COUNTERFACTUAL ANALYSIS UNDER PARTIAL IDENTIFICATION USING LOCALLY ROBUST REFINEMENT

Nathan Canen and Kyungchul Song

University of Houston and University of British Columbia

ABSTRACT. Structural models that admit multiple reduced forms, such as game-theoretic models with multiple equilibria, pose challenges in practice, especially when parameters are set-identified and the identified set is large. In such cases, researchers often choose to focus on a particular subset of equilibria for counterfactual analysis, but this choice can be hard to justify. This paper shows that some parameter values can be more “desirable” than others for counterfactual analysis, even if they are empirically equivalent given the data. In particular, within the identified set, some counterfactual predictions can exhibit more robustness than others, against local perturbations of the reduced forms (e.g. the equilibrium selection rule). We provide a representation of this subset which can be used to simplify the implementation. We illustrate our message using moment inequality models, and provide an empirical application based on a model with top-coded data.

KEY WORDS: Counterfactual Analysis, Multiple Equilibria, Equilibrium Selection Rule, Locally Robust Refinement

JEL CLASSIFICATION: C30, C57

Date: June 11, 2021.
We thank three anonymous referees for their valuable comments. Corresponding author: Kyungchul Song, kysong@mail.ubc.ca, Vancouver School of Economics, University of British Columbia, 6000 Iona Drive, Vancouver, BC, V6t 1Z4, Canada.
1. Introduction

Economists often use a structural model to analyze the effect of a policy that has never been implemented, such as potential mergers or changes to legislation. For such analysis, it is crucial to have a plausible specification of the causal relationship between the policy variable and the outcome variable of interest. However, a plausible specification can come short of determining the causal relationship uniquely. Prominent examples are found in game theoretic models, where the presence of multiple equilibria admits multiple reduced forms. (Tamer (2003) called models with multiple reduced forms incomplete.)\(^1\) The multiplicity of reduced forms is often the source of set-identification of structural parameters.

When a parameter is set-identified, all the values in the identified set are empirically equivalent, in the sense that the observed data cannot be used to distinguish between the values. The identified set represents the empirical content of the model for the parameter (i.e., the amount of information on the value of the parameter that can be extracted from data through the model). However, when a parameter is set-identified with a large identified set, the identified set can be of little use for counterfactual analysis in practice. It is not unusual that predictions from counterfactual analysis based on different equilibria yield self-contradictory results (Aguirregabiria and Nevo (2013), p.111). In this situation, just as in the calibration approach in macroeconomics, the researcher typically uses additional constraints on the parameter values or chooses a specific equilibrium to obtain a meaningful analysis.\(^2\) In other words, the researcher considers a refinement of the identified set, by introducing additional restrictions. However, these restrictions are often only weakly motivated.

---

\(^1\)In this paper, a reduced form refers to a system of equations where the variables on the right hand side are all external variables (i.e. inputs) and the variables on the left hand side are all internal variables (i.e. generated through the system of equations using the inputs) in the sense of Heckman (2010) (p.56). The term “reduced form” in this sense appears in econometrics at least as early as in Koopmans (1949).

\(^2\)For example, in an entry model, Jia (2008) focuses on the equilibrium that is most profitable to one of the firms, and Ackerberg and Gowrisankaran (2006) use an equilibrium selection rule which randomizes between a Pareto-best equilibrium and a Pareto-worst equilibrium with an unknown probability. Another example is Roberts and Sweeting (2013) in the context of auctions, who consider the particular equilibrium with the lowest signal for agents in the model so they enter the auction.
In our view, a desirable approach in this situation is to distinguish between the identification step and the refinement step, and lay out transparently what criterion (or motivation) is used for the refinement step. There is a small body of literature that explores this refinement step. Song (2014) analyzed the problem of choosing a point from the identified interval based on the local asymptotic minimax regret criterion. Aryal and Kim (2013) considered an incomplete English auction model where the seller is ambiguity averse, and employing the Γ-maximin expected utility framework of Gilboa and Schmeidler (1989), suggested a point decision on the optimal reserve price. Jun and Pinkse (2017) considered an incomplete English auction set-up similar to that in Haile and Tamer (2003) and proposed a point decision on the optimal reserve price using a maximum entropy method. Jun and Pinkse (2019) studied a discrete two-player complete information game and established sharp identification of counterfactual predictions of the game. Furthermore, they investigated point decisions on the probability distributions through a maximum entropy method and what they called a Dirichlet approach, and compared the two methods.

The main goal of this paper is to draw attention to the fact that some values in the identified set can give counterfactual predictions that are more robust to the change of certain aspects of model specification than others. Thus, it is reasonable to give priority to those values that produce most “reliable” counterfactual predictions. Generally speaking, predictions that are sensitive to a change of the aspects of the model that the researcher is not sure about cannot be reliable. Given that the researcher is often least sure about how a reduced form is determined among multiple reduced forms, we consider a subset of the identified set which produces counterfactual predictions that are more robust to local perturbations of the reduced forms (e.g. local perturbations of the equilibrium selection rule in a game setting) than others.

More specifically, consider a causal relationship between endogenous variable $y$ and exogenous variables $x, \epsilon$: 

$$y = \rho_{\beta, y}(x, \epsilon; \eta),$$  

(1.1)
where \( \beta \) is the parameter of interest, \( \gamma \) a nuisance parameter, and \( \eta \) is a random variable with distribution \( G \). Without observing \( \eta \) or knowing \( G \), the model admits multiple reduced forms. For example, in the case of games with multiple equilibria, \( G \) plays the role of the equilibrium selection rule.\(^3\)

Suppose that the counterfactual setting of interest involves a counterfactual distribution \( \tilde{F}_X \) for \( X \). Consider the average structural function (ASF) (Blundell and Powell (2004)):

\[
\text{ASF}_{\beta,\gamma,G}(x) = \int \int \rho_{\beta,\gamma}(x, \epsilon; \eta)dG(\eta|x, \epsilon)dF_\epsilon(\epsilon),
\]

where \( F_\epsilon \) denotes the distribution of \( \epsilon \), and \( G(\cdot|x, \epsilon) \) the conditional distribution of \( \eta \) given \( x \) and \( \epsilon \). Then our proposal is that we choose \( \gamma \) that minimizes

\[
\sup_{G' \in B(G; \delta)} \left| \int \text{ASF}_{\beta,\gamma,G}(x)d\tilde{F}_X(x) - \int \text{ASF}_{\beta,\gamma,G'}(x)d\tilde{F}_X(x) \right|,
\]

(1.2)

where \( B(G; \delta) \) denotes a \( \delta \)-neighborhood of \( G \) for some small \( \delta > 0 \). (As we show later, given \( \beta \), the set of \( \gamma \)'s that minimize (1.2) does not depend on the choice of \( \delta \) or \( G \).) We call the resulting set of values \((\beta, \gamma)\) in the identified set the Locally Robust Refinement (LRR). In other words, our refinement is motivated by the desirability of a stable behavior in the counterfactual predictions when the structural parameters are partially identified.

In many set-ups, the LRR is a substantially smaller set than the identified set. Furthermore, the LRR is often very intuitive. For example, as we show in Section 3, the LRR in an entry game picks nuisance parameter values that minimize the region of multiple equilibria. In the example of a switching regression model with interval data, it picks nuisance parameters that minimize a difference across models in switching regression.

However, unlike the identified set, the size of the LRR does not indicate the empirical content of the model. Rather the LRR gives parameter values which produce counterfactual predictions that are most robust to the change of the reduced forms among their multiple

\(^3\)As shown in the paper, our approach applies to causally incomplete models other than game-theoretic models, such as models involving top-coded data.
forms. Due to this robustness property, one may want to pay more attention to those parameter values than others when conducting policy exercises, even if all the values in the identified set are observationally equivalent.

The formulation in (1.2) gives the impression that the LRR involves additional optimization over an infinite dimensional space in estimation, adding to computational cost. Using the linearity of $\text{ASF}_{\beta,\gamma,G}(x)$ in $G$ and Hilbert space geometry, this paper gives a simple characterization of the LRR so that focusing on the LRR does not cause extra computational cost. In fact, due to the additional constraints that come from the LRR, the computational cost is substantially reduced from a benchmark case without using the LRR. Focusing on moment inequality models, this paper presents a method of bootstrap inference on the LRR from the identified set and shows its uniform asymptotic validity. We provide results from Monte Carlo simulations, and illustrate the LRR through an empirical application on top-coded data. These exercises illustrate that the LRR’s predictions are indeed more stable than those with the identified set and that they may affect policymaking in applied contexts (i.e. policies suggested by the identified set could be driven by parameter values that are not robust, and are not part of the LRR).

There has long been interest in studying model sensitivity and robustness to misspecifications in econometric models. Recent contributions in this line include Andrews, Gentzkow, and Shapiro (2020), Armstrong and Kolesár (Forthcoming) and Bonhomme and Weidner (2020) who, among many others, proposed a method of measuring local misspecification and performing inference that is robust to local misspecifications. In the context of instrumental variable models, Nevo and Rosen (2012) permitted the instrumental variable to be correlated with the error term and presented bounds for the parameter of interest under an assumption on the correlation between the instrumental variable and the error term. Oster (2019) proposed a method of extracting information on the omitted variable bias from data in linear regression models. Masten and Poirier (2018) provided an identification analysis of various treatment effect parameters from treatment effect models where the conditional
independence restriction is weakened to a partial dependence condition.\footnote{Robustness to misspecification has also been considered in the macroeconomics literature (e.g. in optimal control theory as in Hansen and Sargent (2001), see Hansen and Marinacci (2016) as well), where one can consider perturbations of the decision maker’s model, but less emphasis is placed on the identification and inference problems in these models. Finally, Andrews, Gentzkow, and Shapiro (2017) focus on the case where the model is point-identified, and present a statistic on how estimates vary when a model’s moments changes.} Christensen and Connault (2019) proposed a computationally tractable method of performing inference on counterfactual quantities that arise in structural models, while permitting a global misspecification of the model. Li (2019) uses a semiparametric restriction (instead of parametric ones) to generate bounds for counterfactuals of interest in moment inequality models.

All these authors are concerned about robustness of inference to misspecifications. Their robust inference often leads to a larger confidence set, and a wider range of policy predictions. In contrast, we consider settings where the identified set is already too large to be useful for a meaningful policy analysis. This can happen even if the model is correctly specified (see Aguirregabiria and Nevo (2013) for a discussion, particularly in the context of multiple equilibria). Instead of enlarging an existing identified set by “robustifying” it as above, we explore a rationale for focusing on a subset of the identified set for policy analysis. Our criterion is based on the stability of counterfactual predictions as we perturb the reduced-form selection rules. One can relate our guiding philosophy to the literature of equilibrium refinements in game theory, where one focuses on a subset of the set of Nash equilibria based on its stability against various perturbations of the game. (See Chapter 5 of Myerson (1991).) Then the idea of local robustness in the paper, i.e., the stability of reduced forms against local perturbations of reduced form selection rules is related to Fudenberg, Kreps, and Levine (1988) who considered robustification of the set of strict Nash equilibria against local perturbations of the games.

The paper is organized as follows. The next section gives a general set-up of a structural model and presents examples that fit our framework. Section 3 presents the approach of LRR and provides its characterization that is useful for implementation in practice. Focusing on generic moment inequality models, Section 4 proposes inference for the LRR and presents
results from Monte Carlo simulations. An empirical application is presented in Section 5. Section 6 concludes. In the Appendix, we provide some mathematical proofs. The Supplemental Note to this paper includes additional information on the data for the empirical application, and presents results on the uniform validity of our inference procedure.

2. A Counterfactual Analysis with Multiple Reduced Forms

2.1. Models with Multiple Reduced Forms

Suppose that data \{\((Y'_i, X'_i)\)\}_{i=1}^{n}
are drawn from the joint distribution \(P\) of \((Y', X')'\) which
obeys the following reduced form: for \(Y \in \mathbb{R}\) and \(W = (X', \epsilon')'\),

\[ Y = \rho(W; \eta), \tag{2.1} \]

for some function \(\rho\), where \(X\) is observed but \(\epsilon\) and \(\eta\) are not. For each value \(\eta\), we regard
the map \(\rho(\cdot; \eta)\) as representing a functional-causal relationship between internal variable
\(Y\) and external variables \(W\).

This relationship is fully determined once the value of \(\eta\) is realized. Typically, \(\epsilon\) has a structural meaning such as unobserved costs, whereas \(\eta\) does not
and is used only as an index for a reduced form. For example, in a game-theoretic model,
\(W = (X', \epsilon')'\) represents observed and unobserved payoff components and \(\eta\) an index for an
equilibrium. The response of \(Y\) to a change in \(W\) is ambiguous unless the value of \(\eta\) is fixed.
In this sense, there is a multiplicity of reduced forms in (2.1). To describe the generation
of observed outcome \(Y\), we assume that Nature selects the value of \(\eta\) from the conditional
distribution \(G(\cdot|w)\) given \(W = w\). In a game setting, \(G\) is often called the equilibrium selection
rule. However, the multiplicity of reduced forms can arise in a setting that has nothing to do

---

5Here we follow the proposal by Heckman (2010) (p.56) and use the terms internal variables and external variables. External variables are specified from outside of the structural equations as inputs, and internal variables are generated through the system of equations from the inputs. External variables are not necessarily exogenous variables. When \(X\) is correlated with \(\epsilon\), the variable \(X\) is called endogenous, although \((X, \epsilon)\) is jointly given externally as inputs, and thus is a vector of external variables. Here a reduced form refers to a system of equations where all the variables on the left hand side are internal variables and all the variables on the right hand side are external variables.
with a game, such as in the case of switching regressions with interval data which we describe below. Hence we call $G$ the reduced-form selection rule in this paper.

2.2. Counterfactual Analysis Under Partial Identification

We focus on a counterfactual experiment where the distribution $F$ of the external vector $W$ is changed to a counterfactual one $\tilde{F}$. Such a counterfactual experiment includes cases where one “shuts off” the effect of a certain covariate, “fixing” of the covariate to a certain value, or changing the distribution of $\epsilon$. Suppose that the counterfactual quantity of interest is generically denoted by $C(\rho, \tilde{F}, G)$. This quantity could be, for example, the prediction of game outcomes or expected payoffs under a counterfactual scenario:

$$C(\rho, \tilde{F}, G) = \int \rho(w; \eta)dG(\eta|w)d\tilde{F}(w).$$

Or in other applications, $C(\rho, \tilde{F}, G)$ could be the average increase in welfare. Let $\mathcal{G}$ be the space of the reduced-form selection rules $G$ and let $\mathcal{R}$ be the identified set for the function $\rho$. Then the identified set for the counterfactual prediction $C(\rho, \tilde{F}, G)$ is given by

$$\{C(\rho, \tilde{F}, G) : G \in \mathcal{G}, \rho \in \mathcal{R}\}.$$

A standard approach to make inference based on this identified set faces multiple challenges in practice. The set of predictions can be of little practical use, when the estimated set is too large. Furthermore, estimation of the identified set itself and statistical inference can be computationally intensive.

Due to these difficulties, it is not unusual that empirical researchers focus on a particular subset of $\mathcal{G}$ (such as focusing on a particular equilibrium) or consider a subset of $\mathcal{R}$ (such as by restricting the admissible values of parameters as in the case of calibration). Such a subset selection approach, while inevitable in practice, is not easy to motivate.
2.3. Examples

For brevity, we consider two examples here to illustrate our set-up: a finite game of complete information and a switching regression with interval data.

2.3.1. Finite Game of Complete Information. Consider the setting of a complete information $N$ player game with finite actions, and let $A$ be the finite action space for each player. Suppose that the payoff for player $i$ playing $\tilde{a} \in A$ when all other players are playing $a_{-i} = (a_j)_{j \neq i}$ is given by $u_i(\tilde{a}, a_{-i}, W)$ for a map $u_i$, where $W = (X, \epsilon)$ is a payoff-relevant characteristic vector of player $i$. For each $a \in A^N$, we define

$$\mathbb{W}(a) = \left\{ w : u_i(a_i, a_{-i}, w) \geq \max_{a' \in A} u_i(a', a_{-i}, w), \forall i = 1, ..., N \right\}. \tag{2.2}$$

Hence if $w$ belongs to $\mathbb{W}(a)$, $a$ is a Nash equilibrium for the game with $w$, so that the set

$$A(w) = \{ a \in A^N : w \in \mathbb{W}(a) \} \tag{2.3}$$

represents the set of equilibria for the game with $w$. Let $\eta$ be a random vector taking values in $A^N$ where the conditional distribution $G(\cdot | w)$ of $\eta$ given $w$ has a support in $A(w)$. Let $Y = (Y_i)_{i=1}^N$ be the observed action profile that comes from a Nash equilibrium $\eta$ of the game after Nature draws $\eta$ from $G(\cdot | w)$. Then, we can write for $i \in N$,

$$Y_i = \rho_i(W; \eta), \tag{2.4}$$

where $\rho_i$ is the $i$-th entry of $\rho$ which is defined as

$$\rho(w; \eta) = \sum_{a \in A(w)} 1\{\eta = a\} a.$$

The conditional distribution $G(\cdot | w)$ over $\eta$ represents the equilibrium selection rule.

---

6A special case is the two player entry game used by Bresnahan and Reiss (1991) and Tamer (2003), among others.
2.3.2. A Switching Regression Model with Interval Outcomes. Consider an interval data set-up similar to one studied by Manski and Tamer (2002). More specifically, we consider top-coded data with threshold variables $Z_1 < Z_2$, where $Z_1, Z_2$ are observed and common across individuals. We observe an outcome random variable which equals a latent $Y$ if $Y \leq Z_1$ and is only known to be within $[Z_1, Z_2]$ otherwise, and a vector of covariates $X_1$. For example, $X_1$ represents a vector of covariates for each worker such as age, education, experience, gender, and $Y$ is the log of the hourly wage.

