Inference in models with partially identified control functions

Andres Aradillas-Lopez*

September 15, 2023

Abstract

In multiple contributions to the literature, James L. Powell and coauthors have developed estimators for semiparametric 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. However, there exist instances where the control functions are unobserved, but we have bounds for them which depend on observable covariates. These bounds can arise directly from the nature of the data available (e.g, with interval data), or they can be derived 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 pairwise functional inequalities. We propose an inferential procedure based on these pairwise inequalities and we analyze its properties.

Keywords: Semiparametric models, 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 "differences out" sources of bias such as selection correction terms, or other sources of endogeneity. Conditional on this matching, a pairwise 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), Blundell and Powell (2004), Honoré and Powell (2005), Aradillas-López, Honoré, and Powell (2007), and Ahn, Ichimura, Powell, and Ruud (2018).

In some instances, not all of the control variables are observable, making it impossible to implement the matching in Powell's methods. In a number of such instances, the nature of the data observed or an auxiliary economic theory may produce *bounds* for these control functions which can be expressed in terms of 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 pairwise, *conditional functional inequalities*. Based on these inequalities, we present an inferential procedure and we describe its properties. Similar to the usual pairwise-difference procedures available when all control variables are observed, our approach eliminates the need to directly estimate nonparametric correction terms, which could be a challenging task when control variables are unobservable. Our working example will be a semiparametric, bivariate sample selection model where some regressors of the selection equation are unobserved, but bounds for the resulting control functions can be characterized in terms of observable covariates, either through the presence of interval-data, or through an auxiliary economic theory.

The paper proceeds as follows. Section 2 describes the bivariate sample selection model that serves as the working example of the paper. We analyze two cases where a subset of control variables in the selection equation are unobservable. The first case is the most natural motivation for our model, and it corresponds to the presence of interval data in the selection equation. By its nature, interval data directly produces bounds for the unobserved regressors. The second case shows that, in some instances, bounds for unobserved control variables can be obtained from an economic model, and these bounds can be expressed as functions of observables. We illustrate this in a model where unobserved beliefs (expectations) by economic agents are among the regressors in the selection equation. We characterize bounds for the resulting control function using different economic models of rationality. In all our examples we show that monotonicity restrictions for some unknown functionals of the model produce pairwise functional inequalities that must be satisfied at the true value of the parameters. Section 3 focuses on a general version of our model, and describes an inferential procedure where we construct a confidence set for the parameters of the model based on the pairwise functional inequalities produced by our restrictions. Section 4 presents Monte Carlo experiments where we implement our approach on designs that involve interval data. Section 5 discusses alternative scenarios to our general model where control functions may also be partially identified. Section 6 concludes. The online appendix includes the proofs of our results, along with additional extensions of our main model.

2 A bivariate sample selection model with censored data

The working example for this paper will be a semiparametric, bivariate sample selection model with censored data. The model consists of a scalar outcome Y_2^*, which is observed if and only if $Y_1^* > 0$, where Y_1^* is a scalar latent variable. We will define

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

As usual, the model specifies that Y_2 is observed (and is equal to Y_2^*) if and only if $Y_1^* > 0$, and otherwise it does not take on any meaningful value. For simplicity, we can code $Y_2 = 0$ to signify the event that $Y_1^* \leq 0$. We assume an outcome equation of the form,

$$Y_2^* = X_2'\beta_{20} + \varepsilon_2, \tag{1}$$

where ε_2 is a latent variable. We will consider different specifications of the selection equation (describing Y_1^*), focusing on situations where some regressors in the selection equation are unobserved. Our first example –and the most natural motivation for our model– is the case where we have interval data in the selection equation. The second example will show that, in the absence of interval data, bounds for unobserved control variables can sometimes be obtained from an auxiliary economic model.

2.1 A model with interval data in the selection equation

Suppose the selection equation is given by,

$$Y_1^* = X_1'\beta_{10} + \varepsilon_1$$

where ε_1 is a latent variable. $X_1 \in \mathbb{R}^{d_1}$ are the regressors of the selection equation. Let us express $X_1 = (X_1^1, \ldots, X_1^{d_1})$. We will assume that there exists at least one regressor $X_1^\ell \in X_1$ that is unobserved, but for which we have observable covariates $(\underline{X}_1^\ell, \overline{X}_1^\ell)$ such that $\underline{X}_1^\ell \leq X_1^\ell \leq \overline{X}_1^\ell$ w.p.1. That is, we only observe interval data for at least one regressor $X_1^\ell \in X_1$. For all the elements $X_1^j \in X_1$

that are observable, we will decree $\underline{X}_1^j = \overline{X}_1^j = X_1^j$. This will allow us to present definitions in terms of $(\underline{X}_1^\ell, \overline{X}_1^\ell)_{\ell=1}^{d_1}$. Interval data is prevalent in many econometric data sets[1], and it has received attention in econometric work. A notable example is Manski and Tamer (2002), who examine inference for the regression functions $E[Y|X,U]$ and $E[U|X]$ when U is an unobserved scalar random variable, and the econometrician observes (Y,X,U_0,U_1), with $U_0 \leq U \leq U_1$ w.p.1. To our knowledge, this paper is the first one to explicitly consider interval data in a multiple-equation semiparametric setting with control functions. Denote,

$$\beta_1' X_{1L} \equiv \min_{\xi \equiv (\xi^1,\ldots,\xi^{d_1})} (\beta_1' \xi) : \xi^\ell \in \{\underline{X}_1^\ell, \overline{X}_1^\ell\} \text{ for } \ell = 1,\ldots,d_1$$
$$\beta_1' X_{1U} \equiv \max_{\xi \equiv (\xi^1,\ldots,\xi^{d_1})} (\beta_1' \xi) : \xi^\ell \in \{\underline{X}_1^\ell, \overline{X}_1^\ell\} \text{ for } \ell = 1,\ldots,d_1$$

If we express $X_{1L} \equiv (X_{1L}^1,\ldots,X_{1L}^{d_1})$ and $X_{1U} \equiv (X_{1U}^1,\ldots,X_{1U}^{d_1})$, then for any given β_1,

$$X_{1L}^\ell = \mathbb{1}\{\beta_1^\ell \geq 0\} \cdot \underline{X}_1^\ell + \mathbb{1}\{\beta_1^\ell < 0\} \cdot \overline{X}_1^\ell \quad \text{and} \quad X_{1U}^\ell = \mathbb{1}\{\beta_1^\ell \geq 0\} \cdot \overline{X}_1^\ell + \mathbb{1}\{\beta_1^\ell < 0\} \cdot \underline{X}_1^\ell.$$

Note that X_{1L} and X_{1U} depend on (the signs of) β_1, but we omit this dependence in our notation to simplify the exposition[2]. By definition, we have

$$\beta_1' X_{1L} \leq \beta_1' X_1 \leq \beta_1' X_{1U} \ \forall\ \beta_1, \text{ w.p.1.} \tag{2}$$

Group $\underline{X}_1 \equiv (\underline{X}_1^1,\ldots,\underline{X}_1^{d_1})$ and $\overline{X}_1 \equiv (\overline{X}_1^1,\ldots,\overline{X}_1^{d_1})$, and let $W_1 \equiv \underline{X}_1 \cup \overline{X}_1$, $V \equiv X_2 \cup W_1$. We will let F denote the underlying distribution that generated $(X_1,V,\varepsilon_1,\varepsilon_2)$, and we will assume that we observe an iid sample $(Y_{1i},Y_{2i},V_i)_{i=1}^n$ generated by F.

Assumption 1 ***(An exclusion and monotonicity restrictions)***

(i) $(\varepsilon_1,\varepsilon_2) \perp (X_1,V)$, *and* $E_F[\varepsilon_2|\varepsilon_1 > c]$ *is well-defined, and nondecreasing in* c *for all* $c \in \mathbb{R}$. *Thus,*

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

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

(ii) $P_F(Y_1 = 1|W_1,X_1) = H_F(X_1'\beta_{10})$, *where* $H_F(\cdot)$ *is unknown but assumed to be nondecreasing.* ■

[1]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]Our notation can be simplified if the signs of β_{10} are assumed to be known ex-ante for the interval-data regressors. However, since this knowledge is not required for our results, we focus on the general case where these signs are not necessarily known ex-ante.

Part (i) presupposes a type of positive stochastic relationship between ε_1 and ε_2 where larger realizations of ε_1 produce larger expected values for ε_2. 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. 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 monotonicity restriction assumed for $\lambda_F(\cdot)$ is stronger than what is typically assumed in existing semiparametric versions of this model[3], which rely only on the exclusion restriction $(\varepsilon_1, \varepsilon_2) \perp (X_1, V)$ that we also imposed (see, e.g, Ahn and Powell (1993)). However, those models rely crucially on the requirement that X_1 is observed. In the presence of interval data for X_1, the additional monotonicity property assumed for $\lambda_F(\cdot)$ will allow us to obtain testable implications for the model. Part (i) of Assumption 1 yields

$$E_F\left[Y_2 | X_1, X_2, Y_1 = 1\right] = X_2'\beta_{20} + \lambda_F\left(X_1'\beta_{10}\right). \tag{3}$$

If X_1 were observable, this would be a special case of the models studied by James L. Powell and coauthors, who have developed semiparametric methods using pairwise comparisons of observations (see Powell (1987), Ahn and Powell (1993), Honoré and Powell (1994), Powell (2001), Blundell and Powell (2004), Honoré and Powell (2005), Aradillas-López, Honoré, and Powell (2007), and Ahn, Ichimura, Powell, and Ruud (2018)). In our case, the pairwise comparisons proposed by Powell's methods would be based on matching[4] $X_{1i}'\beta_{10}$ and $X_{1j}'\beta_{10}$. Group $X \equiv X_1 \cup X_2$, and suppose we have a random sample $(Y_{1i}, Y_{2i}, X_i)_{i=1}^n$ (i.e, X_1 is observable). Take any pair of observations (Y_{1i}, Y_{2i}, X_i) and (Y_{1j}, Y_{2j}, X_j). From (3), we have

$$E_F\left[\left(Y_{2i} - X_{2i}'\beta_{20}\right) - \left(Y_{2j} - X_{2j}'\beta_{20}\right) \,\Big|\, Y_{1i} = Y_{1j} = 1, X_{1i}'\beta_{10} = X_{1j}'\beta_{10}\right] = 0 \quad \text{a.s.} \tag{4}$$

Using existing terminology, $X_1'\beta_{10}$ is a control function that allows us to deal with the selectivity in the outcome equation without having to parameterize or directly estimate $\lambda_F(\cdot)$. Estimation of the slope coefficients[5] of the model could proceed by estimating β_{10} in a first step, and plugging the estimator in a second step to estimate β_{20}, based on the pairwise moment condition (4). The estimator would minimize a kernel-weighted U-statistic, and its asymptotic properties can be derived from the results in Honoré and Powell (2005).

When X_1 is unobserved, the matching in (3) is no longer feasible. Here is where the mono-

[3]In Section 5 we discuss alternative scenarios where control function parameters are partially identified and the monotonicity restrictions for $\lambda_F(\cdot)$ and $H_F(\cdot)$ can be dropped.

[4]If X_1 is observable, we could leave the selection equation nonparametrically specified and simply base our pairwise comparisons on matching X_1 directly. Aggregating X_1 into a lower-dimensional parametric index $X_1'\beta_{10}$ can still be desirable to mitigate the curse of dimensionality. See Honoré and Powell (2005) or Aradillas-López, Honoré, and Powell (2007) for details.

