Inference in models with partially identified control functions

Andres Aradillas-Lopez*

August 4, 2022

Abstract

In multiple contributions to the literature, James L. Powell and coauthors have developed estimators for nonlinear models where sample selectivity and/or endogeneity can be handled through a "control function". Their methods rely on pairwise comparisons of observations which match (asymptotically) the control functions. Conditional on this matching, a moment condition can identify the parameters of the model. There exist data sets for which the control functions are unobserved, but we have bounds for them which depend on observable covariates. These bounds can arise straightforwardly from the nature of the data (e.g, in the case of interval data), or they can be obtained from an economic model. The inability to observe the control functions precludes the matching proposed in Powell's methods. In this paper we show that, under certain conditions, testable implications can still be obtained through pairwise comparisons of observations for which the control-function bounds are *disjoint*. Testable implications now take the form of functional inequalities. We propose an inferential procedure based on these inequalities and we analyze its properties.

Keywords: Control functions, sample selectivity, endogeneity, partial identification, functional inequalities.
JEL classification: C1, C14, C31, C34.

1 Introduction

One of the many contributions of James L. Powell to econometrics has been the development of methods to estimate nonlinear models where sample selectivity and/or endogeneity can be handled through "control functions" or "control variables" which are identifiable functions of observables in the data. Building upon insights from the partially linear regression model (Robinson (1988)) and panel data models with fixed effects (Chamberlain (1984)), the methods proposed

*Department of Economics, Pennsylvania State University, University Park, PA 16802, United States. Email: aaradill@psu.edu

by Powell and his coauthors rely on pairwise comparisons of observations based on matching (asymptotically) the control functions. This matching asymptotically “differences out” sources of bias such as selection correction terms, or other sources of endogeneity. Conditional on this matching, a moment-equality condition is obtained, which can be used to identify and estimate the model’s parameters. This body of work and insights can be traced back to Powell (1987) and Ahn and Powell (1993), and has been expanded to include richer extensions and more general models, for example, in Honoré and Powell (1994), Powell (2001), Honoré and Powell (2005), Aradillas-López, Honoré, and Powell (2007), and Ahn, Ichimura, Powell, and Ruud (2018).

In many applications, not all of the control variables are observable, making it impossible to implement the matching in Powell’s methods. However, in a number of such instances, the nature of the data observed or an underlying economic theory may produce *bounds* for the control functions which depend on observable covariates. In this paper we focus on such a setting and we show conditions under which pairwise comparisons of observations can still produce testable implications. In this case, pairwise comparisons are based on “matching” observations for which the bounds for the control functions are *disjoint*. If the control functions enter the model in a monotonic way, this approach yields testable implications in the form of conditional *functional inequalities*. Based on these inequalities, we present an inferential procedure and we describe its properties. Although the main focus of our paper is a bivariate sample selection model, our results will show how to extend our inferential approach to other models with similar features.

The paper proceeds as follows. Section 2 describes the model that is the focus of the paper. This is a bivariate sample selection model with unobserved control variables. We show conditions under which, if observable bounds for the control functions exist, the model produces conditional functional inequalities based on pairwise comparisons of observations for which these bounds are disjoint. Section 3 presents an inferential procedure based on the observable implications of our model and describes its asymptotic properties. Section 4 includes Monte Carlo experiments. Section 5 describes general features of econometric models that can be analyzed with our approach. Section 6 concludes. The online appendix of the paper[1] includes the econometric proofs.

2 A bivariate sample selection model with censored data

Consider a model with an outcome of interest Y_2^* which is observed if $Y_1^* > 0$, where Y_1^* is a latent variable. Suppose,

$$\begin{aligned} Y_1^* &= g_1(X_1, \beta_{10}) + \varepsilon_1, \\ Y_2^* &= g_2(X_2, \beta_{20}) + \varepsilon_2. \end{aligned} \tag{1}$$

[1] Available online at http://www.personal.psu.edu/aza12/part-identif-control-functions-appendix.pdf

Suppose Y_2^* is observed if and only if $Y_1^* > 0$, and let

$$Y_1 = \begin{cases} 1 & \text{if } Y_1^* > 0 \\ 0 & \text{if } Y_1^* \leq 0 \end{cases} \qquad \text{and} \qquad Y_2 = \begin{cases} Y_2^* & \text{if } Y_1^* > 0 \\ - & \text{if } Y_1^* \leq 0 \end{cases}$$

Suppose both $g_2(\cdot)$ and $g_1(\cdot)$ are parametric functions, whose functional form is known up to the true value of β_{20} and β_{10}. Suppose Y_1, Y_2 and X_2 are observed[2], but *at least a subset of components of X_1 are unobserved*. Assume, however, that there exist observable covariates W_1, and parametric functions $g_{1L}(W_1, \beta_{10})$, $g_{1U}(W_1, \beta_{10})$ with known functional form (up to the true value of β_{10}), such that,

$$g_{1L}(W_1, \beta_{10}) \leq g_1(X_1, \beta_{10}) \leq g_{1U}(W_1, \beta_{10}) \quad \text{w.p.1.} \tag{2}$$

These bounds can arise from the nature of the data (e.g, in the presence of interval data), or they can be obtained from an economic model. We provide examples next.

2.1 Examples of the type of bounds in equation (2)

2.1.1 Interval data

Suppose $g_1(X_1, \beta_{10}) \equiv X_1'\beta_{10}$, so (1) becomes, $Y_1^* = X_1'\beta_{10} + \varepsilon_1$. Suppose not all the elements in X_1 are observed, but we observe X_{1L} and X_{1U} (with $X_{1L} = X_{1U} = X_1$ for all the observable components of X_1), and we have ex-ante knowledge of the signs of the coefficients in β_{10} for the unobserved elements in X_1, so we know that

$$\beta_{10}'X_{1L} \leq \beta_{10}'X_1 \leq \beta_{10}'X_{1U} \quad \text{w.p.1.}$$

In this case, we have $W_1 \equiv (X_{1L}, X_{1U})$, $g_{1L}(W_1, \beta_{10}) \equiv \beta_{10}'X_{1L}$, and $g_{1U}(W_1, \beta_{10}) \equiv \beta_{10}'X_{1U}$. Interval data is prevalent in many econometric data sets[3], and it has received attention in econometric work. A notable example is Manski and Tamer (2002), who examine inference for the regression functions $E[Y|X, V]$ and $E[V|X]$ when V is an unobserved scalar random variable, and the econometrician observes (Y, X, V_0, V_1), with $V_0 \leq V \leq V_1$ w.p.1. Our setting is different from theirs in that it encompasses multiple equations and (potentially) multiple unobserved control variables. We also go beyond the basic interval-data case and consider models where the control-function bounds themselves may be functionals of the parameter of interest. This is illustrated in the following example.

[2] We will discuss an extension where at least a subset of components of X_2 are unobserved in Section 2.3.1.

[3] Well-known examples include surveys such as the Health and Retirement Study (HRS), which measures respondents' wealth in intervals (see Juster and Suzman (1995, page S35)). Other examples may include demand models, where distance between consumers' homes and candidate grocery stores is the desired regressor but we only observe consumers' zip code.

2.1.2 An interactions-based model with unobserved beliefs

Consider a model where the selection equation depends on economic agents' beliefs about the proportion of other agents with their same observable characteristics that will choose $Y_1 = 1$. Specifically, suppose the selection equation in (1) is of the form,

$$Y_1^* = W_1'\beta_{10}^w + \beta_{10}^\pi \pi_1 + \varepsilon_1,$$

where π_1 denotes the agent's subjective expectation for $P(Y_1 = 1|W_1)$ (the probability that an economic agent with characteristics W_1 will select to "participate"). In this case we have $X_1 \equiv (W_1, \pi_1)$, $\beta_{10} \equiv (\beta_{10}^w, \beta_{10}^\pi)$ and $g_1(X_1, \beta_{10}) \equiv W_1'\beta_{10}^w + \beta_{10}^\pi \pi_1$. Suppose beliefs π_1 are unobserved by the econometrician and that we allow two agents with the same characteristics W_1 to have different (and potentially incorrect) beliefs. If we assume that agents share a common prior for the distribution of ε_1, we can place bounds on unobserved beliefs π_1 based on iterated elimination of nonrationalizable beliefs, or on the stronger assumption of Bayesian Nash equilibrium (BNE) beliefs. In what follows, suppose agents have a common parametric prior $H_1(\cdot)$ for the distribution of ε_1, assumed to be known to the econometrician (e.g, $H_1(\cdot)$ can be the standard normal distribution). *The prior H_1 does not have to correspond to the true distribution of ε_1.* For what follows, we only require that the same prior be used by all agents.

Bounds for $g_1(X_1, \beta_{10})$ based on iterated elimination of nonrationalizable beliefs

Suppose economic theory predicts that $\beta_{10}^\pi \geq 0$, so the likelihood of participation increases with the expected proportion of other agents with the same characteristics who will also participate[4]. We can obtain bounds on beliefs by adapting an approach suggested in Aradillas-López and Tamer (2008) in incomplete-information games to this model. Suppose that the prior H_1 is consistent with the assumption that $\varepsilon_1 \perp W_1$. We describe the iterative procedure of elimination of nonrationalizable beliefs next.

Step 1: Since beliefs are probabilities, they must satisfy $\pi_1 \in [0,1]$. Therefore, any set of beliefs consistent with this fact must satisfy,

$$\underbrace{H_1\left(W_1'\beta_{10}^w\right)}_{\equiv \pi_{1L}^1(W_1,\beta_{10})} \leq \pi_1 \leq \underbrace{H_1\left(W_1'\beta_{10}^w + \beta_{10}^\pi\right)}_{\equiv \pi_{1U}^1(W_1,\beta_{10})}.$$

Beliefs outside this range cannot be rationalized if agents know that $\pi_1 \in [0,1]$.

Step 2: Suppose agents assume that everybody else performs at least one step of elimination of nonrationalizable beliefs. In this case, agents know that everyone else's beliefs satisfy $\pi_{1L}^1(W_1, \beta_{10}) \leq$

[4]The bounds that follow can be re-computed if economic theory predicts that $\beta_{10}^\pi \leq 0$.

$\pi_1 \leq \pi^1_{1U}(W_1, \beta_{10})$, where these bounds are described above. Any set of beliefs consistent with this assumption must satisfy,

$$\underbrace{H_1\left(W_1'\beta^w_{10} + \beta^\pi_{10}\pi^1_{1L}(W_1, \beta_{10})\right)}_{\equiv \pi^2_{1L}(W_1, \beta_{10})} \leq \pi_1 \leq \underbrace{H_1\left(W_1'\beta^w_{10} + \beta^\pi_{10}\pi^1_{1U}(W_1, \beta_{10})\right)}_{\equiv \pi^2_{1U}(W_1, \beta_{10})}.$$

Beliefs outside this range cannot be rationalized if agents assume that everybody else performs at least one step of elimination of nonrationalizable beliefs.

Step k: We can extend this construction to $k \geq 3$ steps iteratively as follows. Suppose agents assume that everybody else's beliefs are consistent with at least $k-1$ steps of elimination of nonrationalizable beliefs, so $\pi^{k-1}_{1L}(W_1, \beta_{10}) \leq \pi_1 \leq \pi^{k-1}_{1U}(W_1, \beta_{10})$. Any set of beliefs consistent with this assumption must satisfy,

$$\underbrace{H_1\left(W_1'\beta^w_{10} + \beta^\pi_{10}\pi^{k-1}_{1L}(W_1, \beta_{10})\right)}_{\equiv \pi^k_{1L}(W_1, \beta_{10})} \leq \pi_1 \leq \underbrace{H_1\left(W_1'\beta^w_{10} + \beta^\pi_{10}\pi^{k-1}_{1U}(W_1, \beta_{10})\right)}_{\equiv \pi^k_{1U}(W_1, \beta_{10})}.$$

Beliefs outside this range cannot be rationalized if agents assume that everybody else performs at least $k-1$ steps of elimination of nonrationalizable beliefs.

If we assume that every agent's beliefs is consistent with at least k steps of iterated elimination of nonrationalizable beliefs, we must have

$$\underbrace{W_1'\beta^w_{10} + \beta^\pi_{10}\pi^k_{1L}(W_1, \beta_{10})}_{\equiv g_{1L}(W_1, \beta_{10})} \leq \underbrace{W_1'\beta^w_{10} + \beta^\pi_{10}\pi_1}_{\equiv g_1(X_1, \beta_{10})} \leq \underbrace{W_1'\beta^w_{10} + \beta^\pi_{10}\pi^k_{1U}(W_1, \beta_{10})}_{\equiv g_{1U}(W_1, \beta_{10})}.$$

In this case, we have

$$g_{1L}(W_1, \beta_{10}) \equiv W_1'\beta^w_{10} + \beta^\pi_{10}\pi^{k-1}_{1L}(W_1, \beta_{10}), \quad g_{1U}(W_1, \beta_{10}) \equiv W_1'\beta^w_{10} + \beta^\pi_{10}\pi^{k-1}_{1U}(W_1, \beta_{10}). \tag{3}$$

Given a parametric specification for the prior $H_1(\cdot)$, these bounds have a parametric functional form for any given θ_1.

Bounds for $g_1(X_1, \beta_{10})$ based on the assumption of BNE beliefs

For a given W_1, BNE beliefs are given by any solution in π_1 to the BNE system

$$\pi_1 = H_1(W_1'\beta^w_{10} + \beta^\pi_{10}\pi_1).$$

Assume that H_1 is continuous. Then, existence of a solution follows from Brouwer's fixed point theorem (Mas-Colell, Whinston, and Green (1995, Theorem M.I.1)). If $\beta_{10}^{\pi} > 0$, the BNE system can have multiple solutions. Suppose there exist R solutions, ranked in order as $\pi_{1,1}^*(W_1, \beta_{10}) < \pi_{1,2}^*(W_1, \beta_{10}) < \cdots < \pi_{1,R}^*(W_1, \beta_{10})$. If we make no assumptions about the BNE selection mechanism, we have

$$\underbrace{W_1'\beta_{10}^w + \beta_{10}^{\pi}\pi_{1,1}^*(W_1, \beta_{10})}_{\equiv g_{1L}(W_1,\beta_{10})} \leq \underbrace{W_1'\beta_{10}^w + \beta_{10}^{\pi}\pi_1}_{\equiv g_1(X_1,\beta_{10})} \leq \underbrace{W_1'\beta_{10}^w + \beta_{10}^{\pi}\pi_{1,R}^*(W_1, \beta_{10})}_{\equiv g_{1U}(W_1,\beta_{10})}.$$

In this case, we have

$$g_{1L}(W_1, \beta_{10}) \equiv W_1'\beta_{10}^w + \beta_{10}^{\pi}\pi_{1,1}^*(W_1, \beta_{10}), \quad \text{and} \quad g_{1U}(W_1, \beta_{10}) \equiv W_1'\beta_{10}^w + \beta_{10}^{\pi}\pi_{1,R}^*(W_1, \beta_{10}).$$

Once again, given a parametric specification for $H_1(\cdot)$, these bounds will have a parametric functional form for any given β_1.

2.2 Assumptions leading to testable implications

We will let F denote the underlying distribution that generated $(X_1, X_2, W_1, \varepsilon_1, \varepsilon_2)$. Group $V \equiv (X_2, W_1)$. Assume that we observe an iid sample $(Y_{1i}, Y_{2i}, V_i)_{i=1}^n$ generated by F. This corresponds to a censored sample[5] (see Powell (1986, Section 2)), since V_i is observed for censored and uncensored values of Y_{2i}.

Assumption 1

(i) $(\varepsilon_1, \varepsilon_2) \perp (X_1, V)$ *and* $E_F[\varepsilon_2 | \varepsilon_1 > c]$ *is nondecreasing in* c*. Thus, we have*

$$E_F\left[\varepsilon_2 | \, \varepsilon_1 > -g_1(X_1, \beta_{10})\, , X_1, V\right] \equiv \lambda_F\left(g_1(X_1, \beta_{10})\right), \tag{4}$$

where $\lambda_F(\cdot)$ *is a nonincreasing function.*

(ii) $P_F(Y_1 = 1 | W_1, X_1) = H_F\left(g_1(X_1, \beta_0)\right)$*, where* $H_F(\cdot)$ *is unknown but assumed to be nondecreasing.* ■

If we presupposed instead that $E_F[\varepsilon_2 | \varepsilon_1 > c]$ is nonincreasing in c, the function $\lambda_F(\cdot)$ would be nondecreasing and the methodology we will propose would be modified accordingly. The classic *Type 2 Tobit Model* (see Amemiya (1985, p. 385)), where $(\varepsilon_1, \varepsilon_2)$ are jointly normal with covariance $\sigma_{12} \geq 0$, is a special case of Assumption 1.

[5] We will discuss the case of a truncated sample in Section 2.3.2.

2.2.1 Implications from Assumption 1 when X_1 is observed

The conditions of Assumption 1 yield,

$$E_F[Y_2|X_1, X_2, Y_1] = g_2(X_2, \beta_{20}) + \lambda_F(g_1(X_1, \beta_{10})), \tag{5}$$

If X_1 were observable in the selection equation, ours would be a special case of the models studied by James L. Powell and coauthors who, in a number of seminal contributions to the literature, have developed semiparametric methods using pairwise comparisons of observations (see Powell (1987), Ahn and Powell (1993), Honoré and Powell (1994), Powell (2001), Honoré and Powell (2005), Aradillas-López, Honoré, and Powell (2007), and Ahn, Ichimura, Powell, and Ruud (2018)). In our model, the pairwise comparisons proposed by Powell's methods would be based on matching[6] $g(X_{1i}, \beta_{10})$ and $g(X_{1j}, \beta_{10})$. Group $X \equiv (X_1, X_2)$ and suppose we have a random sample $(Y_{1i}, Y_{2i}, X_i)_{i=1}^n$. Take any pair of observations (Y_{1i}, Y_{2i}, X_i) and (Y_{1j}, Y_{2j}, X_j). From (5), we have

$$E_F\left[\left(Y_{2i} - g_2(X_{2i}, \beta_{20})\right) - \left(Y_{2j} - g_2(X_{2j}, \beta_{20})\right) \Big| X_i, X_j, g_1(X_{1i}, \beta_{10}) = g_1(X_{1j}, \beta_{10})\right] = 0 \quad \text{a.s.}$$

Estimation can proceed by estimating β_{10} in a first step, and plugging in the estimator in a second step to estimate β_{20}, based on the above moment condition. The estimator would minimize a kernel-weighted U-statistic, and its asymptotic properties can be derived from the results in Honoré and Powell (2005). As we will show next, when control functions are unobserved under our assumptions, testable implications can still be derived from pairwise comparisons based on whether the control-function bounds are disjoint.

2.2.2 Implications from Assumption 1 when X_1 is not observed

Let us return to our setting, where X_1 is unobserved but bounds of the form (2) are available. The inability to observe X_1 makes pairwise comparisons based on matching the control functions unfeasible. However, under the conditions of Assumption 1, our model produces testable implications if *pairwise comparisons are based on "matching" pairs of observations for which the bounds in* (2) *are disjoint*. The key is that, under our assumptions, the unobserved control function shifts the conditional expectations in our model *monotonically*. We will show this next. Recall that $V \equiv (X_2, W_1)$. Let $\mu_{2F}(V) \equiv E_F[Y_2|V, Y_1 = 1]$. By Assumption 1, we have

$$\mu_{2F}(V) = g_2(X_2, \beta_{20}) + E_F[\lambda_F(g_1(X_1, \beta_{10})|V].$$

[6] Furthermore, if X_1 is observable, we could leave the index function $g_1(\cdot)$ nonparametrically specified and simply base our pairwise comparisons on matching X_1 directly. Aggregating X_1 into a lower-dimensional parametric index $g_1(X_1, \beta_{10})$ can still be desirable as a way to mitigate the curse of dimensionality if X_1 contains multiple continuously distributed covariates. See Honoré and Powell (2005) or Aradillas-López, Honoré, and Powell (2007) for details.