It is assumed that the latent $Y$ follows:

$$Y = \begin{cases} g_1(\beta_0 D + X'_1 \gamma_0) + \epsilon, & \text{if } \eta = 1, \text{ and} \\ g_2(\beta_0 D + X'_1 \gamma_0) + \epsilon, & \text{if } \eta = 0, \end{cases}$$

where $D$ denotes the policy variable, and $\eta$ the binary indicator of a regime, and $g_1$ and $g_2$ are some known functions. As for the error term $\epsilon$, we assume that $\mathbb{E}[\epsilon|X_1, D] = 0$. We also assume that $Y \leq Z_2$ with probability one. Let $\bar{Y}_1 = \bar{Y}_2 = Y$ if $Y \leq Z_1$, and $\bar{Y}_1 = Z_1$ and $\bar{Y}_2 = Z_2$ if $Y > Z_1$. The econometrician observes $\bar{Y}_1$ and $\bar{Y}_2$. Let $\pi_0(X_1, D) = P\{\eta = 1|X_1, D\}$ and $\Pi$ be the collection of functions that $\pi_0$ is known to belongs to.

It is straightforward that the identified set for $(\beta_0, \gamma_0)$ is given by

$$\Theta_p = \{ (\beta_0, \gamma_0) : \mathbb{E}[\bar{Y}_1|X_1, D] \leq \alpha(X_1, D; \beta_0, \gamma_0, \pi) \leq \mathbb{E}[\bar{Y}_2|X_1, D], \text{a.e. for some } \pi \in \Pi \}, \quad (2.5)$$

where

$$\alpha(X_1, D; \beta_0, \gamma_0, \pi) = g_1(\beta_0 D + X'_1 \gamma_0) \pi(X_1, D) + g_2(\beta_0 D + X'_1 \gamma_0) (1 - \pi(X_1, D)). \quad (2.6)$$

Now we can write

$$Y = \rho(X, \epsilon; \eta),$$
where $X = (X_1, D)$, and

$$
\rho(X, \epsilon; \eta) = g_1(\beta_0 D + X'_1 \gamma_0) \eta + g_2(\beta_0 D + X'_1 \gamma_0)(1 - \eta) + \epsilon.
$$

(2.7)

The distribution $G$ of $\eta$ plays a role analogous to the equilibrium selection rule in a game-theoretic model. Here it selects from a given set of reduced forms.

3. Locally Robust Refinement

3.1. Overview

Suppose that the reduced-form $\rho(\cdot; \eta)$ in (2.1) is parametrized as follows:

$$
Y = \rho_{\beta_0, \gamma_0}(W; \eta),
$$

(3.1)

where $\beta_0$ is a parameter of interest taking values in a set $B$ and $\gamma_0$ is a nuisance parameter taking values in a set $\Gamma$, and $\eta$ belongs to some known compact set denoted by $H$. Let $\Theta_p$ be the identified set for $(\beta_0, \gamma_0)$. The counterfactual experiment involves changing the distribution $F$ of $W$ to a counterfactual distribution $\tilde{F}$. Let us consider the counterfactual average structural outcome:

$$
\text{ASO}_{\beta_0, \gamma_0}(G) = \int \int \rho_{\beta_0, \gamma_0}(w; \eta)dG(\eta|w)d\tilde{F}(w).
$$

(3.2)

This is the average value of outcome $Y$ when a counterfactual change is made to the distribution of $W$ while everything else remains the same. Our focus is on a subset of $\Theta_p$ at which $\text{ASO}_{\beta_0, \gamma_0}(G)$ behaves robustly as we perturb the reduced-form selection rule $G$.

More formally, for any $K > 0$, we introduce a $K$-sensitivity of $\text{ASO}_{\beta_0, \gamma_0}(G)$ as we perturb $G$ around in its $K$-neighborhood:

$$
S_{\beta_0, \gamma_0}(G; K) = \sup_{G' \in G: \delta_G(G', G) \leq K} \frac{|\text{ASO}_{\beta_0, \gamma_0}(G') - \text{ASO}_{\beta_0, \gamma_0}(G)|}{\delta_G(G', G)},
$$

(3.3)
where $\mathcal{G}$ denotes the collection of $G$’s permitted in the model. The quantity $S_{\beta,\gamma}(G;K)$ measures the maximal rate of change of counterfactual average outcome when $G$ is perturbed within its $K$-neighborhood, where this neighborhood is defined using a metric $\delta_{\mathcal{G}}(G', G)$, formally defined below. We search for values of nuisance parameters $\gamma$ which minimize this sensitivity. For each $\beta \in B$, if we let

$$\Gamma_{LRR}(\beta) = \arg\min_{\gamma: (\beta, \gamma) \in \Theta_p} S_{\beta,\gamma}(G;K), \tag{3.4}$$

we define

$$\Theta_{p}^{LRR} = \{ (\beta, \gamma) \in \Theta_p : \gamma \in \Gamma_{LRR}(\beta) \}. \tag{3.5}$$

We call the subset $\Theta_{p}^{LRR}$ the Locally Robust Refinement (LRR) of $\Theta_p$. Hence, the LRR can be viewed as a collection of parameters at which the ASO is stable against local misspecifications of $G$. As we will see later in the characterization of the LRR set, the set is independent of the choice of $K$, $G$ and $\mathcal{G}$. The main reason behind this independence is that the functional $\text{ASO}_{\beta,\gamma}(G)$ is linear in $G$. (This is analogous to that the slope of a linear function is constant on the domain of the function.)\footnote{As long as the true reduced-form selection rule lies in the interior of $\mathcal{G}$, our characterization of the LRR set does not depends on the choice of $\mathcal{G}$. While it is possible to recover some information about the rule $G$ with further specifications, our procedure remains agnostic about $\mathcal{G}$, which is in line with many existing works in the econometrics literature of game-theoretic models - see Ciliberto and Tamer (2009) for one such example.} Note that, as we do not want our concern of local robustification against the reduced-form selection rules to directly affect the parameter of interest, our approach subjects only the nuisance parameters to the refinement.

The inference on $\beta$ based on the Locally Robust Refinement proceeds under the further restriction as follows.

$$(\beta_0, \gamma_0) \in \Theta_{p}^{LRR}. \tag{3.6}$$

We call this restriction the LRR Hypothesis. Once we focus on the LRR set of parameters, this can be used for various exercises of counterfactual analysis, just as one does with the identified set of parameters. Despite the impression that finding the set $\Theta_{p}^{LRR}$ might be computationally
intensive as it involves layers of optimization, it is not, due to our characterization result in the next section.

3.2. Characterization of the Locally Robust Refinement

To derive a characterization of the LRR, let us introduce some geometric structure on the space of reduced-form selection rules \( G \). Let \( \mathcal{G} \) be the collection of all the conditional distributions \( G \) (of \( \eta \) given \( W \)) which are dominated by the uniform distribution \( \mu \) over \( H \), and \( w \in \mathcal{W} \), with \( \mathcal{W} \) denoting the set \( W \) takes values from. Let us introduce the following metric on \( \mathcal{G} \): for \( G, G' \in \mathcal{G} \),

\[
\delta_{\mathcal{G}}(G', G) = \sqrt{\int \int \left( \frac{dG'}{d\mu} - \frac{dG}{d\mu} \right)^2 (\eta | w) d\mu(\eta) d\tilde{F}_p(w)}. \tag{3.7}
\]

Let us define

\[
\Gamma(\beta) = \{ \gamma \in \Gamma : (\beta, \gamma) \in \Theta_p \}, \tag{3.8}
\]

i.e., the set of nuisance parameters that are admitted in the counterfactual analysis together with given \( \beta \). For \( \kappa \geq 0 \), we define the \( \kappa \)-LRR set for \( \gamma \) corresponding to \( \beta \) as

\[
\Gamma^{\text{LRR}}_\kappa(\beta) = \{ \gamma \in \Gamma(\beta) : S^2_{\beta, \gamma}(G; K) \leq S^2_{\beta, \gamma}(G; K) + \kappa, \forall \tilde{\gamma} \in \Gamma(\beta) \}, \tag{3.9}
\]

where \( S_{\beta, \gamma}(G; K) \) is defined in (3.3). The set \( \Gamma^{\text{LRR}}(\beta) \) defined in (3.4) is a special case of \( \Gamma^{\text{LRR}}_\kappa(\beta) \) with \( \kappa = 0 \). However, as we discuss in Section 4.1, it is convenient to use the LRR set with a small positive number \( \kappa > 0 \) which allows us to develop uniform inference. As shown below, this set does not depend on the choice of \( K > 0 \) or that of \( G \in \mathcal{G} \).

We now provide a useful characterization of the set \( \Gamma^{\text{LRR}}_\kappa(\beta) \). For this, we define

\[
\Delta_{\beta, \gamma}(w; \eta) = \rho_{\beta, \gamma}(w; \eta) - \int \rho_{\beta, \gamma}(w; \eta) d\mu(\eta).
\]
The function $\Delta_{\beta,\gamma}(w; \eta)$ is a mean-deviation form of the structural function $\rho_{\beta,\gamma}(w; \eta)$. This function plays a central role in the characterization result. Let us make the following assumption that guarantees that the average variability of this mean-deviation is well defined.

**Assumption 3.1.** For each $\beta \in B$ and $\gamma \in \Gamma(\beta)$, $Q^{LRR}(\beta, \gamma) < \infty$, where

$$Q^{LRR}(\beta, \gamma) = \int \int \Delta^2_{\beta,\gamma}(w; \eta) d\mu(\eta) d\tilde{F}(w). \quad (3.10)$$

In applications where we can compute $\rho_{\beta,\gamma}(w; \eta)$ for each value of $\eta$, we can evaluate $Q^{LRR}(\beta, \gamma)$ easily. (See Section 3.3 below for examples.) The following theorem shows that the set $\Gamma_L^{LRR}(\beta)$ is fully characterized through the function $Q^{LRR}(\beta, \gamma)$.

**Theorem 3.1.** Suppose that Assumption 3.1 holds. Then for each $\beta \in B$ and $\gamma \in \Gamma(\beta)$,

$$S_{\beta,\gamma}(G; K) = \sqrt{Q^{LRR}(\beta, \gamma)}, \quad (3.11)$$

for all $G \in \mathcal{G}$ and all $K > 0$.

Thus the $\kappa$-LRR of $\gamma_0$ can be rewritten as

$$\Gamma_L^{LRR}(\beta) = \{ \gamma \in \Gamma(\beta) : Q^{LRR}(\beta, \gamma) \leq Q^{LRR}(\beta, \bar{\gamma}) + \kappa, \text{ for all } \bar{\gamma} \in \Gamma(\beta) \}. \quad (3.12)$$

The $\kappa$-LRR of the nuisance parameter vector $\gamma_0$ is the set of $\gamma$'s which minimize the average variability of $\rho_{\beta,\gamma}(w; \eta)$ (around its mean) as we perturb $\eta$. Thus, the $\kappa$-LRR for $(\beta_0, \gamma_0)$ is given by

$$\Theta^{LRR}_{\kappa, p} = \{ (\beta, \gamma) \in \Theta_p : \gamma \in \Gamma_L^{LRR}(\beta) \}. \quad (3.13)$$

Sometimes, the target parameter may not be a parameter $\beta$ that determines the reduced form $\rho_{\beta,\gamma}$ but some aggregate quantity such as Average Treatment Effects or aggregate welfare. Suppose that such a target parameter is identified up to $\beta_0$ and $\gamma_0$, so that we can write it as, say $\psi(\beta_0, \gamma_0)$. (We give an example in Section 3.3.2 below.) Then, the LRR set for
ψ(β₀, γ₀) is given by

\[ \Psi_{k}^{\text{LRR}} = \left\{ \psi(\beta, \gamma) : (\beta, \gamma) \in \Theta_{k,p}^{\text{LRR}} \right\}. \]

To build familiarity with our characterization results for the LRR and show how it can be used in practice, we now derive \( Q_{k}^{\text{LRR}} \) for two applied examples.

### 3.3. Examples

#### 3.3.1. Entry Game with Nash Equilibria

Consider the setting of a 2x2 complete information entry games, such as the empirical setting in Bresnahan and Reiss (1991) and Ciliberto and Tamer (2009), where there are two firms, \( i = 1, 2 \), deciding whether to enter a market or not. The entry decisions of the firms can then be formulated as

\[
D_1 = 1\{\beta_{1,0}D_2 + X'\gamma_{1,0} \geq \epsilon_1\}, \quad \text{and} \quad \epsilon_1 \geq 0 \tag{3.15}
\]

\[
D_2 = 1\{\beta_{2,0}D_1 + X'\gamma_{2,0} \geq \epsilon_2\}, \quad \epsilon_2 \geq 0 \tag{3.16}
\]

where \( \beta_{1,0}D_2 + X'\gamma_{1,0} - \epsilon_1 \) represents the payoff difference for firm 1 between entering and not entering the market, \( D_i = 1 \) represents the entry decision by firm \( i \) and \( D_i = 0 \) the decision not to enter by firm \( i \). The random vectors \( X \) and \( \epsilon \) represent respectively the observed and unobserved characteristics of firms 1 and 2 combined. The coefficients \( \beta_{1,0} \) and \( \beta_{2,0} \) capture strategic interactions between firms and they are our coefficients of interest, whereas the coefficients \( \gamma_{1,0} \) and \( \gamma_{2,0} \) measure the roles of firm specific and market specific characteristics and, as such, are nuisance parameters.

[INSERT FIGURE 1 HERE]

In order to find a reduced form as in (2.1), we first put

\[
Y = (D_1, D_2), \quad \text{and} \quad \epsilon = (\epsilon_1, \epsilon_2),
\]
where \( Y \in \mathcal{Y} = \{(0, 0), (1, 1), (0, 1), (1, 0)\} \). Define (with \( \beta_0 = (\beta_{1,0}^{'}, \beta_{2,0}^{'}) \) and \( \gamma_0 = (\gamma_{1,0}^{'}, \gamma_{2,0}^{'}) \))

\[
A_{1,\beta_0,\gamma_0}(X) = \{(e_1, e_2) \in \mathbb{R}^2 : X_1' \gamma_{1,0} < e_1, X_2' \gamma_{2,0} < e_2\}
\]

\[
A_{2,\beta_0,\gamma_0}(X) = \{(e_1, e_2) \in \mathbb{R}^2 : \beta_{1,0} + X_1' \gamma_{1,0} \geq e_1, \beta_{2,0} + X_2' \gamma_{2,0} \geq e_2\}
\]

\[
A_{3,\beta_0,\gamma_0}(X) = \{(e_1, e_2) \in \mathbb{R}^2 : \beta_{1,0} + X_1' \gamma_{1,0} < e_1, X_2' \gamma_{2,0} \geq e_2\}, \text{ and}
\]

\[
A_{4,\beta_0,\gamma_0}(X) = \{(e_1, e_2) \in \mathbb{R}^2 : X_1' \gamma_{1,0} \geq e_1, \beta_{2,0} + X_2' \gamma_{2,0} < e_2\}.
\]

These regions are represented on the left panel of Figure 1.

Let \( \rho_{\beta,\gamma}(x, e; \eta) = (\rho_{\beta,\gamma,1}(x, e; \eta), \rho_{\beta,\gamma,2}(x, e; \eta)) \), where for a random variable \( \eta \in \{0, 1\} \),

\[
\rho_{\beta,\gamma,1}(x, e; \eta) = 1\{(e_1, e_2) \in A_{2,\beta,\gamma}(x)\} + 1\{(e_1, e_2) \in A_{4,3,\beta,\gamma}(x)\}
\]

\[
+ 1\{(e_1, e_2) \in A_{3,4,\beta,\gamma}(x)\} \eta
\]

\[
\rho_{\beta,\gamma,2}(x, e; \eta) = 1\{(e_1, e_2) \in A_{2,\beta,\gamma}(x)\} + 1\{(e_1, e_2) \in A_{3,4,\beta,\gamma}(x)\}
\]

\[
+ 1\{(e_1, e_2) \in A_{3,4,\beta,\gamma}(x)\} \eta,
\]

where \( A_{3,4,\beta,\gamma}(x) = A_{3,\beta,\gamma}(x) \cap A_{4,\beta,\gamma}(x), A_{4,3,\beta,\gamma}(x) = A_{4,\beta_0,\gamma_0}(x) \setminus A_{3,\beta_0,\gamma_0}(x), \) and \( A_{3,4,\beta,\gamma}(x) = A_{3,\beta_0,\gamma_0}(x) \setminus A_{4,\beta_0,\gamma_0}(x) \). Then we can write reduced forms for \( Y_i \), for firm \( i = 1, 2 \) as

\[
Y_i = \rho_{\beta_0,\gamma_0,i}(X, e; \eta).
\]

Let the distribution \( G \) of \( \eta \) be a discrete distribution with support \( \{0, 1\} \), and take \( \mu \) to be a Bernoulli random variable with a mean \( 1/2 \). It follows that for \( i = 1, 2 \),

\[
\rho_{\beta,\gamma,i}(x, e; \eta) = \int \rho_{\beta,\gamma,i}(x, e; \eta) d\mu(\eta) = \left( \eta - \frac{1}{2} \right) 1\{(e_1, e_2) \in A_{3,4,\beta,\gamma}(x)\}.
\]

Hence, by Theorem 3.1, the LRR set based on the ASO of \( Y_1 \) is the same as that based on the ASO of \( Y_2 \). Writing the mean-deviated reduced form on the left hand side of (3.19) as

\footnote{This example focuses on pure-strategy Nash equilibria following the empirical literature on entry-exit games (e.g. Tamer (2003), Jia (2008), Ciliberto and Tamer (2009), among many others).}
\[ \Delta \beta, \gamma (x, \epsilon; \eta), \] we find that

\[ \Delta^2 \beta, \gamma (x, \epsilon; \eta) = \left( \eta - \frac{1}{2} \right)^2 1\{(\epsilon_1, \epsilon_2) \in A_{3\cap 4, \beta, \gamma}(x)\}. \]

Plugging this into (3.10), we find that

\[
Q_{LRR}(\beta, \gamma) = \int \int \left( \eta - \frac{1}{2} \right)^2 d\mu(\eta) 1\{(\epsilon_1, \epsilon_2) \in A_{3\cap 4, \beta, \gamma}(x)\} d\tilde{F}_{X, \epsilon}(x, \epsilon),
\]

\[ = \frac{1}{4} \int 1\{(\epsilon_1, \epsilon_2) \in A_{3\cap 4, \beta, \gamma}(x)\} d\tilde{F}_{X, \epsilon}(x, \epsilon). \tag{3.20} \]

Under the assumption of independence of \((X, \epsilon)\), we can rewrite (3.20) as:

\[
Q_{LRR}(\beta, \gamma) = \frac{1}{4} \int P\{(\epsilon_1, \epsilon_2) \in A_{3\cap 4, \beta, \gamma}(x)\} d\tilde{F}_{X}(x). \tag{3.21} \]

We find the LRR by using this \(Q_{LRR}(\beta, \gamma)\) in (3.12).

