, \hat{P}^{1-\alpha}_{2,L} \leq P_{20} \leq \hat{P}^{1-\alpha}_{2,U}\big) \geq 1-\alpha, \tag{30}$$

where the inequalities should be understood to hold element-wise (we discuss how to construct a rectangular CS for P_0 below).

The idea behind this approach is to replace any moment conditions involving P by inequalities constructed from the rectangular CS. Define the criterion functions

$$\hat{\underline{K}}_{\delta,1-\alpha}(\theta) = \begin{cases} \underline{K}_{\delta,\mathrm{cs}}(\theta; \hat{P}_{1-\alpha}), \\ +\infty, \end{cases} \qquad \hat{\overline{K}}_{\delta,1-\alpha}(\theta) = \begin{cases} \overline{K}_{\delta,\mathrm{cs}}(\theta; \hat{P}_{1-\alpha}) & \text{if } \Delta_{\mathrm{cs}}(\theta; \hat{P}_{1-\alpha}) < \delta, \\ -\infty & \text{if } \Delta_{\mathrm{cs}}(\theta; \hat{P}_{1-\alpha}) \geq \delta, \end{cases}$$

where $\underline{K}_{\delta,\mathrm{cs}}$, $\overline{K}_{\delta,\mathrm{cs}}$, and Δ_{cs} are versions of (13), (14), and (16) formed using

$$\mathbb{E}^F\big[g_1(U,\theta)\big] \leq \hat{P}^{1-\alpha}_{1,U}, \qquad \mathbb{E}^F\big[g_2(U,\theta)\big] \leq \hat{P}^{1-\alpha}_{2,U}, \qquad \mathbb{E}^F\big[-g_2(U,\theta)\big] \leq -\hat{P}^{1-\alpha}_{2,L}, \tag{31}$$

[23] We are grateful to a referee for suggesting this approach.

as well as (1c) and (1d). In these criterions, Λ is replaced by $\Lambda_{\text{cs}} = \mathbb{R}_+^{d_1+2d_2+d_3} \times \mathbb{R}^{d_4}$, g is replaced by $g_{\text{cs}} = (g_1, g_2, -g_2, g_3, g_4)$, P is replaced by $\hat{P}_{1-\alpha} = (\hat{P}_{1,U}^{1-\alpha}, \hat{P}_{2,U}^{1-\alpha}, -\hat{P}_{2,L}^{1-\alpha})$, and λ_{12} denotes the first $d_1 + 2d_2$ elements of λ.

Critical values are computed by optimizing the criterions $\underline{\hat{K}}_{\delta,1-\alpha}$ and $\hat{\overline{K}}_{\delta,1-\alpha}$ with respect to θ:

$$\hat{\underline{\kappa}}_{\delta,1-\alpha} = \inf_{\theta\in\Theta} \underline{\hat{K}}_{\delta,1-\alpha}(\theta), \qquad \hat{\overline{\kappa}}_{\delta,1-\alpha} = \sup_{\theta\in\Theta} \hat{\overline{K}}_{\delta,1-\alpha}(\theta).$$

Lower, upper, and two-sided $100(1-\alpha)\%$ CSs for $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ are then given by

$$\mathrm{CS}_{\delta,L}^{1-\alpha} = [\hat{\underline{\kappa}}_{\delta,1-\alpha}, +\infty), \qquad \mathrm{CS}_{\delta,U}^{1-\alpha} = (-\infty, \hat{\overline{\kappa}}_{\delta,1-\alpha}], \qquad \mathrm{CS}_{\delta}^{1-\alpha} = [\hat{\underline{\kappa}}_{\delta,1-\alpha}, \hat{\overline{\kappa}}_{\delta,1-\alpha}].$$

THEOREM 6.4: *Suppose that Assumptions* Φ *and* M(i), (iii)–(v) *hold and* $\hat{P}_{1-\alpha}$ *satisfies* (30). *Then*

$$\liminf_{n\to\infty} \Pr\big(\underline{\kappa}_\delta \in \mathrm{CS}_{\delta,L}^{1-\alpha}\big) \geq 1-\alpha,$$
$$\liminf_{n\to\infty} \Pr\big(\overline{\kappa}_\delta \in \mathrm{CS}_{\delta,U}^{1-\alpha}\big) \geq 1-\alpha, \qquad \liminf_{n\to\infty} \Pr\big([\underline{\kappa}_\delta, \overline{\kappa}_\delta] \subseteq \mathrm{CS}_{\delta}^{1-\alpha}\big) \geq 1-\alpha.$$