Since $\lambda_F(\cdot)$ is non-increasing, the bounds in (2) imply

$$\lambda_F(g_{1U}(W_1,\beta_{10})) \leq E_F[\lambda_F(g_1(X_1,\beta_{10})|V] \leq \lambda_F(g_{1L}(W_1,\beta_{10})).$$

Let $\mu_{1F}(W_1) \equiv E_F[Y_1|W_1]$. By Assumption 1, we have $\mu_{1F}(W_1) = E_F[H_F(g_1(X_1,\beta_{10}))|W_1]$ and, since $H_F(\cdot)$ is nondecreasing, it follows that,

$$H_F(g_{1L}(W_1,\beta_{10})) \leq \mu_{1F}(W_1) \leq H_F(g_{1U}(W_1,\beta_{10})).$$

Therefore, under Assumption 1,

$$\begin{aligned}&\lambda_F(g_{1U}(W_1,\beta_{10})) \leq \mu_{2F}(V) - g_2(X_2,\beta_{20}) \leq \lambda_F(g_{1L}(W_1,\beta_{10})),\\ &H_F(g_{1L}(W_1,\beta_{10})) \leq \mu_{1F}(W_1) \leq H_F(g_{1U}(W_1,\beta_{10})).\end{aligned} \tag{6}$$

Let (V_i,V_j) be independent draws from F and suppose $g_{1U}(W_{1j},\beta_{10}) \leq g_{1L}(W_{1i},\beta_{10})$. Since $\lambda_F(\cdot)$ is nonincreasing, this implies $\lambda_F(g_{1U}(W_{1j},\beta_{10})) \geq \lambda_F(g_{1L}(W_{1i},\beta_{10}))$ and, from (6), this in turn yields, $\mu_{2F}(V_i) - g_2(X_{2i},\beta_{20}) \leq \mu_{2F}(V_j) - g_2(X_{2j},\beta_{20})$. Therefore, if Assumption 1 holds, we must have,

$$\Big((\mu_{2F}(V_i) - g_2(X_{2i},\beta_{20})) - \Big(\mu_{2F}(V_j) - g_2(X_{2j},\beta_{20})\Big)\Big)\mathbb{1}\Big\{g_{1U}(W_{1j},\beta_{10}) \leq g_{1L}(W_{1i},\beta_{10})\Big\} \leq 0 \quad \text{w.p.1.} \tag{7A}$$

Also from (6), since $H_F(\cdot)$ is nondecreasing, $g_{1U}(W_{1j},\beta_{10}) \leq g_{1L}(W_{1i},\beta_{10})$ implies $\mu_{1F}(W_{1j}) \leq \mu_{1F}(W_{1i})$. Therefore, if Assumption 1 holds, we must also have,

$$\Big(\mu_{1F}(W_{1j}) - \mu_{1F}(W_{1i})\Big)\cdot \mathbb{1}\Big\{g_{1U}(W_{1j},\beta_{10}) \leq g_{1L}(W_{1i},\beta_{10})\Big\} \leq 0 \quad \text{w.p.1.} \tag{7B}$$

2.3 Testable implications for two alternative versions of our model

Our bivariate sample selection model can be modified in various ways. Here we discuss two modifications. A more general description of models suitable to our inferential approach will be included in Section 5.

2.3.1 A bivariate sample selection model with unobserved covariates in the selection and outcome equations

Suppose now that at least a subset of components of X_2 in the outcome equation are also unobserved, and that there exist observable covariates W_2 and parametric functions $g_{2L}(W_2,\beta_{20})$, $g_{2U}(W_2,\beta_{20})$ such that,

$$g_{2L}(W_2,\beta_{20}) \leq g_2(X_2,\beta_{20}) \leq g_{2U}(W_2,\beta_{20}) \quad \text{w.p.1.} \tag{8}$$

As in the examples of Section 2.1, these bounds can be available, e.g, if we have interval-data or they may be obtained from an economic model. Define now $V \equiv (W_1, W_2)$ and suppose we have a random sample $(Y_{1i}, Y_{2i}, V_i)_{i=1}^n$ generated by F. Extending part (i) in Assumption 1, suppose $(\varepsilon_1, \varepsilon_2) \perp (X_1, X_2, V)$. As before, let $\mu_{2F}(V) \equiv E_F\left[Y_2 | V, Y_1 = 1\right]$. We now have,

$$\mu_{2F}(V) = E_F\left[g_2(X_2, \beta_{20}) | V\right] + E_F\left[\lambda_F(g_1(X_1, \beta_{10}) | V\right].$$

Since $\lambda_F(\cdot)$ is nonincreasing and $H_F(\cdot)$ is nondecreasing, the bounds in (2) and (8) yield,

$$\begin{gathered} g_{2L}(W_2, \beta_{20}) + \lambda_F\left(g_{1U}(W_1, \beta_{10})\right) \leq \mu_{2F}(V) \leq g_{2U}(W_2, \beta_{20}) + \lambda_F\left(g_{1L}(W_1, \beta_{10})\right), \\ H_F(g_{1L}(W_1, \beta_{10})) \leq \mu_{1F}(W_1) \leq H_F(g_{1U}(W_1, \beta_{10})). \end{gathered}$$

For a given $\beta \equiv (\beta_1, \beta_2)$, let

$$m_1(V, \beta) \equiv \begin{pmatrix} -g_{2L}(W_2, \beta_2) \\ g_{1U}(W_1, \beta_1) \end{pmatrix} \qquad m_2(V, \beta) \equiv \begin{pmatrix} -g_{2U}(W_2, \beta_2) \\ g_{1L}(W_1, \beta_1) \end{pmatrix}$$

Let (V_i, V_j) be independent draws from F. Since $\lambda_F(\cdot)$ is nonincreasing and $H_F(\cdot)$ is nondecreasing, the model produces the following two functional inequalities,

$$\begin{aligned} \Big(\mu_{2F}(V_i) - \mu_{2F}(V_j)\Big)\, \mathbb{1}\Big\{m_1(V_j, \beta_0) \leq m_2(V_i, \beta_0)\Big\} &\leq 0 \quad \text{w.p.1.} \\ \Big(\mu_{1F}(W_{1j}) - \mu_{1F}(W_{1i})\Big) \cdot \mathbb{1}\Big\{g_{1U}(W_{1j}, \beta_{10}) \leq g_{1L}(W_{1i}, \beta_{10})\Big\} &\leq 0 \quad \text{w.p.1.} \end{aligned} \tag{9}$$

The first functional inequality in (9) replaces (7A) for this version of the model, while the second inequality is the same restriction described in equation (7B). The inequalities in (9) would be the foundation for inference in this case.

2.3.2 A bivariate sample selection model with truncated data

Suppose we have a truncated sample generated by the bivariate sample selection model described in equation (1), with the bounds for the control function $g_1(X_1, \beta_{10})$ given in (2). As before, group $V \equiv (X_2, W_1) \in \mathbb{R}^{L_v}$. Suppose our truncated sample is $(Y_{2i}, V_i)_{i=1}^n$, where $Y_{2i} = Y_{2i}^*$ and $Y_{1i}^* > 0$ for all i. By the truncated nature of our data, the functional inequality in (7B) is no longer useful, since $Y_{1i} = 1$ for all i. However, the inequality in (7A) is still valid and can be used for inference. If we have a truncated sample with unobserved covariates in the selection and outcome equations under the conditions described in the model of Section 2.3.1, the first functional inequality in equation (9) remains valid and, once again, would be the foundation for inference.

3 Inference

In this section we present an inferential method for our main model. Our discussion and analysis will illuminate how to modify our procedure to handle the alternative versions of the model discussed in Section 2.3. Inference for our model will be based on the functional inequalities (7A) and (7B). Thus, we could, in principle, adapt existing inferential methods for conditional moment inequalities (CMIs) to our problem. A partial list of existing methods includes Andrews and Shi (2013), Lee, Song, and Whang (2013), Lee, Song, and Whang (2018), Armstrong (2015), Armstrong (2014), Chetverikov (2017), Armstrong and Chan (2016) and Armstrong (2018). We will opt here to adapt the inferential approach in Aradillas-López and Rosen (2021) to our problem. As we will see, this can be done straightforwardly and it has some theoretical and practical advantages. As we will show, our method will rely on a test-statistic that embeds all the information of the functional inequalities, while adapting asymptotically to the measure of the "contact sets" of values of the conditioning variables at which the inequalities (7A) and/or (7B) are binding. This will allow us to bypass the need to pretest for the slackness of the inequalities, while also avoiding the conservative features of tests that use critical values based on a least favorable configuration in which the inequalities are assume to be binding w.p.1. Furthermore, as our results will show, regularizing the estimator for the asymptotic variance of our statistic will allow us to standardize it in a way that produces asymptotically pivotal properties. Avoiding the need to use resampling methods to compute critical values at each parameter value evaluated can be computationally attractive for practitioners, particularly in richly parameterized models.

3.1 Restricting the parameter space

Our first step is to restrict the parameter to address the fact that, given our testable implications, the original parameters may be impossible to partially identify. Our testable implications revolve around the inequality $g_{1U}(W_{1j}, \beta_{10}) \leq g_{1L}(W_{1i}, \beta_{10})$ and the difference $g_2(X_{2j}, \beta_{20}) - g_2(X_{2i}, \beta_{20})$. Let B denote the parameter space for $\beta \equiv (\beta_1, \beta_2)$. It would be impossible to partially identify β_0 if, for *each* $\beta \in B$, we can find a $\beta' \neq \beta$ in B such that, for all $v \equiv (x_2, w_1)$ and $v' \equiv (x_2', w_1')$,

$$\begin{aligned} &(i) \quad \mathbb{1}\{g_{1U}(w_1', \beta_1) \leq g_{1L}(w_1, \beta_1)\} = \mathbb{1}\{g_{1U}(w_1', \beta_1') \leq g_{1L}(w_1, \beta_1')\} \\ &(ii) \quad g_2(x_2', \beta_2) - g_2(x_2, \beta_2) = g_2(x_2', \beta_2') - g_2(x_2, \beta_2') \end{aligned}$$

If the above is true, at least a subset of parameters in β_0 would be impossible to partially identify for any DGP. We will consider a restriction $\theta(B)$ of the parameter space for which it is possible to *partially* identify $\theta(\beta_0)$ for some DGPs. Depending on the functional forms of g_1 and g_2, the restriction $\theta(B)$ may involve scale and/or location normalizations, or a reduction in the dimension of the parameter space when a subset of parameters in β (e.g, constant terms) drop out from the inequality $g_{1U}(w_1', \beta_1) \leq g_{1L}(w_1, \beta_1)$, or from the difference $g_2(x_2', \beta_2) - g_2(x_2, \beta_2)$. Let $\theta(B)$ be a restriction of the original parameter space such that,

$$
\begin{aligned}
&\exists\ \beta \in B : \forall\ \beta' \neq \beta, \exists\ v \equiv (x_2, w_1),\ v' \equiv (x_2', w_1') : \\
&(i) \quad \mathbb{1}\{g_{1U}(w_1', \theta(\beta_1)) \leq g_{1L}(w_1, \theta(\beta_1))\} \neq \mathbb{1}\{g_{1U}(w_1', \theta(\beta_1')) \leq g_{1L}(w_1, \theta(\beta_1'))\} \\
&(ii) \quad g_2(x_2', \theta(\beta_2)) - g_2(x_2, \theta(\beta_2)) \neq g_2(x_2', \theta(\beta_2')) - g_2(x_2, \theta(\beta_2'))
\end{aligned}
\tag{10}
$$

If (10) holds, it may be possible to partially identify $\theta(\beta_0)$ for DGPs that put positive probability mass over the range of values (v, v') with the features described in (10).

In the interval data example of Section 2.1.1, suppose $g_1(x_1, \beta_1) \equiv \beta_1^c + x_1' \beta_1^x$, with β_1^c being a constant term and β_1^x being the slope coefficients. β_1^c drops out from the inequality $g_{1U}(w_1', \beta_1) \leq g_{1L}(w_1, \beta_1)$, so (10) implies that our restricted parameter space must exclude β_1^c (which is impossible to partially identify), and it *also* requires to normalize the scale of the slope coefficients β_1^x. This can be done in the usual ways, e.g, by fixing one of the elements in β_1^x to ± 1, or by normalizing $\|\beta_1^x\| = 1$. These location and scale restrictions are common in existing econometric models, most notably in rank estimators, whose testable implications involve inequalities between linear indices of parameters (see Han (1987), Sherman (1993), Cavanagh and Sherman (1998)). Bounds for $g_1(x_1, \beta_1)$ that are nonlinear in β_1 would result in different (potentially weaker) restrictions on the parameter space in order to satisfy part (i) of (10). Part (ii) places restrictions on the parameters of the outcome equation. For example, if $g_2(x_2, \beta_2) = \beta_2^c + x_2' \beta_2^x$, our restricted parameter space would exclude the constant term β_2^c, which is impossible to partially identify since it always drops out when we take the difference $g_2(x_2, \beta_2) - g_2(x_2', \beta_2)$. Different (potentially weaker) restrictions on the parameter space would suffice for (10) if $g_2(x_2, \beta_2)$ is nonlinear in β_2.

3.1.1 New notation for the parameters

From now on, we will implicitly consider a restriction $\theta(B)$ of the original parameter space, and we will assume that it satisfies (10). We will denote $\theta_1 \equiv \theta(\beta_1)$, $\theta_2 \equiv \theta(\beta_2)$, and $\theta \equiv (\theta_1, \theta_2)$. We will denote the true parameter values as $\theta_{10} \equiv \theta(\beta_{10})$, $\theta_{20} \equiv \theta(\beta_{20})$, and $\theta_0 \equiv (\theta_{10}, \theta_{20})$. We will denote our (restricted) parameter space as $\Theta \equiv \{\theta : \theta = \theta(\beta), \text{ for some } \beta \in B\}$. We will let k denote the dimension of our restricted parameter θ, so $\Theta \subset \mathbb{R}^k$. We remark that, while our restricted parameter space satisfies (10), we will not assume that the range of values (v, v') that satisfy the conditions described there have a positive probability mass for the underlying data generating process F. If F is such that (10) is violated almost surely, then a subset of parameters may be impossible to partially identify, and our inferential results would reflect that. In this paper we will not impose conditions that lead to identification of θ_0, and we allow for cases where F is such that a subset of parameters is impossible to partially identify. We assume that our restricted parameter space satisfies (10) so that partial identification of θ_0 is possible for *some* DGPs.

3.2 A population statistic for the functional inequalities in (7A) and (7B)

Recall that, in everything that follows, (V_i, V_j) denote two independent draws from F. We will focus on density-weighted versions of the functionals in the inequalities (7A) and (7B). This will be convenient for reasons that will become apparent below. Let $f_V(\cdot)$ denote the density function of V, and let $f_{V,1}(\cdot)$ denote the joint density of (V, Y_1), evaluated at $Y_1 = 1$. That is, $f_{V,1}(V) \equiv P_F(Y_1 = 1|V) \cdot f_V(V)$. Let $f_{W_1}(\cdot)$ denote the density function of W_1. For a given $\theta \in \Theta$, let

$$\begin{aligned}
&\tau_{2F}(V_i, V_j, \theta) \equiv \\
&\Big((\mu_{2F}(V_i) - g_2(X_{2i}, \theta_2)) - \big(\mu_{2F}(V_j) - g_2(X_{2j}, \theta_2)\big)\Big)\mathbb{1}\Big\{g_{1U}(W_{1j}, \theta_1) \le g_{1L}(W_{1i}, \theta_1)\Big\} \cdot f_{V,1}(V_i) f_{V,1}(V_j) \\
&\cdot \phi_2(V_i)^2 \phi_2(V_j)^2, \\
&\tau_{1F}(W_{1i}, W_{1j}, \theta_1) \equiv \\
&\Big(\mu_{1F}(W_{1j}) - \mu_{1F}(W_{1i})\Big) \cdot \mathbb{1}\Big\{g_{1U}(W_{1j}, \theta_1) \le g_{1L}(W_{1i}, \theta_1)\Big\} \cdot f_{W_1}(W_{1i}) f_{W_1}(W_{1j}) \phi_1(W_{1i})^2 \phi_1(W_{1j})^2,
\end{aligned} \tag{11}$$

where $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are weight functions, chosen by the econometrician, which are strictly positive over a pre-specified *inference range* $\mathcal{V} \subseteq \text{Supp}(V)$ and zero everywhere else. That is, $\phi_1(v) > 0$ $\forall\, v \in \mathcal{V}$, $\phi_1(v) = 0$ $\forall\, v \notin \mathcal{V}$, and $\phi_2(w_1) > 0$ $\forall\, w \in \mathcal{V}$, $\phi_2(w_1) = 0$ $\forall\, w \notin \mathcal{V}$. Density-weighting and the use of a testing range $\mathcal{V}$ will be useful in the construction of semiparametric estimators of the functionals described in (11). Let $(A)_+ \equiv A \vee 0$, and denote,

$$\mathcal{T}_{2F}(\theta) \equiv E_F\Big[\big(\tau_{2F}(V_i, V_j, \theta)\big)_+\Big], \quad \mathcal{T}_{1F}(\theta_1) \equiv E_F\Big[\big(\tau_{1F}(W_{1i}, W_{1j}, \theta_1)\big)_+\Big]. \tag{12}$$

By construction, $\mathcal{T}_{2F}(\theta) \ge 0$ $\forall\, \theta$, and $\mathcal{T}_{2F}(\theta) = 0$ if and only if the functional inequality (7A) holds almost surely over our inference range. Similarly, $\mathcal{T}_{1F}(\theta_1) \ge 0$ $\forall\, \theta_1$, and $\mathcal{T}_{1F}(\theta_1) = 0$ if and only if the functional inequality (7B) holds almost surely over our inference range. $\mathcal{T}_{2F}$ and $\mathcal{T}_{1F}$ can be straightforwardly aggregated into a functional that captures whether *both* of the functional inequalities in (7A) and (7B) are satisfied almost surely over our inference range. Consider the functional,

$$\mathcal{T}_F(\theta) \equiv \mathcal{T}_{2F}(\theta) + \mathcal{T}_{1F}(\theta_1). \tag{13}$$

By construction, $\mathcal{T}_F(\theta) \ge 0$ $\forall\, \theta$, and $\mathcal{T}_F(\theta) = 0$ if and only if *both* of the functional inequalities in (7A) and (7B) are satisfied almost surely over our inference range $\mathcal{V}$.