[5]Since $\lambda_F(\cdot)$ is nonparametrically specified in (3), an intercept cannot be identified in the outcome equation. This is a general feature of partially linear models (see Robinson (1988)). Note that an intercept will be differenced out in (4).

tonicity restriction for $\lambda_F(\cdot)$ in Assumption 1 becomes relevant when we have interval data. Recall that $W_1 \equiv \underline{X}_1 \cup \overline{X}_1$ and $V \equiv X_2 \cup W_1$. Let $\mu_{2F}(V) \equiv E_F[Y_2|V, Y_1 = 1]$. By Assumption 1,

$$\mu_{2F}(V) = X_2'\beta_{20} + E_F[\lambda_F(X_1'\beta_{10})|V]. \tag{5}$$

Since $\lambda_F(\cdot)$ is non-increasing, the bounds in (2) imply $\lambda_F(X_{1U}'\beta_{10}) \leq \lambda_F(X_1'\beta_{10}) \leq \lambda_F(X_{1L}'\beta_{10})$ w.p.1. Therefore,

$$\lambda_F(X_{1U}'\beta_{10}) \leq E_F[\lambda_F(X_1'\beta_{10})|V] \leq \lambda_F(X_{1L}'\beta_{10}) \quad \text{w.p.1.} \tag{6}$$

The only restriction we have imposed on $\lambda_F(\cdot)$ is that it is nondecreasing. Thus, without further restrictions, the bounds described in (6) for $E_F[\lambda_F(X_1'\beta_{10})|V]$ are sharp and cannot be improved upon without additional assumptions. Combining (5) and (6),

$$\lambda_F(X_{1U}'\beta_{10}) \leq \mu_{2F}(V) - X_2'\beta_{20} \leq \lambda_F(X_{1L}'\beta_{10}) \quad \text{w.p.1.} \tag{7}$$

Since the bounds in (6) are sharp, so are those in (7). Bounds for other functionals[6] could be obtained if we impose additional (or alternative) restrictions on the joint distribution of $(\varepsilon_1, \varepsilon_2)$. For example, for any given (c_1, c_2) in $\mathbb{R}^2$, let $G_{2|1,F}(c_2|c_1) \equiv P_F(\varepsilon_2, \leq c_2|\varepsilon_1 > c_1)$. Suppose we assume that $G_{2|1,F}(\cdot|c_1) \leq G_{2|1,F}(\cdot|c_1') \;\forall\; c_1 > c_1'$. This first-order stochastic dominance (FOSD) assumption is stronger than our monotonicity restriction in part (i) of Assumption 1. Maintain the assumption that $(\varepsilon_1, \varepsilon_2) \perp (X_1, V)$ (as in part (i) of Assumption 1). Using iterated expectations and the bounds in (2), this FOSD assumption would yield,

$$G_{2|1,F}(c - X_2'\beta_{20}| - X_{1L}'\beta_{10}) \leq P_F(Y_2 \leq c|V, Y_1 = 1) \leq G_{2|1,F}(c - X_2'\beta_{20}| - X_{1U}'\beta_{10}) \;\; \forall\; c \text{ w.p.1.}$$

Similarly, we could obtain bounds for additional functionals if we assumed monotonicity restrictions about $Var_F[\varepsilon_2|\varepsilon_1 > c]$, etc.

Let us go back to our setup, and turn our attention to the selection equation. Let $\mu_{1F}(W_1) \equiv E_F[Y_1|W_1]$. From the exclusion restriction in Assumption 1, we have $\mu_{1F}(W_1) = E_F[H_F(X_1'\beta_{10})|W_1]$. The monotonicity properties of $H_F(\cdot)$ and the bounds in (2) yield,

$$H_F(X_{1L}'\beta_{10}) \leq \mu_{1F}(W_1) \leq H_F(X_{1U}'\beta_{10}). \tag{8}$$

The only restriction we have imposed on $H_F(\cdot)$ is that it is nondecreasing. Thus, without further restrictions, the bounds described in (8) for $\mu_{1F}(W_1)$ are sharp and cannot be improved upon without additional assumptions. Furthermore, since Y_1 is a binary random variable, $\mu_{1F}(W_1)$

[6]Note, for example that, without additional restrictions, we can have $E_F[\lambda_F(X_1'\beta_{10})|X_2, X_{1L}] \gtreqless E_F[\lambda_F(X_1'\beta_{10})|X_2, X_{1U}]$ in our model. Therefore, under our assumptions we can have $E_F[Y_2|X_2, X_{1L}, Y_1 = 1] \gtreqless E_F[Y_2|X_2, X_{1U}, Y_1 = 1]$. Thus, our assumptions do not produce any predictions (e.g, inequalities) for the relationship between $E_F[Y_2|X_2, X_{1L}, Y_1 = 1]$ and $E_F[Y_2|X_2, X_{1U}, Y_1 = 1]$.

fully describes its distribution, so (8) describes sharp bounds for the only relevant functional of Y_1 conditional on observables. Note that our assumptions do not yield any prediction for the relationship between $E_F[Y_1|X_{1L}]$ and $E_F[Y_1|X_{1U}]$. Without additional restrictions, we can have $E_F[H_F(X_1'\beta_{10})|X_{1L}] \gtreqless E_F[H_F(X_1'\beta_{10})|X_{1U}]$, and therefore we can have $E_F[Y_1|X_{1L}] \gtreqless E_F[Y_1|X_{1U}]$, see also footnote 6.

2.1.1 Pairwise-comparison testable implications derived from our functional inequalities

A natural way to exploit the inequalities in (7) and (8) is through pairwise comparisons across observations. However, unlike the usual pairwise comparisons that match pairs of observations based on the control functions when these are observable (see equation 4), our pairwise comparisons would now "match" pairs of observations for which the bounds in (7) and (8) are *disjoint*. Recall that $W_1 \equiv \underline{X}_1 \cup \overline{X}_1$ and $V \equiv X_2 \cup W_1$. Let (V_i, V_j) be independent draws from F and suppose $X_{1Uj}'\beta_{10} \le X_{1Li}'\beta_{10}$. Since $\lambda_F(\cdot)$ is assumed to be nonincreasing, this implies $\lambda_F(X_{1Uj}'\beta_{10}) \ge \lambda_F(X_{1Li}'\beta_{10})$. From the bounds in (7), this in turn implies, $\mu_{2F}(V_i) - X_{2i}'\beta_{20} \le \mu_{2F}(V_j) - X_{2j}'\beta_{20}$. And, since $H_F(\cdot)$ is assumed to be nondecreasing, the bounds in (8) yield $\mu_{1F}(W_{1j}) \le \mu_{1F}(W_{1i})$. Thus, under Assumption 1, we must have

$$\left.\begin{array}{r}\Big(\big(\mu_{2F}(V_i) - X_{2i}'\beta_{20}\big) - \big(\mu_{2F}(V_j) - X_{2j}'\beta_{20}\big)\Big)\mathbb{1}\left\{X_{1Uj}'\beta_{10} \le X_{1Li}'\beta_{10}\right\} \le 0 \\ \big(\mu_{1F}(W_{1j}) - \mu_{1F}(W_{1i})\big)\cdot \mathbb{1}\left\{X_{1Uj}'\beta_{10} \le X_{1Li}'\beta_{10}\right\} \le 0\end{array}\right\} \text{ w.p.1.} \tag{9}$$

From our previous arguments, the pairwise inequalities are derived from sharp bounds for the functionals involved, and cannot be improved upon without additional restrictions. Note that the characterization of our pairwise inequalities bypasses the need to estimate the nonparametric functionals λ_F and H_F, which would require additional assumptions[7] (see, e.g, Das, Newey, and Vella (2003)). This is also an attractive feature of the pairwise-difference methods that are available when control variables are observable.

2.1.2 Identification power of the pairwise inequalities in (9)

In this paper we will propose an inferential procedure for (β_{10}, β_{20}) based on the type of pairwise inequalities described in (9). While point-identification in models with conditional moment inequalities may be possible (see, e.g, Khan and Tamer (2009)), our focus will be on the construction of a confidence set (CS) for (β_{10}, β_{20}). Our methodology will not assume that the pairwise inequalities in (9) point-identify (β_{10}, β_{20}), and the asymptotic properties of our CS will be valid with or without point-identification. However, it is useful to study whether point-identification can be

[7] In Section 5 we advocate to estimate λ_F and H_F when only a lower or an upper bound (but not both) is available for our control function. When lower and upper bounds are available (the main focus of our paper), our pairwise inequalities provide a way to do inference on the parameters of the model without having to estimate these functions.

obtained from our pairwise inequalities, and we do so in this section. We begin by noting that the inequalities in (9) are invariant to the presence of an intercept in β_{20} or β_{10}, as well as any re-scaling of β_{10} by any scalar $c > 0$. Accordingly, our relevant parameter space will include only the slope coefficients in both equations, and the scale of β_{10} will be normalized. These transformations of the parameter space are owed to the nonparametric treatment of λ_F and H_F, and they would still be necessary if the control variables were observed (see footnote 5). To normalize the scale of β_1, we will assume that there exists a regressor $X_1^{\ell} \in X_1$ whose coefficient will be fixed to $\beta_1^{\ell} = \pm 1$ for all $\beta_1 \in \Theta$. Thus, for the rest of this discussion, the parameter space Θ will include only the slope coefficients in β_2 and β_1, and it will fix $\beta_1^{\ell} = \pm 1$ for all $\beta_1 \in \Theta$. Given this transformation of Θ, we will describe a set of conditions under which our pairwise inequalities in (9) can point-identify the parameters of the model.

Recall that $W_1 \equiv \underline{X}_1 \cup \overline{X}_1$ and $V \equiv X_2 \cup W_1$. As we did above, let (V_i, V_j) be independent draws from F. Immediately, we can see that any (β_1, β_2) such that $P_F(X'_{1Uj}\beta_1 \leq X'_{1Li}\beta_1) = 0$ would satisfy the inequalities in (9) regardless of the value of β_2. Thus, any hope to point-identify of (β_{10}, β_{20}) through the inequalities in (9) requires that $P_F(X'_{1Uj}\beta_{10} \leq X'_{1Li}\beta_{10}) > 0$. Let $Supp(\xi)$ denote the support of the marginal distribution ξ, and let $Supp(\xi|\psi)$ denote the support of the distribution of ξ conditional on ψ. Let X_1^{ℓ} be the regressor whose coefficient β_1^{ℓ} is normalized to ± 1. Recall that if X_1^{ℓ} is observable, then $\underline{X}_1^{\ell} = \overline{X}_1^{\ell} = X_1^{\ell}$. Consider the following restriction,

(i) *The marginal distributions of both $\underline{X}_1^{\ell}$ and $\overline{X}_1^{\ell}$ are absolutely continuous with respect to Lebesgue measure conditional on $W_1 \setminus (\underline{X}_1^{\ell}, \overline{X}_1^{\ell})$. Therefore, for each $\beta_1 \in \Theta$, the marginal distributions of both $X'_{1L}\beta_1$ and $X'_{1U}\beta_1$ are absolutely continuous with respect to Lebesgue measure.*

(ii) $\forall\ \beta_1, \widetilde{\beta}_1 \in \Theta : \widetilde{\beta}_1 \neq \beta_1,\ \ Supp(X'_{1L}\beta_1 | X'_{1L}\widetilde{\beta}_1, X'_{1U}\widetilde{\beta}_1) \cap Supp(X'_{1U}\beta_1 | X'_{1L}\widetilde{\beta}_1, X'_{1U}\widetilde{\beta}_1)$ *has nonzero Lebesgue measure.*

(R1)

Intuitively, the above restriction requires that the marginal supports of the lower and upper bounds for the control function have a nonempty overlap. An immediate case where this would happen is when each of these bounds has an unbounded support. Note that (R1) does not assign positive probability to the event that the lower and upper bounds are equal to each other; it is a statement about the comparisons between the marginal (not joint) supports of the bounds. If (R1) holds, we have

$$P_F\Big(X'_{1Li}\beta_1 < X'_{1Uj}\beta_1 < X'_{1Li}\beta_1 + \varepsilon\ ,\ X'_{1Uj}\widetilde{\beta}_1 \leq X'_{1Li}\widetilde{\beta}_1\Big) > 0 \quad \forall\ \beta_1, \widetilde{\beta}_1 \in \Theta\colon \beta_1 \neq \widetilde{\beta}_1,\ \forall\ \varepsilon > 0. \tag{10}$$

The result in (10) can be the first building block towards point-identification of (β_{10}, β_{20}) through the inequalities in (9), but additional conditions are needed. Consider the following restriction,

$$\begin{aligned}
&(i)\ H_F(\cdot) \textit{ is a strictly increasing function over } \mathbb{R}\\
&(ii)\ \forall\ \varepsilon > 0 \textit{ and } F-a.e\ (X_{1L}'\beta_{10}, X_{1U}'\beta_{10}),\\
&\quad P_F\Big(\mu_{1F}(W_1) > H_F(X_{1U}'\beta_{10}) - \varepsilon\ \Big|\ X_{1L}'\beta_{10}, X_{1U}'\beta_{10}\Big) > 0,\\
&\qquad\qquad\qquad\qquad \textit{and}\\
&\quad P_F\Big(\mu_{1F}(W_1) < H_F(X_{1L}'\beta_{10}) + \varepsilon\ \Big|\ X_{1L}'\beta_{10}, X_{1U}'\beta_{10}\Big) > 0
\end{aligned} \tag{R2}$$