When \((\epsilon_1, \epsilon_2)\) fall in the region \(A_{4, \beta, \gamma}(x) \cap A_{3, \beta, \gamma}(x)\), the game exhibits multiple equilibria. Our counterfactuals are most robust if this region is minimal. We cannot altogether disregard this region, because observations in the data may permit this region and it is possible they may arise in the counterfactual setting. Nevertheless, we can consider the value of the nuisance parameter(s) (in their identified set) that minimize the average region of multiple equilibria for the counterfactuals. (See Figure 1 for an illustration of this observation.)

3.3.2. A Switching Regression Model with Interval Outcomes. Let us look for \(Q_{LRR}(\beta, \gamma)\) for the example from Section 2.3.2. We take \(G\) to be a discrete distribution over \(\{0, 1\}\) and \(\mu\) be the uniform distribution over \(\{0, 1\}\). Hence if we write

\[
\rho_{\beta_0, \gamma_0}(X, \gamma_0; \eta) = g_1 (\beta_0 D + X'_1 \gamma_0) \eta + g_2 (\beta_0 D + X'_1 \gamma_0) (1 - \eta) + \epsilon, \tag{3.22} \]
we find that
\[
\Delta_{\beta_0, \gamma_0}(X, \epsilon; \eta) = \rho_{\beta_0, \gamma_0}(X, \epsilon; \eta) - \int \rho_{\beta_0, \gamma_0}(X, \epsilon; \eta) d\mu(\eta)
\]
\[
= \left( \eta - \frac{1}{2} \right) \left( g_1 (\beta_0 D + X'_1 \gamma_0) - g_2 (\beta_0 D + X'_1 \gamma_0) \right).
\]

Therefore, given a counterfactual distribution \( \tilde{F}(x, \epsilon) \) of \( (X, \epsilon) \),
\[
Q^{LRR}(\beta, \gamma) = \int \int \Delta_{\beta, \gamma}^2(x, \epsilon; \eta) d\mu(\eta) d\tilde{F}(x, \epsilon)
\]
\[
= \frac{1}{4} \int \left( g_1 (\beta d + x'_1 \gamma) - g_2 (\beta d + x'_1 \gamma) \right)^2 d\tilde{F}_{X_1, D}(x_1, d),
\]
where \( \tilde{F}_{X_1, D} \) denotes the joint distribution of \( (X_1, D) \) under \( \tilde{F} \). Thus, the \( \kappa \)-LRR of \( \gamma_0 \) can be rewritten as
\[
\Gamma_k^{LRR}(\beta) = \{ \gamma \in \Gamma_p : Q^{LRR}(\beta, \gamma) \leq Q^{LRR}(\beta, \tilde{\gamma}) + \kappa, \text{ for all } \tilde{\gamma} \in \Gamma_p \},
\]
where \( \Gamma_p \) is as defined in (2.5). Then we obtain the LRR set:
\[
\Theta_{k, p}^{LRR} = \{ (\beta, \gamma) : \gamma \in \Gamma_k^{LRR}(\beta), \beta \in B \}.
\]

As mentioned previously, using this LRR set, one may construct the LRR set for other target parameters. Suppose that our ultimate interest is the Average Treatment Effect (ATE) denoted by \( \psi(\beta_0, \gamma_0, \pi_0) \), where
\[
\psi(\beta, \gamma, \pi) = E[\alpha(X_1, 1; \beta, \gamma, \pi) - \alpha(X_1, 0; \beta, \gamma, \pi)],
\]
where \( \alpha(X_1, D; \beta, \gamma, \pi) \) is as defined in (2.6). In this case, the LRR set for \( \psi(\beta_0, \gamma_0, \pi_0) \) is given by
\[
\Psi_{k, p}^{LRR} = \{ \psi(\beta, \gamma, \pi) : (\beta, \gamma) \in \Theta_{k, p}^{LRR}(\pi), \pi \in \Pi \},
\]
where $\Theta^{\text{LRR}}_{k,p}(\pi)$ is the same as $\Theta^{\text{LRR}}_{k,p}$, except that $\Theta_p$ is replaced by

$$
\Theta_p(\pi) = \{ (\beta, \gamma) : \mathbb{E}[\tilde{Y}_1|X_1, D] \leq \alpha(X_1, D; \beta, \gamma, \pi) \leq \mathbb{E}[\tilde{Y}_2|X_1, D], \text{a.e.} \}. 
$$

(3.29)

4. Inference with the Locally Robust Refinement

4.1. Inference from Moment Inequality Models

For the sake of concreteness, we focus on the case where the parameters are identified through moment inequality models and estimated using their sample moments. Let us consider the following moment inequality models:

$$
\mathbb{E}_p[m_j(Z_i; \beta_0, \gamma_0)] \leq 0, \text{ for } j = 1, \ldots, p, \tag{4.1}
$$

for an observed i.i.d. random vector $Z_i$, where $m_j(\cdot; \beta_0, \gamma_0)$'s are moment functions. We assume that the moment restrictions define the identified set $\Theta_p$ so that for all $(\beta, \gamma) \in \Theta_p$, the moment restrictions are satisfied.

Define the sample analogue of the moment in (4.1):

$$
\bar{m}_j(\beta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} m_j(Z_i; \beta, \gamma). \tag{4.2}
$$

Let $\hat{\sigma}^2_j(\beta, \gamma)$ be the estimated variance of $m_j(Z_i; \beta, \gamma)$, i.e.,

$$
\hat{\sigma}^2_j(\beta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} (m_j(Z_i; \beta, \gamma) - \bar{m}_j(\beta, \gamma))^2. \tag{4.3}
$$

\[9\] Our LRR approach does not depend on a particular way in which partial identification of a parameter arises (i.e. whether by moment inequality models or alternative models). Hence, one can combine the LRR approach with other forms of partial identification by refining the identified set using the LRR criterion proposed in this paper. For the sake of concreteness, we focus on the case where partial identification arises from moment inequality restrictions as this is the way partial identification often arises in practice.
We let for $\kappa \geq 0$ (for instance, $\kappa = 0.01, 0.02$ or $0.03$)

$$
\hat{Q}_\kappa(\beta, \gamma) = \sum_{j=1}^{p} \left[ \frac{\bar{m}_j(\beta, \gamma) - \hat{\sigma}_j(\beta, \gamma)}{\hat{\sigma}_j(\beta, \gamma)} - \kappa \right]_+ ,
$$

(4.4)

where $[a]_+ = \max\{a, 0\}, a \in \mathbb{R}$. When $\kappa = 0$, we simply write $\hat{Q}(\beta, \gamma) = \hat{Q}_0(\beta, \gamma)$. Define

$$
\hat{\Gamma}_\kappa(\beta) = \{ \gamma \in \Gamma : \hat{Q}_\kappa(\beta, \gamma) = 0 \} .
$$

(4.5)

Then, we construct $T(\beta, \gamma)$ as

$$
T(\beta, \gamma) = \sqrt{n}\hat{Q}(\beta, \gamma).
$$

(4.6)

As for critical values, we may consider two different approaches. The first approach is based on least favorable configurations. The second approach is a more refined version with enhanced power properties but with a higher computational cost.

As for the first approach, we consider the following:

$$
\hat{Q}^*(\beta, \gamma) = \sum_{j=1}^{p} \left[ \frac{\bar{m}_j^*(\beta, \gamma) - \bar{m}_j^*(\beta, \gamma)}{\hat{\sigma}_j(\beta, \gamma)} \right]_+ ,
$$

(4.7)

where

$$
\bar{m}_j^*(\beta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} m_j(Z^*_i; \beta, \gamma) ,
$$

(4.8)

and $Z^*_i$'s are resampled from the empirical distribution of $Z_i$'s with replacement. Let

$$
\hat{T}^*(\beta, \gamma) = \sqrt{n}\hat{Q}^*(\beta, \gamma).
$$

(4.9)

\(^{10}\)The parameter $\kappa$ makes the inference slightly more conservative, which facilitates the development of the uniform inference (see the Supplemental Note). The choice of a positive constant $\kappa$ does not affect the uniform validity of the inference, as long as it is a fixed constant. In practice, values of $\kappa$ between 0.01 to 0.03 appear to work very well, as we show in the Monte Carlo simulations in Section 4 and the empirical application (Section 5).

\(^{11}\)We take this form of a test statistic for the sake of concreteness. One can use different functionals to produce different test statistics. See Andrews and Shi (2013) for various functionals.
We take critical values $\hat{c}_{1-\alpha}(\beta, \gamma)$ to be the $1-\alpha$ quantile of the bootstrap distribution of $\hat{T}^*(\beta, \gamma)$. Then, we construct the confidence region for $(\beta_0, \gamma_0)$ as follows:

$$C_{1-\alpha}^{\text{LRR}} = \left\{ (\beta, \gamma) \in \Theta : T(\beta, \gamma) \leq \hat{c}_{1-\alpha}(\beta, \gamma), \text{ and } \gamma \in \hat{\Gamma}_{\kappa}^{\text{LRR}}(\beta) \right\},$$

(4.10)

where

$$\hat{\Gamma}_{\kappa}^{\text{LRR}}(\beta) = \left\{ \gamma \in \hat{\Gamma}(\beta) : \hat{Q}_{\text{LRR}}^{\text{LRR}}(\beta, \gamma) \leq \inf_{\tilde{\gamma} \in \hat{\Gamma}_{\kappa}(\beta)} \hat{Q}_{\text{LRR}}^{\text{LRR}}(\beta, \tilde{\gamma}) + 2\kappa \right\},$$

(4.11)

and $\hat{\Gamma}(\beta)$ is as defined in (4.5), and $\hat{Q}_{\text{LRR}}^{\text{LRR}}(\beta, \gamma)$ denotes a consistent estimator of $Q_{\text{LRR}}^{\text{LRR}}(\beta, \gamma)$ defined in (3.10). (For example, in models of entry games or interval observations, this consistent estimator can be obtained by applying the sample analogue principle to the moments in (3.21) and (3.23), once one parametrically specifies the distributions of $\epsilon_1, \epsilon_2,$ and $\epsilon.$)

We take the infimum of an empty set to be infinity. With $\gamma$ restricted to be in $\hat{\Gamma}_{\kappa}^{\text{LRR}}(\beta)$, our confidence interval reflects the focus on the parameter values obeying the LRR hypothesis.

As for the second approach, there are various alternative ways to construct inference that improves upon the one based on least favorable configurations. For example, one may use moment selection procedures as in Hansen (2005), Andrews and Soares (2010) and Andrews and Shi (2013), the contact set method in Linton, Song, and Whang (2010), or a Bonferroni-based procedure as in Romano, Shaikh, and Wolf (2014). Here we adapt the Bonferroni-based procedure to our context.

First, we take $0 < \alpha_1 < \alpha$ (say, $\alpha_1 = 0.005$) and find $\hat{\kappa}_{\alpha_1}(\beta, \gamma)$ using a bootstrap procedure. We define $\hat{\kappa}_{\alpha_1}(\beta, \gamma)$ as the $\alpha_1$ quantile of the bootstrap distribution of

$$\min_{1 \leq j \leq k} \frac{\sqrt{n}(\bar{m}_j(\beta, \gamma) - \bar{m}_j(\beta, \gamma))}{\hat{\sigma}_j(\beta, \gamma)},$$

(4.12)

which uses $\bar{m}_j(\beta, \gamma)$ defined in (4.8). The uniform asymptotic validity of such a bootstrap procedure is well known in the literature. (See also Section A.2 in the Supplemental Note for
details.) Define
\[
\hat{\lambda}_{j,\alpha_1}(\beta, \gamma) = \min \left\{ \overline{m}_j(\beta, \gamma) - \frac{\hat{\sigma}_j(\beta, \gamma) \hat{\kappa}_{\alpha_1}(\beta, \gamma)}{\sqrt{n}}, 0 \right\}.
\] (4.13)

Let
\[
\tilde{Q}^*(\beta, \gamma) = \sum_{j=1}^{p} \left[ \frac{\overline{m}_j^*(\beta, \gamma) - \overline{m}_j(\beta, \gamma) + \hat{\lambda}_{j,\alpha_1}(\beta, \gamma)}{\hat{\sigma}_j(\beta, \gamma)} \right]_+.
\] (4.14)

We construct a bootstrap test statistic as follows:
\[
\tilde{T}^*(\beta, \gamma) = \sqrt{n} \tilde{Q}^*(\beta, \gamma).
\] (4.15)

We take critical values \(\tilde{c}_{1-\alpha+\alpha_1}(\beta, \gamma)\) to be the \(1-\alpha+\alpha_1\) quantile of the bootstrap distribution of \(\tilde{T}^*(\beta, \gamma)\). Then, we construct the confidence region for \((\beta_0, \gamma_0)\) as follows:
\[
\tilde{C}_{LRR}^{1-\alpha} = \{(\beta, \gamma) \in \Theta : T(\beta, \gamma) \leq \tilde{c}_{1-\alpha+\alpha_1}(\beta, \gamma) \text{ and } \gamma \in \hat{\Gamma}_{LRR}(\beta)\}.
\] (4.16)

In the Supplemental Note, we establish the uniform asymptotic validity of the confidence region. In the next section, we describe how to implement this procedure for Example 3.3.2 with both approaches to critical values.

As emphasized in the introduction, the LRR-based confidence region \(\tilde{C}_{LRR}^{1-\alpha}\) should not be confused with one based on the identified set. This LRR-based confidence region does not reflect the empirical content of the model. Rather it shows the confidence region when the true parameter belongs to the LRR set, that is, the true parameter belongs to the set of parameter values at which the ASF is robust to the local perturbations of the equilibrium selection rules. One may use \(\tilde{C}_{LRR}^{1-\alpha}\) as a complement to the confidence region based on the identified set for \((\beta_0, \gamma_0)\) in empirical and policy analysis.
4.2. Monte Carlo Simulations

4.2.1. Simulation Design. In this section, we present results from a Monte Carlo simulation study on the finite sample properties of the inference procedure developed previously. We focus on the switching regression model example considered in Section 3.3.2, which will be the main specification of our empirical application introduced in the next section.

Let $\tilde{Y}_1, \tilde{Y}_2$ be as in Section 2.3.2 and $\alpha(X_1, D; \beta_0, \gamma_0, \pi)$ as in (2.6) with the following specification, also motivated by our empirical application:

$$g_1(\beta_0 D + X'_1 \gamma_0) = \beta_0 D + X'_1 \gamma_0,$$

and

$$g_2(\beta_0 D + X'_1 \gamma_0) = \log(g_1(\beta_0 D + X'_1 \gamma_0)).$$

We take $X_1 = 1$ for simplicity, we draw $\eta$ independently and with equal probability (i.e. $\pi_0 = 0.5$, implying an equal share of $\eta = 1$ and $\eta = 0$ in the population), and treatment is drawn independently for all $i$ with $D_i = 1$ with probability 0.3. The moment inequalities we have are as follows\(^{12}\): for $x_1 = 1$ and $d \in \{0, 1\}$,

$$E[(\tilde{Y}_1 - \alpha(1, d; \beta_0, \gamma_0, \pi_0))1\{D = d\}] \leq 0,$$

and

$$E[(\alpha(1, d; \beta_0, \gamma_0, \pi_0) - \tilde{Y}_2)1\{D = d\}] \leq 0.$$

We consider two cases in our simulations: one where parameters $(\gamma_0, \beta_0, \pi_0)$ are such that the moments are close to equalities, and the other where they are not. For the first case, which we call Specification 1, we set $(\gamma_0, \beta_0, Z_1, Z_2) = (1, 0.15, 2.4, 7)$. The choice of $\beta_0 = 0.15$ implies a return to treatment (for $\eta = 1$) equal to 15%. This specification corresponds to approximately 5% of observations being top-coded.\(^{13}\) The second specification uses parameter values for $(\gamma_0, \beta_0, Z_1, Z_2) = (1, 0.15, 1.5, 7)$ so that the moment inequalities are far from binding. This is done by keeping $(\gamma_0, \beta_0, Z_2)$ the same as the first specification, but decreasing $Z_1$ so that there

---

\(^{12}\)These inequalities can be generalized for a multidimensional $X_1$ by changing the indicator function on the left hand side to accommodate the different values of $X_1$.