3.3 Constructing an estimator for $\mathcal{T}_F$

Denote,

$$\begin{aligned}
R_{2F}(V) &\equiv \mu_{2F}(V) f_{V,1}(V) \phi_2(V), \quad Q_{2F}(V) \equiv f_{V,1}(V) \phi_2(V), \\
R_{1F}(W_1) &\equiv \mu_{1F}(W_1) f_{W_1}(W_1) \phi_1(W_1), \quad Q_{1F}(W_1) \equiv f_{W_1}(W_1) \phi_1(W_1)
\end{aligned} \tag{14}$$

The functionals $\tau_{2F}(V_i, V_j, \theta)$ and $\tau_{1F}(W_{1i}, W_{1j}, \theta_1)$ defined in (11) can be rewritten as,

$$\begin{aligned}
\tau_{2F}(V_i, V_j, \theta) \equiv & \Big(\big(R_{2F}(V_i)Q_{2F}(V_j) - R_{2F}(V_j)Q_{2F}(V_i)\big) - \big(g_2(X_{2i}, \theta_2) - g_2(X_{2j}, \theta_2)\big) Q_{2F}(V_i)Q_{2F}(V_j)\Big) \\
& \cdot \mathbb{1}\Big\{g_{1U}(W_{1j}, \theta_1) \leq g_{1L}(W_{1i}, \theta_1)\Big\} \phi_2(V_i)\phi_2(V_j), \\
\tau_{1F}(W_{1i}, W_{1j}, \theta_1) \equiv & \Big(R_{1F}(W_{1j})Q_{1F}(W_{1i}) - R_{1F}(W_{1i})Q_{1F}(W_{1j})\Big) \\
& \cdot \mathbb{1}\Big\{g_{1U}(W_{1j}, \theta_1) \leq g_{1L}(W_{1i}, \theta_1)\Big\} \phi_1(W_{1i})\phi_1(W_{1j}).
\end{aligned} \tag{15}$$

We will study the asymptotic properties of kernel-based nonparametric estimators for the functionals in (14). Suppose we can partition $V \equiv (V^c, V^d) \in \mathbb{R}^{L_v}$, where V^c and V^d are continuously distributed and discrete, respectively, and let $r \equiv \dim(V^c)$ (the number of continuously distributed covariates in V). Similarly, partition $W_1 \equiv (W_1^c, W_1^d)$, where W_1^c and W_1^d are continuously distributed and discrete, respectively, and let $\ell \equiv \dim(W_1^c)$ (the number of continuously distributed covariates in W_1). Recall that $W_1 \subseteq V$, and therefore $\ell \leq r$. Let $\kappa(\cdot)$ be a real-valued, univariate kernel function and let h_n be a bandwidth sequence. We will use multiplicative kernels where, a given $v \equiv (v^c, v^d)$ and $w_1 \equiv (w_1^c, w_1^d)$,

$$\mathcal{K}\left(\frac{V_i^c - v^c}{h_n}\right) \equiv \prod_{m=1}^{r} \kappa\left(\frac{V_{mi}^c - v_m^c}{h_n}\right), \quad \mathcal{K}\left(\frac{W_{1i}^c - w_1^c}{h_n}\right) \equiv \prod_{m=1}^{\ell} \kappa\left(\frac{W_{1mi}^c - w_{1m}^c}{h_n}\right),$$

$$\Gamma(V_i, v, h_n) \equiv \mathcal{K}\left(\frac{V_i^c - v^c}{h_n}\right) \cdot \mathbb{1}\Big\{V_i^d = v^d\Big\}, \quad \Gamma(W_{1i}, w_1, h_n) \equiv \mathcal{K}\left(\frac{W_{1i}^c - w_1^c}{h_n}\right) \cdot \mathbb{1}\Big\{W_{1i}^d = w_1^d\Big\}.$$

We will describe precise conditions for the kernel $\kappa(\cdot)$ and the bandwidth h_n in Assumption 4, below. We estimate,

$$\begin{aligned}
&\widehat{R}_2(v) \equiv \frac{1}{n \cdot h_n^r} \sum_{i=1}^{n} Y_{2i} Y_{1i} \phi_2(V_i) \Gamma(V_i, v, h_n), \quad \widehat{Q}_2(v) \equiv \frac{1}{n \cdot h_n^r} \sum_{i=1}^{n} Y_{1i} \phi_2(V_i) \Gamma(V_i, v, h_n), \\
&\widehat{R}_1(w_1) \equiv \frac{1}{n \cdot h_n^\ell} \sum_{i=1}^{n} Y_{1i} \phi_1(W_{1i}) \Gamma(W_{1i}, w_1, h_n), \quad \widehat{Q}_1(w_1) \equiv \frac{1}{n \cdot h_n^\ell} \sum_{i=1}^{n} \phi_1(W_{1i}) \Gamma(W_{1i}, w_1, h_n),
\end{aligned} \tag{16}$$

As usual, density-weighting has the theoretical and practical advantage of allowing for estimators that can be expressed as sample averages, without estimated densities in the denominator[7]. Using

[7] The theoretical advantages of density-weighting were also exploited, e.g, in Powell, Stock, and Stoker (1989).

the expressions in (15), we estimate,

$$\begin{aligned}
\widehat{\tau}_2(V_i, V_j, \theta) \equiv &\Big(\big(\widehat{R}_2(V_i)\widehat{Q}_2(V_j) - \widehat{R}_2(V_j)\widehat{Q}_2(V_i)\big) - \big(g_2(X_{2i}, \theta_2) - g_2(X_{2j}, \theta_2)\big)\widehat{Q}_2(V_i)\widehat{Q}_2(V_j)\Big) \\
&\cdot \mathbb{1}\Big\{g_{1U}(W_{1j}, \theta_1) \leq g_{1L}(W_{1i}, \theta_1)\Big\}\phi_2(V_i)\phi_2(V_j), \\
\widehat{\tau}_1(W_{1i}, W_{1j}, \theta_1) \equiv &\Big(\widehat{R}_1(W_{1j})\widehat{Q}_1(W_{1i}) - \widehat{R}_1(W_{1i})\widehat{Q}_1(W_{1j})\Big) \\
&\cdot \mathbb{1}\Big\{g_{1U}(W_{1j}, \theta_1) \leq g_{1L}(W_{1i}, \theta_1)\Big\}\phi_1(W_{1i})\phi_1(W_{1j}).
\end{aligned}$$

From here, for a given $\theta \in \Theta$, our estimators for $\mathcal{T}_{2F}(\theta)$, $\mathcal{T}_{1F}(\theta_1)$ and $\mathcal{T}_F(\theta)$ are,

$$\begin{aligned}
\widehat{\mathcal{T}}_2(\theta) &\equiv \frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\widehat{\tau}_2(V_i, V_j, \theta)\cdot \mathbb{1}\Big\{\widehat{\tau}_2(V_i, V_j, \theta) \geq -b_n\Big\}, \\
\widehat{\mathcal{T}}_1(\theta_1) &\equiv \frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\widehat{\tau}_1(W_{1i}, W_{1j}, \theta_1)\cdot \mathbb{1}\Big\{\widehat{\tau}_1(W_{1i}, W_{1j}, \theta_1) \geq -b_n\Big\}, \\
\widehat{\mathcal{T}}(\theta) &\equiv \widehat{\mathcal{T}}_2(\theta) + \widehat{\mathcal{T}}_1(\theta_1),
\end{aligned} \tag{17}$$

where $b_n > 0$ is a bandwidth sequence converging to zero, whose properties will be described below, in Assumption 4.

Remark 1 *All the results that follow apply exactly as described if we use a bandwidth $b_n^{\tau_2}$ for $\widehat{\mathcal{T}}_2(\theta)$ and a different bandwidth $b_n^{\tau_1}$ for $\widehat{\mathcal{T}}_1(\theta_1)$, as long as both bandwidths satisfy the convergence rate restrictions for b_n that we will describe in Assumption 4.*

3.4 Asymptotic properties of $\widehat{\mathcal{T}}(\theta)$

Next we will describe a series of assumptions involving smoothness, regularity and manageability restrictions, and restrictions for the convergence of the bandwidths h_n and b_n, as well as the kernel functions involved. Combined, these assumptions will ultimately yield a useful asymptotic linear representation result for $\widehat{\mathcal{T}}(\theta)$, which will be the foundation for the construction of a test-statistic for estimating a confidence set for θ_0.

3.4.1 Space of distributions $\mathcal{F}$

In what follows, we will let $\mathcal{F}$ denote the space of distributions that contains F, the distribution generating the sample observed. Our expanded parameter space is $\Theta \times \mathcal{F} \equiv \{(\theta, F) \colon \theta \in \Theta,\ F \in \mathcal{F}\}$. Our goal will be to describe conditions that will yield uniform results over $\Theta \times \mathcal{F}$. We will use the subscript F to explicitly denote functionals of F, except when it makes the notation too cumbersome. In every case case, which objects are functionals of F will be clear from our discussion and definitions. Let $\{m_n(\theta) \colon \theta \in \Theta\}$ be a stochastic process. Following convention, we will use the following terminology,

(i) We say that $\sup_{\theta\in\Theta}\|m_n(\theta)\| = o_p(n^{\lambda})$ *uniformly over* $\mathcal{F}$ if,

$$\sup_{F\in\mathcal{F}} P_F\left(n^{-\lambda}\sup_{\theta\in\Theta}\|m_n(\theta)\| > c\right) \longrightarrow 0 \quad \forall\ c > 0,$$

and we say that $\sup_{\theta\in\Theta}\|m_n(\theta)\| = O_p(n^{\lambda})$ *uniformly over* $\mathcal{F}$ if, for any $\varepsilon > 0$ there exist a finite $\Delta_\varepsilon > 0$ and $n_\varepsilon \in \mathbb{N}$ such that,

$$\sup_{F\in\mathcal{F}} P_F\left(n^{-\lambda}\sup_{\theta\in\Theta}\|m_n(\theta)\| > \Delta_\varepsilon\right) < \varepsilon \quad \forall\ n \geq n_\varepsilon.$$

(ii) Let $\mathcal{C} \subseteq \Theta\times\mathcal{F}$. We say that $\sup_{(\theta,F)\in\mathcal{C}}\|m_n(\theta)\| = o_p(n^{\lambda})$ if,

$$\sup_{F\in\mathcal{F}} P_F\left(n^{-\lambda}\sup_{\theta:(\theta,F)\in\mathcal{C}}\|m_n(\theta)\| > c\right) \longrightarrow 0 \quad \forall\ c > 0,$$

and we say that $\sup_{(\theta,F)\in\mathcal{C}}\|m_n(\theta)\| = O_p(n^{\lambda})$ if, for any $\varepsilon > 0$ there exist a finite $\Delta_\varepsilon > 0$ and $n_\varepsilon \in \mathbb{N}$ such that,

$$\sup_{F\in\mathcal{F}} P_F\left(n^{-\lambda}\sup_{\theta:(\theta,F)\in\mathcal{C}}\|m_n(\theta)\| > \Delta_\varepsilon\right) < \varepsilon \quad \forall\ n \geq n_\varepsilon. \quad \blacksquare$$

The assumptions we will describe will be technical in nature but we will add a discussion to contextualize them and try to describe what they contribute towards our main result.

3.4.2 Some relevant functionals

Our first set of assumptions involve smoothness conditions for a list of relevant functionals. For a given $v \equiv (x_2, w_1)$, let

$$\begin{aligned}
&\eta^{\tau_2}_{a,F}(v,\theta) \equiv \\
&E_F\left[(R_{2F}(V) - (g_2(X_2,\theta_2) - g_2(x_2,\theta_2))\,Q_{2F}(V))\,\mathbb{1}\{g_{1U}(w_1,\theta_1) \leq g_{1L}(W_1,\theta_1)\}\,\mathbb{1}\{\tau_{2F}(V,v,\theta) \geq 0\}\,\phi_2(V)\right], \\
&\eta^{\tau_2}_{b,F}(v,\theta) \equiv \\
&E_F\left[(R_{2F}(V) - (g_2(X_2,\theta_2) - g_2(x_2,\theta_2))\,Q_{2F}(V))\,\mathbb{1}\{g_{1U}(W_1,\theta_1) \leq g_{1L}(w_1,\theta_1)\}\,\mathbb{1}\{\tau_{2F}(v,V,\theta) \geq 0\}\,\phi_2(V)\right], \\
&\eta^{\tau_2}_{c,F}(v,\theta) \equiv \\
&E_F\left[Q_{2F}(V)\mathbb{1}\{g_{1U}(W_1,\theta_1) \leq g_{1L}(w_1,\theta_1)\}\,\mathbb{1}\{\tau_{2F}(v,V,\theta) \geq 0\}\,\phi_2(V)\right], \\
&\eta^{\tau_2}_{d,F}(v,\theta) \equiv \\
&E_F\left[Q_{2F}(V)\mathbb{1}\{g_{1U}(w_1,\theta_1) \leq g_{1L}(W_1,\theta_1)\}\,\mathbb{1}\{\tau_{2F}(V,v,\theta) \geq 0\}\,\phi_2(V)\right].
\end{aligned} \tag{18}$$

And for a given w_1, let

$$\begin{aligned}
\eta^{\tau_1}_{a,F}(w_1,\theta_1) &\equiv E_F\left[R_{1F}(W_1)\mathbb{1}\{g_{1U}(w_1,\theta_1) \le g_{1L}(W_1,\theta_1)\}\mathbb{1}\{\tau_{1F}(W_1,w_1,\theta) \ge 0\}\phi_1(W_1)\right],\\
\eta^{\tau_1}_{b,F}(w_1,\theta_1) &\equiv E_F\left[R_{1F}(W_1)\mathbb{1}\{g_{1U}(W_1,\theta_1) \le g_{1L}(w_1,\theta_1)\}\mathbb{1}\{\tau_{1F}(w_1,W_1,\theta) \ge 0\}\phi_1(W_1)\right],\\
\eta^{\tau_1}_{c,F}(w_1,\theta_1) &\equiv E_F\left[Q_{1F}(W_1)\mathbb{1}\{g_{1U}(W_1,\theta_1) \le g_{1L}(w_1,\theta_1)\}\mathbb{1}\{\tau_{1F}(w_1,W_1,\theta) \ge 0\}\phi_1(W_1)\right],\\
\eta^{\tau_1}_{d,F}(w_1,\theta_1) &\equiv E_F\left[Q_{1F}(W_1)\mathbb{1}\{g_{1U}(w_1,\theta_1) \le g_{1L}(W_1,\theta_1)\}\mathbb{1}\{\tau_{1F}(W_1,w_1,\theta) \ge 0\}\phi_1(W_1)\right]
\end{aligned} \tag{19}$$

Assumption 2 ***(Smoothness, boundedness properties over the inference range $\mathcal{V}$)*** *The weight functions $\phi_2(v)$ and $\phi_1(w_1)$ are M times differentiable w.r.t v^c and w_1^c, respectively, with bounded derivatives for a.e $v \in \mathcal{V}$ and a.e $w \in \mathcal{V}$. Both weight functions are bounded above by a constant $\overline{\phi}$. The inference range $\mathcal{V}$ and the parameter space Θ are compact, and there exists a finite constant $\overline{D}$ such that,*

$$\sup_{\substack{x_2\in\mathcal{V}\\ \theta_2\in\Theta}} |g_2(x_2,\theta_2)| \le \overline{D}, \quad \textit{and} \quad \sup_{v\in\mathcal{V}} \left(f_{V,1}(v) \vee \left|\mu_{2F}(v)\right|\right) \le \overline{D}, \textit{ for each } F \in \mathcal{F}.$$

Let M be as described above. Uniformly over $F \in \mathcal{F}$, the following conditions are satisfied,

(A) For the densities $f_{V,1}$, f_{W_1} and the conditional expectations μ_{2F} and μ_{1F}, the following holds,

- *(i) $f_{V,1}(v)$ and $\mu_{2F}(v)$ are M times differentiable with respect to v^c, with bounded derivatives for a.e $v \in \mathcal{V}$.*
- *(ii) $f_{W_1}(w_1)$ and $\mu_{1F}(w_1)$ are M times differentiable with respect to w_1^c, with bounded derivatives for a.e $w \in \mathcal{V}$.*

(B) For the functionals defined in equations (18) *and* (19)*, the following holds,*

- *(i) Uniformly over Θ, the functionals $\eta^{\tau_2}_{a,F}(v,\theta)$, $\eta^{\tau_2}_{b,F}(v,\theta)$, $\eta^{\tau_2}_{c,F}(v,\theta)$ and $\eta^{\tau_2}_{d,F}(v,\theta)$ are M times differentiable with respect to v^c, with bounded derivatives for a.e $v \in \mathcal{V}$.*
- *(ii) Uniformly over Θ, the functionals $\eta^{\tau_1}_{a,F}(w_1,\theta_1)$, $\eta^{\tau_1}_{b,F}(w_1,\theta_1)$, $\eta^{\tau_1}_{c,F}(w_1,\theta_1)$ and $\eta^{\tau_2}_{d,F}(w_1,\theta_1)$ are M times differentiable with respect to w_1^c, with bounded derivatives for a.e $w \in \mathcal{V}$.* ■

Combined with bias-reduction conditions for our bandwidth h_n and our kernel function, the conditions in Assumption 2 will help ensure that the asymptotic bias of our estimators converges to zero at the appropriate rate, uniformly over $\Theta \times \mathcal{F}$. Next we describe a first set of manageability conditions which will involve some relevant empirical processes and U-processes in our problem. We focus on manageability results that hold for so-called *Euclidean* classes of functions. To have a self-contained discussion, we present the definition of Euclidean classes first.

Definition: Euclidean classes of functions

What follows is taken from Nolan and Pollard (1987, Definition 8), Pakes and Pollard (1989, Definition 2.7), and Sherman (1994, Definition 3). Let $\mathcal{T}$ be a space and d be a pseudometric defined on $\mathcal{T}$. For each $\varepsilon > 0$, define the *packing number* $D(\varepsilon, d, \mathcal{T})$ to be the largest number D for which there exist points $m_1, \ldots, m_D$ in $\mathcal{T}$ such that $d(m_i, m_j) > \varepsilon$ for each $i \neq j$. Packing numbers are a measure of how big $\mathcal{T}$ is with respect to d. Let $\mathscr{G}$ be a class of functions defined on a set $\mathcal{S}_Z^k$. We say that G is an *envelope* for $\mathscr{G}$ is $\sup_{\mathscr{G}}|g(\cdot)| \leq G(\cdot)$. Let μ be a measure on $\mathcal{S}_Z^k$ and denote $\mu h \equiv \int h(z_1, \ldots, z_k) d\mu(z_1, \ldots, z_k)$. We say that the class of functions $\mathscr{G}$ is Euclidean (A, V) for the envelope G if, for any measure μ such that $\mu G^2 < \infty$, we have $D(\varepsilon, d_\mu, \mathscr{G}) \leq A\varepsilon^{-V}$ $\forall$ $0 < \varepsilon \leq 1$, where, for $g_1, g_2 \in \mathscr{G}$, $d_\mu(g_1, g_2) = \left(\mu|g_1 - g_2|^2/\mu G^2\right)^{1/2}$. The constants A and V must not depend on μ. ■

The name "Euclidean" is owed to the fact that $A\varepsilon^{-V}$ is the generic expression of packing numbers for any bounded subset of the *Euclidean* space $\mathbb{R}^V$. Examples of Euclidean classes of functions can be found, in Pollard (1984), Nolan and Pollard (1987), Pakes and Pollard (1989), Pollard (1990), Sherman (1994) and Andrews (1994). Notable examples found in many econometric models include the following.