Part (i) strengthens the monotonicity conditions in Assumption 1 for H_F. Part (ii) assumes that, for F−a.e $(X_{1L}'\beta_{10}, X_{1U}'\beta_{10})$, the support of $\mu_{1F}(W_1)$ spans the entire interval $[H_F(X_{1L}'\beta_{10}),\ H_F(X_{1U}'\beta_{10})]$. This is a special (stronger) case of the inequalities in (8). Combined with (R1), the condition in (R2) yields,

$$P_F\Big(\mu_{1F}(W_{1i}) < \mu_{1F}(W_{1j})\ ,\ X_{1Uj}'\beta_1 \leq X_{1Li}'\beta_1\Big) > 0 \quad \forall\ \beta_1 \in \Theta : \beta_1 \neq \beta_{10}. \tag{11}$$

If (11) holds, the selection-equation parameters β_{10} can be identified from the second inequality in (9). Identification of the entire parameter vector (β_{10}, β_{20}) requires restrictions involving the first inequality in (9). Consider the following restriction,

$$\begin{aligned}
&(i)\ \lambda_F(\cdot) \textit{ is a strictly decreasing, Lipschitz-continuous function over } \mathbb{R}\\
&(ii)\ \forall\ \varepsilon > 0 \textit{ and } F-a.e\ (X_{1L}'\beta_{10}, X_{1U}'\beta_{10}),\\
&\quad P_F\Big(E_F[\lambda_F(X_1'\beta_{10})|V] > \lambda_F(X_{1L}'\beta_{10}) - \varepsilon\ \Big|\ X_{1L}'\beta_{10}, X_{1U}'\beta_{10}\Big) > 0,\\
&\qquad\qquad\qquad\qquad \textit{and}\\
&\quad P_F\Big(E_F[\lambda_F(X_1'\beta_{10})|V] < \lambda_F(X_{1U}'\beta_{10}) + \varepsilon\ \Big|\ X_{1L}'\beta_{10}, X_{1U}'\beta_{10}\Big) > 0
\end{aligned} \tag{R3}$$

Part (i) strengthens the shape restrictions in Assumption 1 for λ_F. Part (ii) states that, for F−a.e $(X_{1L}'\beta_{10}, X_{1U}'\beta_{10})$, the support of $E_F[\lambda_F(X_1'\beta_{10})|V]$ spans the entire interval $[\lambda_F(X_{1U}'\beta_{10}),\ \lambda_F(X_{1L}'\beta_{10})]$. This is a special (stronger) case of the inequalities in (6). Combined with (R1), the condition in (R3) yields,

$$\begin{aligned}
&P_F\Big(E_F[\lambda_F(X_{1i}'\beta_0)|V_i] > E_F[\lambda_F(X_{1j}'\beta_0)|V_j]\ ,\ X_{1Uj}'\beta_1 \leq X_{1Li}'\beta_1\Big) > 0 \quad \forall\ \beta_1 \in \Theta : \beta_1 \neq \beta_{10},\\
&P_F\Big(\big|E_F[\lambda_F(X_{1i}'\beta_0)|V_i] - E_F[\lambda_F(X_{1j}'\beta_0)|V_j]\big| < \varepsilon\ ,\ X_{1Uj}'\beta_{10} \leq X_{1Li}'\beta_{10}\Big) > 0 \quad \forall\ \varepsilon > 0.
\end{aligned} \tag{12}$$

The first result in (12) implies that (β_1, β_{20}) would violate the first inequality in (9) with strictly positive probability for any $\beta_1 \neq \beta_{10}$. Combined with an *exclusion restriction between* W_1 *and* X_2, the two results in (12) can lead to identification of β_{20}. Recall that $X_1^{\ell} \in X_1$ is the regressor whose coefficient β_1^ℓ has been normalized to $\beta_1^\ell = \pm 1$. Consider the following restriction,

$$\begin{aligned}&(i)\ (\underline{X}_1^\ell, \overline{X}_1^\ell) \notin X_2\\ &(ii)\ \forall\ \delta_2 \neq 0, \forall\ \beta_1 \in \Theta, \textit{ and } F\text{–}a.e\ W_1,\ Supp(X_2'\delta_2 \mid \beta_1' X_{1L}) \cap Supp(X_2'\delta_2 \mid \beta_1' X_{1U}) \textit{ is nonempty,}\\ &\quad \textit{and includes at least two points.}\end{aligned} \tag{R4}$$

If the condition in (R4) holds, we have,

$$\left.\begin{array}{c} P_F\Big(X_{2j}'\delta_2 > X_{2i}'\delta_2\ ,\ X_{1Uj}'\beta_1 \leq X_{1Li}'\beta_1\Big) > 0 \\ and \\ P_F\Big(X_{2j}'\delta_2 < X_{2i}'\delta_2\ ,\ X_{1Uj}'\beta_1 \leq X_{1Li}'\beta_1\Big) > 0 \end{array}\right\}\ \forall\ \beta_1 \in \Theta, \forall\ \delta_2 \neq 0. \tag{13}$$

Combined with (12), the result in (13) leads to point-identification of (β_{10}, β_{20}) through the first inequality in (9). In the online appendix (Section A1.1) we show that an exclusion restriction between X_2 and W_1 is a necessary condition for the result in (13). If the conditions in (R1), (R3) and (R4) hold, we have

$$P_F\Big(E_F[\lambda_F(X_{1i}'\beta_{10})|V_i] + X_{2i}'\delta_2 > E_F[\lambda_F(X_{1j}'\beta_{10})|V_j] + X_{2j}'\delta_2\ ,\ X_{1Uj}'\beta_1 \leq X_{1Li}'\beta_1\Big) > 0\ \forall\ \delta_2 \neq 0, \forall\ \beta_1 \in \Theta.$$

Since $\mu_{2F}(V) - X_2'\beta_2 = E_F[\lambda_F(X_1'\beta_{10})|V] + X_2'(\beta_{20} - \beta_2)$ for any β_2, the previous result implies that, if the restrictions in (R1), (R3) and (R4) hold,

$$P_F\Big(\mu_{2F}(V_i) - X_{2i}'\beta_2 > \mu_{2F}(V_j) - X_{2j}'\beta_2\ ,\ X_{1Uj}'\beta_1 \leq X_{1Li}'\beta_1\Big) > 0 \quad \forall\ (\beta_1, \beta_2) : \beta_2 \neq \beta_{20}. \tag{14}$$

(12) and (14) imply that any $(\beta_1, \beta_2) \neq (\beta_{10}, \beta_{20})$ will violate the first inequality in (9) with strictly positive probability, leading to point-identification of the parameters through our inequalities. We summarize our results next.

Result 1 *Suppose restriction* (R1) *holds. Then,*

(i) If restriction (R2) *holds, any* (β_1, β_2) *such that* $\beta_1 \neq \beta_{10}$ *will violate the second inequality in* (9) *with strictly positive probability.*

(iii) If restrictions (R3) *and* (R4) *hold, any* $(\beta_1, \beta_2) \neq (\beta_{10}, \beta_{20})$ *will violate the first inequality in* (9) *with strictly positive probability. This is true whether or not* (R2) *is satisfied.*

(iv) If restrictions (R2) *and* (R3) *hold, and if the condition in part (ii) of restriction* (R4) *holds for* β_{10} *(but not necessarily for every* $\beta_1 \in \Theta$*), then any* $(\beta_1, \beta_2) \neq (\beta_{10}, \beta_{20})$ *will violate either the first or the second inequalities in* (9) *with strictly positive probability.*

Thus, we conclude that if restrictions (R1) *and* (R2) *hold, the inequalities in* (9) *identify* β_{10}*, and if restrictions* (R1), (R3) *and* (R4) *hold, the inequalities in* (9) *identify* (β_{10}, β_{20})*. If the restriction in* (R4) *is satisfied only for* β_{10}*, our inequalities can still point-identify* (β_{10}, β_{20}) *if* (R1), (R2) *and* (R3) *hold.*

Proof: The statements in Result 1 are simply a summary of the results described in equations (10), (11), (12), (13) and (14). Thus, we can prove Result 1 by showing that the results in those equations follow from restrictions (R1), (R2), (R3) and (R4). We do this in the online appendix. ■

In summary, if the marginal supports of the lower and upper bounds for the control function have a nonempty overlap, and if the range of values that the control function can take spans the entire interval given by these bounds, the pairwise inequalities in (9) can point-identify the parameters in the model if there exists a continuously distributed regressor in the selection equation that is excluded in the outcome equation. Our Monte Carlo experiments in Section 4 will revisit Result 1 using designs that satisfy its conditions. There, we will compute the probabilities of violations of our inequalities for parameter values other than (β_{10}, β_{20}). In practice, whether the restrictions leading to Result 1 are reasonable would depend on the specific application. For this reason, our inferential approach will not rely on these restrictions, and we will not presuppose that our inequalities point-identify the parameters of the model.

2.2 A model with unobserved regressors which can be bounded by an economic model

Interval data can be viewed as the most natural motivation for our model. In the absence of interval data, bounds for unobserved control variables can, in some instances, be obtained from an auxiliary economic theory. We illustrate this in a model where the selection equation depends on economic agents' unobserved beliefs. Leave the outcome equation as described in (1), and consider a selection equation 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"). Group $X_1 \equiv (W_1, \pi_1)$. 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 concept 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; we only require that the same prior be used by all agents.

2.2.1 Bounds for $W_1'\beta_{10}^w + \beta_{10}^\pi \pi_1$ 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[8]. We can obtain bounds on beliefs by adapting an approach suggested in Aradillas-López and Tamer (2008) in incomplete-information games. Suppose that the prior H_1 is consistent with the assumption that $\varepsilon_1 \perp W_1$. We describe the procedure of iterated elimination of nonrationalizable beliefs next.

Step 1: Since beliefs are probabilities, they must satisfy $0 \leq \pi_1 \leq 1$. 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 $\pi_{1L}^1(W_1,\beta_{10}) \leq \pi_1 \leq \pi_{1U}^1(W_1,\beta_{10})$, where these bounds are described above. Any set of beliefs consistent with this assumption must satisfy,

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

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 $\pi_{1L}^{k-1}(W_1,\beta_{10}) \leq \pi_1 \leq \pi_{1U}^{k-1}(W_1,\beta_{10})$. Any set of beliefs consistent with this assumption must satisfy,

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

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

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

Suppose that the conditions in Assumption 1 hold, with $X_1 \equiv (W_1, \pi_1)$. In this case, our control function is now $W_1'\beta_{10}^w + \beta_{10}^\pi \pi_1$ (instead of $X_1'\beta_{10}$, in the interval-data model). Since π_1 is unobserved, the control function $W_1'\beta_{10}^w + \beta_{10}^\pi \pi_1$ is unobserved. If we assume that agents' beliefs are consistent with at least k steps of iterated elimination of nonrationalizable beliefs, we must have,

$$\underbrace{W_1'\beta_{10}^w + \beta_{10}^\pi \pi_{1L}^k(W_1, \beta_{10})}_{\equiv g_{1L}(W_1,\beta_{10})} \leq W_1'\beta_{10}^w + \beta_{10}^\pi \pi_1 \leq \underbrace{W_1'\beta_{10}^w + \beta_{10}^\pi \pi_{1U}^k(W_1, \beta_{10})}_{\equiv g_{1U}(W_1,\beta_{10})}. \tag{15}$$

These bounds for the control function will replace those in (2) in the interval-data model.

2.2.2 Bounds for $W_1'\beta_{10}^w + \beta_{10}^\pi \pi_1$ based on the assumption of BNE beliefs

For a given W_1, Bayesian-Nash equilibrium (BNE) beliefs are given by any solution in π_1 to the BNE system

$$\pi_1 = H_1(W_1'\beta_{10}^w + \beta_{10}^\pi \pi_1).$$

Assume that H_1 is continuous. Then, existence[9] 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

$$\underbrace{W_1'\beta_{10}^w + \beta_{10}^\pi \pi_{1,1}^*(W_1, \beta_{10})}_{\equiv g_{1L}(W_1,\beta_{10})} \leq W_1'\beta_{10}^w + \beta_{10}^\pi \pi_1 \leq \underbrace{W_1'\beta_{10}^w + \beta_{10}^\pi \pi_{1,R}^*(W_1, \beta_{10})}_{\equiv g_{1U}(W_1,\beta_{10})}. \tag{16}$$

Under the assumptions of BNE beliefs, these bounds for the control function replace those in (2) in the interval-data model.