\(^{13}\)As can be seen by the identified set (2.5), setting one moment as an equality leaves the other moment as a strict inequality. Therefore, we do not consider a “least-favorable” set-up with all moments holding as equalities.
are now approximately 10% of observations top-coded, since we have an increased range of outcomes that are unobserved. In general, the identified set from the second specification will be larger than the first.

For both cases, we show results for the identified set and for using LRR with and without the Bonferroni-based method of Romano, Shaikh, and Wolf (2014). For the LRR, we set $\kappa = 0.03$ throughout. The confidence interval follows the computation in (4.16). We set the nominal size at 0.05, the simulation number to 500, and the bootstrap number to 999 under a sample size of $n = 500$. In this set-up, our parameter space is such that $\beta_0 \geq 0$ and the ATE given in (3.27) equals 0.145.

4.2.2. Results. The results across specifications and procedures are shown in Table 1. They show the coverage of both the identified set and the LRR, as well as the length of the confidence sets. From the results, it is immediate that the LRR provides a smaller set of ATE $\psi(\beta_0, \gamma_0, \pi_0)$, implying less variation on predicted (counterfactual) outcomes $Y_i$. (The identified set, on the other hand, presents all possible values consistent with the data.) The coverage of the theoretical LRR is appropriate (at least 95%) for values in the interior of the set, calculated using equations (3.23) and (3.25).

We also find that the Bonferroni-based method of Romano, Shaikh, and Wolf (2014) is effective at reducing the length of the confidence intervals, particularly for the identified set. It is most effective when the DGP is away from the configuration that is closest to moment equalities. However, this modification comes at the computational cost of calculating the term $\hat{\kappa}_{\alpha_1}$ via a bootstrap procedure for every value of $(\beta, \gamma)$ in a first step. This step could be time consuming when the dimension of $\gamma$ is large.

---

14In simulations available with the authors, the results improve as $n$ grows, as can be expected by our theoretical results. However, the computational cost also grows in these cases, particularly for the Bonferroni-based modification.
4.3. Predictions Using the LRR and the Identified Set

We conduct two additional Monte Carlo exercises to provide additional insights on the benefits of using the Locally Robust Refinement.

Our first exercise compares the predictions for our parameter of interest ($\psi$) when using the LRR compared to using the identified set across Specifications 1 and 2. It illustrates when the LRR's predictions are closest to the true parameter values and when they are most informative (i.e. when the predictions vary the least).

To do so, we draw $s = 1, \ldots, 1000$ values $\psi^s$ from the average (conservative) 95% confidence set for the LRR or the identified set described in Table 1. These sets would be the predictions reported by an empirical researcher, who does not observe or point-identify the true $\psi(\beta_0, \gamma_0, \pi_0)$ due to top-coding. We draw these vectors in 2 different ways: (i) uniformly (i.e. any parameter within the appropriate set is drawn with equal probability) and (ii) furthest-case (i.e. we draw the parameters in each set that generate average predictions furthest away from the true values). To showcase the results, in Figure 2, we report the distribution of the absolute deviation $|\psi^s - \psi(\beta_0, \gamma_0, \pi_0)|$: i.e. how do the predictions using the LRR and the identified set differ from the true value.

As we can see, the LRR's predictions vary less than those with the identified set (they are often at least 30% shorter), regardless of the specification or how the parameters are drawn. The use of the LRR seems particularly favorable when (i) the identified set is large (Specification 2), and (ii) when there is a significant chance of drawing extreme values of parameters from the identified set. This difference is most salient when the furthest-care scenario is particularly important to the researcher (for instance, when a bad policy could be very costly in terms of welfare). Afterall, the LRR's objective is not about mean predictions, but rather as minimizing the worst case.

15These sampling schemes represent different scenarios that illustrate the performance of our refinement. They may be thought of as different researcher priors on the true parameter. One can also perform the exercise under intermediate weighting schemes, obtaining similar conclusions.
The specifications above satisfy the LRR hypothesis, since the true \( \psi(\beta_0, \gamma_0, \pi_0) \) is in the LRR set. Does the LRR still preserve these favorable properties even if the LRR hypothesis is not satisfied? We investigate this question in a last exercise. We set \((\beta_0, Z_1, Z_2)\) to the same values as Specification 1, but we increase \(\gamma_0\) to 2.5. As a result, \(\gamma_0\) is no longer in the LRR. We then compare the performance of the LRR and of the identified set in predicting the Average Structural Outcome (ASO) obtained using (2.7).\(^{16}\) This outcome is unobserved, but it can be predicted based on the LRR and the identified set. The results are shown in Figure 3. In contrast to Figure 2, some draws of the identified set can outpredict the LRR. After all, the former contains the true parameters which are now absent in the LRR. Nevertheless, the LRR still generates more stable predictions, and its gains are still most noticeable and relevant when the identified set is large and/or worst-case outcomes are likely. Hence, while our results rely on the simulation designs motivated by our empirical application, the approach of LRR seems to generates predictions with stable performance regardless of whether the LRR hypothesis is satisfied or not.

5. Empirical Illustration

We provide an empirical application to illustrate how our procedure works with data. The application considers the case of top-coded data, as in the Monte Carlo simulations. We focus on the returns to college education, a topic subject to a large body of work in economics (see Oreopoulos and Petronijevic (2013) for a survey). For the purpose of illustration, we choose a simplified approach and focus on the ATE defined in (3.27).\(^{17}\) Depending on the question of interest, such a parameter could allow the evaluation of policies meant to increase enrollment,

\(^{16}\)In this setting, \( \psi(\beta_0, \gamma_0, \pi_0) \) will generally satisfy the LRR hypothesis due to the parametrizations in this example. Hence as a target parameter that is not in the LRR set in this design, we consider the ASO in this exercise.

\(^{17}\)One could extend our empirical analysis to incorporate potential endogeneity of the policy variable by using instrumental variables, and considering a Local Average Treatment Effect (LATE) which is the returns of college education on wages for the subgroup of individuals who are induced to attend college by a change in the instrument. Or alternatively, one could focus on the Marginal Treatment Effect or a Marginal Policy Relevant Treatment Effect as emphasized by Carneiro, Heckman, and Vytlacil (2011).
such as changing taxes or subsidies for college (see Dynarski and Scott-Clayton (2017) for an overview of different policies in place in the U.S. and their associated costs). These policies could have large fiscal and welfare consequences.

In this set-up, the outcome ($Y_i$) is the log of the hourly wages, the policy treatment is attending college (i.e. $D_i \in \{0, 1\}$, where $D_i = 1$ represents that $i$ attended college and 0 otherwise) and $X_{1,i}$ are individual covariates such as gender and/or race. Finally, the decision to attend college may depend on unobserved types: some types may have higher wages and/or benefit more from college. We model this by the random variable $\eta$, which is assumed to be binary: the returns from treatment $D_i = 1$ may depend on whether unobserved type is high ($\eta = 1$) or low ($\eta = 0$) and this difference can exist throughout the distribution of $X_{1,i}$.\(^{18}\)

Since top-coding is prevalent in wage data (it is present in such widely used datasets as the NLSY and the Current Population Survey - CPS) and such top-coding is not random (for example, college educated individuals are more likely to be top-coded since they have higher wages on average), this set-up is nested within Example 3.3.2. Hence, the identified set is given in (2.5) and the LRR in (3.13) can be used with the expressions defined in and above (3.25).

We follow Romano and Shaikh (2010) and use data from the 2000 Annual Demographic Supplement of the Current Population Survey (CPS), keeping individuals aged 20 to 24, white, with a primary source of income from wages and salaries and that worked at least 2 hours per week on average. Akin to their work, our covariates $X_{1,i}$ are a constant and a binary variable for Male/Female. We differ from their sample by keeping both college graduates and non-college graduates, since that is our counterfactual of interest. This yields a sample of 5816 individuals, of which 826 (14%) are treated ($D_i = 1$). Details on the data construction are provided in the Supplemental Note.\(^{19}\) We then explore the effects of different amounts of top-coding on the results.

---

\(^{18}\)For illustration purposes, we assume $\eta$ is binary and independent of $X_1, D$. Both can be generalized, but that is beyond the scope of our exercise.

\(^{19}\)Romano and Shaikh (2010) took the empirical distribution of the data as a “true” probability distribution to perform Monte Carlo simulations. Here, as in typical empirical research, we regard the data as a sample from an unknown population.
Following the large literature on Mincerian equations (see Heckman, Humphries, and Veramendi (2018) for a discussion), we use a linear specification for $g_1$:

$$g_1(\beta_0D + X_1'\gamma_0) = \beta_0D + X_1'\gamma_0.$$  

For the reduced form for $\eta = 0$, we parametrize $g_2(\beta_0D + X_1'\gamma_0) = \log(g_1(\beta_0D + X_1'\gamma_0))$.\(^{20}\)

We present confidence sets for both the identified set for the ATE, $\psi_0 = \psi(\beta_0, \gamma_0, \pi_0)$, and the LRR for $\psi_0$ using both conservative and a Bonferroni-correction based inference. We present results for this data under different top-coding amounts, imposing either 5% or 10% top-coding in the sample, under a value of top possible wages of $Z_2 = 10^8$.\(^{21}\) The results are shown below. As standard, we report the returns to a year of college (i.e. $\psi_0/4$).

[INSERT TABLE 2 HERE]

Our estimates are largely consistent with the literature, including the ATE found in works as Carneiro, Heckman, and Vytlacil (2011), despite the different sample. While the identified set is large, including ATE close to 0 and others close to 25%, the LRR is significantly shorter, between 3 and 11% (for 5% top-coding) and between 0 and 10% (for 10% top-coding). These results are also robust to different choices of $\kappa$.\(^{22}\)

To see the usefulness of the refinement in this context, consider one of the policies motivating this exercise. This could be a change in subsidies to college or expanding college loans with the purpose of a large increase in college attendance. In the presence of top-coding, the identified set cannot reject large rates of return for college (e.g. above 12% a year) - suggesting a rationale for financially costly policies. However, the LRR rejects these larger values, finding rates of return that are lower than 11% and more compatible with those observed

---

\(^{20}\)While this is an illustration and other specifications could be used, this nonlinearity allows for increasing differentials across types, while its nonseparability allows for heterogeneous returns to college based on observable characteristics (see Barrow and Rouse (2005) for one such study).

\(^{21}\)We also set the upper bound on $\beta_0$ to 100% (or 25% per year), above all estimates typically found in the literature (see Oreopoulous and Petronijevic (2013)). This is a natural restriction using the literature’s results, and helps speed up computation as we do not search for a priori unreasonable values of $\beta_0$.

\(^{22}\)We used $\kappa = 0.03$ which was shown to work well in the Monte Carlo simulations. When setting $\kappa = 0.01$, the LRR confidence sets are very similar (and shorter, since a larger $\kappa$ is more conservative): in the conservative inference case, they are $[0.061, 0.091]$ for 5% top-coding and $[0, 0.091]$ for 10% top-coding.
in the literature. The LRR suggests that more costly counterfactual policies should not be undertaken. This rejection is not due to the informational content of the model, which is still captured by the identified set, but rather from the robustness of the counterfactual predictions. The larger returns to college found in the identified set seem to be driven by features of the model the researcher is less confident about.

6. Conclusion

This paper explores methods to deal with models with multiple reduced forms for counterfactual analysis. As mentioned by Eizenberg (2014), there is an inherent tradeoff within the partial identification literature: while researchers may prefer not to impose strong identifying assumptions, they must confront the issue of counterfactuals with larger identified sets when doing policy analysis.

This paper points out that not all the values of the parameter are equally desirable for counterfactual analysis despite their empirical relevance. In particular, this paper focuses on the robustness property of the parameter values when the reduced form selection rule is locally perturbed. Thus the refinement looks at the subset of counterfactual scenarios that give most reliable predictions against a range of perturbations of the reduced forms (such as through changing equilibrium selection rules).

There could be other ways to discriminate different values of the identified set depending on the purpose of the counterfactual analysis in the application. For example, one may consider local misspecification of causal relationships between certain variables. This is plausible when the researcher is not sure about the direction of causality between certain variables. Similarly as we do in this paper, one may want to find a refinement that is robust to this type of misspecification. A fruitful formulation of this approach is left for future research.
References

ACKERBERG, D. A., AND G. GOWRISANKARAN (2006): “Quantifying Equilibrium Network Externalities in the ACH Banking Industry,” *RAND Journal of Economics*, 37, 738–761.

AGUIRREGABIRIA, V., AND A. NEVO (2013): “Recent Developments in Empirical IO: Dynamic Demand and Dynamic Games,” *Advances in economics and econometrics*, 3, 53–122.

ANDREWS, D. W., AND G. SOARES (2010): “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica*, 78(1), 119–157.

ANDREWS, D. W. K., AND X. SHI (2013): “Inference Based on Conditional Moment Inequalities,” *Econometrica*, 81, 609–666.

ANDREWS, I., M. GENTZKOW, AND J. M. SHAPIRO (2017): “Measuring the Sensitivity of Parameter Estimates to Estimation Moments,” *The Quarterly Journal of Economics*, 132(4), 1553–1592.

——— (2020): “On the Informativeness of Descriptive Statistics for Structural Estimates,” *Econometrica*, 88(6), 2231–2258.

ARMSTRONG, T. B., AND M. KOLESÁR (Forthcoming): “Sensitivity Analysis Using Approximate Moment Condition Models,” *Quantitative Economics*.

ARYAL, G., AND D.-H. KIM (2013): “A Point Decision for Partially Identified Auction Models,” *Journal of Business and Economic Statistics*, 31, 384–397.

BARROW, L., AND C. E. ROUSE (2005): “Do Returns to Schooling Differ by Race and Ethnicity?,” *American Economic Review*, 95(2), 83–87.

BLUNDELL, R. W., AND J. L. POWELL (2004): “Endogeneity in Semiparametric Binary Response Models,” *Review of Economic Studies*, 71, 655–679.

BONHOMME, S., AND M. WEIDNER (2020): “Minimizing Sensitivity to Model Misspecification,” *arXiv preprint arXiv:1807.02161*.

BRESNAHAN, T. F., AND P. C. REISS (1991): “Empirical Models of Discrete Games,” *Journal of Econometrics*, 48, 57–81.
CARNEIRO, P., J. J. HECKMAN, AND E. J. VYTLACIL (2011): “Estimating Marginal Returns to Education,” American Economic Review, 101(6), 2754–81.

CHRISTENSEN, T., AND B. CONNAULT (2019): “Counterfactual Sensitivity and Robustness,” arXiv preprint arXiv:1904.00989.

CILIBERTO, F., AND E. Tamer (2009): “Market Structure and Multiple Equilibria in Airline Markets,” Econometrica, 77, 1791–1828.

DYNARSKI, S., AND J. SCOTT-CLAYTON (2017): “Tax Benefits for College Attendance,” in The Economics of Tax Policy, ed. by A. Auerbach, and K. Smetters. Oxford University Press.

EIZENBERG, A. (2014): “Upstream Innovation and Product Variety in the United States Home PC Market,” Review of Economic Studies, 81, 1003–1045.

FUDENBERG, D., D. M. KREPS, AND D. K. LEVINE (1988): “On the Robustness of Equilibrium Refinements,” Journal of Economic Theory, 44.

GILBOA, I., AND D. SCHMEIDLER (1989): “Maximin Expected Utility with Non-Unique Prior,” Journal of Mathematical Economics, 18, 141 – 153.

HAILE, P. A., AND E. Tamer (2003): “Inference with an Incomplete Model of English Auctions,” Journal of Political Economy, 111, 1–51.

HANSEN, B. E. (2005): “Challenges for Econometric Model Selection,” Econometric Theory, 21(1), 60–68.

HANSEN, L., AND T. J. SARGENT (2001): “Robust Control and Model Uncertainty,” American Economic Review, 91(2), 60–66.

HANSEN, L. P., AND M. MARINACCI (2016): “Ambiguity Aversion and Model Misspecification: An Economic Perspective,” Statistical Science, 31(4), 511–515.

HECKMAN, J. J. (2010): “Causal Parameters and Policy Analysis in Economists: A Twentieth Century Restrospective,” Quarterly Journal of Economics, 115, 45–97.

HECKMAN, J. J., J. E. HUMPHRIES, AND G. VERAMENDI (2018): “Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking,” Journal of Political Economy, 126(S1), S197–S246.
JIA, P. (2008): “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry,” *Econometrica*, 76, 1263–1316.

JUN, S. J., AND J. PINKSE (2017): “An Information-Theoretic Approach to Partially Identified Auction Models,” *Working Paper*.

——— (2019): “Counterfactual Prediction in Complete Information Games: Point Prediction Under Partial Identification,” *Journal of Econometrics, Forthcoming*.

KOOPMANS, T. C. (1949): “Identification Problems in Economic Model Construction,” *Econometrica*, 17, 125–144.

Li, L. (2019): “Identification of Structural and Counterfactual Parameters in a Large Class of Structural Econometric Models,” Discussion paper, Johns Hopkins University.

LINTON, O., K. SONG, AND Y.-J. WHANG (2010): “An Improved Bootstrap Test of Stochastic Dominance,” *Journal of Econometrics*, 154, 186–202.

MANSKI, C. F., AND E. TAMER (2002): “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica*, 70(2), 519–546.

MASTEN, M. A., AND A. POIRIER (2018): “Identification of Treatment Effects under Conditional Partial Independence,” *Econometrica*, 86(1), 317–351.

MYERSON, R. B. (1991): *Game Theory*. Harvard University Press, USA.

NEVO, A., AND A. M. ROSEN (2012): “Identification with Imperfect Instruments,” *Review of Economics and Statistics*, 94(3), 659–671.

OREOPOULOS, P., AND U. PETRONIJEVIC (2013): “Making College Worth It: A Review of the Returns to Higher Education.,” *Future of Children*, 23(1), 41–65.