(i) (Pakes and Pollard (1989, Lemma 2.13)) Let $\mathscr{G} = \{g(\cdot, t) : t \in T\}$ be a class of functions on $\mathscr{X}$ indexed by a bounded subset T of $\mathbb{R}^d$. If there exists an $\alpha > 0$ and a $\phi(\cdot) \geq 0$ such that $|g(x,t) - g(x,t')| \leq \phi(x) \cdot \|t - t'\|^\alpha$ for $x \in \mathscr{X}$ and $t, t' \in T$. Then $\mathscr{G}$ is Euclidean for the envelope $G \equiv |g(\cdot, t_0)| + M\phi(\cdot)$, where $t_0 \in T$ is an arbitrary point and $M \equiv (2\sqrt{d}\sup_T \|t - t_0\|)^\alpha$.

(ii) (Nolan and Pollard (1987, Lemma 22), Pakes and Pollard (1989, Example 10)) Let $\lambda(\cdot)$ be a real-valued function of bounded variation on $\mathbb{R}$. The class $\mathscr{G}$ of all functions on $\mathbb{R}^d$ of the form $x \rightarrow \lambda(\alpha' x + \beta)$, with α ranging over $\mathbb{R}^d$ and β ranging over $\mathbb{R}$ is Euclidean for the constant envelope $G \equiv \sup|\lambda|$.

(iii) (Pakes and Pollard (1989, p. 1033)) Classes of indicator functions over VC classes of sets are Euclidean for the constant envelope 1.

(iv) Type I, II and III classes of functions described in Andrews (1994) are special cases of Euclidean classes.

Pointwise algebraic operations such as products, linear combinations, minima and maxima allow us to combine Euclidean classes and preserve the Euclidean property (see Pakes and Pollard (1989, Lemma 2.14)). Empirical processes and U-processes produced by Euclidean classes of functions satisfy the *Pollard's entropy condition* (see Andrews (1994, Definition 4.2)) and *manageability* (see Pollard (1990, Definition 7.9), Andrews (1994, Equation A.1)). From here, maximal inequality results follow (Sherman (1994)), which will be useful in obtaining our main asymptotic results. The following assumption describes our first set of conditions involving Euclidean classes.

Assumption 3 ***(Manageability, integrability)*** *There exists a finite constant $\overline{D}_4$ such that $E_F[|Y_2|^4] \leq \overline{D}_4$ for all $F \in \mathcal{F}$. There exist constants (A_2, V_2) and an envelope G_2 such that the class of functions*

$$\mathscr{G}_2 \equiv \{m(x_2) = g_2(x_2, \theta_2) \text{ for some } \theta_2 \in \Theta\}$$

is Euclidean (A_2, V_2) for the envelope G_2, and there exists a constant $\overline{C}_4 < \infty$ such that $E_F\left[G_2(X_2)^4\right] \leq \overline{C}_4$ for all $F \in \mathcal{F}$. Take the functionals $\eta^{\tau_2}_{\ell,F}$ defined in (18) for $\ell \in \{a,b,c,d\}$. There exists a finite constant $\overline{G}$ such that, for each $\ell \in \{a,b,c,d\}$ and for any $\theta, \theta' \in \Theta$,

$$\sup_{v \in \mathcal{V}} \left|\eta^{\tau_2}_{\ell,F}(v,\theta) - \eta^{\tau_2}_{\ell,F}(v,\theta')\right| \leq \overline{G} \cdot \left\|\theta - \theta'\right\| \quad \forall\, F \in \mathcal{F}.$$

(i) The class of sets

$$\mathscr{C} \equiv \left\{(w_1, w_1) \in \mathbb{R}^{d_U} \times \mathbb{R}^{d_L} : g_{1U}(w_1, \theta_1) \leq g_{1L}(w_1, \theta_1) \text{ for some } \theta_1 \in \Theta\right\}$$

is a VC class with VC dimension $\overline{V}_C$.

(ii) There exists a $c_0 > 0$ and a finite constant $\overline{V}_D$ such that, for each $F \in \mathcal{F}$, the classes of sets

$$\mathscr{S}^{\tau_2}_F \equiv \left\{(v_1, v_2) \in \mathbb{R}^{L_v} \times \mathbb{R}^{L_v} : \tau_{2F}(v_1, v_2, \theta) \geq c \text{ for some } c \in [-c_0, 0] \text{ and } \theta \in \Theta\right\},$$
$$\mathscr{S}^{\tau_1}_F \equiv \left\{(w_1, w_2) \in \mathbb{R}^{L_w} \times \mathbb{R}^{L_w} : \tau_{1F}(w_1, w_2, \theta) \geq c \text{ for some } c \in [-c_0, 0] \text{ and } \theta \in \Theta\right\}$$

are VC classes of sets with VC dimension bounded above by $\overline{V}_D$. ■

VC classes of sets are defined, e.g, in Pakes and Pollard (1989, Definition 2.2) and Kosorok (2008, Section 9.1.1). Verifiable criteria that suffice for a class of sets to have the VC property can be found, e.g, in Pollard (1984, Section II.4), Dudley (1984, Section 9), or Kosorok (2008, Section 9.1.1). An example commonly encountered in econometric models (Pakes and Pollard (1989, Lemma 2.4) is the class of sets of the form $\{g \geq t\}$ or $\{g > t\}$, with $g \in \mathscr{G}$ and $t \in \mathbb{R}$, where $\mathscr{G}$ is a finite dimensional vector space of real-valued functions. This class encompasses econometric models where the parameters of interest enter through linear indices. Combining VC classes of sets through a finite number of Boolean operations (e.g, unions, intersections and/or complements) preserves the VC property (Pakes and Pollard (1989, Lemma 2.5)). Assumption 3 implies that the following is a VC class of sets for each F, with VC dimension uniformly bounded over $\mathcal{F}$ by a finite constant $\overline{V}_D$,

$$\mathscr{D}^{\tau_2}_{1,F} \equiv \left\{(v_1, v_2) \in \mathbb{R}^{L_v} \times \mathbb{R}^{L_v} : \tau_{2F}(v_1, v_2, \theta) \geq 0 \text{ for some } \theta \in \Theta\right\}$$

And, by VC-preserving properties of Boolean operations described, e.g, in Pakes and Pollard (1989, Lemma 2.5), Assumption 3 implies that, for each $F \in \mathcal{F}$, the following class of sets is also a

VC class, with VC dimension uniformly bounded over $\mathcal{F}$ by a finite constant,

$$\mathscr{D}_{2,F}^{\tau_2} \equiv \Big\{(v_1,v_2) \in \mathbb{R}^{L_v} \times \mathbb{R}^{L_v} : -c \leq \tau_{2F}(v_1,v_2,\theta) < 0 \text{ for some } 0 < c \leq c_0 \text{ and } \theta \in \Theta\Big\}.$$

Indicator functions for these classes of sets are relevant in our problem. The VC properties in Assumption 3 will lead us to invoke maximal inequality results since indicator functions over VC classes of sets are Euclidean classes of functions (Pakes and Pollard (1989, p. 1033)). The next assumption describes our restrictions for tuning parameters (bandwidths and kernels).

Assumption 4 ***(Kernels and bandwidths)*** *Let M be the integer described in Assumption 2.*

(i) We use a multiplicative kernel K such that, for any $\psi \equiv (\psi_1,\ldots,\psi_D)'$, we have $K(\psi) = \prod_{d=1}^{D}\kappa(\psi_d)$, where $\kappa(\cdot)$ is a bias-reducing kernel of order M with support of the form $[-S,S]$ (with $\kappa(S) = \kappa(-S) = 0$, $\kappa(v) = 0\ \forall\ v \notin (-S,S)$, with $\int_{-S}^{S}\kappa(v)dv = 1$, $\int_{-S}^{S}v^j\kappa(v)dv = 0$ for $j = 1,\ldots,M-1$ and $\int_{-S}^{S}|v|^M\kappa(v)dv < \infty$) and symmetric around zero (i.e, $\kappa(v) = \kappa(-v)$ for all v). $\kappa(\cdot)$ is a function of bounded variation, satisfying $|\kappa(\cdot)| \leq \overline{\kappa}$ for a constant $\overline{\kappa} < \infty$.

(ii) The bandwidth sequences $h_n > 0$ and $b_n > 0$ are such that there exists $0 < \epsilon < 1/2$ such that $n^{1/2-\epsilon}\cdot h_n^{2r} \longrightarrow \infty$ and $n^{1/2-\epsilon}\cdot h_n^{r}\cdot b_n \longrightarrow \infty$, while $n^{1/2+\epsilon}\cdot b_n^2 \longrightarrow 0$, and $n^{1/2+\epsilon}\cdot h_n^{M} \longrightarrow 0$. ■

Since $r \equiv \dim(V^c)$, $\ell \equiv \dim(W_1^c)$, Assumption 4 ensures that $n^{1/2-\epsilon}\cdot h_n^{2\ell} \longrightarrow \infty$ and $n^{1/2-\epsilon}\cdot h_n^{\ell}\cdot b_n \longrightarrow \infty$. Section 3.7 includes a discussion about bandwidth and kernel selection. Next we present a regularity condition related to the stochastic properties of our functionals in a neighborhood of the form $[-b_0, 0)$.

Assumption 5 ***((Behavior of $\boldsymbol{\tau_{2F}(V_i,V_j,\theta)}$ and $\boldsymbol{\tau_{1F}(W_{1i},W_{1j},\theta_1)}$ at zero from below))*** *Let V_i, V_j be independent draws from F. There exist $b_0 > 0$ and $\overline{m} < \infty$ such that, $\forall\ b \in [-b_0, 0)$ and $\forall\ F \in \mathcal{F}$,*

$$(i) \qquad \sup_{\theta\in\Theta} E_F\Big[\mathbb{1}\Big\{-b \leq \tau_{2F}(V_i,V_j,\theta) < 0\Big\}\Big] \leq \overline{m}\cdot b \quad \forall\ 0 < b \leq b_0,$$

$$\sup_{\theta_1\in\Theta} E_F\Big[\mathbb{1}\Big\{-b \leq \tau_{1F}(W_{1i},W_{1j},\theta_1) < 0\Big\}\Big] \leq \overline{m}\cdot b \quad \forall\ 0 < b \leq b_0,$$

$$(ii) \qquad \sup_{\substack{\theta\in\Theta\\ v\in\mathcal{V}}} E_F\left[\mathbb{1}\{-b \leq \tau_{2F}(V_i,v,\theta) < 0\}\right] \leq \overline{m}\cdot b \quad \forall\ 0 < b \leq b_0,$$

$$\sup_{\substack{\theta\in\Theta\\ v\in\mathcal{V}}} E_F\left[\mathbb{1}\{-b \leq \tau_{2F}(v,V_i,\theta) < 0\}\right] \leq \overline{m}\cdot b \quad \forall\ 0 < b \leq b_0,$$

$$(iii) \qquad \sup_{\substack{\theta_1\in\Theta\\ w_1\in\mathcal{V}}} E_F\left[\mathbb{1}\{-b \leq \tau_{1F}(W_{1i},w_1,\theta_1) < 0\}\right] \leq \overline{m}\cdot b \quad \forall\ 0 < b \leq b_0,$$

$$\sup_{\substack{\theta_1\in\Theta\\ w_1\in\mathcal{V}}} E_F\left[\mathbb{1}\{-b \leq \tau_{1F}(w_1,W_{1i},\theta_1) < 0\}\right] \leq \overline{m}\cdot b \quad \forall\ 0 < b \leq b_0. \quad \blacksquare$$

Assumption 5 is a mild requirement, which amounts to assuming that our functionals have a finite density in a neighborhood of the form $[-b_0, 0)$. Let us illustrate this. For a given θ, let $F_{\tau_2}(\cdot|\theta)$ and $F_{\tau_1}(\cdot|\theta_1)$ denote the distribution functions of $\tau_{2F}(V_i, V_j, \theta)$ and $\tau_{1F}(W_{1i}, W_{1j}, \theta_1)$, respectively, and let $f_{\tau_2}(\cdot|\theta)$ and $f_{\tau_1}(\cdot|\theta_1)$ denote the corresponding density functions. Part (i) in Assumption 5 presupposes that $F_{\tau_2}(\cdot|\theta)$ and $F_{\tau_1}(\cdot|\theta_1)$ are continuous in the interval $[-b_0, 0)$, and that $f_{\tau_2}(\cdot|\theta)$ and $f_{\tau_1}(\cdot|\theta_1)$ are bounded above by $\overline{m}$ over $[-b_0, 0)$, uniformly over $\Theta \times \mathcal{F}$. That is,

$$\sup_{\substack{\theta\in\Theta \\ b\in[-b_0,0)}} f_{\tau_2}(b|\theta) \leq \overline{m}, \quad \text{and} \quad \sup_{\substack{\theta_1\in\Theta \\ b\in[-b_0,0)}} f_{\tau_1}(b|\theta_1) \leq \overline{m}, \quad \forall\, F \in \mathcal{F}.$$

A mean-value argument yields the condition in part (i) of Assumption 5. Parts (ii) and (iii) of Assumption 5 impose analogous restrictions on the densities of the functionals described there. Note that the conditions in Assumption 5 *allow for each one of the functionals described there to have a point-mass at zero*, since these conditions focus on an interval of the form $[-b_0, 0)$, which excludes zero. A point mass at zero occurs when the inequalities are binding with positive probability.

3.4.3 An asymptotic linear representation result for $\widehat{\mathcal{T}}(\theta)$

Equipped with the previous set of assumptions we can present the main result in this section.

Theorem 1 *Group all the observable covariates in the model as $Z \equiv (Y_1, Y_2, V)$. In the results that follow, let $\epsilon > 0$ be the constant described in Assumption 4.*

(A) Let (V_i, V_j) be two independent draws from F and let

$$H_{1F}^{\mathcal{T}_2}(V_i,\theta) \equiv \frac{1}{2}\cdot\left(E_F\left[\left(\tau_{2F}(V_i,V_j,\theta)\right)_+\middle|V_i\right] + E_F\left[\left(\tau_{2F}(V_i,V_j,\theta)\right)_+\middle|V_i\right]\right) - \mathcal{T}_{2F}(\theta)$$

Note that $E_F[H_{1F}^{\mathcal{T}_2}(V_i,\theta)] = 0\ \forall\ \theta$. Next, take the functionals defined in (18) *and let,*

$$\begin{aligned}H_{2F}^{\mathcal{T}_2}(Z_i,\theta) \equiv &\left(\left(\eta_{a,F}^{\tau_2}(V_i,\theta) - \eta_{b,F}^{\tau_2}(V_i,\theta)\right)\cdot Y_{1i} + \left(\eta_{c,F}^{\tau_2}(V_i,\theta) - \eta_{d,F}^{\tau_2}(V_i,\theta)\right)\cdot Y_{2i}Y_{1i}\right)\cdot f_V(V_i)\cdot\phi_2(V_i)^2 \\ &- E_F\left[\left(\left(\eta_{a,F}^{\tau_2}(V_i,\theta) - \eta_{b,F}^{\tau_2}(V_i,\theta)\right)\cdot Y_{1i} + \left(\eta_{c,F}^{\tau_2}(V_i,\theta) - \eta_{d,F}^{\tau_2}(V_i,\theta)\right)\cdot Y_{2i}Y_{1i}\right)\cdot f_V(V_i)\cdot\phi_2(V_i)^2\right].\end{aligned}$$

Note that $E_F[H_{2F}^{\mathcal{T}_2}(Y_{2i},V_i,\theta)] = 0\ \forall\ \theta$. Now let $\psi_F^{\mathcal{T}_2}(Z_i,\theta) \equiv 2\cdot H_{1F}^{\mathcal{T}_2}(V_i,\theta) + H_{2F}^{\mathcal{T}_2}(Z_i,\theta)$, and note that $E_F[\psi_F^{\mathcal{T}_2}(Z_i,\theta)] = 0\ \forall\ \theta$. If Assumptions 1-5 hold,

$$\begin{aligned}&\widehat{\mathcal{T}}_2(\theta) = \mathcal{T}_{2F}(\theta) + \frac{1}{n}\sum_{i=1}^{n}\psi_F^{\mathcal{T}_2}(Z_i,\theta) + \xi_n^{\mathcal{T}_2}(\theta), \quad \textit{where} \\ &\sup_{\theta\in\Theta}\left|\xi_n^{\mathcal{T}_2}(\theta)\right| = o_p\left(\frac{1}{n^{1/2+\epsilon}}\right), \quad \textit{uniformly over } \mathcal{F},\end{aligned}$$

(B) Suppose (W_{1i}, W_{1j}) are two independent draws from F and let

$$H^{\mathcal{T}_1}_{1F}(W_{1i},\theta_1) \equiv \frac{1}{2}\cdot\Big(E_F\Big[\big(\tau_{1F}(W_{1i},W_{1j},\theta_1)\big)_+\Big|W_{1i}\Big] + E_F\Big[\big(\tau_{1F}(W_{1j},W_{1i},\theta_1)\big)_+\Big|W_{1i}\Big]\Big) - \mathcal{T}_{1F}(\theta_1)$$

Note that $E_F[H^{\mathcal{T}_1}_{1F}(W_{1i},\theta_1)] = 0\ \forall\ \theta_1$. Next, take the functionals defined in (19) and let,

$$\begin{aligned}H^{\mathcal{T}_1}_{2F}(Z_i,\theta_1) \equiv &\Big(\big(\eta^{\tau_1}_{a,F}(W_{1i},\theta_1) - \eta^{\tau_1}_{b,F}(W_{1i},\theta_1)\big) + \big(\eta^{\tau_1}_{c,F}(W_{1i},\theta_1) - \eta^{\tau_1}_{d,F}(W_{1i},\theta_1)\big)\cdot Y_{1i}\Big)\cdot f_{W_1}(W_{1i})\cdot\phi_1(W_{1i})^2\\ &- E_F\Big[\Big(\big(\eta^{\tau_1}_{a,F}(W_{1i},\theta_1) - \eta^{\tau_1}_{b,F}(W_{1i},\theta_1)\big) + \big(\eta^{\tau_1}_{c,F}(W_{1i},\theta_1) - \eta^{\tau_1}_{d,F}(W_{1i},\theta_1)\big)\cdot Y_{1i}\Big)\cdot f_{W_1}(W_{1i})\cdot\phi_1(W_{1i})^2\Big].\end{aligned}$$

Note that $E_F[H^{\mathcal{T}_1}_{2F}(Z_i,\theta_1)] = 0\ \forall\ \theta_1$. Now let, $\psi^{\mathcal{T}_1}_F(Z_i,\theta_1) \equiv 2\cdot H^{\mathcal{T}_1}_{1F}(W_{1i},\theta_1) + H^{\mathcal{T}_1}_{2F}(Z_i,\theta_1)$, and note that $E_F[\psi^{\mathcal{T}_1}_F(Z_i,\theta_1)] = 0\ \forall\ \theta_1$. If Assumptions 1-5 hold,

$$\begin{aligned}&\widehat{\mathcal{T}}_1(\theta_1) = \mathcal{T}_{1F}(\theta_1) + \frac{1}{n}\sum_{i=1}^n \psi^{\mathcal{T}_1}_F(Z_i,\theta_1) + \xi^{\mathcal{T}_1}_n(\theta_1), \quad \textit{where}\\ &\sup_{\theta_1\in\Theta}\left|\xi^{\mathcal{T}_1}_n(\theta_1)\right| = o_p\left(\frac{1}{n^{1/2+\epsilon}}\right), \quad \textit{uniformly over } \mathcal{F}.\end{aligned}$$