To construct a rectangular CS for P_0 satisfying (30), suppose $\sqrt{n}(\hat{P} - P_0) \to_d N(0, \Sigma)$ and we have a consistent estimator $\hat{\Sigma}$ of Σ. Let $\hat{\sigma}$ denote the vector formed by taking the square root of each diagonal entry of $\hat{\Sigma}$. Partition $\hat{\sigma}$ conformably as $\hat{\sigma} = (\hat{\sigma}_{(1)}, \hat{\sigma}_{(2)})$ and set

$$\hat{P}_{1,U}^{1-\alpha} = \hat{P}_1 + n^{-1/2}\hat{c}_{1-\alpha,1}\hat{\sigma}_{(1)}, \qquad \hat{P}_{2,L}^{1-\alpha} = \hat{P}_2 - n^{-1/2}\hat{c}_{1-\alpha,2}\hat{\sigma}_{(2)},$$
$$\hat{P}_{2,U}^{1-\alpha} = \hat{P}_2 + n^{-1/2}\hat{c}_{1-\alpha,2}\hat{\sigma}_{(2)},$$

where the (scalar) critical values $\hat{c}_{1-\alpha,1}$ and $\hat{c}_{1-\alpha,2}$ solve

$$\Pr\Big(\max_{1\leq i\leq d_1} Z_i/\hat{\sigma}_i \leq \hat{c}_{1-\alpha,1}, \max_{d_1+1\leq i\leq d_2} |Z_i/\hat{\sigma}_i| \leq \hat{c}_{1-\alpha,2}\Big) = 1-\alpha, \quad Z \sim N(0, \hat{\Sigma}).$$

If $d_2 = 0$, then $\hat{c}_{1-\alpha,1}$ is the $(1-\alpha)$-quantile of $\max_{1\leq i\leq d_1} Z_i/\hat{\sigma}_i$; similarly, if $d_1 = 0$, then $\hat{c}_{2,1-\alpha}$ is the $(1-\alpha)$-quantile of $\max_{1\leq i\leq d_2} |Z_i/\hat{\sigma}_i|$.

7. CONCLUSION

This paper introduced a framework for analyzing the sensitivity of counterfactuals to parametric assumptions about the distribution of latent variables in structural models. In particular, we derived bounds on the set of counterfactuals obtained as the distribution of latent variables spans nonparametric neighborhoods of a given parametric specification while other "structural" model features are maintained. We illustrated our procedure with empirical applications to matching models and dynamic discrete choice.

REFERENCES

AGUIRREGABIRIA, VICTOR, AND PEDRO MIRA (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75 (1), 1–53. [263]

Allen, Roy, and John Rehbeck (2019): "Identification With Additively Separable Heterogeneity," *Econometrica*, 87 (3), 1021–1054. [265,267]

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *The Quarterly Journal of Economics*, 132 (4), 1553–1592. [265]

——— (2020): "On the Informativeness of Descriptive Statistics for Structural Estimates," *Econometrica*, 6, 2231–2258. [265]

Armstrong, Timothy B., and Michal Kolesár (2021): "Sensitivity Analysis Using Approximate Moment Condition Models," *Quantitative Economics*, 12 (1), 77–108. [265]

Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75 (5), 1331–1370. [263]

Beresteanu, Arie, Ilya Molchanov, and Francesca Molinari (2011): "Sharp Identification Regions in Models With Convex Moment Predictions," *Econometrica*, 79 (6), 1785–1821. [265,268]

Berger, James O. (1984): "The Robust Bayesian Viewpoint (With Discussion)," in *Robustness of Bayesian Analyses*, ed. by Joseph B. Kadane. North-Holland. [265]

Berry, Steven T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60 (4), 889–917. [263,267,268]

Berry, Steven T., and Philip A. Haile (2010): "Nonparametric Identification of Multinomial Choice Demand Models With Heterogeneous Consumers," Cowles Foundation Discussion Paper No. 1718. [267]

——— (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82 (5), 1749–1797. [267]

Bonhomme, Stephane, and Martin Weidner (2022): "Minimizing Sensitivity to Model Misspecification," *Quantitative Economics*, 13 (3), 907–954. [265]

Bresnahan, Timothy F., and Peter C. Reiss (1990): "Entry in Monopoly Market," *The Review of Economic Studies*, 57 (4), 531–553. [267]

——— (1991): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48 (1), 57–81. [267]

Chamberlain, Gary, and Edward E. Leamer (1976): "Matrix Weighted Averages and Posterior Bounds," *Journal of the Royal Statistical Society. Series B (Methodological)*, 38 (1), 73–84. [265]