2.2.3 Testable implications

Maintain the exclusion and monotonicity restrictions in Assumption 1, with $X_1 \equiv (W_1, \pi_1)$ in this case. Then, the bounds in equations (6) and (8) are obtained once again, with $X_1'\beta_{10}^w + \beta_{10}^\pi \pi_1$ replacing $X_1'\beta_{10}$, and with $g_{1L}(W_1, \beta_{10})$ and $g_{1U}(W_1, \beta_{10})$ replacing $X_{1L}'\beta_{10}$ and $X_{1U}'\beta_{10}$, respectively. By the same arguments we presented in Section 2.1, these bounds are sharp given Assumption 1 and cannot be improved upon without further restrictions. The same steps leading to (9) now

[9] Regarding cardinality, the Index Theorem (Mas-Colell, Whinston, and Green (1995, Proposition 17.D.2)) can be used to show that this system has a finite number (in fact, an odd number) of *regular* solutions. These are solutions for which the Jacobian of the BNE system with respect to π_1 is non-zero.

yield the pairwise inequalities,

$$\left.\begin{array}{r}\Big(\big(\mu_{2F}(V_i)-X_{2i}'\beta_{20}\big)-\big(\mu_{2F}(V_j)-X_{2j}'\beta_{20}\big)\Big)\mathbb{1}\Big\{g_{1U}(W_{1j},\beta_{10})\leq g_{1L}(W_{1i},\beta_{10})\Big\}\leq 0\\ \big(\mu_{1F}(W_{1j})-\mu_{1F}(W_{1i})\big)\mathbb{1}\Big\{g_{1U}(W_{1j},\beta_{10})\leq g_{1L}(W_{1i},\beta_{10})\Big\}\leq 0\end{array}\right\}\text{ w.p.1.} \tag{17}$$

The expressions of the control-function bounds $g_{1L}(W_1,\beta_{10})$ and $g_{1U}(W_1,\beta_{10})$ would depend on the behavioral model assumed. Note that this approach can be a novel way to do inference in models of strategic interaction (games) with multiple equilibria or an uncertain solution concept.

We will use the pairwise inequalities in (17) as a general expression of the testable implications of our bivariate sample-selection model, since they include the interval-data example (equation 9) as a special case.

3 Inference

In this section we propose an inferential procedure for the parameters of the model based on the type of pairwise functional inequalities described in (17). We begin by presenting a general description of a bivariate sample-selection model with censored data that encompasses both examples in Section 2 as special cases. The pairwise inequalities produced by our general model can be expressed exactly as in equation (17).

3.1 A general description of our model

We will maintain that the outcome equation is described by (1), with observable regressors[10] X_2, and we will assume a general expression for the selection equation of the form,

$$Y_1^* = g_1(X_1,\beta_{10}) + \varepsilon_1 \tag{18}$$

$g_1(X_1,\beta_1)$ has known functional form, but a subset of regressors in X_1 are unobserved. We assume that, either by the nature of the data available, or through an auxiliary economic theory, we have bounds[11] for $g_1(X_1,\beta_{10})$,

$$g_{1L}(W_1,\beta_{10}) \leq g_1(X_1,\beta_{10}) \leq g_{1U}(W_1,\beta_{10}) \quad \text{w.p.1,} \tag{19}$$

where W_1 are observable covariates, and where $g_{1L}(W_1,\beta_1)$ and $g_{1U}(W_1,\beta_1)$ have known functional form. This includes the examples from Section 2 as special cases. Group $V \equiv X_2 \cup W_1$, and maintain the exclusion and monotonicity restrictions described in Assumption 1. These

[10]Extensions where X_2 is unobserved are included in the online appendix.

[11]Cases where we only have a lower or an upper bound (but not both) for the control function will be discussed in Section 5. However, our results and methodology will focus on the setting described in (19).

restrictions once again yield that, if we denote $E_F[\varepsilon_2|\ \varepsilon_1 > -g_1(X_1, \beta_{10}), X_1, V] \equiv \lambda_F(g_1(X_1, \beta_{10}))$, and $P_F(Y_1 = 1|W_1, X_1) = H_F(g_1(X_1, \beta_{10}))$, then $\lambda_F(\cdot)$ is nonincreasing, and $H_F(\cdot)$ is nondecreasing. Denoting, once again, $\mu_{2F}(V) \equiv E_F[Y_2|V, Y_1 = 1]$, Assumption 1 yields, $\mu_{2F}(V) = X_2'\beta_{20} + E_F[\lambda_F(g_1(X_1, \beta_{10}))|V]$, which is a generalized version of equation (5). From here, using the same arguments as in Section 2, we obtain a generalized version of the bounds described in equations (6)-(8),

$$\left.\begin{array}{l} \lambda_F(g_{1U}(W_1, \beta_{10})) \leq E_F[\lambda_F(g_1(X_1, \beta_{10}))|V] \leq \lambda_F(g_{1L}(W_1, \beta_{10})) \\ \lambda_F(g_{1U}(W_1, \beta_{10})) \leq \mu_{2F}(V) - X_2'\beta_{20} \leq \lambda_F(g_{1L}(W_1, \beta_{10})) \\ H_F((g_{1L}(W_1, \beta_{10}))) \leq \mu_{1F}(W_1) \leq H_F((g_{1U}(W_1, \beta_{10}))) \end{array}\right\} \text{w.p.1.} \qquad (20)$$

By the same arguments described in Section 2, these bounds are sharp given the conditions in Assumption 1, and cannot be improved upon without further restrictions. From here, the model produces the pairwise functional inequalities given by the expressions in (17).

3.2 Variations and extensions of our model

Our bivariate sample selection model can be modified in various ways. In the online appendix (Section A2), we explore two modifications/extensions. The first one describes the case where we have unobserved covariates in both the selection and outcome equations, with bounds that depend on observables. The second modification discusses the truncated-data case, where our data consists only of observations where $Y_{1i}^* > 0$. In each case we describe the pairwise functional inequalities that arise, which are the counterpart versions of (17) in our main model. Equipped with the corresponding pairwise inequalities, inference would be carried out by modifying the procedure we will describe below.

Section 5 focuses on more significant departures from our main model, and analyzes alternative scenarios where the control functions may be partially identified. There, we discuss models where all regressors are observed, but the parameters of the control function are nevertheless partially identified. This could happen, for example, when all the regressors in the selection equation are discrete. We also discuss cases where control variables are unobserved (as in our main model), but only lower or upper bounds (but not both) are available. In each case, we outline the inferential strategies we would pursue. However, the main results in the paper will focus on the case described above, where control functions are partially identified because a subset of control variables are unobserved, and where (lower and upper) bounds which depend on observable covariates are available, either directly from the data (e.g, with interval data), or from an auxiliary economic theory.

3.3 On the possibility that (17) may point-identify β_0

In Section 2.1.2 we described conditions under which the pairwise inequalities in the interval-data example would be capable of point-identifying the parameters of the model. Those restrictions are described in equations (R1)-(R4) and they yield the identification properties in Result 1. At a high-level, these restrictions can be straightforwardly extended to our general model. The first basic requirement is that the marginal supports of the control-function bounds $g_{1L}(W_1, \beta_1)$ and $g_{1U}(W_1, \beta_1)$ have a nonempty overlap for each $\beta_1 \in \Theta$. From here, point-identification can follow if: (i) the supports of the functionals $E_F[\lambda_F(g_1(X_1, \beta_{10}))]$ and $\mu_{1F}(W_1)$ span the entire intervals given by the bounds in (20), and (ii) if there exists a regressor in W_1 with rich enough support, excluded from X_2, such that the type of conditions described in restriction (R4) are satisfied. These are high-level conditions, which were examined in detail in the interval-data example. Bringing them down to more basic restrictions in our general model would depend on the specific parameterizations of the selection equation and of the control-function bounds.

If one is willing to impose the type of point-identification restrictions leading to Result 1, an estimator for (β_{10}, β_{20}) could be constructed using a sample objective function based on the pairwise inequalities in (17). An example of an extremum estimator based on conditional moment inequalities can be found in Khan and Tamer (2009). While their estimator is based on using a space of instrument functions (an approach suggested by Dominguez and Lobato (2004) in conditional GMM models, and further generalized in Andrews and Shi (2013)), the nature of our problem would require the use of nonparametric estimators for the functionals involved. Having said this, in this paper we will not assume (or rule out) that our pairwise inequalities point-identify the parameters of the model, since the restrictions required may not be realistic in many data sets. Instead, we will focus on the construction of confidence sets with asymptotic properties that are valid whether or not our pairwise inequalities can point-identify[12], since the restrictions required may not be realistic in many data sets. Instead, we will focus on the construction of confidence sets with asymptotic properties that are valid whether or not our pairwise inequalities can point-identify the parameters of the model.

3.4 Preliminary transformations to the parameter space

As we pointed out in Section 2.1.2, our nonparametric treatment of λ_F and H_F will necessitate some preliminary transformations to the parameter space. First, note that our inequalities are invariant to the presence of an intercept in β_{20}, which is always differenced out in our inequalities. Second, we need a generalization of the location and scale normalizations of β_{10} in the interval data example. This can be done as follows. Let $d_{w_1} \equiv \dim(W_1)$. In what follows, we will assume

[12] One could also use a simple data-generating process; for example, one where all regressors are discrete (but still observed in intervals), to obtain the population identified set produced by our inequalities, and have a good sense of its size and informativeness.

that the parameter space Θ has been modified, so that,

$$\exists\ \beta_1, \beta_1' \in \Theta, \text{ and } \exists\ (w_1, \widetilde{w}_1) \in \mathbb{R}^{d_{w_1}} \times \mathbb{R}^{d_{w_1}} : g_{1U}(w_1, \beta_1) \leq g_{1L}(\widetilde{w}_1, \beta_1), \text{ and } g_{1U}(w_1, \beta_1') > g_{1L}(\widetilde{w}_1, \beta_1') \tag{21}$$

This rules out a parameter space where *for every* β_1 there exists $\beta_1' \neq \beta_1$ that is observationally equivalent in terms of our pairwise inequalities. Any original features of the parameter space that must be modified in order to satisfy (21) are impossible to partially identify to any extent through our pairwise inequalities. In the interval-data example when the control-function bounds are linear indices of β_{10}, the condition in (21) requires that the parameter space exclude an intercept for β_{10}, and that the scale of the slope coefficients in β_{10} be normalized. These were exactly the modifications to the parameter space we made in Section 2.1.2.

The need to modify the parameter space is owed to our nonparametric treatment of $\lambda_F(\cdot)$ and $H_F(\cdot)$, and it would also be necessary even if the control functions were perfectly observed (see footnote 5). Location and scale restrictions are common in semiparametric models, including the pairwise-differencing methods of Powell and coauthors, as well as rank estimators, whose testable implications involve inequalities between linear indices of parameters (see Han (1987), Sherman (1993), Cavanagh and Sherman (1998)), and our model shares similarities with both. To avoid introducing new notation, we will keep denoting our parameters as $(\beta_{10}, \beta_{20}) \equiv \beta_0$, and our parameter space as Θ, keeping in mind the preliminary modifications to the parameter space described above.

3.5 A population statistic for the pairwise inequalities (17)

We observe an iid sample $(Y_{1i}, Y_{2i}, V_i)_{i=1}^n$ generated by F, where $V \equiv X_2 \cup W_1$. In what follows, for any $i \neq j$, (Y_{1i}, Y_{2i}, V_i) and (Y_{1j}, Y_{2j}, V_j) denote two independent draws from F. We will focus on *density-weighted* versions of the functionals in the pairwise inequalities (17). Density-weighting 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 $\beta \in \Theta$, let

$$\begin{aligned}
&\tau_{2F}(V_i, V_j, \beta) \equiv \\
&\Big(\big(\mu_{2F}(V_i) - X_{2i}'\beta_2\big) - \big(\mu_{2F}(V_j) - X_{2j}'\beta_2\big)\Big)\mathbb{1}\Big\{g_{1U}(W_{1j}, \beta_1) \leq g_{1L}(W_{1i}, \beta_1)\Big\} \cdot f_{V,1}(V_i) f_{V,1}(V_j) \\
&\cdot \phi_2(V_i)^2 \phi_2(V_j)^2, \\
&\tau_{1F}(W_{1i}, W_{1j}, \beta_1) \equiv \\
&\big(\mu_{1F}(W_{1j}) - \mu_{1F}(W_{1i})\big) \cdot \mathbb{1}\Big\{g_{1U}(W_{1j}, \beta_1) \leq g_{1L}(W_{1i}, \beta_1)\Big\} \cdot f_{W_1}(W_{1i}) f_{W_1}(W_{1j}) \phi_1(W_{1i})^2 \phi_1(W_{1j})^2,
\end{aligned} \tag{22}$$

where $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are weight functions, chosen by the econometrician[13] , which are strictly positive over a pre-specified *inference range* $\mathcal{V} \subseteq \text{Supp}(V)$ and zero everywhere else. Our inference procedure will be based on testing the pairwise inequalities (17) over the inference range $\mathcal{V}$. Let $(A)_+ \equiv A \vee 0$, and denote,

$$\mathcal{T}_{2F}(\beta) \equiv E_F\left[\left(\tau_{2F}(V_i, V_j, \beta)\right)_+\right], \quad \mathcal{T}_{1F}(\beta_1) \equiv E_F\left[\left(\tau_{1F}(W_{1i}, W_{1j}, \beta_1)\right)_+\right]. \tag{23}$$

By construction, $\mathcal{T}_{2F}(\beta) \geq 0 \; \forall \; \beta$, and $\mathcal{T}_{2F}(\beta) = 0$ if and only if the first inequality in (17) holds almost surely over our inference range. Similarly, $\mathcal{T}_{1F}(\beta_1) \geq 0 \; \forall \; \beta_1$, and $\mathcal{T}_{1F}(\beta_1) = 0$ if and only if the second inequality in (17) holds almost surely over our inference range. $\mathcal{T}_{2F}$ and $\mathcal{T}_{1F}$ can be straightforwardly aggregated into a statistic that captures whether both of the functional inequalities in (17) are satisfied almost surely over our inference range. Consider the population statistic,

$$\mathcal{T}_F(\beta) \equiv \mathcal{T}_{2F}(\beta) + \mathcal{T}_{1F}(\beta_1). \tag{24}$$

By construction, $\mathcal{T}_F(\beta) \geq 0 \; \forall \; \beta$, and $\mathcal{T}_F(\beta) = 0$ if and only if *both* of the functional inequalities in (17) are satisfied almost surely over our inference range. We can generalize (24) to assign different weights to $\mathcal{T}_{2F}$ and $\mathcal{T}_{1F}$ based, for example, on the scale of the covariates in the outcome and selection equations, respectively[14]. Our results can be straightforwardly extended to such cases. We focus on (24) for simplicity.