(C) Let $\psi^{\mathcal{T}}_F(Z_i,\theta) \equiv \psi^{\mathcal{T}_2}_F(Z_i,\theta) + \psi^{\mathcal{T}_1}_F(Z_i,\theta_1)$). From (A) and (B), we have $E_F[\psi^{\mathcal{T}}_F(Z_i,\theta)] = 0\ \forall\ \theta$ and, if Assumptions 1-5 hold,

$$\begin{aligned}&\widehat{\mathcal{T}}(\theta) = \mathcal{T}_F(\theta) + \frac{1}{n}\sum_{i=1}^n \psi^{\mathcal{T}}_F(Z_i,\theta) + \xi^{\mathcal{T}}_n(\theta), \quad \textit{where}\\ &\sup_{\theta\in\Theta}\left|\xi^{\mathcal{T}}_n(\theta)\right| = o_p\left(\frac{1}{n^{1/2+\epsilon}}\right), \quad \textit{uniformly over } \mathcal{F}. \quad \blacksquare\end{aligned}$$

The proof of Theorem 1 is included in Section A1 of the online appendix[8]. We will summarize the main steps next, focusing on part (A) of the theorem since part (B) uses analogous steps and part (C) follows immediately from (A) and (B). Let

$$\widetilde{\mathcal{T}}_2(\theta) \equiv \frac{1}{n\cdot(n-1)}\sum_{i=1}^n\sum_{j\neq i}\widehat{\tau}_2(V_i,V_j,\theta)\cdot\mathbb{1}\Big\{\tau_{2F}(V_i,V_j,\theta)\geq 0\Big\}.$$

$\widetilde{\mathcal{T}}_2(\theta)$ takes $\widehat{\mathcal{T}}_2(\theta)$ and replaces the indicator function $\mathbb{1}\Big\{\widehat{\tau}_2(V_i,V_j,\theta)\geq -b_n\Big\}$ with $\mathbb{1}\Big\{\tau_{2F}(V_i,V_j,\theta)\geq 0\Big\}$. Let $r^{\mathcal{T}_2}_n(\theta) \equiv \widehat{\mathcal{T}}_2(\theta) - \widetilde{\mathcal{T}}_2(\theta)$. The first series of steps of the proof lead to the result,

$$\sup_{\theta\in\Theta}\left|r^{\mathcal{T}_2}_n(\theta)\right| = o_p\left(\frac{1}{n^{1/2+\epsilon}}\right), \quad \text{uniformly over } \mathcal{F}.$$

[8] Available online at `http://www.personal.psu.edu/aza12/part-identif-control-functions-appendix.pdf`

where $\epsilon > 0$ is the constant described in Assumption 4. From here, re-expressing,

$$
\begin{aligned}
\widetilde{\mathcal{T}}_2(\theta) &= \frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\Big(\tau_{2F}(V_i,V_j,\theta)\Big)_+ \\
&+ \frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\Big(\widehat{\tau}_2(V_i,V_j,\theta)-\tau_{2F}(V_i,V_j,\theta)\Big)\cdot\mathbb{1}\Big\{\tau_{2F}(V_i,V_j,\theta)\geq 0\Big\} \\
&= \mathcal{T}_{2F}(\theta) + \frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\Big(\Big(\tau_{2F}(V_i,V_j,\theta)\Big)_+ - \mathcal{T}_{2F}(\theta)\Big) \\
&+ \frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\Big(\widehat{\tau}_2(V_i,V_j,\theta)-\tau_{2F}(V_i,V_j,\theta)\Big)\cdot\mathbb{1}\Big\{\tau_{2F}(V_i,V_j,\theta)\geq 0\Big\},
\end{aligned}
$$

the next series of steps is to show that the above expression becomes,

$$
\begin{aligned}
&\widetilde{\mathcal{T}}_2(\theta) = \mathcal{T}_{2F}(\theta) + \frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\Big(\Big(\tau_{2F}(V_i,V_j,\theta)\Big)_+ - \mathcal{T}_{2F}(\theta)\Big) + \frac{(n-2)}{n}\cdot\frac{1}{h_n^{r_v}}\cdot U_{a,n}(\theta,h_n) + \xi_{b,n}^{\widetilde{\mathcal{T}}_2}(\theta), \\
&\text{where } \sup_{\theta\in\Theta}\left|\xi_{b,n}^{\widetilde{\mathcal{T}}_2}(\theta)\right| = o_p\left(\frac{1}{n^{1/2+\epsilon}}\right), \text{ uniformly over } \mathcal{F}.
\end{aligned}
$$

Where $\epsilon > 0$ is the constant described in Assumption 4, and $\{U_{a,n}(\theta,h)\colon \theta\in\Theta, h>0\}$ is a U-process of order 2. The final step to obtain the result in part (A) of the theorem is to compute the Hoeffding decomposition (see Serfling (1980, pages 177-178) or Sherman (1994, equations (6)-(7))) of $U_{a,n}(\theta,h_n)$ and apply the maximal inequality results in Sherman (1994, Corollary 4A). All the step-by-step details are included in Section A1 of the online appendix.

3.5 A statistic based on Theorem 1

We will rely on the linear representation result in Theorem 1 to build a statistic that will be used to estimate a CS for θ_0. The following subsets of $\Theta\times\mathcal{F}$ will be relevant for our analysis. First, let,

$$
\begin{aligned}
&\Lambda_{\Theta,\mathcal{F}} \equiv \Big\{(\theta,F)\in\Theta\times\mathcal{F}: \\
&P_F\Big(\Big(\big(\mu_{2F}(V_i)-g_2(X_{2i},\theta_2)\big)-\big(\mu_{2F}(V_j)-g_2(X_{2j},\theta_2)\big)\Big)\mathbb{1}\{g_{1U}(W_{1j},\theta_1)\leq g_{1L}(W_{1i},\theta_1)\}\leq 0 \;\Big|\; V_i,V_j\in\mathcal{V}\Big)=1, \\
&P_F\Big(\big(\mu_{1F}(W_{1j})-\mu_{1F}(W_{1i})\big)\mathbb{1}\{g_{1U}(W_{1j},\theta_1)\leq g_{1L}(W_{1i},\theta_1)\}\leq 0 \;\Big|\; W_{1i},W_{1j}\in\mathcal{V}\Big)=1\Big\}
\end{aligned}
$$

$\Lambda_{\Theta,\mathcal{F}}$ is the collection of all (θ,F) that satisfy the functional inequalities (7A) and (7B) almost surely over our inference range. Before proceeding, let us formalize the notion of *contact sets*.

Contact sets.- For a given θ, the contact sets are defined as the collection of all values of Z for which at least one of the functional inequalities (either (7A), or (7B)) are binding. From now on, we will denote $\sigma_F^2(\theta) \equiv E_F[\psi_F^{\mathcal{T}}(Z,\theta)^2]$. Note from (20) that $\sigma_F^2(\theta) = 0$ if the contact sets for θ have measure zero with respect to F. As our results will highlight, $\sigma_F^2(\theta)$ *will be the relevant measure of the contact sets in our methodology.* ∎

Consider the following subset of $\Lambda_{\Theta,\mathcal{F}}$,

$$\overline{\Lambda}_{\Theta,\mathcal{F}} \equiv \Big\{(\theta,F) \in \Theta \times \mathcal{F} :$$
$$P_F\Big(\Big((\mu_{2F}(V_i) - g_2(X_{2i},\theta_2)) - (\mu_{2F}(V_j) - g_2(X_{2j},\theta_2))\Big)\mathbb{1}\{g_{1U}(W_{1j},\theta_1) \le g_{1L}(W_{1i},\theta_1)\} < 0 \;\Big|\; V_i, V_j \in \mathcal{V}\Big) = 1,$$
$$P_F\Big((\mu_{1F}(W_{1j}) - \mu_{1F}(W_{1i}))\mathbb{1}\{g_{1U}(W_{1j},\theta_1) \le g_{1L}(W_{1i},\theta_1)\} < 0 \;\Big|\; W_{1i}, W_{1j} \in \mathcal{V}\Big) = 1\Big\}$$

$\overline{\Lambda}_{\Theta,\mathcal{F}}$ is the collection of all (θ,F) that satisfy the functional inequalities (7A) and (7B) *strictly*, almost surely over our inference range. Thus, the contact sets have $F-$measure zero for all $(\theta,F) \in \overline{\Lambda}_{\Theta,\mathcal{F}}$, while $\Lambda_{\Theta,\mathcal{F}} \setminus \overline{\Lambda}_{\Theta,\mathcal{F}}$ contains all (θ,F) that (i) satisfy the functional inequalities, and (ii) have *contact sets* with nonzero $F-$measure. By definition, if V_i, V_j are independent draws from F, we have $\mathbb{1}\{\tau_{2F}(V_i,V_j,\theta) \ge 0\} = 0$ and $\mathbb{1}\{\tau_{2F}(V_i,V_j,\theta) \ge 0\} = 0$ $F-$a.s for each $(\theta,F) \in \overline{\Lambda}_{\Theta,\mathcal{F}}$. From here, inspecting the structure of the influence function $\psi_F^{\mathcal{T}}(Z_i,\theta)$ in Theorem 1, we can see that,

$$\begin{aligned} &(i) \quad E_F[\psi_F^{\mathcal{T}}(Z_i,\theta)] = 0 \;\forall\; (\theta,F) \in \Theta \times \mathcal{F}. \\ &(ii) \quad \psi_F^{\mathcal{T}}(Z_i,\theta) = 0 \;\; F\text{–a.s} \;\;\forall\; (\theta,F) \in \overline{\Lambda}_{\Theta,\mathcal{F}}. \end{aligned} \tag{20}$$

Note that $\sigma_F^2(\theta) = 0$ for all $(\theta,F) \in \overline{\Lambda}_{\Theta,\mathcal{F}}$, which illustrates why $\sigma_F^2(\theta)$ is the relevant measure of the contact sets in our approach. Combined with the linear representations in Theorem 1, the result in (20) implies,

$$\sup_{(\theta,F)\in\overline{\Lambda}_{\Theta,\mathcal{F}}} \left|\widehat{\mathcal{T}}(\theta)\right| = \sup_{(\theta,F)\in\overline{\Lambda}_{\Theta,\mathcal{F}}} \left|\xi_n^{\mathcal{T}}(\theta)\right| = o_p\left(\frac{1}{n^{1/2+\epsilon}}\right) \quad \Longrightarrow \quad \sup_{(\theta,F)\in\overline{\Lambda}_{\Theta,\mathcal{F}}} \left|n^{1/2}\cdot\widehat{\mathcal{T}}(\theta)\right| = o_p(1), \tag{21}$$

where $\epsilon > 0$ is the constant described in Assumption 4. This result will be useful in the construction of a test-statistic. Next, let us focus on $(\Theta \times \mathcal{F}) \setminus \overline{\Lambda}_{\Theta,\mathcal{F}}$, the collection of all (θ,F) such that at least one of the inequalities is binding or violated with positive probability. We will allow for $\sigma_F^2(\theta)$ to become arbitrarily close to zero over $(\Theta \times \mathcal{F}) \setminus \overline{\Lambda}_{\Theta,\mathcal{F}}$ as long as the following integrability condition is satisfied.

Assumption 6 ***(A sufficient condition for a uniform Berry-Esseen bound)*** *There exists a $B < \infty$ such that,*

$$\frac{E_F[|\psi_F^{\mathcal{T}}(Z_i,\theta)|^3]}{\sigma_F^3(\theta)} < B \quad \forall\; (\theta,F) \in (\Theta \times \mathcal{F}) \setminus \overline{\Lambda}_{\Theta,\mathcal{F}} \quad \blacksquare$$

Note that Assumption 6 allows for $\sigma_F^2(\theta)$ to become arbitrarily close to zero over $(\Theta\times\mathcal{F})\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}$. By the Berry-Esseen Theorem (Lehmann and Romano (2005, Theorem 11.2.7)), the condition in Assumption 6 is sufficient to ensure the existence of a $C>0$ such that

$$\sup_{(\theta,F)\in(\Theta\times\mathcal{F})\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}}\ \sup_{d}\left|P_F\left(\frac{1}{\sqrt{n}}\frac{\sum_{i=1}^{n}\psi_F^{\mathcal{T}}(Z_i,\theta)}{\sigma_F(\theta)}\le d\right)-\Phi(d)\right|\le\frac{C}{n^{1/2}} \tag{22}$$

where Φ denotes, as usual, the standard normal cdf.

A regularized statistic

Let $\kappa>0$ be an arbitrarily small, but strictly positive constant, and define

$$t_n(\theta)\equiv\frac{\sqrt{n}\cdot\widehat{\mathcal{T}}(\theta)}{(\sigma_F(\theta)\vee\kappa)}=\begin{cases}\frac{\sqrt{n}\cdot\xi_n^{\mathcal{T}}(\theta)}{(\sigma_F(\theta)\vee\kappa)} & \forall\,(\theta,F)\in\overline{\Lambda}_{\Theta,\mathcal{F}},\\ \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^{n}\psi_F^{\mathcal{T}}(Z_i,\theta)}{(\sigma_F(\theta)\vee\kappa)}+\frac{\sqrt{n}\cdot\xi_n^{\mathcal{T}}(\theta)}{(\sigma_F(\theta)\vee\kappa)} & \forall\,(\theta,F)\in\Lambda_{\Theta,\mathcal{F}}\setminus\overline{\Lambda}_{\Theta,\mathcal{F}},\\ \frac{\sqrt{n}\cdot\mathcal{T}_F(\theta)}{(\sigma_F(\theta)\vee\kappa)}+\frac{1}{\sqrt{n}}\frac{\sum_{i=1}^{n}\psi_F^{\mathcal{T}}(Z_i,\theta)}{(\sigma_F(\theta)\vee\kappa)}+\frac{\sqrt{n}\cdot\xi_n^{\mathcal{T}}(\theta)}{(\sigma_F(\theta)\vee\kappa)} & \forall\,(\theta,F)\in(\Theta\times\mathcal{F})\setminus\Lambda_{\Theta,\mathcal{F}}\end{cases} \tag{23}$$

The purpose of $\kappa>0$ in $t_n(\theta)$ is to *regularize* the asymptotic standard deviation of $\widehat{\mathcal{T}}(\theta)$, which is equal to zero over $\overline{\Lambda}_{\Theta,\mathcal{F}}$. Since $\widehat{\mathcal{T}}(\theta)$ is a scalar, regularization can be done in a straightforward way. Note from the result in Theorem 1 that,

$$\sup_{(\theta,F)\in\Theta\times\mathcal{F}}\left|\frac{\sqrt{n}\cdot\xi_n^{\mathcal{T}}(\theta)}{(\sigma_F(\theta)\vee\kappa)}\right|=o_p\left(\frac{1}{n^{\epsilon}}\right), \tag{24}$$

where $\epsilon>0$ is the constant described in Assumption 4. The pivotal in properties in (22) and the asymptotic results in (23) can be the foundation for the construction of a confidence set (CS) for θ_0 based on $t_n(\theta)$. Fix $\alpha\in(0,1)$ and let $z_{1-\alpha}$ be the standard normal $(1-\alpha)^{th}$ quantile. If Assumptions 1-6 hold, Theorem 1 and the resulting properties in (21)-(24) yield,

$$\begin{aligned}&(i)\quad\lim_{n\to\infty}\ \sup_{(\theta,F)\in\overline{\Lambda}_{\Theta,\mathcal{F}}}P_F\left(t_n(\theta)>z_{1-\alpha}\right)=0,\\ &(ii)\quad\lim_{n\to\infty}\ \sup_{\substack{(\theta,F)\in\Lambda_{\Theta,\mathcal{F}}\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}:\\ \sigma_F(\theta)\ge\kappa}}\left|P_F\left(t_n(\theta)>z_{1-\alpha}\right)-\alpha\right|=0,\\ &(iii)\quad\limsup_{n\to\infty}\ \sup_{\substack{(\theta,F)\in\Lambda_{\Theta,\mathcal{F}}\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}:\\ \sigma_F(\theta)<\kappa}}P_F\left(t_n(\theta)>z_{1-\alpha}\right)\le\alpha.\end{aligned} \tag{25}$$

From (25), we have

$$\liminf_{n\to\infty}\ \inf_{(\theta,F)\in\Lambda_{\Theta,\mathcal{F}}}P_F\left(t_n(\theta)\le z_{1-\alpha}\right)\ge 1-\alpha \tag{26}$$

Next note from (22) that, for any $(\theta, F) \in (\Theta \times \mathcal{F}) \setminus \Lambda_{\Theta,\mathcal{F}}$, and any given c,

$$\lim_{n\to\infty} P_F\left(\frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n \psi_F^{\mathcal{T}}(Z_i,\theta)}{(\sigma_F(\theta) \vee \kappa)} + \underbrace{\frac{\sqrt{n}\cdot \mathcal{T}_F(\theta)}{(\sigma_F(\theta) \vee \kappa)}}_{\to\infty} > c \right) = 1$$

Combined with (23)-(24), this yields $\lim_{n\to\infty} P_F(t_n(\theta) > z_{1-\alpha}) = 1$ for each $(\theta, F) \in (\Theta \times \mathcal{F}) \setminus \Lambda_{\Theta,\mathcal{F}}$. More generally, take any sequence in $(\theta_n, F_n) \in (\theta, F) \setminus \Lambda_{\Theta,\mathcal{F}}$ such that $\mathcal{T}_{F_n}(\theta_n) \geq \delta_n n^{-1/2} D$ for some fixed $D > 0$ and some sequence of positive constants $\delta_n \to \infty$. The results in (22), (23) and (24) yield,

$$\lim_{n\to\infty} P_{F_n}(t_n(\theta_n) > z_{1-\alpha}) = 1. \tag{27}$$

Next, we can describe local power properties related to the statistic t_n. Take any sequence $(\theta_n, F_n) \in (\theta, F) \setminus \Lambda_{\Theta,\mathcal{F}}$ such that

$$\lim_{n\to\infty} \frac{\left(\sigma_{F_n}(\theta_n) \vee \kappa\right)}{\sigma_{F_n}(\theta_n)} = s_1, \quad \text{and} \quad \lim_{n\to\infty} \frac{\sqrt{n}\cdot \mathcal{T}_{F_n}(\theta_n)}{\left(\sigma_{F_n}(\theta_n) \vee \kappa\right)} = s_2.$$

Note that $s_1 \geq 1$ and $s_2 \geq 0$. From the results in (22), (23) and (24), we have,

$$\lim_{n\to\infty} P_{F_n}(t_n(\theta_n) > z_{1-\alpha}) = 1 - \Phi(s_1 \cdot z_{1-\alpha} - s_1 \cdot s_2) \tag{28}$$

Having $\lim_{n\to\infty} P_{F_n}(t_n(\theta_n) > z_{1-\alpha}) > \alpha$ corresponds to the notion of *nontrivial asymptotic power* (see Lee, Song, and Whang (2018, Definition 3)) for the test underlying the construction of our CS. From (28), this test will have nontrivial asymptotic power for the type of sequences described above iff $s_1 \cdot z_{1-\alpha} - s_1 \cdot s_2 < z_{1-\alpha}$.

3.6 Construction of a confidence set for θ_0

We propose to construct a confidence set (CS) for θ_0 based on the properties of the statistic $t_n(\theta)$ described in (26)-(28). Before proceeding, we need to construct an estimator for $\sigma_F^2(\theta)$.