Chen, Xiaohong, Timothy M. Christensen, and Elie Tamer (2018): "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*, 86 (6), 1965–2018. [268]

Chen, Xiaohong, Elie Tamer, and Alexander Torgovitsky (2011): "Sensitivity Analysis in Semiparametric Likelihood Models," Cowles Foundation Discussion Paper No. 1836. [265]

Chesher, Andrew, and Adam M. Rosen (2017): "Generalized Instrumental Variable Models," *Econometrica*, 85 (3), 959–989. [265]

Chiappori, Pierre-André, Bernard Salanié, and Yoram Weiss (2017): "Partner Choice, Investment in Children, and the Marital College Premium," *American Economic Review*, 107 (8), 2109–2167. [265,283]

Choo, Eugene, and Aloysius Siow (2006): "Who Marries Whom and Why," *Journal of Political Economy*, 114 (1), 175–201. [263,265,283]

Christensen, Timothy, and Benjamin Connault (2022): "Counterfactual Sensitivity and Robustness," arxiv preprint, arXiv:1904.00989 [econ.em]. [266,272,273,276,291]

——— (2023): "Supplement to 'Counterfactual Sensitivity and Robustness'," *Econometrica Supplemental Material*, 91, https://doi.org/10.3982/ECTA17232. [266]

Ciliberto, Federico, and Elie Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77 (6), 1791–1828. [263,268]

Csiszár, Imre (1967): "Information-Type Measures of Difference of Probability Distributions and Indirect Observations," *Studia Scientiarum Mathematicarum Hungarica*, 2, 299–318. [271]

Csiszár, Imre, and Paul C. Shields (2004): "Information Theory and Statistics: A Tutorial," *Communications and Information Theory*, 1 (4), 417–528. [282]

Cuturi, Marco (2013): "Sinkhorn Distances: Lightspeed Computation of Optimal Transport," *Advances in Neural Information Processing Systems*, 26, 2292–2300. [276]

Dagsvik, John K. (2000): "Aggregation in Matching Markets," *International Economic Review*, 41 (1), 27–58. [263,283]

Duchi, John C., and Hongseok Namkoong (2021): "Learning Models With Uniform Performance via Distributionally Robust Optimization," *The Annals of Statistics*, 49 (3), 1378–1406. [264]

Ekeland, Ivar, Alfred Galichon, and Marc Henry (2010): "Optimal Transportation and the Falsifiability of Incompletely Specified Economic Models," *Economic Theory*, 42 (2), 355–374. [265]

Fang, Zheng, and Andres Santos (2019): "Inference on Directionally Differentiable Functions," *Review of Economic Studies*, 86 (1), 377–412. [293,294]

GALICHON, ALFRED, AND MARC HENRY (2011): "Set Identification in Models With Multiple Equilibria," *The Review of Economic Studies*, 78 (4), 1264–1298. [265]

GALICHON, ALFRED, AND BERNARD SALANIÉ (2022): "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," *The Review of Economic Studies*, 89 (5), 2600–2629. [283,284]

GIACOMINI, RAFFAELLA, TORU KITAGAWA, AND HARALD UHLIG (2019): "Estimation Under Ambiguity," Technical report, UCL and Chicago. [265]

GRIECO, PAUL L. E. (2014): "Discrete Games With Flexible Information Structures: An Application to Local Grocery Markets," *The RAND Journal of Economics*, 45 (2), 303–340. [268]

GUALDANI, CRISTINA, AND SHRUTI SINHA (2022): "Partial Identification in Matching Models for the Marriage Market," *Journal of Political Economy* (forthcoming). [265,286]

HANSEN, LARS PETER, AND THOMAS J. SARGENT (2001): "Robust Control and Model Uncertainty," *The American Economic Review*, 91 (2), 60–66. [264,271]

——— (2008): *Robustness*. Princeton. [264]

HO, PAUL (2022): "Global Robust Bayesian Analysis in Large Models," *Journal of Econometrics* (forthcoming). [265]

KALOUPTSIDI, MYRTO, YUICHI KITAMURA, LUCAS LIMA, AND EDUARDO SOUZA-RODRIGUES (2020): "Partial Identification and Inference for Dynamic Models and Counterfactuals," NBER Working Paper No. 26761. [265]

KALOUPTSIDI, MYRTO, PAUL T. SCOTT, AND EDUARDO SOUZA-RODRIGUES (2021): "Identification of Counterfactuals in Dynamic Discrete Choice Models," *Quantitative Economics*, 12 (2), 351–403. [265]

KITAMURA, YUICHI, TAISUKE OTSU, AND KIRILL EVDOKIMOV (2013): "Robustness, Infinitesimal Neighborhoods, and Moment Restrictions," *Econometrica*, 81 (3), 1185–1201. [265]