OSTER, E. (2019): “Unobservable Selection and Coefficient Stability: Theory and Evidence,” *Journal of Business & Economic Statistics*, 37(2), 187–204.

ROBERTS, J. W., AND A. SWEETING (2013): “When Should Sellers Use Auctions?,” *American Economic Review*, 103, 1830–1861.

ROMANO, J. P., AND A. M. SHAIKH (2010): “Inference for the Identified Set in Partially Identified Econometric Models,” *Econometrica*, 78, 169–211.
ROMANO, J. P., A. M. SHAIKH, AND M. WOLF (2014): “A Practical Two-Step Method for Testing Moment Inequalities,” *Econometrica*, 82, 1979–2002.

SONG, K. (2014): “Point Decisions for Interval-Identified Parameters,” *Econometric Theory*, 30, 334–356.

TAMER, E. (2003): “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria,” *Review of Economic Studies*, 70, 147–165.

Tables and Figures

![Diagram](image_url)

**Figure 1.** Locally Robust Refinement for the Entry Game Example

Notes: The figure is based on Figure 1 of *Ciliberto and Tamer (2009)* and presents a graphical analysis of the LRR for the entry game example. In the data, the region in light blue is the region of multiple equilibria which is the set $A_{3,\beta,\gamma}(X) \cap A_{4,\beta,\gamma}(X)$. In our counterfactual setting involving $X$ changed to, say, $\tilde{X}$, this region becomes $A_{3,\beta,\gamma}(\tilde{X}) \cap A_{4,\beta,\gamma}(\tilde{X})$ depicted in dark blue in the right hand side figure. The LRR chooses nuisance parameters in their identified set that minimize the probability mass in the dark blue region.
FIGURE 2. Distribution of the Absolute Difference Between Predicted ATEs and the True ATE

Notes: The figure presents the absolute deviation of the predicted statistic, $\psi_s$ drawn from the 95% confidence set for the LRR or Identified set from the specifications in Section 4.2, relative to the parameter of interest, $\psi(\beta_0, \gamma_0, \pi_0)$, used for policy analysis. We present the uniformly-drawn parameters case, where we draw $\psi_s$ with equal probability from their respective sets $S = 1000$ times. Drawing the furthest away parameter from the true values always yield the same parameter draw (so it is not pictured above). In the latter case, the mean absolute deviation is 0.521 for the LRR and 0.895 for the identified set for Specification 1, and 1.012 and 1.492 respectively for Specification 2.

FIGURE 3. Distribution of the Absolute Difference Between Predicted ASOs and the True ASO

Notes: The figure presents the absolute deviation of the predicted ASO, $\text{ASO}' = E\pi' (\gamma' + \beta'D) + (1 - \pi') \log(\gamma' + \beta'D)$, to the ASO under the true parameters $(\beta_0, \pi_0, \gamma_0) = (0.15, 0.5, 2.5)$ across draws $s = 1, ..., 1000$. For the predicted outcome, we set $\pi'$ at either 0.25 or 0.5 (true value), and draw the remaining parameters $S = 1000$ times from the LRR or the identified set (calculated numerically).
TABLE 1. The Empirical Coverage Probability and Average Length of Confidence Intervals for $\psi(\beta_0, \gamma_0, \pi_0)$ at 95% Nominal Level.

| Coverage Probability | Specification 1 | Specification 2 |
|----------------------|-----------------|-----------------|
|                      | Least Favorable | RSW             | Least Favorable | RSW             |
| Identified Set       | $n = 500$       | 1               | 0.996          | 1               | 1               |
| LRR                  | $n = 500$       | 0.95            | 0.95           | 1               | 1               |

| Average Length of CI | Specification 1 | Specification 2 |
|----------------------|-----------------|-----------------|
|                      | Least Favorable | RSW             | Least Favorable | RSW             |
| Identified Set       | $n = 500$       | 1.040           | 0.923          | 1.637           | 1.477           |
| LRR                  | $n = 500$       | 0.666           | 0.666          | 1.157           | 1.157           |

Notes: The first half of the table reports the empirical coverage probability of the confidence interval for $\psi$ computed with the bootstrap procedure described in the main text, and the second half reports its average length for the specifications described in the main text. The first inference procedure, denoted “LF”, follows the conservative approach defined in the main text, while the second is denoted “RSW” and uses the Bonferroni-based modification of Romano, Shaikh, and Wolf (2014).

TABLE 2. Confidence Intervals for $\psi$ (ATE), Parameter of Interest

|                      | 5% top-coding | 10% top-coding |
|----------------------|---------------|---------------|
| Identified Set - LF  | [0.023, 0.245] | [0, 0.245]    |
| Identified Set - RSW | [0.041, 0.245] | [0, 0.245]    |
| LRR - LF             | [0.034, 0.110] | [0, 0.091]    |
| LRR - RSW            | [0.041, 0.110] | [0, 0.091]    |

Notes: The Table presents the results for the empirical application described in the main text, where we use the data and top-code it at different values (5% or 10%), assuming the top wage is $Z_2 = 10^8$. The first inference procedure, denoted “LF”, follows the conservative approach defined in the main text - see equation (4.10), while the second is denoted “RSW” and uses the Bonferroni-based modification of Romano, Shaikh, and Wolf (2014), see equation (4.16).
7. Appendix: Mathematical Proofs

7.1. Proof of Theorem 3.1

**Proof of Theorem 3.1:** Let $U$ and $W$ be the sets from which $\eta$ and $W = (X, \epsilon)$ take values respectively. Let $\mathcal{H}$ be the collection of measurable maps $h : W \times U \to \mathbb{R}$ such that

$$\int h(w, \eta)d\mu(\eta) = 0, \tilde{F}\text{-a.s., and } \int \int h^2(w, \eta)d\mu(\eta)d\tilde{F}(w) < \infty. \quad (7.1)$$

We endow $\mathcal{H}$ with the inner-product $\langle \cdot, \cdot \rangle$ as follows: for $h_1, h_2 \in \mathcal{H}$,

$$\langle h_1, h_2 \rangle = \int \int h_1(w, \eta)h_2(w, \eta)d\mu(\eta)d\tilde{F}(w). \quad (7.2)$$

Then $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is a Hilbert space (up to an equivalence class). Recall that all distributions $G \in \mathcal{G}$ are dominated by the uniform distribution $\mu$ over $H$. Define $\|h\|^2 = \langle h, h \rangle$ and write

$$\delta_{\mathcal{G}}(G', G) = \left\| \frac{dG'}{d\mu} - \frac{dG}{d\mu} \right\|,$$

and $\text{ASO}_{\beta, \gamma}(G) = \int \int \rho_{\beta, \gamma}(w; \eta)dG(\eta|w)d\tilde{F}(w)$. If we define

$$\Delta_{\beta, \gamma}(\eta, w) = \rho_{\beta, \gamma}(w; \eta) - \int \rho_{\beta, \gamma}(w; \eta)d\mu(\eta), \quad (7.4)$$

we can write

$$\sup_{G' \in \mathcal{G}: \delta_{\mathcal{G}}(G', G) \leq K} \frac{|\text{ASO}_{\beta, \gamma}(G') - \text{ASO}_{\beta, \gamma}(G)|}{\delta_{\mathcal{G}}(G', G)} = \sup_{h' \in \mathcal{H}: ||h'|| = K} \frac{\langle \rho_{\beta, \gamma}, h' \rangle}{||h'||} = \sup_{h' \in \mathcal{H}: ||h'|| = K} \frac{\langle \Delta_{\beta, \gamma}, h' \rangle}{||h'||},$$

where the first equality comes because $dG'/d\mu - dG/d\mu \in \mathcal{H}$ (from the fact that each density integrates to one) and the second equality comes from the fact that for all $h \in \mathcal{H}$, $\int h(w, \eta)d\mu(\eta) = 0$. Since $\Delta_{\beta, \gamma} \in \mathcal{H}$ by Assumption 3.1, using (7.1) and Cauchy-Schwarz inequality, we can achieve the last supremum by taking $h'$ in the supremum to be $K\Delta_{\beta, \gamma}/||\Delta_{\beta, \gamma}||$. The result is nothing but $\sqrt{Q_{\mathcal{LRR}}^{\frac{1}{2}}(\beta, \gamma)}$, completing the proof. \footnote{We note that the right hand side is not a function of $\delta$ or $K$. This is because the ASO is linear in $G$, and the dependence on those terms is removed by the difference operator defined on the left hand side.}
Supplemental Note to “Counterfactual Analysis under Partial Identification Using Locally Robust Refinement”

Nathan Canen and Kyungchul Song
University of Houston and University of British Columbia

This note consists of three parts. In the first part (Appendix A), we present the proof of the uniform asymptotic validity of our inference, beginning with various auxiliary results. In the second part (Appendix B), we provide further details on the data used for the empirical application in Section 5. In Appendix C, we provide a notation table as a convenient reference.

Appendix A. Uniform Asymptotic Validity of the Confidence Region Under the LRR Hypothesis

A.1. Uniform Asymptotic Validity. This section establish the uniform validity of the inference on LRR. To make explicit their dependence on $P$, we write $\hat{F}$, $Q^{LRR}(\beta, \gamma)$, and $\Gamma^{LRR}_k$ as $\tilde{F}_P$, $Q^{LRR}_P(\beta, \gamma)$, and $\Gamma^{LRR}_{k,p}$ in this section. We first, make the following assumption.

Assumption A.1. For each $\varepsilon > 0$,

$$\limsup_{n \to \infty} \sup_{P \in \mathcal{P}} \left\{ \sup_{(\beta, \gamma) \in \Theta} \left| \hat{Q}^{LRR}(\beta, \gamma) - Q^{LRR}_P(\beta, \gamma) \right| > \varepsilon \right\} = 0. \quad (A.1)$$

Assumption A.1 requires uniform consistency of the sample criterion function for the LRR (uniform in $(\beta, \gamma) \in \Theta$ and $P \in \mathcal{P}$). Typically, $Q^{LRR}_P(\beta, \gamma)$ takes the following form:

$$Q^{LRR}_P(\beta, \gamma) = \int q(x; \beta, \gamma) d\tilde{F}_P(x),$$
for some map \( q \) and a distribution \( \tilde{F}_P \). Let \( \hat{F}(x) \) and \( \hat{q}(x; \beta, \gamma) \) be the estimated versions of \( \tilde{F}_P(x) \) and \( q(x; \beta, \gamma) \) using which we construct

\[
\hat{Q}^{LRR}(\beta, \gamma) = \int \hat{q}(x; \beta, \gamma) d\hat{F}(x).
\]

(Note that when the researcher knows fully \( \tilde{F}_P \) under the counterfactual scenario of focus, she can simply take \( \hat{F} \) to be \( \tilde{F}_P \).) Thus, we can find lower level conditions for (A.1) in terms of \( \hat{F}(x) \) and \( \hat{q}(x; \beta, \gamma) \) using standard arguments.

We let

\[
m_{j,p}(\beta, \gamma) = E_P[m_j(Z_i; \beta, \gamma)], \quad \text{and} \quad \sigma_{j,p}^2(\beta, \gamma) = \text{Var}_P(m_j(Z_i; \beta, \gamma)), \quad (A.2)
\]

where \( \text{Var}_P \) denotes the variance under \( P \). Let

\[
\mu_{i,j,p}(\beta, \gamma) = \frac{m_j(Z_i; \beta, \gamma)}{\sigma_{j,p}(\beta, \gamma)}. \quad (A.3)
\]

Let us make the following assumption.

**Assumption A.2.** (i) There exist \( C > 0 \) and \( q \geq 2 \) such that

\[
\max_{1 \leq j \leq k} \sup_{\beta, \gamma \in \Theta} \left( \sup_{\mu_{i,j,p}(\beta, \gamma)^{q+1}} \right) < C. \quad (A.4)
\]

(ii) There exists \( C > 0 \) such that for any \( \delta > 0 \) and any \( \theta = (\beta, \gamma) \in \Theta \),

\[
\max_{1 \leq j \leq k} \sup_{\beta, \gamma \in \Theta} \left( \sup_{\theta' \in \Theta : \|\theta - \theta'\| \leq \delta} \left| \mu_{i,j,p}(\theta) - \mu_{i,j,p}(\theta') \right|^2 \right) < C\delta^2. \quad (A.5)
\]

(iii) \( \Theta \) is compact.

Assumption A.2(i) is a uniform moment condition. Assumption A.2(ii) imposes the condition of \( L_2 \)-equicontinuity on the function \( \mu_{i,j,p}(\beta, \gamma) \). This condition, similar in spirit to Chen, Linton, and Van Keilegom (2003), is used to verify the asymptotic equicontinuity of the empirical process involved in \( \hat{Q}(\beta, \gamma) \) and its bootstrap version.
We now introduce an assumption on the limit Gaussian process of the empirical process. First, let us introduce some notation. Let \( \{ \nu_{j,p}(\beta, \gamma) : (\beta, \gamma) \in \Theta \} \) be a Gaussian process having covariance function equal to \( \text{Cov}_P(\mu_{i,j,p}(\beta, \gamma), \mu_{i,j,p}(\tilde{\beta}, \tilde{\gamma})) \). Let for \( M > 0 \),

\[
\zeta_{\text{sum}}(\beta, \gamma; r) = \sum_{j=1}^{p} \min \{ [\nu_{j,p} + r_j](\beta, \gamma), M \}, \quad \text{and} \quad (A.6)
\]

\[
\zeta_{\text{min}}(\beta, \gamma) = \min_{1 \leq j \leq p} \nu_{j,p}(\beta, \gamma), \quad (A.7)
\]

where \( r = (r_j)_{j=1}^{p} \) is a vector of real measurable maps on \( \Theta \). Define

\[
R_M = \{ (r_j)_{j=1}^{p} : r_j \text{ is measurable, and } -M \leq r_j(\cdot) \leq 0, \text{ for all } j = 1, \ldots, p \}. \quad (A.8)
\]

**Assumption A.3.** There exists \( M_0 > 0 \) such that the conditions below hold for \( \zeta_{M,p}(\beta, \gamma; r) = \zeta_{\text{sum}}(\beta, \gamma; r) \) or \( \zeta_{M,p}(\beta, \gamma; r) = -\zeta_{\text{min}}(\beta, \gamma) \).

(i) For any \( M > M_0 \) and any \( 0 < c < pM \),

\[
\sup_{P \in \mathcal{P}} \sup_{r \in R_M} \sup_{(\beta, \gamma) \in \Theta_P} P\{ c - \varepsilon \leq \zeta_{M,p}(\beta, \gamma; r) \leq c + \varepsilon \} \to 0, \quad (A.9)
\]

as \( \varepsilon \to 0 \).

(ii) For any \( M > M_0 \), there exists \( \varepsilon_M > 0 \) such that for any \( 0 < c < pM \), the slope of the CDF of \( \zeta_{M,p}(\beta, \gamma; r) \) at \( c \) is bounded from below by \( \varepsilon_M \) for all \( (\beta, \gamma) \in \Theta_P, r \in R_M, \) and \( P \in \mathcal{P} \).

Assumption A.3(i) is a mild regularity condition for the Gaussian process \( \nu_{p} \), which is generically satisfied because the Gaussian process has an unbounded support. (Note that \( R_M \) is a collection of vectors of functions that are bounded between \(-M \) and \( 0 \).) The assumption can be violated when \( \nu_{j,p}(\beta, \gamma) \) has zero variance at some point \( (\beta, \gamma) \in \Theta_P \). Assumption A.3(ii) is a technical condition which requires that \( \zeta_{M,p}(\beta, \gamma; r) \) has CDF strictly increasing at any \( c \in (0, pM) \) with slope bounded away from zero uniformly over \( (\beta, \gamma) \in \Theta_P, r \in R_M, \) and \( P \in \mathcal{P} \). Note that when the Gaussian process \( \nu_{p} \) has an unbounded support for all \( (\beta, \gamma) \in \Theta_P \).
and \( P \in \mathcal{P} \), the support of \( \zeta_{M,p}(\beta, \gamma; r) \) is \([0, pM]\) for all \((\beta, \gamma) \in \Theta_p, r \in R_M, \) and \( P \in \mathcal{P} \). Thus Assumption A.3(ii) is not implausible.

The following theorem gives the uniform asymptotic validity of the confidence region \( \tilde{C}_{1-\alpha}^{LRR} \) defined in (4.16).

**Theorem A.1.** Suppose that Assumptions A.1 and A.2 hold. Then,

\[
\lim_{n \to \infty} \inf_{P \in \mathcal{P}} \inf_{(\beta, \gamma) \in \Theta_p} P \{ (\beta, \gamma) \in \tilde{C}_{1-\alpha}^{LRR} \} \geq 1 - \alpha. \tag{A.10}
\]

For the proof of uniform asymptotic validity, we approximate the bootstrap distribution by a truncated version of a functional of a Gaussian process. This facilitates the use of the continuous mapping theorem in Linton, Song, and Whang (2010) which does not require that the continuous functionals have a well defined limit.

The rest of Appendix A is devoted to proving Theorem A.1.

**A.2. Auxiliary Results I: General Results on Empirical Processes.** We introduce a theorem that gives the weak convergence of a stochastic process to a tight limit along a given sequence of probabilities. This result is based on Section 2.8.3 of van der Vaart and Wellner (1996). We modify the results slightly because in our case the class of functions that index an empirical process depends on the underlying distribution \( P \in \mathcal{P} \). The notation in this subsection is self-contained.