3.6.1 An estimator for $\sigma_F^2(\theta)$

Using the structure of the influence function $\psi_F^{\mathcal{T}}(z,\theta)$ in Theorem 1, we can construct an estimator for $\sigma_F^2(\theta) \equiv E_F[\psi_F^{\mathcal{T}}(Z,\theta)^2]$. Our estimator is

$$\widehat{\sigma}^2(\theta) \equiv \frac{1}{n}\sum_{i=1}^n \widehat{\psi}^{\mathcal{T}}(Z_i,\theta)^2, \quad \text{where} \quad \widehat{\psi}^{\mathcal{T}}(z,\theta) \equiv \widehat{\psi}^{\mathcal{T}_2}(z,\theta) + \widehat{\psi}^{\mathcal{T}_1}(z,\theta_1),$$

$\widehat{\psi}^{\mathcal{T}_2}(Z_i,\theta)$ and $\widehat{\psi}^{\mathcal{T}_1}(Z_i,\theta_1)$ are estimators of $\psi_F^{\mathcal{T}_2}(Z_i,\theta)$ and $\psi_F^{\mathcal{T}_1}(Z_i,\theta_1)$, constructed as follows. We estimate $\psi_F^{\mathcal{T}_2}(Z_i,\theta)$ with $\widehat{\psi}^{\mathcal{T}_2}(Z_i,\theta) \equiv 2\cdot \widehat{H}_1^{\mathcal{T}_2}(V_i,\theta) + \widehat{H}_2^{\mathcal{T}_2}(Z_i,\theta)$ where, for a given $v \equiv (x_2, w_1)$ and $z \equiv (y_1, y_2, v)$, based on the expressions in part (A) of Theorem 1, we estimate

$$\widehat{H}_1^{\mathcal{T}_2}(v,\theta) \equiv \frac{1}{2}\cdot\frac{1}{n}\sum_{j=1}^{n}\Big[\widehat{\tau}_2(v,V_j,\theta)\mathbb{1}\Big\{\widehat{\tau}_2(v,V_j,\theta)\geq -b_n\Big\} + \widehat{\tau}_2(V_j,v,\theta)\mathbb{1}\Big\{\widehat{\tau}_2(V_j,v,\theta)\geq -b_n\Big\}\Big] - \widehat{\mathcal{T}}_2(\theta)$$

$$\begin{aligned}\widehat{H}_2^{\mathcal{T}_2}(z,\theta) \equiv& \left(\left(\widehat{\eta}_a^{\tau_2}(v,\theta) - \widehat{\eta}_b^{\tau_2}(v,\theta)\right)\cdot y_1 + \left(\widehat{\eta}_c^{\tau_2}(v,\theta) - \widehat{\eta}_d^{\tau_2}(v,\theta)\right)\cdot y_2 y_1\right)\cdot \widehat{f}_V(v)\cdot \phi_2(v)^2 \\ &- \frac{1}{n}\sum_{j=1}^{n}\left[\left(\left(\widehat{\eta}_a^{\tau_2}(V_j,\theta) - \widehat{\eta}_b^{\tau_2}(V_j,\theta)\right)\cdot Y_{1j} + \left(\widehat{\eta}_c^{\tau_2}(V_j,\theta) - \widehat{\eta}_d^{\tau_2}(V_j,\theta)\right)\cdot Y_{2j}Y_{1j}\right)\cdot \widehat{f}_V(V_j)\cdot \phi_2(V_j)^2\right].\end{aligned}$$

Using the definitions in (18), the estimators on the right-hand side of the previous expressions are,

$$\begin{aligned}\widehat{\eta}_a^{\tau_2}(v,\theta) \equiv& \frac{1}{n}\sum_{j=1}^{n}\left(\widehat{R}_2(V_j) - \left(g_2(X_{2j},\theta_2) - g_2(x_2,\theta_2)\right)\widehat{Q}_2(V_j)\right)\mathbb{1}\Big\{g_{1U}(w_1,\theta_1)\leq g_{1L}(W_{1j},\theta_1)\Big\}\phi_2(V_j) \\ &\cdot \mathbb{1}\Big\{\widehat{\tau}_2(V_j,v,\theta)\geq -b_n\Big\}, \\ \widehat{\eta}_b^{\tau_2}(v,\theta) \equiv& \frac{1}{n}\sum_{j=1}^{n}\left(\widehat{R}_2(V_j) - \left(g_2(X_{2j},\theta_2) - g_2(x_2,\theta_2)\right)\widehat{Q}_2(V_j)\right)\mathbb{1}\Big\{g_{1U}(W_{1j},\theta_1)\leq g_{1L}(w_1,\theta_1)\Big\}\phi_2(V_j) \\ &\cdot \mathbb{1}\Big\{\widehat{\tau}_2(v,V_j,\theta)\geq -b_n\Big\}, \\ \widehat{\eta}_c^{\tau_2}(v,\theta) \equiv& \frac{1}{n}\sum_{j=1}^{n}\widehat{Q}_2(V_j)\mathbb{1}\Big\{g_{1U}(W_{1j},\theta_1)\leq g_{1L}(w_1,\theta_1)\Big\}\phi_2(V_j)\mathbb{1}\Big\{\widehat{\tau}_2(v,V_j,\theta)\geq -b_n\Big\}, \\ \widehat{\eta}_d^{\tau_2}(v,\theta) \equiv& \frac{1}{n}\sum_{j=1}^{n}\widehat{Q}_2(V_j)\mathbb{1}\Big\{g_{1U}(w_1,\theta_1)\leq g_{1L}(W_{1j},\theta_1)\Big\}\phi_2(V_j)\mathbb{1}\Big\{\widehat{\tau}_2(V_j,v,\theta)\geq -b_n\Big\},\end{aligned}$$

with $\widehat{R}_2$ and $\widehat{Q}_2$ as described in (16), and and $\widehat{f}_V(v) \equiv \frac{1}{h_n^r}\cdot\frac{1}{n}\sum_{i=1}^{n}\Gamma(V_i,v,h_n)$. Next, our estimator for $\psi_F^{\mathcal{T}_1}(Z_i,\theta_1)$ is given by $\widehat{\psi}^{\mathcal{T}_1}(Z_i,\theta_1) \equiv 2\cdot\widehat{H}_1^{\mathcal{T}_1}(W_{1i},\theta_1) + \widehat{H}_2^{\mathcal{T}_1}(Z_i,\theta_1)$ where, for a given $v \equiv (x_2, w_1)$ and $z \equiv (y_1, y_2, v)$, based on the expressions in part (B) of Theorem 1, we estimate

$$\begin{aligned}&\widehat{H}_1^{\mathcal{T}_1}(w_1,\theta_1) \equiv \\ &\frac{1}{2}\cdot\frac{1}{n}\sum_{j=1}^{n}\Big[\widehat{\tau}_1(w_1,W_{1j},\theta_1)\mathbb{1}\Big\{\widehat{\tau}_1(w_1,W_{1j},\theta_1)\geq -b_n\Big\} + \widehat{\tau}_1(W_{1j},w_1,\theta_1)\mathbb{1}\Big\{\widehat{\tau}_1(W_{1j},w_1,\theta_1)\geq -b_n\Big\}\Big] - \widehat{\mathcal{T}}_1(\theta_1),\end{aligned}$$

$$\begin{aligned}\widehat{H}_2^{\mathcal{T}_1}(z,\theta_1) \equiv& \left(\left(\widehat{\eta}_a^{\tau_1}(w_1,\theta_1) - \widehat{\eta}_b^{\tau_1}(w_1,\theta_1)\right) + \left(\widehat{\eta}_c^{\tau_1}(w_1,\theta_1) - \widehat{\eta}_d^{\tau_1}(w_1,\theta_1)\right)\cdot y_1\right)\cdot \widehat{f}_{W_1}(w_1)\cdot\phi_1(w_1)^2 \\ &- \frac{1}{n}\sum_{j=1}^{n}\left[\left(\left(\widehat{\eta}_a^{\tau_1}(W_{1j},\theta_1) - \widehat{\eta}_b^{\tau_1}(W_{1j},\theta_1)\right) + \left(\widehat{\eta}_c^{\tau_1}(W_{1j},\theta_1) - \widehat{\eta}_d^{\tau_1}(W_{1j},\theta_1)\right)\cdot Y_{1j}\right)\cdot\widehat{f}_{W_1}(W_{1j})\cdot\phi_1(W_{1j})^2\right].\end{aligned}$$

Using the definitions in (19), the estimators on the right-hand side of the previous expressions are,

$$\widehat{\eta}_a^{\tau_1}(w_1,\theta_1) \equiv \frac{1}{n}\sum_{j=1}^{n} \widehat{R}_1(W_{1j})\mathbb{1}\left\{g_{1U}(w_1,\theta_1) \le g_{1L}(W_{1j},\theta_1)\right\}\mathbb{1}\left\{\widehat{\tau}_1(W_{1j},w_1,\theta) \ge -b_n\right\}\phi_1(W_{1j}),$$

$$\widehat{\eta}_b^{\tau_1}(w_1,\theta_1) \equiv \frac{1}{n}\sum_{j=1}^{n} \widehat{R}_1(W_{1j})\mathbb{1}\left\{g_{1U}(W_{1j},\theta_1) \le g_{1L}(w_1,\theta_1)\right\}\mathbb{1}\left\{\widehat{\tau}_1(w_1,W_{1j},\theta) \ge -b_n\right\}\phi_1(W_{1j}),$$

$$\eta_{c,F}^{\tau_1}(w_1,\theta_1) \equiv \frac{1}{n}\sum_{j=1}^{n} \widehat{Q}_1(W_{1j})\mathbb{1}\left\{g_{1U}(W_{1j},\theta_1) \le g_{1L}(w_1,\theta_1)\right\}\mathbb{1}\left\{\widehat{\tau}_1(w_1,W_{1j},\theta) \ge -b_n\right\}\phi_1(W_{1j}),$$

$$\eta_{d,F}^{\tau_1}(w_1,\theta_1) \equiv \frac{1}{n}\sum_{j=1}^{n} \widehat{Q}_1(W_{1j})\mathbb{1}\left\{g_{1U}(w_1,\theta_1) \le g_{1L}(W_{1j},\theta_1)\right\}\mathbb{1}\left\{\widehat{\tau}_1(W_{1j},w_1,\theta) \ge -b_n\right\}\phi_1(W_{1j})$$

In Section A2 of the online appendix we show that, under the conditions of Theorem 1, we have

$$\sup_{\theta\in\Theta}\left|\widehat{\sigma}^2(\theta) - \sigma_F^2(\theta)\right| = o_p(1), \quad \text{uniformly over } \mathcal{F}. \tag{29}$$

3.6.2 Confidence set

By continuity, (29) implies $\sup_{\theta\in\Theta}|(\widehat{\sigma}(\theta)\vee\kappa) - (\sigma_F(\theta)\vee\kappa)| = o_p(1)$, uniformly over $\mathcal{F}$. Let

$$\widehat{t}_n(\theta) \equiv \frac{\sqrt{n}\cdot\widehat{\mathcal{T}}(\theta)}{(\widehat{\sigma}(\theta)\vee\kappa)}.$$

We wish to construct a CS that contains θ_0 with asymptotic target coverage probability $1-\alpha$ (recall that $(\theta_0,F)\in\Lambda_{\Theta,\mathcal{F}}$). Based on the results in (26)-(29), we construct our CS as,

$$\widehat{CS}_{1-\alpha} \equiv \left\{\theta\in\Theta : \widehat{t}_n(\theta) \le z_{1-\alpha}\right\}.$$

From our previous results, the properties of $\widehat{CS}_{1-\alpha}$ are summarized in the following theorem.

Theorem 2 *Suppose Assumptions 1-6 hold. Then $\widehat{CS}_{1-\alpha}$ has the following asymptotic properties.*

(i) Uniform asymptotic coverage: $\liminf_{n\to\infty}\inf_{(\theta,F)\in\Lambda_{\Theta,\mathcal{F}}} P_F(\theta\in\widehat{CS}_{1-\alpha}) \ge 1-\alpha$.

(ii) Consistency of the associated test for $(\theta,F)\in\Lambda_{\Theta,\mathcal{F}}$: For any $(\theta_n,F_n)\in(\theta,F)\setminus\Lambda_{\Theta,\mathcal{F}}$ such that $\mathcal{T}_{F_n}(\theta_n) \ge \delta_n n^{-1/2}D$ for some fixed $D>0$ and some sequence of positive constants $\delta_n\to\infty$, we have $\lim_{n\to\infty}P_{F_n}(\theta_n\in\widehat{CS}_{1-\alpha}) = 0$.

(iii) Nontrivial local power of the associated test for $(\theta,F)\in\Lambda_{\Theta,\mathcal{F}}$: Take any sequence $(\theta_n,F_n)\in(\theta,F)\setminus$

$\Lambda_{\Theta,\mathcal{F}}$ *such that* $\lim_{n\to\infty} \frac{(\sigma_{F_n}(\theta_n)\vee\kappa)}{\sigma_{F_n}(\theta_n)} = s_1$ *and* $\lim_{n\to\infty} \frac{\sqrt{n}\cdot\mathcal{T}_{F_n}(\theta_n)}{(\sigma_{F_n}(\theta_n)\vee\kappa)} = s_2$ *(note that* $s_1 \geq 1$ *and* $s_2 \geq 0$*). We have* $\lim_{n\to\infty} P_{F_n}(\theta_n \in \widehat{CS}_{1-\alpha}) < 1-\alpha$ *if* $s_1 \cdot z_{1-\alpha} - s_1 \cdot s_2 < z_{1-\alpha}$. ■

The properties in Theorem 2 follow from the results we obtained in equations (26)-(28) and the uniform consistency of our estimator $\widehat{\sigma}(\theta)$, described in equation (29) and shown in Section A2 of the online appendix. Combined with the result in (29), part (i) of Theorem 2 follows from (26), while parts (ii) and (iii) follow from (27) and (28), respectively.

3.6.3 A stronger version of Assumption 6

Assumption 6 allows for $\sigma_F^2(\theta)$ (the relevant measure of the contact sets in our problem) to become arbitrarily close to zero over $(\Theta\times\mathcal{F})\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}$. If we strengthen Assumption 6 to assume now that $\sigma_F^2(\theta)$ is bounded away from zero uniformly over $(\Theta\times\mathcal{F})\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}$, we can turn our regularization parameter κ into a sequence that vanishes asymptotically. Suppose we replace Assumption 6 with the following stronger restriction.

Assumption 6' ***(A stronger version of Assumption 6)*** *There exist a* $B < \infty$ *and* $\underline{C} > 0$ *such that,*

$$E_F[|\psi_F^{\mathcal{T}}(Z_i,\theta)|^3] \leq B, \quad \textit{and} \quad \sigma_F^2(\theta) \geq \underline{C} \quad \forall\ (\theta,F)\in(\Theta\times\mathcal{F})\setminus\overline{\Lambda}_{\Theta,\mathcal{F}} \quad ■$$

The Berry-Esseen condition produced by Assumption 6, and the results in Theorem 2 still hold under the stronger restrictions of Assumption 6', but we now also have the following result. Take any positive sequence $\kappa_n \to 0$ such that $\kappa_n \cdot n^{\epsilon} \to \infty$, with $\epsilon > 0$ being the constant described in Assumption 4. Note from (24) that,

$$\sup_{(\theta,F)\in\Theta\times\mathcal{F}} \left|\frac{n^{1/2}\cdot\xi_n^{\mathcal{T}}(\theta)}{(\sigma_F(\theta)\vee\kappa_n)}\right| = o_p\left(\frac{1}{\kappa_n\cdot n^{\epsilon}}\right) = o_p(1). \tag{30}$$

If Assumption 6' holds, then for n large enough we have $(\sigma_F(\theta)\vee\kappa_n) = \sigma_F(\theta)\ \forall\ (\theta,F)\in(\Theta\times\mathcal{F})\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}$. Thus, if we replace the constant regularization parameter $\kappa > 0$ with a sequence $\kappa_n \to 0$ such that $\kappa_n\cdot n^{\epsilon}\to\infty$ and define now,

$$t_n(\theta) \equiv \frac{\sqrt{n}\cdot\widehat{\mathcal{T}}(\theta)}{(\sigma_F(\theta)\vee\kappa_n)}.$$

If we replace Assumption 6 with Assumption 6', the results in equation (25) are strengthened to the following,

$$\begin{aligned} &(i) \quad \lim_{n\to\infty} \sup_{(\theta,F)\in\overline{\Lambda}_{\Theta,\mathcal{F}}} P_F\left(t_n(\theta) > z_{1-\alpha}\right) = 0, \\ &(ii) \quad \lim_{n\to\infty} \sup_{(\theta,F)\in\Lambda_{\Theta,\mathcal{F}}\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}} \left|P_F\left(t_n(\theta) > z_{1-\alpha}\right) - \alpha\right| = 0. \end{aligned} \tag{25'}$$

Thus, the test based on $t_n(\theta)$ would no longer be conservative if (θ, F) are such that $\sigma_F(\theta) < \kappa$ when κ is a constant regularization parameter instead of a sequence vanishing to zero. All the remaining results regarding the construction of our confidence set remain valid.

3.6.4 Asymptotic adaptation to the measure of the contact sets and the role of b_n

Several papers have proposed methods to detect how close moment inequalities come to being binding for the purpose of obtaining critical values. These include generalized moment selection as developed by Andrews and Soares (2010) and Andrews and Shi (2013), adaptive inequality selection as in Chernozhukov, Lee, and Rosen (2013), the refined moment selection method proposed in Chetverikov (2017), and the use of contact set estimators proposed by Lee, Song, and Whang (2018). All of these methods use tuning parameters. In our case, how close the functional inequalities come to binding is related to the measure of the contact sets. The properties of the sequence b_n are designed to ensure that the estimators $\widehat{\mathcal{T}}(\theta)$ and $\widehat{\sigma}(\theta)$ adapt asymptotically to the measure of the contact sets, captured in our case by $\sigma_F^2(\theta)$.