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

Denote,

$$\begin{aligned} R_{2F}(V) &\equiv \mu_{2F}(V) f_{V,1}(V)\phi_2(V), \quad Q_{2F}(V) \equiv f_{V,1}(V)\phi_2(V), \\ R_{1F}(W_1) &\equiv \mu_{1F}(W_1) f_{W_1}(W_1)\phi_1(W_1), \quad Q_{1F}(W_1) \equiv f_{W_1}(W_1)\phi_1(W_1) \end{aligned} \tag{25}$$

The functionals $\tau_{2F}(V_i, V_j, \beta)$ and $\tau_{1F}(W_{1i}, W_{1j}, \beta_1)$ defined in (22) can be rewritten as,

$$\begin{aligned} \tau_{2F}(V_i, V_j, \beta) \equiv &\Big(\big(R_{2F}(V_i)Q_{2F}(V_j) - R_{2F}(V_j)Q_{2F}(V_i)\big) - \big(X_{2i}'\beta_2 - X_{2j}'\beta_2\big) Q_{2F}(V_i)Q_{2F}(V_j)\Big) \\ &\cdot \mathbb{1}\left\{g_{1U}(W_{1j}, \beta_1) \leq g_{1L}(W_{1i}, \beta_1)\right\} \phi_2(V_i)\phi_2(V_j), \\ \tau_{1F}(W_{1i}, W_{1j}, \beta_1) \equiv &\Big(R_{1F}(W_{1j})Q_{1F}(W_{1i}) - R_{1F}(W_{1i})Q_{1F}(W_{1j})\Big) \\ &\cdot \mathbb{1}\left\{g_{1U}(W_{1j}, \beta_1) \leq g_{1L}(W_{1i}, \beta_1)\right\} \phi_1(W_{1i})\phi_1(W_{1j}). \end{aligned} \tag{26}$$

We will focus on kernel-based nonparametric estimators for the functionals in (25). 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,

[13] The reason why ϕ_1 and ϕ_2 enter squared into our expressions is for convenience, as it will allow us to rewrite them in the way we will describe below, in equations (25) and (26).

[14] Our results will be based on a statistic that will be constructed using a properly normalized estimator of $\mathcal{T}_F$. The normalization will be based on a (regularized) estimator of a standard deviation that will reflect the scale of $\mathcal{T}_F$.

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 $w_1 \equiv (w_1^c, w_1^d)$,

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

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

We will describe restrictions for the kernel $\kappa(\cdot)$ and the bandwidth h_n below. We estimate,

$$\begin{aligned} \widehat{R}_2(v) &\equiv \frac{1}{n \cdot h_n^r} \sum_{i=1}^{n} Y_{2i} Y_{1i} \phi_2(V_i) \Gamma(V_i, v, h_n), \quad \widehat{Q}_2(v) \equiv \frac{1}{n \cdot h_n^r} \sum_{i=1}^{n} Y_{1i} \phi_2(V_i) \Gamma(V_i, v, h_n), \\ \widehat{R}_1(w_1) &\equiv \frac{1}{n \cdot h_n^\ell} \sum_{i=1}^{n} Y_{1i} \phi_1(W_{1i}) \Gamma(W_{1i}, w_1, h_n), \quad \widehat{Q}_1(w_1) \equiv \frac{1}{n \cdot h_n^\ell} \sum_{i=1}^{n} \phi_1(W_{1i}) \Gamma(W_{1i}, w_1, h_n), \end{aligned} \tag{27}$$

Density-weighting has the advantage of producing estimators in the form of sample averages, without estimated densities in the denominator[15]. Using the expressions in (26), we estimate,

$$\begin{aligned} \widehat{\tau}_2(V_i, V_j, \beta) &\equiv \left(\left(\widehat{R}_2(V_i)\widehat{Q}_2(V_j) - \widehat{R}_2(V_j)\widehat{Q}_2(V_i)\right) - \left(X_{2i}'\beta_2 - X_{2j}'\beta_2\right)\widehat{Q}_2(V_i)\widehat{Q}_2(V_j)\right) \\ &\quad \cdot \mathbb{1}\left\{g_{1U}(W_{1j}, \beta_1) \leq g_{1L}(W_{1i}, \beta_1)\right\} \phi_2(V_i)\phi_2(V_j), \\ \widehat{\tau}_1(W_{1i}, W_{1j}, \beta_1) &\equiv \left(\widehat{R}_1(W_{1j})\widehat{Q}_1(W_{1i}) - \widehat{R}_1(W_{1i})\widehat{Q}_1(W_{1j})\right) \\ &\quad \cdot \mathbb{1}\left\{g_{1U}(W_{1j}, \beta_1) \leq g_{1L}(W_{1i}, \beta_1)\right\} \phi_1(W_{1i})\phi_1(W_{1j}). \end{aligned}$$

For a given $\beta \in \Theta$, our estimators for $\mathcal{T}_{2F}(\beta)$, $\mathcal{T}_{1F}(\beta_1)$ and $\mathcal{T}_F(\beta)$ are,

$$\begin{aligned} \widehat{\mathcal{T}}_2(\beta) &\equiv \frac{1}{n \cdot (n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \widehat{\tau}_2(V_i, V_j, \beta) \cdot \mathbb{1}\left\{\widehat{\tau}_2(V_i, V_j, \beta) \geq -b_n\right\}, \\ \widehat{\mathcal{T}}_1(\beta_1) &\equiv \frac{1}{n \cdot (n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \widehat{\tau}_1(W_{1i}, W_{1j}, \beta_1) \cdot \mathbb{1}\left\{\widehat{\tau}_1(W_{1i}, W_{1j}, \beta_1) \geq -b_n\right\}, \\ \widehat{\mathcal{T}}(\beta) &\equiv \widehat{\mathcal{T}}_2(\beta) + \widehat{\mathcal{T}}_1(\beta_1), \end{aligned} \tag{28}$$

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

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

Remark 1 All the results that follow will hold exactly as described if we use a bandwidth $b_n^{\tau_2}$ for $\widehat{\mathcal{T}}_2(\beta)$ and a different bandwidth $b_n^{\tau_1}$ for $\widehat{\mathcal{T}}_1(\beta_1)$, as long as both bandwidths satisfy the convergence rate restrictions for b_n that we will describe in Assumption 4. ■

3.7 Asymptotic properties of $\widehat{\mathcal{T}}(\beta)$

Next, we will describe a series of assumptions involving conditions such as smoothness, manageability and regularity for the functionals in our model, along with restrictions for our tuning parameters. Combined, these assumptions will yield a linear representation result for $\widehat{\mathcal{T}}(\beta)$, which will be the foundation for the construction of a test-statistic for estimating a confidence set for β_0. The assumptions we will describe are technical in nature, but we will add an intuitive discussion to describe how each one of them contributes towards our main result.

We will let $\mathcal{F}$ denote the space of distributions that contains F, the distribution generating the sample observed. Our expanded parameter space is then $\Theta \times \mathcal{F} \equiv \{(\beta, F)\colon \beta \in \Theta, F \in \mathcal{F}\}$. Our goal will be to describe conditions that will yield uniform asymptotic 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, figuring out which objects are functionals of F will be clear from our discussion and definitions. Let $\{m_n(\beta)\colon \beta \in \Theta\}$ be a stochastic process. Following convention, we will use the following terminology. We say that $\sup_{\beta\in\Theta}\left\|m_n(\beta)\right\| = o_p(n^{\lambda})$ *uniformly over* $\mathcal{F}$ if,

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

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

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

Similarly, consider a set $\mathcal{C} \subseteq (\Theta \times \mathcal{F})$ of the form $\mathcal{C} \equiv \{(\beta, F) \in \Theta \times \mathcal{F}\colon \beta \in \mathcal{C}(F)\}$. For example, we can have $\beta \in \mathcal{C}(F) \Longleftrightarrow E_F[G_F(Z,\beta)] = 0$ for some $G_F(\cdot)$. We say that $\sup_{(\beta,F)\in\mathcal{C}}\left\|m_n(\beta)\right\| = o_p(n^{\lambda})$ if,

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

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

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

Clarifying the above notation will facilitate the interpretation of our results.

3.7.1 Assumptions leading to an asymptotic linear representation result

Our first set of assumptions involve smoothness conditions for a collection of functionals that show up in our linear approximations. For a given $v \equiv (x_2, w_1)$ and β, let,

$$\begin{aligned}
\eta^{\tau_2}_{a,F}(v,\beta) &\equiv E_F\left[(R_{2F}(V) - \beta_2'(X_2 - x_2)\,Q_{2F}(V))\,\mathbb{1}\{g_{1U}(w_1,\beta_1) \leq g_{1L}(W_1,\beta_1)\}\,\mathbb{1}\{\tau_{2F}(V,v,\beta) \geq 0\}\,\phi_2(V)\right],\\
\eta^{\tau_2}_{b,F}(v,\beta) &\equiv E_F\left[(R_{2F}(V) - \beta_2'(X_2 - x_2)\,Q_{2F}(V))\,\mathbb{1}\{g_{1U}(W_1,\beta_1) \leq g_{1L}(w_1,\beta_1)\}\,\mathbb{1}\{\tau_{2F}(v,V,\beta) \geq 0\}\,\phi_2(V)\right],\\
\eta^{\tau_2}_{c,F}(v,\beta) &\equiv E_F\left[Q_{2F}(V)\mathbb{1}\{g_{1U}(W_1,\beta_1) \leq g_{1L}(w_1,\beta_1)\}\,\mathbb{1}\{\tau_{2F}(v,V,\beta) \geq 0\}\,\phi_2(V)\right],\\
\eta^{\tau_2}_{d,F}(v,\beta) &\equiv E_F\left[Q_{2F}(V)\mathbb{1}\{g_{1U}(w_1,\beta_1) \leq g_{1L}(W_1,\beta_1)\}\,\mathbb{1}\{\tau_{2F}(V,v,\beta) \geq 0\}\,\phi_2(V)\right].
\end{aligned} \tag{29}$$