Let \( \mathcal{T} \) be a given set and let \( f_{j,p}(\cdot; \tau) : \mathbb{R}^d \to \mathbb{R}, j = 1, \ldots, k, \) be a function indexed by \( \tau \in \mathcal{T} \), potentially depending on \( P \in \mathcal{P} \). Define for each \( j = 1, \ldots, k, \)

\[
\mathcal{F}_{j,p} = \{ f_{j,p}(\cdot; \tau) : \tau \in \mathcal{T} \}, \quad \text{and} \quad \mathcal{F}_p = \{ f_p(\cdot; \tau) : \tau \in \mathcal{T} \}, \tag{A.11}
\]

where \( f_p(\cdot; \tau) = [f_{1,p}(\cdot; \tau), \ldots, f_{k,p}(\cdot; \tau)]' \). Given i.i.d. random vectors \( X_i \), we introduce a stochastic process on \( \mathcal{T} \):

\[
\nu_{n,p}(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f_p(X_i; \tau) - E_p[f_p(X_i; \tau)]).
\]
We let $\ell_\infty(T)$ be the set of bounded $\mathbb{R}^k$-valued functions on $T$, and $\| \cdot \|_\infty$ denote the sup norm on $\ell_\infty(T)$, i.e., $\|v\|_\infty = \sup_{\tau \in T} \|v(\tau)\|$, $v \in \ell_\infty(T)$, where $\| \cdot \|$ denotes the Euclidean norm in $\mathbb{R}^k$. We introduce a semimetric $\rho_p$ on $T$ as follows:

$$\rho_p(\tau_1, \tau_2) = \sqrt{\mathbb{E}_p[\|f_p(X_i; \tau_1) - f_p(X_i; \tau_2)\|^2]}.$$ 

Define the class of bounded Lipschitz functionals:

$$\text{BL}_1 = \left\{ \rho : \ell_\infty(T) \to \mathbb{R} : |\rho(v_1) - \rho(v_2)| \leq \|v_1 - v_2\|_\infty, \sup_{v \in \ell_\infty(T)} |\rho(v)| \leq 1 \right\}. \quad (A.12)$$

Write the $L_2(P)$ norm by $\|f\|_{L_2} = \sqrt{\mathbb{E}_p\|f(X_i)\|^2}$. Given a class $\mathcal{F}_p$ of $\mathbb{R}^k$-valued measurable functions, let $N_{[\epsilon]}(\epsilon, \mathcal{F}_p, \| \cdot \|_{L_2})$ be the $\epsilon$-bracketing entropy of $\mathcal{F}_p$ with respect to the $L_2(P)$ norm, which is the smallest number $N$ of the brackets $\{[\ell_j, u_j]\}_{j=1}^N$ such that each $f \in \mathcal{F}_p$ satisfies $\ell_j \leq f \leq u_j$ for some $[\ell_j, u_j]$ and $\|\ell_j - u_j\|_{L_2} \leq \epsilon$. (The inequalities required of the brackets are point-wise inequalities when $f$ is $\mathbb{R}^k$-valued.)

We also define $P^*$ and $\mathbb{E}_p^*$ to be the outer-probability and the outer-expectation under $P$. (See van der Vaart and Wellner (1996), Section 1.2 for details.) The following theorem formulates conditions for the weak convergence of $\nu_{n, p}$ to a tight Gaussian limit uniformly over $P \in \mathbb{P}$.

**Theorem A.2.** Suppose that the two conditions are satisfied.

(i) For each $j = 1, \ldots, k$, there exists a measurable envelope $F_{j, p}$ for $\mathcal{F}_{j, p}$ such that

$$\sup_{P \in \mathbb{P}} \mathbb{E}_p[F_{j, p}^2(X_i)] < \infty. \quad (A.13)$$

(ii)

$$\int_0^\infty \sup_{P \in \mathbb{P}} \sqrt{\log N_{[\epsilon]}(\epsilon, \mathcal{F}_{j, p}, \| \cdot \|_{L_2})} d\epsilon < \infty. \quad (A.14)$$
Then, we have

$$\sup_{f \in \mathcal{H}_1} \sup_{P \in \mathcal{P}} |E^*_P[f(\nu_{n,p})] - E_P[f(\nu_P)]| \to 0, \text{ as } n \to \infty,$$

(A.15)

and

$$\limsup_{\delta \downarrow 0} E_P \left[ \sup_{\rho_P(\tau_1, \tau_2) < \delta} \| \nu_P(\tau_1) - \nu_P(\tau_2) \| \right] = 0,$$

(A.16)

where $\nu_P = [\nu_{1,P}, ..., \nu_{p,P}]'$ is an $\mathbb{R}^p$-valued Gaussian process with covariance kernel given by

$$E[\nu_{j,P}(\tau_1) \nu_{\ell,P}(\tau_2)] = \text{Cov}_P (f_{j,P}(X_i; \tau_1), f_{\ell,P}(X_i; \tau_2)), \quad \tau_1, \tau_2 \in \mathcal{T},$$

for $j, \ell = 1, ..., p$.

**Proof:** By following the proof of Theorem 2.8.4 of van der Vaart and Wellner (1996), we find that Conditions (i) and (ii) give the asymptotic equicontinuity of the following form: for every $\varepsilon > 0$,

$$\lim_{n \to \infty} \sup_{P \in \mathcal{P}} P^* \left\{ \sup_{\rho_P(\tau_1, \tau_2) < \delta} \| \nu_{n,P}(\tau_1) - \nu_{n,P}(\tau_2) \| > \varepsilon \right\} = 0,$$

(A.17)

and $(\mathcal{T}, \rho_P)$ is totally bounded uniformly over $P \in \mathcal{P}$ in the sense that for any $\varepsilon > 0$, and any $P \in \mathcal{P}$ there exists an $m_\varepsilon$ number of $\varepsilon$-open balls (with respect to $\rho_P$) that cover $\mathcal{T}$ where $m_\varepsilon$ does not depend on $P$. Then both the results of (A.15) and (A.16) follow by Theorem 2.6 of Giné and Zinn (1991).

We consider a bootstrap version of Theorem A.2. Given a sample $\{X_1, ..., X_n\}$, we let $\mathcal{G}_n$ be the $\sigma$-field generated by this sample. Let $\hat{P}$ be the empirical measure of $\{X_1, ..., X_n\}$, and $\{X_1^*, ..., X_n^*\}$ i.i.d. draws from $\hat{P}$. First, define

$$\nu_{n,P}^*(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (f_P(X_i^*; \tau) - E_P[f_P(X_i^*; \tau)])$$

which is the bootstrap counterpart of $\nu_{n,P}$.
Theorem A.3. Suppose that the conditions of Theorem A.2 are satisfied. Then for any $\epsilon > 0$, 

$$\sup_{P \in \mathcal{P}} \left\{ \sup_{f \in \mathcal{BL}_1} \left| \mathbb{E}[f(\nu^*_n, P)] - \mathbb{E}_P[f(\nu_P)] \right| > \epsilon \right\} \rightarrow 0,$$  \hspace{1cm} (A.18)

as $n \rightarrow \infty$, where $\nu_P$ is the Gaussian process in Theorem A.2.

Proof: The proof basically follows that of Lemma A.2 of Linton, Song, and Whang (2010). The only difference here is that the class $\mathcal{F}_P$ depends on $P$. As seen in the proof of Theorem A.2, $(\mathcal{T}, \rho_P)$ is totally bounded uniformly over $P \in \mathcal{P}$. Then $(\mathcal{T} - \mathcal{T}, \rho_P)$ is totally bounded uniformly over $P \in \mathcal{P}$ as well, where $\mathcal{T} - \mathcal{T} = \{ \tau_1 - \tau_2 : \tau_1, \tau_2 \in \mathcal{T} \}$. For any given $\epsilon > 0$, there is a map $\pi_\epsilon : \mathcal{T} \rightarrow \mathcal{T}$ which takes $m_\epsilon$ values and $\rho_P(\pi_\epsilon \tau, \tau) \leq \epsilon$. Take any $f \in \mathcal{BL}_1$, and consider

$$\left| \mathbb{E}[f(\nu^*_n, P)] - \mathbb{E}_P[f(\nu_P)] \right| \leq \left| \mathbb{E}[f(\nu^*_n, P)] - \mathbb{E}[f(\nu^*_n \circ \pi_\epsilon)] \right| \hspace{1cm} (A.19)$$

$$+ \left| \mathbb{E}[f(\nu^*_n \circ \pi_\epsilon)] - \mathbb{E}_P[f(\nu_P \circ \pi_\epsilon)] \right|$$

$$+ \left| \mathbb{E}_P[f(\nu_P \circ \pi_\epsilon)] - \mathbb{E}_P[f(\nu_P)] \right|.$$

The last term vanishes uniformly in $f \in \mathcal{BL}_1$ as $\epsilon \rightarrow 0$ by (A.16). By using the proof of Lemma 2.1 in Giné and Zinn (1991), p.763, and the moment condition (A.13) to handle the remainder part of the process that is left over from truncation, we can show that the second term on the right hand side vanishes in probability uniformly over $P \in \mathcal{P}$ as $n \rightarrow \infty$. We deal with the first term using Le Cam’s Poissonization lemma and following the proof of Theorem 2.2 of Giné (1997), p.104.\textsuperscript{24} More specifically, we define $\mathcal{T}(\epsilon, \rho_P) = \{ \tau_1 - \tau_2 \in \mathcal{T} - \mathcal{T}$ :

\textsuperscript{24}The reference Theorem 2.2 of Giné and Zinn (1991) on page 196 of Linton, Song, and Whang (2010) should be Theorem 2.2 of Giné (1997), p.104.
\( \rho_p(\tau_1, \tau_2) < \varepsilon \), and note that
\[
\mathbb{E}_p \left[ \mathbb{E} \left[ \sup_{\tau \in \mathcal{T}(\epsilon, \rho_p)} \| \nu^*_n(\tau) \| G_n \right] \right]
\]
\[
\leq \frac{e}{e-1} \left\{ \mathbb{E}_p \left[ \sup_{\tau \in \mathcal{T}(\epsilon, \rho_p)} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (N_i - 1)(f_p(X_i; \tau) - \mathbb{E}_p[f_p(X_i; \tau)]) \right\| \right] \right\}
\]
\[
+ \mathbb{E}_p \left[ \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (N_i - 1) \sup_{\tau \in \mathcal{T}(\epsilon, \rho_p)} \left\| \frac{1}{n} \sum_{i=1}^{n} f_p(X_i; \tau) \right\| \right\] ,
\]
where \( N_i \) is i.i.d. Poisson random variables with mean 1 independent of all the other random variables. We apply Theorem A.2 to the class \( \mathcal{T}(\epsilon, \rho_p) \), and find that both terms on the right hand, after taking supremum over \( P \in \mathcal{P} \), vanish as \( n \to \infty \) and \( \epsilon \to 0 \). ■

The following lemma is the bootstrap version of Lemma A.1 of Linton, Song, and Whang (2010). The proof of the lemma can be proceeded similarly as in the proof of Lemma A.1 of Linton, Song, and Whang (2010) using (A.18) and is omitted.

**Lemma A.1.** Suppose that the bootstrap empirical process \( \nu^*_n \) satisfies (A.18) for any \( \epsilon > 0 \). Let \( f_n, p \) be a continuous functional on \( \ell^\infty(\mathcal{T}) \) that is potentially \( G_n \)-measurable such that for some possibly \( G_n \)-measurable real-valued map \( C_{n,p} \),
\[
|f_n, p(\nu_1) - f_n, p(\nu_2)| \leq C_{n,p} \sup_{\tau \in \mathcal{T}} |\nu_1(\tau) - \nu_2(\tau)| ,
\]
for any \( \nu_1, \nu_2 \in \ell^\infty(\mathcal{T}) \), and for each \( c \in A \) for some set \( A \subset \mathbb{R} \),
\[
\lim_{\epsilon \to 0} \lim_{n \to \infty} \sup_{P \in \mathcal{P}} P \left\{ |f_n, p(\nu_p) - c| \leq C_{n,p} \epsilon \right\} = 0.
\]
Then, for each \( c \in A \) and each \( \epsilon > 0 \),
\[
\lim_{n \to \infty} \sup_{P \in \mathcal{P}} P \left\{ \left| P \left\{ f_n, p(\nu^*_n) \leq c \right\} - P \left\{ f_n, p(\nu_p) \leq c \right\} \right| > \epsilon \right\} = 0.
\]

A.3. **Auxiliary Results II: Convergence of Stochastic Processes in Test Statistics.** We apply the general results on the empirical processes in the previous section to the convergence
of stochastic processes in test statistics in this paper. For this subsection, we assume that Assumptions A.1-A.3 hold.

For \( j = 1, \ldots, p \), let
\[
\hat{\nu}_n, j, P(\beta, \gamma) = \frac{\sqrt{n}(m_j(\beta, \gamma) - m_{j, p}(\beta, \gamma))}{\hat{\sigma}_j(\beta, \gamma)}, \quad \text{and} \quad (A.24)
\]
and let \( \hat{\nu}_n, P = [\hat{\nu}_1, P, \ldots, \hat{\nu}_p, P]' \), and \( \tilde{\nu}_n, P(\beta, \gamma) = \frac{p_n(m_j(\beta, \gamma) - m_{j, p}(\beta, \gamma))}{\sigma_{j, p}(\beta, \gamma)} \), and (A.24)
\[
\tilde{\nu}_n, j, P(\beta, \gamma) = \frac{p_n(m_j(\beta, \gamma) - m_{j, p}(\beta, \gamma))}{\sigma_{j, p}(\beta, \gamma)}, \quad \text{and} \quad \tilde{\nu}_n, P = [\tilde{\nu}_{n, 1, P}, \ldots, \tilde{\nu}_{n, p, P}]'.
\]

Similarly, we also take for \( j = 1, \ldots, p \),
\[
\hat{\nu}^*_n(\beta, \gamma) = \frac{\sqrt{n}(m^*_j(\beta, \gamma) - \overline{m}_j(\beta, \gamma))}{\hat{\sigma}_j(\beta, \gamma)}, \quad \text{and} \quad \hat{\nu}^*_n, j, P(\beta, \gamma) = \frac{\sqrt{n}(m^*_j(\beta, \gamma) - m_{j, p}(\beta, \gamma))}{\hat{\sigma}_j(\beta, \gamma)} \quad \text{and} \quad \hat{\nu}^*_n, P = [\hat{\nu}^*_{n, 1, P}, \ldots, \hat{\nu}^*_{n, p, P}]'.
\]
and let \( \hat{\nu}_n, P = [\hat{\nu}_1, P, \ldots, \hat{\nu}_p, P] \) and \( \tilde{\nu}_n, P = [\tilde{\nu}_{n, 1, P}, \ldots, \tilde{\nu}_{n, p, P}]' \). Let us first present the weak convergence results of the processes, \( \hat{\nu}_{n, j, P}(\beta, \gamma) \) and \( \tilde{\nu}_{n, j, P}(\beta, \gamma) \). Let \( \mathcal{G}_n \) be the \( \sigma \)-field generated by \( \{Z_i\}_{i=1}^n \).

Let \( BL_1 \) be the same as \( BL_1 \) in (A.12) except that \( \mathcal{T} \) is replaced by \( \Theta \).

**Lemma A.2.** There exists a tight Borel measurable Gaussian process \( \nu_p \) on \( \Theta \) such that
\[
\sup_{f \in BL_1} \sup_{P \in \mathcal{P}} \left| E_p[f(\tilde{\nu}_n, P)] - E_p[f(\nu_p)] \right| \to 0, \quad \text{as} \quad n \to \infty, \quad (A.26)
\]
and for each \( \epsilon > 0 \),
\[
\sup_{P \in \mathcal{P}} \left\{ \sup_{f \in BL_1} \left| E[f(\tilde{\nu}_n)|\mathcal{G}_n] - E_p[f(\nu_p)] \right| > \epsilon \right\} \to 0, \quad \text{as} \quad n \to \infty. \quad (A.27)
\]

**Proof:** For the first statement, we use Theorem A.2, and verify the conditions there. To map the current set-up to that of Theorem A.2, we take \( \mathcal{T} = \Theta \) and identify \( \tau = (\beta, \gamma) \in \Theta, X_i = Z_i, \) and \( f_{j, p}(Z_i, \tau) = m_j(Z_i; \beta, \gamma)/\sigma_{j, p}(\beta, \gamma) \). Condition (i) of the theorem follows by Assumption A.2(i), and Condition (ii) by Assumption A.2(ii) and the assumption that \( \Theta \) is compact. Thus (A.26) follows by Theorem A.2. By Theorem A.3, the second statement follows similarly.
Lemma A.3. For all $\epsilon > 0$,

$$\max_{1 \leq j \leq p} \sup_{P \in \mathcal{P}} \left\{ \sup_{(\beta, \gamma) \in \Theta} \left| \frac{\sigma_{j, P}(\beta, \gamma)}{\hat{\sigma}_j(\beta, \gamma)} - 1 \right| > \epsilon \right\} \rightarrow 0,$$

as $n \rightarrow \infty$.

**Proof:** We can show this using Assumption A.2(i) and (ii) and following the same steps in the proof of Lemma D.2 of Bugni, Canay, and Shi (2015). Details are omitted. ■

Now combining Lemmas A.2 and A.3, we obtain the following result as a corollary.