3.7 On the choice of tuning parameters

Our procedure uses three tuning parameters: the bandwidth sequences h_n and b_n, along with the regularization constant κ. While we leave the exploration of a general theory of how to choose these tuning parameters for future work, we can provide recommendations based on the results of our Monte Carlo experiments in Section 4. First, we consider covariate-specific bandwidths for each continous covariate V_m^c of the form $h_n = c_h \cdot \|\widehat{\sigma}(V_m^c)\| \cdot n^{-\alpha_h}$, where $\alpha_h > 0$ denotes the rate of convergence of h_n, which will be set to satisfy the conditions in Assumption 4. Next, we set b_n and the regularization tuning parameter κ to be proportional to a measure of the scale of the functionals $\tau_{2F}(V_i, V_j, \theta)$ and $\tau_{1F}(W_{1i}, W_{1j}, \theta_1)$. Denote $\overline{g}_2(x_2) \equiv \sup_{\theta_2 \in \Theta} |g_2(x_2, \theta_2)|$, and let

$$
\begin{aligned}
\overline{\tau}_{2F}(V_i, V_j) \equiv & \Big(\Big(\big|R_{2F}(V_i)Q_{2F}(V_j)\big| + \big|R_{2F}(V_j)Q_{2F}(V_i)\big|\Big) + \big(\overline{g}_2(X_{2i}) + \overline{g}_2(X_{2j})\big)\big|Q_{2F}(V_i)Q_{2F}(V_j)\big|\Big) \\
& \cdot \phi_2(V_i)\phi_2(V_j), \qquad (31) \\
\overline{\tau}_{1F}(W_{1i}, W_{1j}) \equiv & \Big(\big|R_{1F}(W_{1j})Q_{1F}(W_{1i})\big| + \big|R_{1F}(W_{1i})Q_{1F}(W_{1j})\big|\Big) \cdot \phi_1(W_{1i})\phi_1(W_{1j}).
\end{aligned}
$$

By construction, we have $|\tau_{2F}(V_i, V_j, \theta)| \le \overline{\tau}_{2F}(V_i, V_j)$ and $|\tau_{1F}(W_{1i}, W_{1j})| \le \overline{\tau}_{1F}(W_{1i}, W_{1j})$ $\forall\ \theta \in \Theta$. Fix $q \in (0,1)$ and let $\overline{\tau}_{2F_{(q)}}$ and $\overline{\tau}_{1F_{(q)}}$ denote the q^{th} quantiles of $\overline{\tau}_{2F}(V_i, V_j)$ and $\overline{\tau}_{1F}(W_{1i}, W_{1j})$, respectively. We propose using a tuning parameter b_n specific to each of the two functionals[9], (see Remark 1), with $b_n^{\tau_2} = c_b \cdot \widehat{\overline{\tau}}_{2_{(q)}} \cdot n^{-\alpha_b}$ for $\widehat{\tau}_2(V_i, V_j, \theta)$, and $b_n^{\tau_1} = c_b \cdot \widehat{\overline{\tau}}_{1_{(q)}} \cdot n^{-\alpha_b}$ for $\widehat{\tau}_1(W_{1i}, W_{1j}, \theta_1)$. Both $b_n^{\tau_2}$ and $b_n^{\tau_1}$ have the same convergence rate, $\alpha_b > 0$, which will be set to satisfy the restrictions in

[9] In our Monte Carlo experiments we use $q = 0.50$ (i.e, the median).

Assumption 4. We estimate,

$$\widehat{\mathcal{T}}_2(\theta) \equiv \frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\widehat{\tau}_2(V_i,V_j,\theta)\cdot\mathbb{1}\left\{\widehat{\tau}_2(V_i,V_j,\theta)\geq -b_n^{\tau_2}\right\},$$
$$\widehat{\mathcal{T}}_1(\theta_1) \equiv \frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\widehat{\tau}_1(W_{1i},W_{1j},\theta_1)\cdot\mathbb{1}\left\{\widehat{\tau}_1(W_{1i},W_{1j},\theta_1)\geq -b_n^{\tau_1}\right\},$$
$$\widehat{\mathcal{T}}(\theta) \equiv \widehat{\mathcal{T}}_2(\theta)+\widehat{\mathcal{T}}_1(\theta_1),$$

We propose a regularization parameter κ of the form $\kappa = c_\kappa\cdot(\widehat{\overline{\tau}}_{2_{(q)}}\wedge\widehat{\overline{\tau}}_{1_{(q)}})$. Recall that r denotes the number of continuously distributed covariates in V. Fix $\epsilon>0$. We will describe conditions under which this ϵ satisfies the restrictions described in Assumption 4. Take any $\delta>2\epsilon$ such that $\epsilon+\delta<\frac{1}{2}$. Consider the convergence rates $\alpha_h=\frac{1}{4r}-\frac{\epsilon+\delta}{2r}$ and $\alpha_b=\frac{1}{4}+\Delta_b$, where $\frac{\epsilon}{2}<\Delta_b<\frac{\delta-\epsilon}{2}$. It is easy to verify that α_h and α_b satisfy the bandwidth convergence restrictions in Assumption 4 if $M>2r\left(\frac{1+2\epsilon}{1-2(\epsilon+\delta)}\right)$. The lower bound for M is $2r+1$, which is attained if ϵ and δ are chosen to be small enough such that $\frac{2\epsilon+\delta}{1-2(\epsilon+\delta)}<\frac{1}{4r}$.

Regarding the choice of the constants c_h, c_κ and c_b, in our Monte Carlo experiments we proceeded as follows. First, we chose values of c_h close to Silverman's so-called "rule of thumb" (Silverman (1986, p. 45)), which would be $c_h=1.06$. Specifically, we computed our results for $c_h\in\{0.80,1.00,1.20\}$ and found that our results were robust throughout. Similarly, we computed our results for $c_b\in\{0.001,0.01,0.10\}$. Choosing $c_b=0.001$ brings our bandwidth b_n very close to zero, while $c_b=0.10$ produces a more conservative choice. To isolate the effect of b_n, we chose $c_\kappa=10^{-10}$ for our regularization constant κ. Even though there was some tradeoff between size and power for the underlying test, our estimated CS displayed finite-sample properties consistent with the asymptotic predictions in Theorem 2 for all the tuning parameter values analyzed. The detils are included in our Monte Carlo experiments in Section 4. Overall, we find that choosing h_n in a "conventional" way (e.g, following "rule of thumb" choices from the literature) and choosing small, nonzero values for b_n and κ will produce inferential results with an underlying test that has good power and size properties in finite samples.

4 Monte Carlo experiments

We apply our inferential method to an interactions-based model with unobserved beliefs like the one we studied in Section 2.1.2. We have a selection equation of the form,

$$\begin{aligned}&Y_1^* = \beta_{10}^0+\beta_{10}^1\cdot W_1^1+\beta_{10}^2\cdot W_1^2+\beta_{10}^\pi\pi_1+\varepsilon_1,\\ &\text{where}\quad \beta_{10}^0=0.10,\quad \beta_{10}^1=0.125,\quad \beta_{10}^2=-0.25,\quad \beta_{10}^\pi=0.40.\end{aligned} \tag{32}$$

As in the interactions-based model from Section 2.1.2, π_1 denotes the agent's subjective expectation for $P(Y_1 = 1|W_1)$, with $W_1 \equiv (W_1^1, W_1^2)$. The observable covariates in (32) were generated as $W_1^1 \sim$ logistic and $W_1^2 \sim \mathcal{N}(0,1)$, both independent of each other and independent of $\varepsilon_1 \sim \mathcal{N}(0,1)$. The unobserved shock ε_1 was generated as a standard normal random variable, independent of W_1. We generated data assuming that agents use Bayesian Nash equilibrium (BNE) beliefs, which are a solution (in π_1) to the BNE system,

$$\pi_1 = H_1(\beta_{10}^0 + \beta_{10}^1 \cdot W_1^1 + \beta_{10}^2 \cdot W_1^2 + \beta_{10}^{\pi} \pi_1),$$

where $H_1(\cdot)$ denotes the standard normal cdf. If there exist multiple BNE solutions, we used an equilibrium selection mechanism that chose the closest solution to 1/2. Beliefs are the unobserved control variable for the econometrician, and we use the iterated dominance approach described in Section 2.1.2 to construct bounds for them, with $k = 2$ steps of iterated elimination of nonrationalizable beliefs. Since $\beta_{10}^{\pi} > 0$, the construction of our bounds follows the steps described in Section 2.1.2. Our bounds for beliefs are nonlinear functions of θ_{10}, *including the intercept term* β_{10}^0. The outcome equation in our experiments is given by,

$$Y_2^* = \underbrace{\beta_{20}^0 + \beta_{20}^1 \cdot X_2^1 + \beta_{20}^2 \cdot X_2^2}_{g_2(X_2, \beta_{20})} + \varepsilon_2, \quad \text{with} \quad \beta_{20}^0 = 0.50, \quad \beta_{20}^1 = 0.40, \quad \beta_{20}^2 = -0.50. \tag{33}$$

The observable covariates in (33) were generated as $X_2^1 \sim \chi_1^2$ and $X_2^2 \sim$ exponential$(\lambda = 1)$, with the unobserved shock $\varepsilon_2 \sim \mathcal{N}(0,2)$, and

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \right), \quad \text{and} \quad (\varepsilon_1, \varepsilon_2) \perp (W_1^1, W_1^2, X_2^1, X_2^2).$$

Thus, the correlation between $\rho(\varepsilon_1, \varepsilon_2) \approx 0.70$. The intercept term β_{20}^0 in (33) is impossible to partially identify since it drops from the difference $g_2(X_{2i}, \beta_{20}) - g_2(X_{2j}, \beta_{20})$. Thus, the parameter of interest in the outcome equation is $\theta_{20} \equiv (\beta_{20}^1, \beta_{20}^2)$, and the parameter vector we will attempt to do inference on is

$$\theta_0 \equiv (\underbrace{\beta_{10}^0, \beta_{10}^1, \beta_{10}^2, \beta_{10}^{\pi}}_{\theta_{10}}, \underbrace{\beta_{20}^1, \beta_{20}^2}_{\theta_{20}}).$$

Our restricted parameter space assumes $\beta_1^{\pi} > 0$. We consider a setting where the signs of the slope coefficients in both equations are assumed to be known ex-ante –a realistic scenario in many applications–. The sign of the intercept β_{10}^0 is left unknown. We set $||\theta|| \leq 2 \ \ \forall \ \theta \in \Theta$.

4.1 Inference range, kernels, and bandwidths

Our experiments include $r \equiv 4$ continuously distributed covariates $V \equiv (X_2^1, X_2^2, W_1^1, W_1^2)$, which we will denote as $V \equiv (V_1, V_2, V_3, V_4)$. We chose as our inference range

$$\mathcal{V} = \Big\{ v \equiv (v_1, v_2, v_3, v_4) \in \mathbb{R}^4 : V_{m_{(0.01)}} \leq v_m \leq V_{m_{(0.99)}} \text{ for } m = 1, \ldots, 4 \Big\}.$$

That is, the range of values of V such that each component is between the 0.01 and 0.99 quantiles. As our weight functions ϕ_2 and ϕ_1, we simply chose $\phi_2(V_i) \equiv \mathbb{1}\{V_i \in \mathcal{V}\}$ and $\phi_1(W_{1i}) \equiv \mathbb{1}\{W_{1i} \in \mathcal{V}\}$.

We used a bias-reducing kernel of order $M = 12$. As we described in Section 3.7, this satisfies the restrictions in Assumption 4 combined with our choices for bandwidth convergence rates (to be described below). Furthermore, this value of M satisfies our theoretical smoothness restrictions since our functionals have derivatives of every order. Our kernel is multiplicative, $K(v) = \prod_{m=1}^4 \kappa(v_m)$, with $\kappa(z) = \sum_{\ell=1}^6 c_\ell \cdot (S^2 - z^2)^{2\ell} \cdot \mathbb{1}\{|z| \leq S\}$. By construction, $\kappa(z)$ is symmetric around zero, with support $[-S, S]$. In our experiments we used $S = 10$. The coefficients c_ℓ, are chosen to satisfy the conditions of a bias-reducing kernel of order $M = 12$. We used covariate-specific bandwidths, $h_n = c_h \cdot \|\widehat{\sigma}(V_m)\| \cdot n^{-\alpha_h}$, for $m = 1, \ldots, 4$. Our bandwidth design followed the approach described in Section 3.7. Since Θ was such that $\|\theta\| \leq 2$, we have $\sup_{\theta_2 \in \Theta} \left|\theta_2^1 \cdot X_{2i}^1 + \theta_2^2 \cdot X_{2i}^2\right| \leq 2 \cdot \left(|X_{2i}^1| + |X_{2i}^2|\right)$. From here, using the definition in (31), we used

$$\begin{aligned}
\widehat{\overline{\tau}}_2(V_i, V_j) &\equiv \left(\left(\left|\widehat{R}_2(V_i)\widehat{Q}_2(V_j)\right| + \left|\widehat{R}_2(V_j)\widehat{Q}_2(V_i)\right|\right) + 2\left(|X_{2i}^1| + |X_{2i}^2| + |X_{2j}^1| + |X_{2j}^2|\right)\left|\widehat{Q}_2(V_i)\widehat{Q}_2(V_j)\right|\right) \\
&\quad \cdot \phi_2(V_i)\phi_2(V_j), \\
\widehat{\overline{\tau}}_1(W_{1i}, W_{1j}) &\equiv \left(\left|\widehat{R}_1(W_{1j})\widehat{Q}_1(W_{1i})\right| + \left|\widehat{R}_1(W_{1i})\widehat{Q}_1(W_{1j})\right|\right) \cdot \phi_1(W_{1i})\phi_1(W_{1j}).
\end{aligned}$$

These are estimated measures of the scale of our functionals. Let $\widehat{\overline{\tau}}_{2_{(0.5)}}$ and $\widehat{\overline{\tau}}_{1_{(0.5)}}$ denote the sample medians of $\widehat{\overline{\tau}}_2(V_i, V_j)$ and $\widehat{\overline{\tau}}_1(W_{1i}, W_{1j})$, respectively. As described in Section 3.7, our tuning parameter b_n was specific to each of the two functional inequalities (see Remark 1), with $b_n^{\tau_2} = c_b \cdot \widehat{\overline{\tau}}_{2_{(0.5)}} \cdot n^{-\alpha_b}$ for $\widehat{\tau}_2(V_i, V_j, \theta)$, and $b_n^{\tau_1} = c_b \cdot \widehat{\overline{\tau}}_{1_{(q)}} \cdot n^{-\alpha_b}$ for $\widehat{\tau}_1(W_{1i}, W_{1j}, \theta_1)$. The rates α_h and α_b were given by the formulas in Section 3.7, with $r = 4$, $\epsilon = 0.001$, $\delta = 0.003$ and $\Delta_b = 0.0006$.

Regarding the choices of the bandwidth constants c_h and c_b, we ran our experiments for $c_h \in \{0.80, 1.00, 1.20\}$ and $c_b \in \{0.001, 0.01, 0.10\}$. The values chosen for c_h are close to Silverman's so-called "rule of thumb" (≈ 1.06, see Silverman (1986, p. 45)). Our goal in analyzing different values of c_h, was to analyze how the tradeoff between variance and bias of our kernel estimators impacts our results as we go from "under-smoothing" (with $c_h = 0.80$ in our case) to "over-smoothing" (with $c_h = 1.20$ in our case). We used values close to the so-called "rule of thumb" constant as a reference point. The values chosen for c_b are meant to explore the tradeoff between size and power for the underlying test as we change the magnitude of b_n, going from a more "aggressive" choice ($c_b = 0.001$) to a more conservative one ($c_b = 0.10$). In order to focus on the effects of h_n and b_n, we set our regularization parameter c_κ to $c_\kappa = 10^{-10}$.

4.2 Results

We constructed a CS for with target coverage probability 95%. We used samples of size $n = 500$, $n = 1000$ and $n = 2000$. We generated 1000 simulated samples in each case.

4.2.1 Coverage of the true parameter value θ_0

Given the way the data was generated, our functional inequalities are almost surely never binding, which, by Theorem 2, leads to an asymptotic coverage probability of θ_0 that converges to 1. However, for relatively large values of the covariates, the lower and upper bounds for beliefs become very close to each other, and thus the inequalities come numerically close to binding. Thus, while the asymptotic coverage probability of θ_0 is 1, in finite samples it may be close to the target 95%. This conjecture is supported by our results, which are included in Table 1. There, we describe the frequency with which θ_0 was included in our CS for different values of c_b and c_h. Consistent with with our asymptotic predictions, the coverage frequency was at least 95$ in all cases.

Larger (i.e, more conservative) values of c_b led, in all cases, to a higher probability of including θ_0 in our CS. Larger values of c_b produce larger values of $b_n^{\tau_1}$ and $b_n^{\tau_2}$, leading to more conservative results. We were particularly interested in whether more aggressive (i.e, smaller) choices of c_b might lead to over-rejecting θ_0 from our CS. We find in our experiments that this was not the case, as even when c_b was set to 0.001, our coverage probability remained above the asymptotic target in all cases. Thus, our experiments suggest that choosing aggressive, small values of b_n can lead to more power of the associated test without producing over-rejections of the true parameter value. This conjecture will be further corroborated below.

Regarding the choice of the bandwidth h_n (reflected in our choice of c_h), our results reflected the effects of the variance/bias tradeoffs associated with bandwidth selection in kernel-based estimators. Smaller values of c_h ($c_h = 0.80$ in our case) resulted in an increase in the estimated variance of our test-statistic which, in turn, reduced the probability of rejecting θ_0. Conversely, larger values of h_n led to a smaller variance, but eventually (for $c_h = 1.2$ in our experiments), over-smoothing led to an increase in the bias of our estimators which, eventually, led once again to a decrease in the probability of rejecting θ_0. Overall, the best tradeoff between bias and variance within the values considered was achieved when we chose $c_h = 1$. Overall, even though our results were robust for all values analyzed, the best results were obtained when $c_b = 0.001$ and $c_h = 1$, and we chose these values for our remaining experiments.

Table 1: Monte Carlo results. Coverage frequency for θ_0

	$c_h = 0.80$		
	$c_b = 0.001$	$c_b = 0.01$	$c_b = 0.10$
$n = 500$	0.985	0.989	0.993
$n = 1,000$	0.992	0.997	0.999
$n = 2,000$	0.990	0.992	0.999
	$c_h = 1.00$		
	$c_b = 0.001$	$c_b = 0.01$	$c_b = 0.10$
$n = 500$	0.963	0.975	0.985
$n = 1,000$	0.983	0.990	0.995
$n = 2,000$	0.977	0.988	0.993
	$c_h = 1.20$		
	$c_b = 0.001$	$c_b = 0.01$	$c_b = 0.10$
$n = 500$	0.976	0.982	0.988
$n = 1,000$	0.989	0.991	0.995
$n = 2,000$	0.985	0.990	0.996

• 1000 Monte Carlo simulations in each case.

4.2.2 Some features of our estimated CS

Next, we describe some properties of our estimated confidence set, which was constructed using a random grid search over the parameter space described above.

How informative are our results?

We begin by exploring how the width or volume of our CS changes with the sample size. This is an indication of how informative our inference is. We begin with a proposed measure of the "volume" of our CS. Let

$$\widehat{\rho}^{\theta_0}_{H,1-\alpha} \equiv \sup_{\theta \in \widehat{CS}_{1-\alpha}} ||\theta - \theta_0|| \Big/ \sup_{\theta \in \Theta} ||\theta - \theta_0||.$$

This is the Hausdorff distance between our confidence set and the true parameter value, normalized to lie in $[0,1]$. It is meant to represent a measure of the volume of our estimated CS around the true parameter value. Recall that $\theta_1 \equiv (\beta_1^0, \beta_1^1, \beta_1^2, \beta_1^\pi)$ represents the parameters of the selection equation (including the intercept), and $\theta_2 \equiv (\beta_2^1, \beta_2^2)$ represents the slope coefficients of the outcome equation. Let $\beta_1 \equiv (\beta_1^1, \beta_1^2, \beta_1^\pi)$ denote the slope coefficients of the selection equation. Let

$$\widehat{\rho}_{H,1-\alpha}^{\theta_{10}} \equiv \sup_{\theta_1 \in \widehat{CS}_{1-\alpha}} \|\theta_1 - \theta_{10}\| \Big/ \sup_{\theta_1 \in \Theta} \|\theta_1 - \theta_{10}\|$$

$$\widehat{\rho}_{H,1-\alpha}^{\beta_{10}} \equiv \sup_{\beta_1 \in \widehat{CS}_{1-\alpha}} \left\|\beta_1 - \beta_{10}\right\| \Big/ \sup_{\beta_1 \in \Theta} \left\|\beta_1 - \beta_{10}\right\|$$

$$\widehat{\rho}_{H,1-\alpha}^{\theta_{20}} \equiv \sup_{\theta_2 \in \widehat{CS}_{1-\alpha}} \|\theta_2 - \theta_{20}\| \Big/ \sup_{\theta_2 \in \Theta} \|\theta_2 - \theta_{20}\|$$

These are the corresponding volume measures for θ_1, β_1, and θ_2. Since they are all normalized to $[0,1]$ and they are all centered around the corresponding true parameter value, these measures are comparable. Table 2 includes the median values of these volume measures across our simulations. The results suggest that the true value of the intercept in the selection equation was hard to partially identify, while our results were informative for all the slope coefficients in the model. Notably, our volume measure for the CS of the slope coefficients in the outcome equation, $\widehat{\rho}_{H,0.95}^{\theta_{20}}$, reduced sharply (by approximately 50%) when we went from $n = 1000$ to $n = 2000$. The reductions were less dramatic, but still noticeable, for the slope coefficients of the outcome equation, as illustrated by the behavior of $\widehat{\rho}_{H,0.95}^{\beta_{10}}$. The volume of our overall CS reduced slowly with n reflecting the difficulty to partially identify the intercept coefficient in the selection equation.