And for a given w_1 and β_1, let,

$$\begin{aligned}
\eta^{\tau_1}_{a,F}(w_1,\beta_1) &\equiv E_F\left[R_{1F}(W_1)\mathbb{1}\{g_{1U}(w_1,\beta_1) \leq g_{1L}(W_1,\beta_1)\}\,\mathbb{1}\{\tau_{1F}(W_1,w_1,\beta) \geq 0\}\,\phi_1(W_1)\right],\\
\eta^{\tau_1}_{b,F}(w_1,\beta_1) &\equiv E_F\left[R_{1F}(W_1)\mathbb{1}\{g_{1U}(W_1,\beta_1) \leq g_{1L}(w_1,\beta_1)\}\,\mathbb{1}\{\tau_{1F}(w_1,W_1,\beta) \geq 0\}\,\phi_1(W_1)\right],\\
\eta^{\tau_1}_{c,F}(w_1,\beta_1) &\equiv E_F\left[Q_{1F}(W_1)\mathbb{1}\{g_{1U}(W_1,\beta_1) \leq g_{1L}(w_1,\beta_1)\}\,\mathbb{1}\{\tau_{1F}(w_1,W_1,\beta) \geq 0\}\,\phi_1(W_1)\right],\\
\eta^{\tau_1}_{d,F}(w_1,\beta_1) &\equiv E_F\left[Q_{1F}(W_1)\mathbb{1}\{g_{1U}(w_1,\beta_1) \leq g_{1L}(W_1,\beta_1)\}\,\mathbb{1}\{\tau_{1F}(W_1,w_1,\beta) \geq 0\}\,\phi_1(W_1)\right]
\end{aligned} \tag{30}$$

Assumption 2 ***(Smoothness and 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,*

$$\sup_{\substack{x_2\in\mathcal{V}\\ \beta_2\in\Theta}} \left|x_2'\beta_2\right| \leq \overline{D}, \quad \textit{and} \quad \sup_{v\in\mathcal{V}} \left(f_{V,1}(v) \vee \left|\mu_{2F}(v)\right|\right) \leq \overline{D}, \ \textit{for each } F \in \mathcal{F}.$$

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* (29) *and* (30), *the following holds,*

(i) *Uniformly over* Θ, *the functionals* $\eta^{\tau_2}_{a,F}(v,\beta)$, $\eta^{\tau_2}_{b,F}(v,\beta)$, $\eta^{\tau_2}_{c,F}(v,\beta)$ *and* $\eta^{\tau_2}_{d,F}(v,\beta)$ *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* $\eta^{\tau_1}_{a,F}(w_1,\beta_1)$, $\eta^{\tau_1}_{b,F}(w_1,\beta_1)$, $\eta^{\tau_1}_{c,F}(w_1,\beta_1)$ *and* $\eta^{\tau_2}_{d,F}(w_1,\beta_1)$ *are M times differentiable with respect to* w_1^c, *with bounded derivatives for a.e* $w \in \mathcal{V}$. ■

We will describe below a lower bound for M (the degree of smoothness of our functionals). Combined with bias-reduction properties for our tuning parameters, the conditions in Assumption 2 will help ensure that the asymptotic bias of our estimators converges to zero at the appropriate rate, uniformly over $\Theta \times \mathcal{F}$.

To obtain uniform asymptotic results, we will maintain assumptions that produce manageability properties (see Pollard (1990, Definition 7.9)) for a collection of empirical processes and U-processes that are relevant to our problem. Specifically, we will maintain assumptions that produce *Euclidean* classes of functions, leading to a special case of manageable processes. Euclidean classes of functions are defined, for example, in Nolan and Pollard (1987, Definition 8), Pakes and Pollard (1989, Definition 2.7), and Sherman (1994, Definition 3). They include, among others, classes of linear indices (and transformations of bounded variation of linear indices), indicator functions over VC classes of sets, as well as the Type I, II and III classes of functions described in Andrews (1994). Euclidean classes encompass, in particular, the parametric functions in the two examples in Section 2. In the online appendix we present the definition of Euclidean classes, along with detailed examples. Having Euclidean classes will allow us to invoke convenient maximal inequality results in Sherman (1994) on our path to our main asymptotic result.

Assumption 3 ***(Manageability, integrability)*** *There exist* $\overline{D}_4 < \infty$ *and* $\overline{C}_4 < \infty$ *such that* $E_F[|Y_2|^4] \leq \overline{D}_4$ *and* $E_F[\|X_2\|^4] \leq \overline{C}_4$ *for all* $F \in \mathcal{F}$. *Take the functionals* $\eta^{\tau_2}_{\ell,F}$ *defined in* (29) *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* $\beta, \beta' \in \Theta$,

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

(i) The class of sets

$$\mathscr{C} \equiv \left\{ (w_1, w_1) \in \mathbb{R}^{d_U} \times \mathbb{R}^{d_L} : g_{1U}(w_1,\beta_1) \leq g_{1L}(w_1,\beta_1) \;\; \textit{for some } \beta_1 \in \Theta \right\}$$

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

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

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). Sufficient conditions for 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), and are also discussed in the online appendix. Due to the linear-index nature of the parametric functions involved, the VC property in part (i) of Assumption 3 holds in our interval-data example by Pakes and Pollard (1989, Lemma 2.4). The fundamental reason for our VC property assumptions is that indicator functions over VC classes of sets are themselves Euclidean classes of functions. 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. 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 $n^{1/2+\epsilon} \cdot h_n^M \longrightarrow 0$. ■

Section 3.10 includes a practical discussion about bandwidth selection. There, we show that if our bandwidths are of the form $h_n \propto n^{-\alpha_h}$ and $b_n \propto n^{-\alpha_b}$, then the smallest value of M that can satisfy the restrictions in Assumption 4 is $M = 2r + 1$. Next, we present a *regularity* condition.

Assumption 5 ***((Behavior of $\boldsymbol{\tau_{2F}(\cdot)}$ and $\boldsymbol{\tau_{1F}(\cdot)}$ 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* $\forall\ F \in \mathcal{F}$,

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

$$\sup_{\beta_1\in\Theta} E_F\left[\mathbb{1}\left\{-b \leq \tau_{1F}(W_{1i}, W_{1j}, \beta_1) < 0\right\}\right] \leq \overline{m}\cdot b \quad \forall\ 0 < b \leq b_0,$$

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

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

$$(iii) \qquad \sup_{\substack{\beta_1\in\Theta\\ w_1\in\mathcal{V}}} E_F\left[\mathbb{1}\left\{-b \leq \tau_{1F}(W_{1i}, w_1, \beta_1) < 0\right\}\right] \leq \overline{m}\cdot b \quad \forall\ 0 < b \leq b_0,$$

$$\sup_{\substack{\beta_1\in\Theta\\ w_1\in\mathcal{V}}} E_F\left[\mathbb{1}\left\{-b \leq \tau_{1F}(w_1, W_{1i}, \beta_1) < 0\right\}\right] \leq \overline{m}\cdot b \quad \forall\ 0 < b \leq b_0. \quad \blacksquare$$

Assumption 5 is a mild requirement, amounting to the restriction that, uniformly over $\Theta\times\mathcal{F}$, the functionals described there have a *finite density* in a neighborhood of the form $[-b_0, 0)$. Let us illustrate this for part (i). For a given β, let $F_{\tau_2}(\cdot|\beta)$ and $F_{\tau_1}(\cdot|\beta_1)$ denote the distribution functions of $\tau_{2F}(V_i, V_j, \beta)$ and $\tau_{1F}(W_{1i}, W_{1j}, \beta_1)$, respectively, and let $f_{\tau_2}(\cdot|\beta)$ and $f_{\tau_1}(\cdot|\beta_1)$ denote the corresponding density functions. Part (i) of Assumption 5 presupposes that there exists an interval $[-b_0, 0)$ and a finite constant $\overline{m}$ such that $F_{\tau_2}(\cdot|\beta)$ and $F_{\tau_1}(\cdot|\beta_1)$ are continuous, and $f_{\tau_2}(\cdot|\beta_2)$ and $f_{\tau_1}(\cdot|\beta_1)$ are bounded above by $\overline{m}$, uniformly over $\Theta\times\mathcal{F}$. That is,

$$\sup_{\substack{\beta\in\Theta\\ b\in[-b_0,0)}} f_{\tau_2}(b|\beta) \leq \overline{m}, \quad \text{and} \quad \sup_{\substack{\beta_1\in\Theta\\ b\in[-b_0,0)}} f_{\tau_1}(b|\beta_1) \leq \overline{m}, \quad \forall\ F \in \mathcal{F}.$$

If the above is true, 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. Combined with the rate-of-convergence restrictions on b_n, the conditions in Assumption 5 will ensure that the expectation of a particular process vanishes with n at the appropriate rate. The details and intermediate steps can be found in the online appendix.

3.7.2 An asymptotic linear representation result for $\widehat{\mathcal{T}}(\beta)$

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

Lemma 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*

$$H^{\mathcal{T}_2}_{1F}(V_i,\beta) \equiv \frac{1}{2}\cdot\Big(E_F\Big[\big(\tau_{2F}(V_i,V_j,\beta)\big)_+\big|V_i\Big] + E_F\Big[\big(\tau_{2F}(V_i,V_j,\beta)\big)_+\big|V_i\Big]\Big) - \mathcal{T}_{2F}(\beta)$$

Note that $E_F[H^{\mathcal{T}_2}_{1F}(V_i,\beta)] = 0\ \forall\ (\beta,F)$. *Next, take the functionals defined in* (29) *and let,*

$$\begin{aligned}H^{\mathcal{T}_2}_{2F}(Z_i,\beta) \equiv &\Big(\big(\eta^{\tau_2}_{a,F}(V_i,\beta) - \eta^{\tau_2}_{b,F}(V_i,\beta)\big)\cdot Y_{1i} + \big(\eta^{\tau_2}_{c,F}(V_i,\beta) - \eta^{\tau_2}_{d,F}(V_i,\beta)\big)\cdot Y_{2i}Y_{1i}\Big)\cdot f_V(V_i)\cdot\phi_2(V_i)^2\\ &- E_F\Big[\Big(\big(\eta^{\tau_2}_{a,F}(V_i,\beta) - \eta^{\tau_2}_{b,F}(V_i,\beta)\big)\cdot Y_{1i} + \big(\eta^{\tau_2}_{c,F}(V_i,\beta) - \eta^{\tau_2}_{d,F}(V_i,\beta)\big)\cdot Y_{2i}Y_{1i}\Big)\cdot f_V(V_i)\cdot\phi_2(V_i)^2\Big].\end{aligned}$$

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

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

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

$$H^{\mathcal{T}_1}_{1F}(W_{1i},\beta_1) \equiv \frac{1}{2}\cdot\Big(E_F\Big[\big(\tau_{1F}(W_{1i},W_{1j},\beta_1)\big)_+\big|W_{1i}\Big] + E_F\Big[\big(\tau_{1F}(W_{1j},W_{1i},\beta_1)\big)_+\big|W_{1i}\Big]\Big) - \mathcal{T}_{1F}(\beta_1)$$

Note that $E_F[H^{\mathcal{T}_1}_{1F}(W_{1i},\beta_1)] = 0\ \forall\ (\beta_1,F)$. *Next, take the functionals defined in* (30) *and let,*

$$\begin{aligned}H^{\mathcal{T}_1}_{2F}(Z_i,\beta_1) \equiv &\Big(\big(\eta^{\tau_1}_{a,F}(W_{1i},\beta_1) - \eta^{\tau_1}_{b,F}(W_{1i},\beta_1)\big) + \big(\eta^{\tau_1}_{c,F}(W_{1i},\beta_1) - \eta^{\tau_1}_{d,F}(W_{1i},\beta_1)\big)\cdot Y_{1i}\Big)\cdot f_{W_1}(W_{1i})\cdot\phi_1(W_{1i})^2\\ &- E_F\Big[\Big(\big(\eta^{\tau_1}_{a,F}(W_{1i},\beta_1) - \eta^{\tau_1}_{b,F}(W_{1i},\beta_1)\big) + \big(\eta^{\tau_1}_{c,F}(W_{1i},\beta_1) - \eta^{\tau_1}_{d,F}(W_{1i},\beta_1)\big)\cdot Y_{1i}\Big)\cdot f_{W_1}(W_{1i})\cdot\phi_1(W_{1i})^2\Big].\end{aligned}$$

Note that $E_F[H^{\mathcal{T}_1}_{2F}(Z_i,\beta_1)] = 0\ \forall\ (\beta_1,F)$. *Now let,* $\psi^{\mathcal{T}_1}_F(Z_i,\beta_1) \equiv 2\cdot H^{\mathcal{T}_1}_{1F}(W_{1i},\beta_1) + H^{\mathcal{T}_1}_{2F}(Z_i,\beta_1)$, *and note that* $E_F[\psi^{\mathcal{T}_1}_F(Z_i,\beta_1)] = 0\ \forall\ (\beta_1,F)$. *If Assumptions 1-5 hold,*

$$\begin{aligned}&\widehat{\mathcal{T}}_1(\beta_1) = \mathcal{T}_{1F}(\beta_1) + \frac{1}{n}\sum_{i=1}^{n}\psi^{\mathcal{T}_1}_F(Z_i,\beta_1) + \xi^{\mathcal{T}_1}_n(\beta_1), \quad \textit{where}\\ &\sup_{\beta_1\in\Theta}\left|\xi^{\mathcal{T}_1}_n(\beta_1)\right| = o_p\Big(\frac{1}{n^{1/2+\epsilon}}\Big), \quad \textit{uniformly over } \mathcal{F}.\end{aligned}$$