Lemma A.4. There exists a tight Borel measurable Gaussian process $\nu_p$ on $\Theta$ such that

$$\sup_{f \in \mathcal{B}L_1} \sup_{P \in \mathcal{P}} \left| \mathbb{E}_P[f(\hat{\nu}_n, P)] - \mathbb{E}_P[f(\nu_p)] \right| \rightarrow 0, \text{ as } n \rightarrow \infty, \quad (A.28)$$

and for each $\epsilon > 0$,

$$\sup_{P \in \mathcal{P}} \left\{ \sup_{f \in \mathcal{B}L_1} \left| \mathbb{E}_P[f(\hat{\nu}_n)|G_n] - \mathbb{E}_P[f(\nu_p)] \right| > \epsilon \right\} \rightarrow 0, \text{ as } n \rightarrow \infty. \quad (A.29)$$

**Proof:** Note that

$$\hat{\nu}_{n,p}(\beta, \gamma) = \tilde{\nu}_{n,p}(\beta, \gamma) + \tilde{\nu}_{n,p}(\beta, \gamma) \left( \frac{\sigma_{j, P}(\beta, \gamma)}{\hat{\sigma}_j(\beta, \gamma)} - 1 \right).$$

By Lemma A.2, the process $\tilde{\nu}_{n,p}$ is asymptotically uniformly tight uniformly over $P \in \mathcal{P}$. By Lemma A.3, the last term on the right hand side is $o_P(1)$ uniformly over $(\beta, \gamma) \in \Theta$ and over $P \in \mathcal{P}$. Lemma A.2 then yields the first result of the lemma. The proof for the second result is similar. ■

Lemma A.5.

$$\inf_{P \in \mathcal{P}} P \left\{ \Gamma_{\kappa, P}^{\text{LRR}}(\beta) \subset \hat{\Gamma}_{\kappa}^{\text{LRR}}(\beta), \forall \beta \in B \right\} \rightarrow 1, \quad (A.30)$$

as $n \rightarrow \infty$. 
Proof: In light of Assumption A.1, it suffices to show that for all $\varepsilon > 0$,

$$\sup_{P \in \mathcal{P}} \left\{ \sup_{(\beta, \gamma) \in \Theta} |\hat{Q}(\beta, \gamma) - Q_P(\beta, \gamma)| > \varepsilon \right\} \to 0,$$

as $n \to \infty$, where

$$Q_P(\beta, \gamma) = \sum_{j=1}^{p} \left[ \frac{m_{j,p}(\beta, \gamma)}{\sigma_{j,p}(\beta, \gamma)} \right]_+.$$

(A.31)

First note that $\hat{Q}(\beta, \gamma)$ is a continuous map of $m_j(\beta, \gamma)/\sigma_j(\beta, \gamma)$. We use Assumption A.2(ii) and the fact that $\Theta$ is compact to apply Glivenko-Cantelli lemma (i.e., the uniform law of large numbers) to the process $m_j(\beta, \gamma)/\sigma_j(\beta, \gamma)$. Then, the final result is obtained from Lemma A.3. As the arguments are standard, details are omitted. ■

A.4. Auxiliary Results III: Approximation by Truncated Functionals. The test statistics in this paper are functionals of shifted empirical processes. As $n$ grows large, the shift can grow to infinity, which makes it cumbersome to apply the Continuous Mapping Theorem. To deal with this difficulty, we use the tightness of the limit Gaussian processes and approximate the functionals by truncated ones.

We define

$$q_{n,j,p}(\beta, \gamma) = \frac{\sqrt{n}m_{j,p}(\beta, \gamma)}{\sigma_{j,p}(\beta, \gamma)}$$

and

$$\hat{q}_{n,j,p}(\beta, \gamma) = \frac{\sqrt{n}m_{j,p}(\beta, \gamma)}{\hat{\sigma}_{j,p}(\beta, \gamma)}.$$

(A.32)

Given $\nu \in \ell^\infty(\Theta)$, let

$$h_n(\nu)(\beta, \gamma) = \sum_{j=1}^{p} [\nu_j + q_{n,j,p}]_+(\beta, \gamma),$$

(A.33)

and

$$\hat{h}_n(\nu)(\beta, \gamma) = \sum_{j=1}^{p} [\nu_j + \hat{q}_{n,j,p}]_+(\beta, \gamma),$$
and for some constant $M > 0$, define

$$\tilde{h}_n^M(\nu)(\beta, \gamma) = \sum_{j=1}^p \min\{[\nu_j + \max\{q_{n,j,p}, -M\}]_+\}(\beta, \gamma),$$

(A.34)

$$\hat{h}_n^M(\nu)(\beta, \gamma) = \sum_{j=1}^p \min\{[\nu_j + \max\{\hat{q}_{n,j,p}, -M\}]_+\}(\beta, \gamma),$$

(A.35)

From here on, the constant $M > 0$ is taken to be such that $M > M_0$ with the constant $M_0 > 0$ in Assumption A.3. The following lemma delivers the approximation of functional $h_n$ by a truncated version.

**Lemma A.6.** For any $\nu \in \ell^\infty(\Theta)$, and $n \geq 1$,

$$\sup_{(\beta, \gamma) \in \Theta_p} |h_n(\nu)(\beta, \gamma) - \tilde{h}_n^M(\nu)(\beta, \gamma)| \leq 3M \sum_{j=1}^p A_{j,M}(\nu),$$

where $A_{j,M}(\nu) = 1\{\sup_{(\beta, \gamma) \in \Theta} |v_j(\beta, \gamma)| \geq M\}$.

**Proof:** Let

$$h_n^M(\nu)(\beta, \gamma) = \sum_{j=1}^p \min\{[\nu_j + \max\{q_{n,j,p}, -M\}]_+\}(\beta, \gamma),$$

(A.35)

$$\hat{h}_n^M(\nu)(\beta, \gamma) = \sum_{j=1}^p \min\{[\nu_j + \max\{\hat{q}_{n,j,p}\}]_+\}(\beta, \gamma).$$

(A.36)

Let us also define

$$h_n^{M\Delta}(\nu)(\beta, \gamma) = |h_n(\nu)(\beta, \gamma) - h_n^M(\nu)(\beta, \gamma)|$$

(A.37)

$$\tilde{h}_n^{M\Delta}(\nu)(\beta, \gamma) = |h_n^M(\nu)(\beta, \gamma) - \tilde{h}_n^M(\nu)(\beta, \gamma)|,$$

and

$$\hat{h}_n^{M\Delta}(\nu)(\beta, \gamma) = |\tilde{h}_n^M(\nu)(\beta, \gamma) - \hat{h}_n^M(\nu)(\beta, \gamma)|.$$

Since we can bound

$$|h_n(\nu)(\beta, \gamma) - \tilde{h}_n^M(\nu)(\beta, \gamma)| \leq h_n^{M\Delta}(\nu)(\beta, \gamma) + \tilde{h}_n^{M\Delta}(\nu)(\beta, \gamma),$$
it suffices to show that

\[ \sup_{(\beta, \gamma) \in \Theta_p} \tilde{h}_n^{M\Delta}(\nu)(\beta, \gamma) \leq M \sum_{j=1}^{p} A_{j,M}(\nu), \quad (A.38) \]

\[ \sup_{(\beta, \gamma) \in \Theta_p} \tilde{h}_n^{M\Delta}(\nu)(\beta, \gamma) \leq M \sum_{j=1}^{p} A_{j,M}(\nu), \] and

\[ \sup_{(\beta, \gamma) \in \Theta_p} |h_n^{M\Delta}(\nu)(\beta, \gamma) - \tilde{h}_n^{M\Delta}(\nu)(\beta, \gamma)| \leq M \sum_{j=1}^{p} A_{j,M}(\nu). \]

For each \( j = 1, \ldots, p \) and \((\beta, \gamma) \in \Theta_p\), the absolute difference

\[ |\min\{\nu_j + q_{n,j,p}\}_+(\beta, \gamma, M) - \min\{\nu_j + \max\{q_{n,j,p}, -M\}\}_+(\beta, \gamma, M)| \quad (A.39) \]

is zero if \( q_{n,j,p}(\beta, \gamma) \geq -M \). Suppose that \( q_{n,j,p}(\beta, \gamma) < -M \). Then the difference is again zero if \( \nu_j(\beta, \gamma) \leq M \). Thus, we obtain the first statement in (A.38). As for the second statement, note that for each \( j = 1, \ldots, p \) and \((\beta, \gamma) \in \Theta_p\), the difference

\[ |\nu_j + \max\{q_{n,j,p}, -M\}_+ - \min\{\nu_j + \max\{q_{n,j,p}, -M\}\}_+| \quad (A.40) \]

is not zero only if \( \nu_j + \max\{q_{n,j,p}, -M\}_+ \geq M \). Since \( \max\{q_{n,j,p}(\beta, \gamma), -M\} \leq 0 \) for each \((\beta, \gamma) \in \Theta_p\), the latter inequality implies that \( \nu_j(\beta, \gamma) \geq M \). Thus we obtain the second statement. As for the third statement in (A.38), the difference between

\[ |\nu_j + q_{n,j,p}_+ - \nu_j + \max\{q_{n,j,p}, -M\}_+| \text{ and } (A.41) \]

\[ |\min\{\nu_j + q_{n,j,p}_+, (\beta, \gamma, M) - \min\{\nu_j + \max\{q_{n,j,p}, -M\}_+, (\beta, \gamma, M)\}|, \]

is zero if \( q_{n,j,p}(\beta, \gamma) > -M \). Assume that \( q_{n,j,p}(\beta, \gamma) \leq -M \). Then the last difference is not zero only if \( \nu_j + q_{n,j,p}_+ \geq M \) or

\[ \nu_j + \max\{q_{n,j,p}, -M\}_+ \geq M. \quad (A.42) \]

The union of the latter two events, under \( q_{n,j,p}(\beta, \gamma) \leq -M \), is contained in the event \( \nu_j - M \geq M \) which implies that \( \nu_j(\beta, \gamma) > M \) or that \( |\nu_j(\beta, \gamma)| > M \). ■
Lemma A.7. The following holds for each $\epsilon, \eta > 0$, as $n \to \infty$ and $M \to \infty$.

(i) 
\[
\sup_{P \in \mathcal{P}} \sup_{(\beta, \gamma) \in \Theta} P \left\{ |\hat{h}_n(\hat{\nu}_{n,P})(\beta, \gamma) - \hat{h}_n^M(\hat{\nu}_{n,P})(\beta, \gamma)| > \epsilon \right\} \to 0. 
\]

(ii) 
\[
\sup_{P \in \mathcal{P}} \sup_{(\beta, \gamma) \in \Theta} P \left\{ P \left\{ |\hat{h}_n(\hat{\nu}_{n,P})(\beta, \gamma) - \hat{h}_n^M(\hat{\nu}_{n,P})(\beta, \gamma)| > \epsilon \right\} > \eta \right\} \to 0. 
\]

Proof: (i) First, we show that
\[
\sup_{P \in \mathcal{P}} \sup_{(\beta, \gamma) \in \Theta} P \left\{ |\hat{h}_n(\hat{\nu}_{n,P})(\beta, \gamma) - \hat{h}_n^M(\hat{\nu}_{n,P})(\beta, \gamma)| > \epsilon \right\} \to 0. \tag{A.43}
\]

By the first statement of Lemma A.4, we find that
\[
\sup_{P \in \mathcal{P}} \left| P \left\{ \sup_{(\beta, \gamma) \in \Theta} |\hat{\nu}_{n,j,P}(\beta, \gamma)| \geq M \right\} - P \left\{ \sup_{(\beta, \gamma) \in \Theta} |\nu_{j,P}(\beta, \gamma)| \geq M \right\} \right| \to 0,
\]
where $\hat{\nu}_{n,j,P}(\beta, \gamma)$ denotes the $j$-th component of $\hat{\nu}_{n,P}(\beta, \gamma)$. By Lemma A.6, it suffices to show that for any $j = 1, \ldots, p$,
\[
\sup_{P \in \mathcal{P}} MP \left\{ \sup_{(\beta, \gamma) \in \Theta} |\nu_{j,P}(\beta, \gamma)| \geq M \right\} \to 0,
\]
as $M \to \infty$. However, the last probability is bounded by
\[
M^{-1} E \left[ \sup_{(\beta, \gamma) \in \Theta} |\nu_{j,P}(\beta, \gamma)|^2 \right] \leq \frac{4}{M} \left\| \sup_{(\beta, \gamma) \in \Theta} |\nu_{j,P}(\beta, \gamma)| \right\|^2_{\psi_2} \tag{A.44}
\]
for some $C > 0$, where $\| \cdot \|_{\psi_2}$ denotes the Orlicz norm with $\psi_2(x) = \exp(x^2) - 1$. The first inequality follows because for any random variable $X$ and $p \geq 1$, $\|X\|_{2p} \leq p!\|X\|_{\psi_2}$. The last inequality follows from the proof of Corollary 2.2.8 of van der Vaart and Wellner (1996) and the bracketing entropy bound which comes from Assumption A.2 and the assumption that $\Theta$
is compact. The last term in (A.44) vanishes as \( M \to \infty \). Thus we obtain the convergence in (A.43). Then the desired result follows from this convergence and Lemma A.3.

(ii) The second statement follows similarly by using the second statement of Lemma A.4, and following similar arguments in the proof of (i) using Lemma A.3.

**Lemma A.8.** For each \( M > M_0 \) and each \( c \in (0, pM) \),

\[
\lim_{\varepsilon \to 0} \lim_{n \to \infty} \sup_{P \in \mathcal{P}} \sup_{(\beta, \gamma) \in \Theta_P} \sup_{|\nu_p| \leq p} \{ |\hat{h}_n^M(\nu_p)(\beta, \gamma) - c| \leq \varepsilon \} = 0.
\]  

(A.45)

**Proof:** Let \( \bar{r}_{j,p}(\beta, \gamma) = \max\{q_{n,j,p}(\beta, \gamma), -M\} \mathbb{1}\{(\beta, \gamma) \in \Theta_P\}, (\beta, \gamma) \in \Theta \). Then \( \bar{r}_{j,p} \in R_M \) for each \( j = 1, \ldots, p \), and the desired result follows by Assumption A.3.

**Lemma A.9.** For each \( M > M_0 \) and \( \eta \in (0, pM/2) \),

\[
\sup_{P \in \mathcal{P}} \sup_{(\beta, \gamma) \in \Theta_P} \sup_{\eta \leq |\nu_p| \leq pM - \eta} \left| P\left\{ \hat{h}_n^M(\hat{\nu}_{n,p})(\beta, \gamma) \leq c \right\} - P\left\{ \bar{h}_n^M(\nu_p)(\beta, \gamma) \leq c \right\} \right| \to 0,
\]

and, for each \( \varepsilon > 0 \),

\[
\sup_{P \in \mathcal{P}} \sup_{(\beta, \gamma) \in \Theta_P} \left\{ \sup_{\eta \leq |\nu_p| \leq pM - \eta} \left| P\left\{ \hat{h}_n^M(\hat{\nu}_{n,p})(\beta, \gamma) \leq c \right\} - P\left\{ \bar{h}_n^M(\nu_p)(\beta, \gamma) \leq c \right\} \right| > \varepsilon \right\} \to 0,
\]

as \( n \to \infty \).

**Proof:** For each fixed \( M > 0 \), the functional \( \bar{h}_n^M \) is bounded Lipschitz with the Lipschitz coefficient equal to \( p \). By Lemmas A.1, A.4, and A.8, we have

\[
\sup_{P \in \mathcal{P}} \sup_{(\beta, \gamma) \in \Theta_P} \left| P\left\{ \hat{h}_n^M(\hat{\nu}_{n,p})(\beta, \gamma) \leq c \right\} - P\left\{ \bar{h}_n^M(\nu_p)(\beta, \gamma) \leq c \right\} \right| \to 0,
\]  

(A.46)

for each \( c \in [\eta, pM - \eta] \). Due to Assumption A.3(i), we can elevate this pointwise convergence with uniformity over \( c \in [\eta, pM - \eta] \) by using the same arguments in the proof of Lemma 2.11 of van der Vaart (1998), p.12. Using Lemmas A.2 and A.1, we can follow the same arguments to obtain the second statement for the bootstrap empirical process.
Lemma A.10. For any \( \varepsilon > 0 \), there exists \( M_\varepsilon > 0 \) such that

\[
\limsup_{n \to \infty} \sup_{P \in P} P \left\{ h_n(\hat{v}_{n,P})(\beta, \gamma) \geq M_\varepsilon \right\} < \varepsilon,
\]

and

\[
\limsup_{n \to \infty} \sup_{P \in P} P \left\{ P \left\{ h_n(\hat{v}_{n,P})(\beta, \gamma) \geq M_\varepsilon \right\} > \varepsilon \right\} = 0.
\]

Proof: Note that

\[
\sup_{(\beta, \gamma) \in \Theta} h_n(\nu)(\beta, \gamma) = \sup_{(\beta, \gamma) \in \Theta} \sum_{j=1}^{p} [\nu_j + q_{n,j,P}](\beta, \gamma) \geq \sup_{(\beta, \gamma) \in \Theta} \sum_{j=1}^{p} |\nu_j(\beta, \gamma)|,
\]

because \( q_{n,j,P}(\beta, \gamma) \leq 0 \) for \((\beta, \gamma) \in \Theta\). By Lemma A.4, \( h_n(\hat{v}_{n,P})(\beta, \gamma) \) is uniformly tight uniformly over \( P \in P \), and thus we obtain the first result.