KLINE, BRENDAN, AND ELIE TAMER (2016): "Bayesian Inference in a Class of Partially Identified Models," *Quantitative Economics*, 7 (2), 329–366. [268]

KOMUNJER, IVANA, AND GIUSEPPE RAGUSA (2016): "Existence and Characterization of Conditional Density Projections," *Econometric Theory*, 32 (4), 947–987. [272]

LAFFÉRS, LUKÁŠ (2019): "Bounding Average Treatment Effects Using Linear Programming," *Empirical Economics*, 57 (3), 727–767. [265]

LEAMER, EDWARD E. (1982): "Sets of Posterior Means With Bounded Variance Priors," *Econometrica*, 50 (3), 725–736. [265]

——— (1985): "Sensitivity Analyses Would Help," *The American Economic Review*, 75 (3), 308–313. [264]

LI, LIXIONG (2018): "Identification of Structural and Counterfactual Parameters in a Large Class of Structural Econometric Models," Technical report, Pennsylvania State University. [265]

LIESE, FRIEDRICH, AND IGOR VAJDA (1987): *Convex Statistical Distances*. Leipzig: Teubner. [281]

MACCHERONI, FABIO, MASSIMO MARINACCI, AND ALDO RUSTICHINI (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74 (6), 1447–1498. [271]

MANSKI, CHARLES F. (2007): "Partial Identification of Counterfactual Choice Probabilities," *International Economic Review*, 48 (4), 1393–1410. [265]

——— (2014): "Identification of Income–Leisure Preferences and Evaluation of Income Tax Policy," *Quantitative Economics*, 5 (1), 145–174. [265]

MCFADDEN, DANIEL L. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*. Academic Press, 105–142. [263]

——— (1978): "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*. Amsterdam: North Holland, 75–96. [267]

MILGROM, PAUL, AND ILYA SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70 (2), 583–601. [276]

MUKHIN, YAROSLAV (2018): "Sensitivity of Regular Estimators," arxiv preprint, arXiv:1805.08883 [econ.em]. [265]

NORETS, ANDRIY, AND XUN TANG (2014): "Semiparametric Inference in Dynamic Binary Choice Models," *The Review of Economic Studies*, 81 (3), 1229–1262. [265,279]

OWEN, ART B. (2017): "A Randomized Halton Algorithm in R," arxiv preprint, arXiv:1706.02808 [stat.co]. [277]

QIAO, YU, AND NOBUAKI MINEMATSU (2010): "A Study on Invariance of f-Divergence and Its Application to Speech Recognition," *IEEE Transactions on Signal Processing*, 58 (7), 3884–3890. [281]

ROCKAFELLAR, R. TYRRELL (1970): *Convex Analysis*. Princeton University Press. [270]

RUST, JOHN (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55 (5), 999–1033. [263,265,268,279,287,288,290]

SCHENNACH, SUSANNE M. (2014): "Entropic Latent Variable Integration via Simulation," *Econometrica*, 82 (1), 345–385. [265]

SHAPIRO, ALEXANDER (1990): "On Concepts of Directional Differentiability," *Journal of Optimization Theory and Applications*, 66 (3), 477–487. [292]

——— (2017): "Distributionally Robust Stochastic Programming," *SIAM Journal on Optimization*, 27 (4), 2258–2275. [264]

SU, CHE-LIN, AND KENNETH L. JUDD (2012): "Constrained Optimization Approaches to Estimation of Structural Models," *Econometrica*, 80 (5), 2213–2230. [277]

TAMER, ELIE (2003): "Incomplete Simultaneous Discrete Response Model With Multiple Equilibria," *The Review of Economic Studies*, 70 (1), 147–165. [267]

——— (2015): "Sensitivity Analysis in Some Econometric Models," Cowles Lecture, Econometric Society World Congress, August 17–21, 2015, Montréal. [264]

TEBALDI, PIETRO, ALEXANDER TORGOVITSKY, AND HANBIN YANG (2022): "Nonparametric Estimates of Demand in the California Health Insurance Exchange," NBER working paper. [265]

TORGOVITSKY, ALEXANDER (2019a): "Nonparametric Inference on State Dependence in Unemployment," *Econometrica*, 87 (5), 1475–1505. [265]

——— (2019b): "Partial Identification by Extending Subdistributions," *Quantitative Economics*, 10 (1), 105–144. [265,286]

Co-editor Aviv Nevo handled this manuscript.

Manuscript received 15 April, 2019; final version accepted 18 July, 2022; available online 14 September, 2022.