Table 2: Monte Carlo results. Median values across our simulations of measures of the volume of our estimated CS

$\widehat{\rho}_{H,0.95}^{\theta_0}$			$\widehat{\rho}_{H,0.95}^{\theta_{10}}$		
$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$
0.482	0.479	0.473	0.561	0.557	0.551
$\widehat{\rho}_{H,0.95}^{\beta_{10}}$			$\widehat{\rho}_{H,0.95}^{\theta_{20}}$		
$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$
0.579	0.523	0.475	0.717	0.663	0.304

- Bandwidths computed with $c_h = 1$, $c_b = 0.001$.
- 1000 Monte Carlo simulations in each case.

Next, we constructed confidence intervals (CIs) for each individual parameter by taking projections from our CS in the usual way. For each parameter $\theta_\ell \in \theta$, we projected a 95% CI given by an interval $(\underline{\theta}_{\ell,0.95}, \overline{\theta}_{\ell,0.95})$, where

$$\underline{\theta}_{\ell,0.95} = \min\left\{\theta_\ell \colon \theta_\ell \in \widehat{CS}_{0.95}\right\}, \quad \overline{\theta}_{\ell,0.95} = \max\left\{\theta_\ell \colon \theta_\ell \in \widehat{CS}_{0.95}\right\}.$$

Table 3 presents median values across our simulations for the lower and upper bounds of the individual CIs. Consistent with the findings in Table 2, the intercept coefficient in the selection equation was difficult to partially identify, with CIs that reduced in width slowly with n, including a vast portion of the parameter space for this coefficient (the interval $[-2, 2]$). In contrast, the CIs for all the remaining parameters became increasingly more informative as n increased. The rate at which this occurred varied across each parameter. We conjecture that these differences reflect the different distributions used to generate each observable covariate. Importantly, note that our results are informative for β_1^{π}, *the slope coefficient of the unobserved control variable in our model*, and that the width of the CI for this parameter reduces noticeably as n grows.

Table 3: Monte Carlo results. Median values across our simulations for the lower bound $\underline{\theta}_{\ell,0.95}$, and for the upper bound $\overline{\theta}_{\ell,0.95}$ of individual CIs

Selection equation				
	β_1^0	β_1^1	β_1^2	β_1^{π}
$n = 500$	$(-1.974, -1.976)$	$(0.013, 1.646)$	$(-1.982, -0.036)$	$(0.012, 1.983)$
$n = 1000$	$(-1.953, -1.941)$	$(0.030, 0.533)$	$(-1.881, -0.077)$	$(0.023, 1.865)$
$n = 2000$	$(-1.929, -1.910)$	$(0.055, 0.253)$	$(-0.601, -0.134)$	$(0.038, 1.636)$

- True values: $\beta_{10}^0 = 0.10$, $\beta_{10}^1 = 0.125$, $\beta_{10}^2 = -0.25$, $\beta_{10}^{\pi} = 0.40$
- Bandwidths computed with $c_h = 1$, $c_b = 0.001$.
- 1000 Monte Carlo simulations in each case.

Outcome equation		
	β_2^1	β_2^2
$n = 500$	$(0.011, 1.972)$	$(-1.984, -0.015)$
$n = 1000$	$(0.020, 1.777)$	$(-1.808, -0.028)$
$n = 2000$	$(0.029, 1.025)$	$(-1.102, -0.037)$

- True values: $\beta_{20}^1 = 0.40$, $\beta_{20}^2 = -0.50$
- Bandwidths computed with $c_h = 1$, $c_b = 0.001$.
- 1000 Monte Carlo simulations in each case.

Power of the associated test

Finally, we looked at some examples of parameter values $\theta \neq \theta_0$ to examine the ability of our test to *exclude* these points from our CS. In each case, we took the true parameter value and perturbed only one or two parameters, leaving the rest fixed at their true values. We focused on the slope coefficients of the selection and outcome equations since our previous results showed that it is difficult to partially identify the intercept of the selection equation. Table 4 presents results for various examples of parameter values.

Table 4: Monte Carlo results. Frequency with which some examples of parameter values $\theta \neq \theta_0$ were excluded from our confidence set

Parameters in the selection equation			
	$\boldsymbol{\beta_1^1 = -0.1, \beta_1^2 = -0.2}$	$\boldsymbol{\beta_1^1 = 0.5, \beta_1^2 = -0.1}$	$\boldsymbol{\beta_1^1 = 1.2, \beta_1^2 = -1}$
$n = 500$	0.913	0.442	0.053
$n = 1000$	0.930	0.859	0.218
$n = 2000$	0.969	0.931	0.704
	$\boldsymbol{\beta_1^\pi = 1.2}$	$\boldsymbol{\beta_1^\pi = 1.4}$	$\boldsymbol{\beta_1^\pi = 1.6}$
$n = 500$	0.030	0.043	0.044
$n = 1000$	0.182	0.229	0.248
$n = 2000$	0.472	0.521	0.545
Parameters in the outcome equation			
	$\boldsymbol{\beta_2^1 = 1.2, \beta_2^2 = -1}$	$\boldsymbol{\beta_2^1 = 1.4, \beta_2^2 = -1.4}$	$\boldsymbol{\beta_2^1 = 1.9, \beta_2^2 = -1.6}$
$n = 500$	0.050	0.134	0.295
$n = 1000$	0.379	0.608	0.820
$n = 2000$	0.811	0.860	0.871

- Bandwidths were computed using $c_h = 1$ and $c_b = 0.001$.
- In each case, the remaining parameters were set at their true values.
- True parameter values: $\beta_{10}^1 = 0.125$, $\beta_{10}^2 = -0.25$, $\beta_{10}^\pi = 0.40$, $\beta_{20}^1 = 0.40$, $\beta_{20}^2 = -0.50$.

Our test showed good power properties, particularly for moderately large sample sizes ($n = 1000$ and above). Sufficiently large perturbations or perturbations with where parameters had the wrong sign were rejected with probability as high as 91% even for samples of size $n = 500$. In line with our previous results, power varied across different subsets of parameters, but for samples of size $n = 2000$ (a realistic sample size in many applications), every parameter value analyzed was rejected with probability at least 50%. Once again, we are particularly encouraged by the fact that our procedure was informative for β_1^π, the coefficient of the unobserved control variable in our experiments.

5 General features of econometric models suitable to our approach

Our bivariate sample selection model, and the two variations described in Section 2.3, are special cases of a class of models that can be described as follows. We have a collection of R observable, scalar random variables $(Y^r)_{r=1}^R$, and we have a pair of random variables (X, V) such that, letting F denote the underlying DGP, these models predict that,

$$E_F[Y^r|X,V] = S^r(V,\beta_0) + \phi_F^r(g(X,\beta_0)), \quad r = 1,\ldots,R. \tag{34}$$

$\beta_0 \in B$ is a finite-dimensional parameter, which is the object of interest to the econometrician. The function $S^r(\cdot)$ has a known, parametric functional form[10] but $\phi_F^r(\cdot)$ (which can be a functional of F in these models) is not necessarily known, except for monotonicity features that will be summarized in equation (35), below. $g(X,\beta_0)$ is a vector of control functions. Some elements of X may be included in V, but *at least a subset of the control variables X is unobserved*, so the control function $g(X,\beta_0)$ is unobservable even if β_0 were known. Group $Y \equiv (Y^1,\ldots,Y^R)$. Our data consists of a random sample $(Y_i,V_i)_{i=1}^n$. The properties of $\phi_F^r(\cdot)$ and V in these models are such that, for each r, there exists a pair of (possibly vector-valued) functions[11] $m_1^r(V,\beta_0)$ and $m_2^r(V,\beta_0)$, with known parametric functional form[12], such that, for any pair of observations $i \neq j$ in our data,

$$m_1^r(V_i,\beta_0) \geq m_2^r(V_j,\beta_0) \implies \phi_F^r(g(X_i,\beta_0)) \leq \phi_F^r(g(X_j,\beta_0)) \text{ w.p.1.} \tag{35}$$

The functions $m_1^r(V,\beta_0)$ and $m_2^r(V,\beta_0)$ consist of bounds for components of the unobserved control function $g(X,\beta_0)$. Let $\mu_{rF}(V) \equiv E_F[Y^r|V]$. From (34), $\mu_{rF}(V) = S^r(V,\beta_0) + E_F[\phi_F^r(g(X,\beta_0))|V]$. Thus, combining (34) and (35), for any pair of observations $i \neq j$ in the data, we have

$$\Big(\big(\mu_{rF}(V_i) - S^r(V_i,\beta_0)\big) - \big(\mu_{rF}(V_j) - S^r(V_j,\beta_0)\big)\Big)\mathbb{1}\Big\{m_2^r(V_j,\beta_0) \leq m_1^r(V_i,\beta_0)\Big\} \leq 0 \quad \text{w.p.1.} \tag{36}$$

From the functional inequality in (36), we can conduct inference adapting the approach we studied in Section 3. As we did in our analysis, we can focus on density-weighted functionals,

$$\begin{aligned}
&\tau_{rF}(V_i,V_j,\beta) \equiv \\
&\Big(\big(\mu_{rF}(V_i) - S^r(V_i,\beta)\big) - \big(\mu_{rF}(V_j) - S^r(V_j,\beta)\big)\Big)\mathbb{1}\Big\{m_2^r(V_j,\beta) \leq m_1^r(V_i,\beta)\Big\} \cdot f_V(V_i) f_V(V_j) \cdot \phi_r(V_i)^2 \phi_r(V_j)^2 \\
&\text{for } r = 1,\ldots,R.
\end{aligned}$$

Where each $\phi_r(\cdot)$ is a weight function which is nonzero over a pre-specified inference range $\mathcal{V} \subseteq \text{Supp}(V)$ and $f_V(\cdot)$ denotes the density function of V. As we defined in (12), let

$$\mathcal{T}_{rF}(\beta) \equiv E_F\Big[\Big(\tau_{rF}(V_i,V_j,\beta)\Big)_+\Big], \quad \text{for } r = 1,\ldots,R.$$

The functional inequalities in (36) imply that $\mathcal{T}_{rF}(\beta_0) = 0$ for each r. As we did in our model, we can aggregate these functionals as,

$$\mathcal{T}_F(\beta) \equiv \sum_{r=1}^R \mathcal{T}_{rF}(\beta)$$

[10] We can extend our framework to cases where $S^r(v,\beta)$ does not have a known functional form but can be consistently estimated in the data for any v in our inference range and any β in our parameter space.

[11] Note that in the version of our model described in Section 2.3.1, the bounds $m_1(V,\beta_0)$ and $m_2(V,\beta_0)$ are vector-valued.

[12] Our approach can be potentially extended to cases where the functional forms of $m_1^r(v,\beta)$ and $m_2^r(v,\beta)$ are unknown but can be consistently estimated in the data for any v in our inference range and any β in our parameter space.

We can assign different positive weights to each one of the inequalities in the above sum, or we could aggregate them in other ways. By construction, $\mathcal{T}_F(\beta) \geq 0$ for all β and $\mathcal{T}_F(\beta_0) = 0$. Inference can proceed from here by adapting the steps and the assumptions in Section 3 to the specific features of the model. Our previous discussion and the details of the proofs in the online Appendix provide a useful roadmap of how to proceed.

5.1 More general models

The general features in equations (34) and (35) can encompass many models with sample selectivity and/or endogeneity, including multiple types of Tobit models or game-theoretic models, among others. We can extend and adapt our approach to an even wider class of models, for example, by relaxing the additive separability in $S^r(V, \beta_0)$ and $\phi_F^r(g(X, \beta_0))$ in equation (34). However, any model suitable to an adaptation of the method proposed in this paper must satisfy the following restrictions.

1.– We must have bounds for the unobserved control functions that depend on observable covariates. These bounds must have either a known functional form (up to a finite-dimensional parameter) or they must be estimable (for any given parameter value).

2.– The way in which the control functions enter the structural model must satisfy some known (or assumed to be known) monotonicity conditions.

3.– These monotonicity conditions must be such that they produce inequalities which depend on observable covariates when we make pairwise comparisons of observations for which the bounds are disjoint. These inequalities would be the basis for inference.

The development of inferential methods for models that fail to satisfy either of the three features described above is beyond the scope of this paper.

6 Concluding remarks

Control function methods to estimate models with sample selection or endogeneity in have been studied in multiple settings. The general approach proposed by James L. Powell and coauthors consists of making pairwise comparisons in the data based on matching (asymptotically) these control functions. Conditional on the matching, these models produce moment conditions that allow us to identify and estimate the parameter of interest. In a number of applications, control functions may be unobserved because at least a subset of control variables is unobservable. This makes pairwise matching impossible. However, in many cases, we may have bounds for the unobserved control functions which depend on observables. These bounds may result from the presence of interval data or they may be obtained from economic theory. This paper shows that if these bounds are available, and if the way the control functions enter the econometric model satisfies monotonicity conditions, inference can still proceed by making pairwise comparisons. In this case, the "matching" is based on identifying pairs of observations for which the bounds for the control functions are disjoint. Focusing on a bivariate sample selection model, we showed that these pairwise comparisons produce functional inequalities that depend on observable covariates and on the structural parameter of interest. Using this result, we proposed an inferential procedure that embeds all the information of the functional inequalities, while adapting asymptotically to the measure of the contact sets of values of the conditioning variables at which the inequalities are binding. Regularizing the estimator for the asymptotic variance of our statistic allows us to standardize it in a way that produces asymptotically pivotal properties. From here, a confidence set (CS) for the true parameter value was constructed. We described conditions under which our CS has uniform asymptotic coverage properties, satisfies consistency and has nontrivial local power. Our Monte Carlo results showed that our procedure has finite-sample properties aligned with its asymptotic predictions. Finally, we provided a general characterization of models suitable to our approach. The existence of observable bounds for the control functions, and the presence of monotonicity features for the way in which these control functions shift the structural model are basic features of any such model.

References

Ahn, H., H. Ichimura, J. Powell, and P. Ruud (2018). Simple estimators for invertible index models. *Journal of Business and Economic Statistics 36*(1), 1–10.

Ahn, H. and J. Powell (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics 58*, 3–29.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Andrews, D. (1994). Empirical process methods in econometrics. In R. Engle and D. McFadden (Eds.), *The Handbook of Econometrics*, Volume 4, Chapter 37, pp. 2247–2294. North-Holland.

Andrews, D. W. K. and X. Shi (2013). Inference for parameters defined by conditional moment inequalities. *Econometrica 81*(2), 609–666.

Andrews, D. W. K. and G. Soares (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica 78*(1), 119–157.

Aradillas-López, A., B. Honoré, and J. Powell (2007). Pairwise difference estimation with nonparametric control variables. *International Economic Review 48*(4), 1119–1158.

Aradillas-López, A. and A. Rosen (2021). Inference in ordered response games with complete information. *Journal of Econometrics 226*(2), 451–476.

Aradillas-López, A. and E. Tamer (2008). The identification power of equilibrium in simple games. *Journal of Business and Economic Statistics 26*(3), 261–310.

Armstrong, T. B. (2014). Weighted ks statistics for inference on conditional moment inequalities. *Journal of Econometrics 181*(2), 92–116.

Armstrong, T. B. (2015). Asymptotically exact inference in conditional moment inequality models. *Journal of Econometrics 186*(1), 51–65.

Armstrong, T. B. (2018). On the choice of test statistic for conditional moment inequalities. *Journal of Econometrics 203*(1), 241–255.

Armstrong, T. B. and H. P. Chan (2016). Multiscale adaptive inference on conditional moment inequalities. *Journal of Econometrics 194*(1), 24–43.

Cavanagh, C. and R. Sherman (1998). Rank estimators for monotonic index models. *Journal of Econometrics 84*(2), 351–381.

Chamberlain, G. (1984). Panel data. In Z. Griliches and M. Intriligator (Eds.), *The Handbook of Econometrics*, Volume 2, pp. 1247–1318. North-Holland.

Chernozhukov, V., S. Lee, and A. Rosen (2013). Intersection bounds, estimation and inference. *Econometrica 81*(2), 667–737.

Chetverikov, D. (2017). Adaptive test of conditional moment inequalities. *Econometric Theory 34*(1), 186–227.

Dudley, R. (1984). A course on empirical processes. *Ecole d'Été de Probabilitiés de Saint-Flour, XII-1982. Lecture Notes in Math 1097*, 1–142.

Han, A. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics 35*(2), 303–316.

Honoré, B. and J. Powell (1994). Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics 64*(2), 241–278.

Honoré, B. and J. Powell (2005). Pairwise difference estimators for nonlinear models. In D. Andrews and J. Stock (Eds.), *Identification and Inference for Econometric Models*, pp. 520–553. Cambridge.

Juster, F. T. and R. Suzman (1995). An overview of the health and retirement study. *The Journal of Human Resources 30*, S7–S56.

Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

Lee, S., K. Song, and Y.-J. Whang (2013). Testing functional inequalities. *Journal of Econometrics 172*(1), 14–32.

Lee, S., K. Song, and Y.-J. Whang (2018). Testing for a general class of functional inequalities. *Econometric Theory* (34), 1018–1064.

Lehmann, E. and J. Romano (2005). *Testing Statistical Hypotheses*. Springer.

Manski, C. F. and E. Tamer (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica 70*(2), 519–546.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995). *Microeconomic Theory*. Oxford University Press.

Nolan, D. and D. Pollard (1987). U-processes: Rates of convergence. *Annals of Statistics 15*, 780–799.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer Verlag.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics.

Powell, J. (1986). Symmetrically trimmed least squares estimation for tobit models. *Econometrica 54*(6), 1435–1460.

Powell, J., J. Stock, and T. Stoker (1989). Semiparametric estimation of index coefficients. *Econometrica 57*, 1403–1430.

Powell, J. L. (1987). Semiparametric estimation of bivariate latent variable models. SSRI Working Paper No. 8704.

Powell, J. L. (2001). Semiparametric estimation of censored selection models. In C. Hsiao, K. Morimune, and J. Powell (Eds.), *Nonlinear Statistical Modeling*. Cambridge University Press.

Robinson, P. (1988). Root-n consistent semiparametric regression. *Econometrica 56*, 931–954.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley. New York, NY.

Sherman, R. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica 61*(1), 123–137.

Sherman, R. (1994). Maximal inequalities for degenerate u-processes with applications to optimization estimators. *Annals of Statistics 22*, 439–459.

Silverman, B. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall/CRC.