(C) Let $\psi_F^{\mathcal{T}}(Z_i,\beta) \equiv \psi_F^{\mathcal{T}_2}(Z_i,\beta) + \psi_F^{\mathcal{T}_1}(Z_i,\beta_{1)})$. From (A) and (B), we have $E_F[\psi_F^{\mathcal{T}}(Z_i,\beta)] = 0 \ \forall\ (\beta,F)$. If Assumptions 1-5 hold,

$$\widehat{\mathcal{T}}(\beta) = \mathcal{T}_F(\beta) + \frac{1}{n}\sum_{i=1}^{n}\psi_F^{\mathcal{T}}(Z_i,\beta) + \xi_n^{\mathcal{T}}(\beta), \quad \textit{where}$$

$$\sup_{\beta\in\Theta}\left|\xi_n^{\mathcal{T}}(\beta)\right| = o_p\left(\frac{1}{n^{1/2+\epsilon}}\right), \quad \textit{uniformly over } \mathcal{F}. \quad \blacksquare$$

The proof of Lemma 1 is included in Section A3 of the online appendix. We will summarize the main steps next, focusing on part (A). Part (B) uses analogous steps, and part (C) follows immediately from (A) and (B). Let

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

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

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

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

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

then the next series of steps show that the above expression becomes,

$$\widetilde{\mathcal{T}}_2(\beta) = \mathcal{T}_{2F}(\beta) + \frac{1}{n\cdot(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\Big(\Big(\tau_{2F}(V_i,V_j,\beta)\Big)_+ - \mathcal{T}_{2F}(\beta)\Big) + \frac{(n-2)}{n}\cdot\frac{1}{h_n^{r_v}}\cdot U_{a,n}(\beta,h_n) + \xi_{b,n}^{\widetilde{\mathcal{T}}_2}(\beta),$$

where $\sup_{\beta\in\Theta}\left|\xi_{b,n}^{\widetilde{\mathcal{T}}_2}(\beta)\right| = o_p\left(\frac{1}{n^{1/2+\epsilon}}\right)$, uniformly over $\mathcal{F}$.

Where $\epsilon > 0$ is the constant described in Assumption 4, and $\{U_{a,n}(\beta, h)\colon \beta \in \Theta,\ h > 0\}$ is a U-process of order 2. The final step to obtain the result in part (A) of the lemma is to compute the Hoeffding decomposition (see Serfling (1980, pages 177-178) or Sherman (1994, equations 6-7)) of $U_{a,n}(\beta, h_n)$ and apply the maximal inequality results in Sherman (1994, Corollary 4A). All the step-by-step details are included in Section A3 of the online appendix.

3.8 A statistic based on Lemma 1

We will rely on the linear representation result in Lemma 1 to build a statistic that will be used to estimate a CS for β_0. The following subsets of $\Theta \times \mathcal{F}$ will be relevant for our analysis. First, let,

$$
\begin{aligned}
&\Lambda_{\Theta,\mathcal{F}} \equiv \Big\{(\beta, F) \in \Theta \times \mathcal{F} : \\
&P_F\Big(\big((\mu_{2F}(V_i) - X_{2i}'\beta_2) - (\mu_{2F}(V_j) - X_{2j}'\beta_2)\big)\mathbb{1}\{g_{1U}(W_{1j}, \beta_1) \leq g_{1L}(W_{1i}, \beta_1)\} \leq 0 \;\Big|\; V_i, V_j \in \mathcal{V}\Big) = 1, \\
&P_F\Big((\mu_{1F}(W_{1j}) - \mu_{1F}(W_{1i}))\mathbb{1}\{g_{1U}(W_{1j}, \beta_1) \leq g_{1L}(W_{1i}, \beta_1)\} \leq 0 \;\Big|\; W_{1i}, W_{1j} \in \mathcal{V}\Big) = 1\Big\}
\end{aligned}
$$

$\Lambda_{\Theta,\mathcal{F}}$ is the collection of all (β, F) that satisfy both of the functional inequalities in (17) $F-$almost surely over our inference range. Before proceeding, let us formalize the notion of *contact sets*.

Contact sets.- For a given β, the contact sets are defined as the collection of all values of V for which *at least one of the functional inequalities in* (17) *is binding*. Contact sets can have positive $F-$measure, for example, if the lower and upper bounds for our control function are equal to each other with strictly positive probability. ■

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

$$
\begin{aligned}
&\overline{\Lambda}_{\Theta,\mathcal{F}} \equiv \Big\{(\beta, F) \in \Theta \times \mathcal{F} : \\
&P_F\Big(\big((\mu_{2F}(V_i) - X_{2i}'\beta_2) - (\mu_{2F}(V_j) - X_{2j}'\beta_2)\big)\mathbb{1}\{g_{1U}(W_{1j}, \beta_1) \leq g_{1L}(W_{1i}, \beta_1)\} < 0 \;\Big|\; V_i, V_j \in \mathcal{V}\Big) = 1, \\
&P_F\Big((\mu_{1F}(W_{1j}) - \mu_{1F}(W_{1i}))\mathbb{1}\{g_{1U}(W_{1j}, \beta_1) \leq g_{1L}(W_{1i}, \beta_1)\} < 0 \;\Big|\; W_{1i}, W_{1j} \in \mathcal{V}\Big) = 1\Big\}
\end{aligned}
$$

$\overline{\Lambda}_{\Theta,\mathcal{F}}$ is the collection of all (β, F) that satisfy both of the functional inequalities in (17) *strictly* (i.e, as strict inequalities), $F-$almost surely over our inference range. Thus, the contact sets have $F-$measure zero for all $(\beta, F) \in \overline{\Lambda}_{\Theta,\mathcal{F}}$. Accordingly, $\Lambda_{\Theta,\mathcal{F}} \setminus \overline{\Lambda}_{\Theta,\mathcal{F}}$ contains all (β, F) that (i) satisfy the functional inequalities, and (ii) have contact sets with strictly positive $F-$measure. Inspecting the structure of the influence function $\psi_F^{\mathcal{T}}(Z_i, \beta)$ in Lemma 1, we can see that,

$$
\begin{aligned}
&(i) \quad E_F[\psi_F^{\mathcal{T}}(Z_i, \beta)] = 0 \ \forall\ (\beta, F) \in \Theta \times \mathcal{F}. \\
&(ii) \quad \psi_F^{\mathcal{T}}(Z_i, \beta) = 0 \ \ F\text{–a.s} \ \ \forall\ (\beta, F) \in \overline{\Lambda}_{\Theta,\mathcal{F}}.
\end{aligned}
\tag{31}
$$

Denote $\sigma_F^2(\beta) \equiv E_F[\psi_F^{\mathcal{T}}(Z,\beta)^2]$. Then, $\sigma_F^2(\beta) = 0$ for all $(\beta,F) \in \overline{\Lambda}_{\Theta,\mathcal{F}}$, and $\sigma_F^2(\beta) > 0$ for all $\Lambda_{\Theta,\mathcal{F}} \setminus \overline{\Lambda}_{\Theta,\mathcal{F}}$. Thus, $\sigma_F^2(\beta)$ will be the relevant measure of the contact sets in our construction. From (31), the linear representation result in Lemma 1 yields,

$$\sup_{(\beta,F)\in\overline{\Lambda}_{\Theta,\mathcal{F}}} \left|\widehat{\mathcal{T}}(\beta)\right| = \sup_{(\beta,F)\in\overline{\Lambda}_{\Theta,\mathcal{F}}} \left|\xi_n^{\mathcal{T}}(\beta)\right| = o_p\left(\frac{1}{n^{1/2+\epsilon}}\right), \text{ therefore, } \sup_{(\beta,F)\in\overline{\Lambda}_{\Theta,\mathcal{F}}} \left|n^{1/2}\cdot\widehat{\mathcal{T}}(\beta)\right| = o_p(1), \quad (32)$$

where $\epsilon > 0$ is the constant described in Assumption 4. Thus, the statistic $\widehat{\mathcal{T}}(\beta)$ vanishes in probability faster than $n^{-1/2}$, uniformly over all (β,F) for which our inequalities are satisfied and the contact sets have F–measure zero. 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 (β,F) such that at least one of the inequalities is binding or violated with positive probability. We allow for $\sigma_F^2(\beta)$ 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,*

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

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

$$\sup_{(\beta,F)\in(\Theta\times\mathcal{F})\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}} \sup_d \left|P_F\left(\frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n \psi_F^{\mathcal{T}}(Z_i,\beta)}{\sigma_F(\beta)} \le d\right) - \Phi(d)\right| \le \frac{C}{n^{1/2}} \quad (33)$$

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

3.8.1 A regularized statistic and its asymptotic properties

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

$$t_n(\beta) \equiv \frac{\sqrt{n}\cdot\widehat{\mathcal{T}}(\beta)}{(\sigma_F(\beta)\vee\kappa)} = \begin{cases} \frac{\sqrt{n}\cdot\xi_n^{\mathcal{T}}(\beta)}{(\sigma_F(\beta)\vee\kappa)} & \forall\,(\beta,F)\in\overline{\Lambda}_{\Theta,\mathcal{F}}, \\ \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n \psi_F^{\mathcal{T}}(Z_i,\beta)}{(\sigma_F(\beta)\vee\kappa)} + \frac{\sqrt{n}\cdot\xi_n^{\mathcal{T}}(\beta)}{(\sigma_F(\beta)\vee\kappa)} & \forall\,(\beta,F)\in\Lambda_{\Theta,\mathcal{F}}\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}, \\ \frac{\sqrt{n}\cdot\mathcal{T}_F(\beta)}{(\sigma_F(\beta)\vee\kappa)} + \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n \psi_F^{\mathcal{T}}(Z_i,\beta)}{(\sigma_F(\beta)\vee\kappa)} + \frac{\sqrt{n}\cdot\xi_n^{\mathcal{T}}(\beta)}{(\sigma_F(\beta)\vee\kappa)} & \forall\,(\beta,F)\in(\Theta\times\mathcal{F})\setminus\Lambda_{\Theta,\mathcal{F}} \end{cases} \quad (34)$$

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

$$\sup_{(\beta,F)\in\Theta\times\mathcal{F}} \left|\frac{\sqrt{n}\cdot\xi_n^{\mathcal{T}}(\beta)}{(\sigma_F(\beta)\vee\kappa)}\right| = o_p\left(\frac{1}{n^{\epsilon}}\right), \quad (35)$$

where $\epsilon > 0$ is the constant described in Assumption 4. Thus, our regularized statistic vanishes in probability, uniformly over all (β, F) for which our inequalities are satisfied and the contact sets have $F-$measure zero. This result, along with the pivotal in properties in (33) can be the foundation for the construction of a CS for β_0 based on $t_n(\beta)$. Fix $\alpha \in (0,1)$ and let $z_{1-\alpha}$ be the $(1-\alpha)^{th}$ quantile of the $\mathcal{N}(0,1)$ distribution. If Assumptions 1-6 hold, Lemma 1 and the resulting properties in (32)-(35) yield,

$$\begin{aligned}
&(i) \quad \lim_{n\to\infty} \sup_{(\beta,F)\in\overline{\Lambda}_{\Theta,\mathcal{F}}} P_F\left(t_n(\beta) > z_{1-\alpha}\right) = 0, \\
&(ii) \quad \lim_{n\to\infty} \sup_{\substack{(\beta,F)\in\Lambda_{\Theta,\mathcal{F}}\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}: \\ \sigma_F(\beta)\geq\kappa}} \left| P_F\left(t_n(\beta) > z_{1-\alpha}\right) - \alpha \right| = 0, \\
&(iii) \quad \limsup_{n\to\infty} \sup_{\substack{(\beta,F)\in\Lambda_{\Theta,\mathcal{F}}\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}: \\ \sigma_F(\beta)<\kappa}} P_F\left(t_n(\beta) > z_{1-\alpha}\right) \leq \alpha.
\end{aligned} \tag{36}$$

From (36), we have

$$\liminf_{n\to\infty} \inf_{(\beta,F)\in\Lambda_{\Theta,\mathcal{F}}} P_F\left(t_n(\beta) \leq z_{1-\alpha}\right) \geq 1-\alpha \tag{37}$$