As for the second result, note that

\[
\sup_{P \in P} \sup_{(\beta, \gamma) \in \Theta} P \left\{ P \left\{ h_n(\hat{v}_{n,P})(\beta, \gamma) \geq M_\varepsilon \right\} > \varepsilon \right\}
\leq \sup_{P \in P} \sup_{(\beta, \gamma) \in \Theta} P \left\{ |P \left\{ h_n(\hat{v}_{n,P})(\beta, \gamma) \geq M_\varepsilon \right\} - P \left\{ h_n(\nu_P)(\beta, \gamma) \geq M_\varepsilon \right\}| > \varepsilon/2 \right\}
+ \sup_{P \in P} \sup_{(\beta, \gamma) \in \Theta} 1 \{ P \left\{ h_n(\nu_P)(\beta, \gamma) \geq M_\varepsilon \right\} > \varepsilon/2 \}.
\]

The first supremum on the right hand side goes to zero as \( n \to \infty \) by Lemma A.4. By (A.47) and the tightness of the Gaussian process \( \nu_P \), we can choose a large \( M_\varepsilon \) such that the last term becomes zero from some large \( n \) on. ■

Lemma A.11. For each \( M > M_0 \) and \( \eta \in (0, pM/2) \),

\[
\sup_{\beta, \gamma \in \Theta} \sup_{\eta \leq c \leq pM-\eta} \left| P \left\{ \hat{h}_n(\hat{v}_{n,P})(\beta, \gamma) \leq c \right\} - P \left\{ h_n(\nu_P)(\beta, \gamma) \leq c \right\} \right| \to 0,
\]
and, for each $\epsilon > 0$,

$$\sup_{\eta \leq \epsilon \leq \eta M - \eta} \sup_{P \in \mathcal{P}} P \left\{ \sup_{\eta \leq \epsilon \leq \eta M - \eta} \left| P \left\{ \hat{h}_n(\hat{v}_n^s)(\beta, \gamma) \leq c|\mathcal{G}_n \} - P \left\{ h_n(\nu_p)(\beta, \gamma) \leq c \right\} \right| > \epsilon \right\} \to 0,$$

as $n \to \infty$.

**Proof:** Since $\nu_p$ is a tight Gaussian process and $q_{n,j,p}(\beta, \gamma) \leq 0$ for $(\beta, \gamma) \in \Theta_p$, for each $\epsilon > 0$, there exists $M_\epsilon > 0$ such that

$$\sup_{\eta \leq \epsilon \leq \eta M - \eta} \sup_{P \in \mathcal{P}} P \left\{ h_n(\nu_p)(\beta, \gamma) \geq M_\epsilon \right\} < \epsilon.$$

The desired results follow from Lemmas A.7, A.9 and A.10. ■

A.5. **Proof of Theorem A.1.** First note that we can write

$$T(\beta, \gamma) = \hat{h}_n(\hat{v}_{n,p})(\beta, \gamma).$$

Let us define

$$\tilde{T}^*(\beta, \gamma) = \hat{h}_n(\hat{v}_n^s)(\beta, \gamma).$$

(A.48)

Note that from the definition of $\hat{\lambda}_{j,\alpha_1}(\beta, \gamma)$ in (4.13),

$$\frac{\sqrt{n}\hat{\lambda}_{j,\alpha_1}(\beta, \gamma)}{\hat{\sigma}(\beta, \gamma)} = \min \left\{ \hat{v}_{j,n}(\beta, \gamma) + \hat{q}_{j,n,p}(\beta, \gamma) - \hat{k}_{\alpha_1}(\beta, \gamma), 0 \right\},$$

where $\hat{q}_{n,j,p}$ is as defined in (A.32). If we define

$$\hat{h}_{\alpha_1,n}(\nu)(\beta, \gamma) = \sum_{j=1}^{p} \left[ \nu_j(\beta, \gamma) + \min \{ \hat{v}_{j,n}(\beta, \gamma) + \hat{q}_{j,n,p}(\beta, \gamma) - \hat{k}_{\alpha_1}(\beta, \gamma), 0 \} \right]_+, $$

we can write

$$\tilde{T}^*(\beta, \gamma) = \hat{h}_{\alpha_1,n}(\hat{v}_n^s)(\beta, \gamma).$$
Let \( \hat{c}_{1-\alpha_1}(\beta, \gamma) \) be the \( 1 - \alpha + \alpha_1 \) quantile of the bootstrap distribution of \( \tilde{T}^*(\beta, \gamma) \), and recall that \( \hat{c}_{1-\alpha_1}(\beta, \gamma) \) is the \( 1 - \alpha + \alpha_1 \) quantile of the bootstrap distribution of \( \tilde{T}^*(\beta, \gamma) \). We compare \( \hat{c}_{1-\alpha_1}(\beta, \gamma) \) and \( \tilde{c}_{1-\alpha_1}(\beta, \gamma) \). For this, we first prepare the following lemma.

**Lemma A.12.**

\[
\lim_{n \to \infty} \inf_{P \in \mathcal{P}} \inf_{(\beta, \gamma) \in \Theta_P} \left\{ \min_{1 \leq j \leq k} \hat{\nu}_{n,j,p}(\beta, \gamma) \geq \hat{\kappa}_{\alpha_1}(\beta, \gamma) \right\} \geq 1 - \alpha_1,
\]

as \( n \to \infty \).

**Proof:** By the second statement of Lemma A.4 and Assumption A.3(ii), there exists \( \kappa_{\alpha_1}(\beta, \gamma) \) such that for any \( \epsilon > 0 \),

\[
sup_{P \in \mathcal{P}} \sup_{(\beta, \gamma) \in \Theta_P} P \left\{ |\hat{\kappa}_{\alpha_1}(\beta, \gamma) - \kappa_{\alpha_1}(\beta, \gamma)| > \epsilon \right\} = o(1), \tag{A.49}
\]

as \( n \to \infty \), and the density of \( \min_{1 \leq j \leq k} \nu_{j,p}(\beta, \gamma) \) is bounded away from zero in a neighborhood of \( \kappa_{\alpha_1}(\beta, \gamma) \). By using the definition of \( \hat{\kappa}_{\alpha_1}(\beta, \gamma) \) and the second statement of Lemma A.4, and the standard arguments of approximating an indicator function by a sequence of bounded Lipschitz functions (e.g., see the proof of Lemma A.1 of Linton, Song, and Whang (2010)), we have

\[
1 - \alpha_1 \leq \inf_{(\beta, \gamma) \in \Theta_P} P \left\{ \min_{1 \leq j \leq k} \hat{\nu}_{n,j}(\beta, \gamma) \geq \hat{\kappa}_{\alpha_1}(\beta, \gamma) \right\}
\]

\[
\leq \inf_{(\beta, \gamma) \in \Theta_P} P \left\{ \min_{1 \leq j \leq k} \nu_{j,p}(\beta, \gamma) \geq \kappa_{\alpha_1}(\beta, \gamma) \right\} + o_p(1),
\]

uniformly over \( P \in \mathcal{P} \) by Lemma A.5. By the first statement of Lemma A.4 and Lemma A.1 of Linton, Song, and Whang (2010), the last infimum is equal to

\[
\inf_{(\beta, \gamma) \in \Theta_P} P \left\{ \min_{1 \leq j \leq k} \hat{\nu}_{n,j}(\beta, \gamma) \geq \hat{\kappa}_{\alpha_1}(\beta, \gamma) \right\} + o(1)
\]

\[
\leq \inf_{(\beta, \gamma) \in \Theta_P} P \left\{ \min_{1 \leq j \leq k} \nu_{j,p}(\beta, \gamma) \geq \kappa_{\alpha_1}(\beta, \gamma) \right\} + o(1),
\]

uniformly over \( P \in \mathcal{P} \), by Lemma A.5 and (A.49). \( \blacksquare \)
The lemma below compares the bootstrap critical values $\tilde{c}_{1-a+a_1}(\beta, \gamma)$ and $\tilde{c}_{1-a+a_1}(\beta, \gamma)$.

**Lemma A.13.** Suppose that the conditions of Theorem A.1 hold. Then,

$$\lim_{n \to \infty} \inf_{\beta, \gamma} P \{ \tilde{c}_{1-a+a_1}(\beta, \gamma) \geq \tilde{c}_{1-a+a_1}(\beta, \gamma) \} \geq 1 - \alpha_1.$$  \hspace{1cm} (A.50)

**Proof:** Denote by $E_n(\beta, \gamma)$ the event that for all $t \geq 0$,

$$P \{ \hat{h}_{a_1,n}(\hat{\nu}_{n}^*) = t \} \leq \hat{c}_{a_1}(\beta, \gamma).$$

and denote by $A_n(\beta, \gamma)$ the event that

$$\min_{1 \leq j \leq p} \hat{\nu}_{n,j,p}(\beta, \gamma) \geq \hat{c}_{a_1}(\beta, \gamma).$$

Then, since $E_n(\beta, \gamma)$ contains $A_n(\beta, \gamma)$,

$$\lim_{n \to \infty} \inf_{\beta, \gamma} P E_n(\beta, \gamma) \geq \lim_{n \to \infty} \inf_{\beta, \gamma} P A_n(\beta, \gamma) \geq 1 - \alpha_1,$$

by Lemma A.12. This completes the proof because $P E_n(\beta, \gamma)$ is bounded by the probability in (A.30). □

**Proof of Theorem A.1:** Note that

$$\limsup_{n \to \infty} \sup_{\beta, \gamma} P \{ T(\beta, \gamma) > \tilde{c}_{1-a+a_1}(\beta, \gamma) \}$$

$$\leq \limsup_{n \to \infty} \sup_{\beta, \gamma} P \{ T(\beta, \gamma) > \tilde{c}_{1-a+a_1}(\beta, \gamma) \} + \alpha_1 \text{ (by Lemma A.13)}.$$

Hence the proof is complete once we show that the last limsup is bounded by $\alpha - \alpha_1$. However, this latter result comes from Lemma A.11 and Assumption A.3(ii). The uniform validity of the confidence interval follows from this result and Lemma A.5. □
Appendix B. Additional Details on the Data for the Empirical Application

Our empirical application follows Romano and Shaikh (2010) in using data from the 2000 Annual Demographic Supplement of the Current Population Survey (CPS) to study the relationship between covariates and wages in the presence of top-coding. The CPS contains information on both individual characteristics (e.g. race, education, gender) and wages (from the primary earning job as well as those from other sources), making it appropriate for our analysis. However, the exercise in Romano and Shaikh (2010) only covered college graduates. By contrast, our interest in the effects of college requires a sample with both college graduates and non-graduates. As a result, we must expand their sample to cover the latter group. Here, we describe our steps in this regard.

First, we used the IPUMS (Integrated Public Use Microdata Series) version of the CPS data - see Flood, King, Rodgers, Ruggles, and Warren (2020). While the dataset is the same, this version integrates multiple years of the data and clarifies the original dictionary. As in Romano and Shaikh (2010) p.186, we keep individuals who are between 20 and 24 of age (using the variable AGE in the dataset), white (using the variable RACE), with a primary source of income given by wages and salaries (using the variable SRCEARN, which stands for “Source of earnings from longest job”) and that worked at least 2 hours per week on average (using the variable UHRSWORKLY). For wages, we used the variable INCWAGE which measures the total wage and salary income. This variable can be decomposed into “Earnings from longest job” (variable INCLONGJ in the CPS) and “Earnings from other work included wage and salary earnings” (variable OINCWAGE). From their code and data, it appears that Romano and Shaikh (2010) use the sum of the latter two.

To code college graduates and non-college graduates, we use the standard definition from the CPS - and, to the best of our understanding - the one used by Romano and Shaikh (2010). That is, we code college graduates as those who have an “Associate's Degree, Academic Program”, a “Bachelor's Degree”, a “Master's Degree”, a “Professional Degree” or a “Doctorate Degree”. Non-graduates are those with “Associate's Degree, Occupational/Vocational Program”,
“Some College/No Degree”, or those whose highest education is a higher school degree (or the equivalent) or lower. We dropped individuals with missing data in one of these variables. This results in a final sample of 5816 individuals, of which 826 are college graduates (14.2% of the sample), and 4990 non-graduates (85.8% of the sample). For our empirical application, we code the treatment variable $D_i$ as equal to 1 for those in the first group, and 0 for the latter.\footnote{We note that our resulting sample size for college graduates is 826, which is much larger than the 305 in Romano and Shaikh (2010). This is the case even though this subsample imposes the same restrictions outlined in their work. Unfortunately, we could not find details in their published version or in their Supplementary Material, such as further sample restrictions, which could explain these differences. We then extensively investigated the dataset to find why such differences could exist. First, we found that all of their original (distinct) observations are present in our sample of 826. We conclude that our sample comprehensively covers theirs and that they must have imposed additional restrictions. We extensively probed different variables to check for the latter, but we could not find a clear restriction. For instance, we could obtain the same number of observations, 305, by further restricting our sample to the first respondents in the household ($\text{PERNUM}$ equal to 1) and by dropping observations those whose households lack a CPS identifier across surveys (based on the variable $\text{CPSID}$). However, these restrictions dropped over 40% of the observations in the original Romano and Shaikh (2010) sample, so we did not pursue it. We repeated our exercises with the non-IPUMS version of the CPS, available on the NBER webpage, to the same conclusions.}
| Notation | Description | Place of Def. |
|----------|-------------|--------------|
| $B_P$ | Identified set for $\beta$, the parameter of interest, given $P \in P$ | |
| $\beta_0$ | The parameter of interest. | |
| $C^{\text{LRR}}_{1-\alpha}$ | A bootstrap confidence region for $\beta_0$ under the LRR hypothesis. (based on the least favorable configuration) | (4.10) |
| $\tilde{C}^{\text{LRR}}_{1-\alpha}$ | A bootstrap confidence region for $\beta_0$ under the LRR hypothesis. (based on the approach of Romano, Shaikh, and Wolf (2014)) | (4.16) |
| $\hat{c}_{1-\alpha}(\beta)$ | The $1 - \alpha$ quantile bootstrap critical value. | Above (4.10) |
| $\hat{c}_{1-\alpha+\alpha_1}(\beta)$ | The $1 - \alpha + \alpha_1$ quantile bootstrap critical value. | Above (4.16) |
| $F$ | The distribution of $W = (X, \epsilon)$ in the DGP for the data. | |
| $\tilde{F}$ | The counterfactual distribution of $W = (X, \epsilon)$. | |
| $G(\cdot | w)$ | The conditional distribution of $\eta$ given $W = w$, as the reduced-form selection rule. | |
| $G_n$ | The $\sigma$-field generated by the observations $\{(Y_i,X'_i)'\}_{i=1}^n$. | |
| $\Gamma(\beta)$ | Identified set for $\gamma_0$, given $\beta$ and $P \in P$ | (3.8) |
| $\Gamma^{\text{LRR}}(\beta)$ | The $\kappa$-LRR set. | (3.12) |
| $\hat{\Gamma}(\beta)$ | The estimated version of $\Gamma_P(\beta)$. | (4.5) |
| $\hat{\Gamma}^{\text{LRR}}(\beta)$ | The estimated upper LRR set | (4.11) |
| $\gamma_0$ | The nuisance parameter. | |
| $h_n, \bar{h}_n, \bar{h}_M$ | Functionals used to define various statistics. | (A.33), (A.34) |
| $m_{j\cdot}(\beta, \gamma)$ | The expected moment function in the moment inequality model | (A.2) |
| $\bar{m}_{j\cdot}(\beta, \gamma)$ | The average moment function in the moment inequality model | (4.2) |
| $\hat{\psi}_{n,j,p}(\beta, \gamma)$ | Empirical processes normalized by $\hat{\sigma}_j(\beta, \gamma)$ | (A.24) |
| $\tilde{\psi}_{n,j,p}(\beta, \gamma)$ | Bootstrapped empirical processes normalized by $\sigma_j(\beta, \gamma)$ | (A.25) |
| $\tilde{Q}_p(\beta, \gamma)$ | The population obj. function used to set-identify $(\beta_0, \gamma_0)$ | (A.31) |
| $\hat{Q}(\beta, \gamma)$ | The sample obj. function from the moment inequality model. | (4.4) |
| $\hat{Q}^*(\beta, \gamma)$ | The bootstrap sample objective function. (based on the least favorable configuration.) | (4.7) |
| $Q^{\text{LRR}}(\beta, \gamma)$ | The population criterion for LRR. | (3.10) |
| $T(\beta, \gamma)$ | The test statistic. | (4.6) |
| $\bar{T}^*(\beta, \gamma)$ | The bootstrap test statistic. (based on the approach of Romano, Shaikh, and Wolf (2014)) | (4.15) |
| $\Theta_p$ | Identified set for $(\beta_0, \gamma_0)$ given $P \in P$ | Section 3.1 |
| $\Theta^{\text{LRR}}_p$ | The set of $(\beta_0, \gamma_0) \in \Theta_p$ under the LRR hypothesis. | (3.13) |
References

BUGNI, F. A., I. A. CANAY, AND X. SHI (2015): “Specification Tests for Partially Identified Models Defined by Moment Inequalities,” *Journal of Econometrics*, 185, 259–282.

CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003): “Estimation of Semiparametric Models When the Criterion Function is Not Smooth,” *Econometrica*, 71, 1591–1608.

FLOOD, S., M. KING, R. RODGERS, S. RUGGLES, AND J. R. WARREN (2020): *Integrated Public Use Microdata Series, Current Population Survey: Version 8.0* [dataset], IPUMS, 2020, Minneapolis, MN.

GINÉ, E. (1997): “Lecture Notes on Some Aspects of the Bootstrap,” in *Ecole de 'Ete de Calcul de Probabilit'es de Saint-Flour. Lecture Notes in Mathematics*, vol. 1665.

GINÉ, E., AND J. ZINN (1991): “Gaussian Characterization of Uniform Donsker Classes of Functions,” *Annals of Probability*, 19, 758–782.

LINTON, O., K. SONG, AND Y.-J. WHANG (2010): “An Improved Bootstrap Test of Stochastic Dominance,” *Journal of Econometrics*, 154, 186–202.

ROMANO, J. P., AND A. M. SHAIKH (2010): “Inference for the Identified Set in Partially Identified Econometric Models,” *Econometrica*, 78, 169–211.

ROMANO, J. P., A. M. SHAIKH, AND M. WOLF (2014): “A Practical Two-Step Method for Testing Moment Inequalities,” *Econometrica*, 82, 1979–2002.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press, New York, USA.

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer, New York, USA.

Department of Economics, University of Houston, 3623 Cullen Boulevard - Office 221-C Houston, TX, 77204, USA

Vancouver School of Economics, University of British Columbia, 6000 Iona Drive, Vancouver, BC, V6T 1L4, Canada