Each $(\beta, F) \in (\Theta \times \mathcal{F}) \setminus \Lambda_{\Theta,\mathcal{F}}$ violates our inequalities with strictly positive probability (according to F) over our inference range. Our previous results allow us to study the consistency and power properties of $t_n(\beta)$ as a statistic that can test for violations to our inequalities. Note from (33) that, for any $(\beta, F) \in (\Theta \times \mathcal{F}) \setminus \Lambda_{\Theta,\mathcal{F}}$, and any given c,

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

Combined with (34)-(35), this yields $\lim_{n\to\infty} P_F\left(t_n(\beta) > z_{1-\alpha}\right) = 1$ for each $(\beta, F) \in (\Theta \times \mathcal{F}) \setminus \Lambda_{\Theta,\mathcal{F}}$. More generally, take any sequence in $(\beta_n, F_n) \in (\Theta \times \mathcal{F}) \setminus \Lambda_{\Theta,\mathcal{F}}$ such that $\mathcal{T}_{F_n}(\beta_n) \geq \delta_n n^{-1/2} D$ for some fixed $D > 0$ and some sequence of positive constants $\delta_n \to \infty$. The results in (33), (34) and (35) yield,

$$\lim_{n\to\infty} P_{F_n}\left(t_n(\beta_n) > z_{1-\alpha}\right) = 1. \tag{38}$$

The result in (38) establishes *consistency* of the statistic $t_n(\beta)$ for detecting violations to our inequalities. To characterize its *local power* properties, take any sequence $(\beta_n, F_n) \in (\Theta \times \mathcal{F}) \setminus \Lambda_{\Theta,\mathcal{F}}$ such that

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

Note that $s_1 \geq 1$ and $s_2 \geq 0$. From the results in (33), (34) and (35), we have,

$$\lim_{n\to\infty} P_{F_n}\left(t_n(\beta_n) > z_{1-\alpha}\right) = 1 - \Phi\left(s_1 \cdot z_{1-\alpha} - s_1 \cdot s_2\right) \tag{39}$$

Having $\lim_{n\to\infty} P_{F_n}\left(t_n(\beta_n) > z_{1-\alpha}\right) > \alpha$ corresponds to the notion of *nontrivial asymptotic power* (see Lee, Song, and Whang (2018, Definition 3)) for detecting violations to our inequalities. (39) implies that we will have nontrivial asymptotic power for the type of sequences described above iff $s_1 \cdot z_{1-\alpha} - s_1 \cdot s_2 < z_{1-\alpha}$. Equipped with the above results, we are almost ready to construct a CS for β_0. The final step is to obtain an estimator for $\sigma_F^2(\beta)$.

3.9 Construction of a confidence set for β_0

We propose to construct a confidence set (CS) for β_0 based on the properties of the statistic $t_n(\beta)$ described in (37)-(39). Before proceeding, we need to construct an estimator for $\sigma_F^2(\beta)$.

3.9.1 An estimator for $\sigma_F^2(\beta)$

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

$$\widehat{\sigma}^2(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n} \widehat{\psi}^{\mathcal{T}}(Z_i,\beta)^2, \quad \text{where} \quad \widehat{\psi}^{\mathcal{T}}(z,\beta) \equiv \widehat{\psi}^{\mathcal{T}_2}(z,\beta) + \widehat{\psi}^{\mathcal{T}_1}(z,\beta_1),$$

$\widehat{\psi}^{\mathcal{T}_2}(Z_i,\beta)$ and $\widehat{\psi}^{\mathcal{T}_1}(Z_i,\beta_1)$ are estimators of $\psi_F^{\mathcal{T}_2}(Z_i,\beta)$ and $\psi_F^{\mathcal{T}_1}(Z_i,\beta_1)$. We estimate $\psi_F^{\mathcal{T}_2}(Z_i,\beta)$ with $\widehat{\psi}^{\mathcal{T}_2}(Z_i,\beta) \equiv 2 \cdot \widehat{H}_1^{\mathcal{T}_2}(V_i,\beta) + \widehat{H}_2^{\mathcal{T}_2}(Z_i,\beta)$ 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 Lemma 1, we estimate

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

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

Using the definitions in (29), the estimators on the right-hand side of the previous expressions are,

$$\widehat{\eta_a^{\tau_2}}(v,\beta) \equiv \frac{1}{n}\sum_{j=1}^{n}\left(\widehat{R}_2(V_j) - \left(X_{2j}'\beta_2 - x_2'\beta_2\right)\widehat{Q}_2(V_j)\right)\mathbb{1}\left\{g_{1U}(w_1,\beta_1) \le g_{1L}(W_{1j},\beta_1)\right\}\phi_2(V_j)\mathbb{1}\left\{\widehat{\tau_2}(V_j,v,\beta) \ge -b_n\right\},$$

$$\widehat{\eta_b^{\tau_2}}(v,\beta) \equiv \frac{1}{n}\sum_{j=1}^{n}\left(\widehat{R}_2(V_j) - \left(X_{2j}'\beta_2 - x_2'\beta_2\right)\widehat{Q}_2(V_j)\right)\mathbb{1}\left\{g_{1U}(W_{1j},\beta_1) \le g_{1L}(w_1,\beta_1)\right\}\phi_2(V_j)\mathbb{1}\left\{\widehat{\tau_2}(v,V_j,\beta) \ge -b_n\right\},$$

$$\widehat{\eta_c^{\tau_2}}(v,\beta) \equiv \frac{1}{n}\sum_{j=1}^{n}\widehat{Q}_2(V_j)\mathbb{1}\left\{g_{1U}(W_{1j},\beta_1) \le g_{1L}(w_1,\beta_1)\right\}\phi_2(V_j)\mathbb{1}\left\{\widehat{\tau_2}(v,V_j,\beta) \ge -b_n\right\},$$

$$\widehat{\eta_d^{\tau_2}}(v,\beta) \equiv \frac{1}{n}\sum_{j=1}^{n}\widehat{Q}_2(V_j)\mathbb{1}\left\{g_{1U}(w_1,\beta_1) \le g_{1L}(W_{1j},\beta_1)\right\}\phi_2(V_j)\mathbb{1}\left\{\widehat{\tau_2}(V_j,v,\beta) \ge -b_n\right\},$$

with $\widehat{R}_2$ and $\widehat{Q}_2$ as described in (27), 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 $\psi_F^{\mathcal{T}_1}(Z_i,\beta_1)$ is given by $\widehat{\psi^{\mathcal{T}_1}}(Z_i,\beta_1) \equiv 2\cdot\widehat{H}_1^{\mathcal{T}_1}(W_{1i},\beta_1) + \widehat{H}_2^{\mathcal{T}_1}(Z_i,\beta_1)$ 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 Lemma 1, we estimate

$$\widehat{H}_1^{\mathcal{T}_1}(w_1,\beta_1) \equiv$$
$$\frac{1}{2}\cdot\frac{1}{n}\sum_{j=1}^{n}\left[\widehat{\tau_1}(w_1,W_{1j},\beta_1)\mathbb{1}\left\{\widehat{\tau_1}(w_1,W_{1j},\beta_1) \ge -b_n\right\} + \widehat{\tau_1}(W_{1j},w_1,\beta_1)\mathbb{1}\left\{\widehat{\tau_1}(W_{1j},w_1,\beta_1) \ge -b_n\right\}\right] - \widehat{\mathcal{T}_1}(\beta_1),$$

$$\widehat{H}_2^{\mathcal{T}_1}(z,\beta_1) \equiv \left(\left(\widehat{\eta_a^{\tau_1}}(w_1,\beta_1) - \widehat{\eta_b^{\tau_1}}(w_1,\beta_1)\right) + \left(\widehat{\eta_c^{\tau_1}}(w_1,\beta_1) - \widehat{\eta_d^{\tau_1}}(w_1,\beta_1)\right)\cdot y_1\right)\cdot\widehat{f_{W_1}}(w_1)\cdot\phi_1(w_1)^2$$
$$-\frac{1}{n}\sum_{j=1}^{n}\left[\left(\left(\widehat{\eta_a^{\tau_1}}(W_{1j},\beta_1) - \widehat{\eta_b^{\tau_1}}(W_{1j},\beta_1)\right) + \left(\widehat{\eta_c^{\tau_1}}(W_{1j},\beta_1) - \widehat{\eta_d^{\tau_1}}(W_{1j},\beta_1)\right)\cdot Y_{1j}\right)\cdot\widehat{f_{W_1}}(W_{1j})\cdot\phi_1(W_{1j})^2\right].$$

Using the definitions in (30), the estimators on the right-hand side of the previous expressions are,

$$\widehat{\eta_a^{\tau_1}}(w_1,\beta_1) \equiv \frac{1}{n}\sum_{j=1}^{n}\widehat{R}_1(W_{1j})\mathbb{1}\left\{g_{1U}(w_1,\beta_1) \le g_{1L}(W_{1j},\beta_1)\right\}\mathbb{1}\left\{\widehat{\tau_1}(W_{1j},w_1,\beta) \ge -b_n\right\}\phi_1(W_{1j}),$$

$$\widehat{\eta_b^{\tau_1}}(w_1,\beta_1) \equiv \frac{1}{n}\sum_{j=1}^{n}\widehat{R}_1(W_{1j})\mathbb{1}\left\{g_{1U}(W_{1j},\beta_1) \le g_{1L}(w_1,\beta_1)\right\}\mathbb{1}\left\{\widehat{\tau_1}(w_1,W_{1j},\beta) \ge -b_n\right\}\phi_1(W_{1j}),$$

$$\widehat{\eta_c^{\tau_1}}(w_1,\beta_1) \equiv \frac{1}{n}\sum_{j=1}^{n}\widehat{Q}_1(W_{1j})\mathbb{1}\left\{g_{1U}(W_{1j},\beta_1) \le g_{1L}(w_1,\beta_1)\right\}\mathbb{1}\left\{\widehat{\tau_1}(w_1,W_{1j},\beta) \ge -b_n\right\}\phi_1(W_{1j}),$$

$$\widehat{\eta_d^{\tau_1}}(w_1,\beta_1) \equiv \frac{1}{n}\sum_{j=1}^{n}\widehat{Q}_1(W_{1j})\mathbb{1}\left\{g_{1U}(w_1,\beta_1) \le g_{1L}(W_{1j},\beta_1)\right\}\mathbb{1}\left\{\widehat{\tau_1}(W_{1j},w_1,\beta) \ge -b_n\right\}\phi_1(W_{1j})$$

In Section A4 of the online appendix we show that, under the conditions of Lemma 1, we have

$$\sup_{\beta\in\Theta}\left|\widehat{\sigma}^2(\beta)-\sigma_F^2(\beta)\right| = o_p(1), \quad \text{uniformly over } \mathcal{F}. \tag{40}$$

From here, we take the asymptotic properties of $t_n(\beta)$ and we replace $\sigma_F(\beta)$ with $\widehat{\sigma}(\beta)$ to construct a CS.

3.9.2 Confidence set

Let,

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

We wish to construct a CS that contains β_0 with asymptotic target coverage probability $1-\alpha$. Based on the results in (37)-(40), we construct our CS as,

$$\widehat{CS}_{1-\alpha} \equiv \left\{\beta\in\Theta : \widehat{t}_n(\beta)\leq z_{1-\alpha}\right\},$$

where $z_{1-\alpha}$ is the $(1-\alpha)^{th}$ quantile of the $\mathcal{N}(0,1)$ distribution. From our previous results, the properties of $\widehat{CS}_{1-\alpha}$ are summarized in the following theorem.

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

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

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

(iii) Nontrivial local power of the associated test for $(\beta,F)\in\Lambda_{\Theta,\mathcal{F}}$: Take any sequence $(\beta_n,F_n)\in(\Theta\times\mathcal{F})\setminus\Lambda_{\Theta,\mathcal{F}}$ such that $\lim_{n\to\infty}\frac{(\sigma_{F_n}(\beta_n)\vee\kappa)}{\sigma_{F_n}(\beta_n)}=s_1$ *and* $\lim_{n\to\infty}\frac{\sqrt{n}\cdot\mathcal{T}_{F_n}(\beta_n)}{(\sigma_{F_n}(\beta_n)\vee\kappa)}=s_2$ *(note that $s_1\geq 1$ and $s_2\geq 0$). We have* $\lim_{n\to\infty}P_{F_n}(\beta_n\in\widehat{CS}_{1-\alpha})<1-\alpha$ *if $s_1\cdot z_{1-\alpha}-s_1\cdot s_2<z_{1-\alpha}$.* ■

Proof: Combined with the result in (40), part (i) of Theorem 1 follows from (37), while parts (ii) and (iii) follow from (38) and (39), respectively. ■

3.9.3 Asymptotic adaptation to 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). Like ours, all of these methods rely on tuning parameters. In our case, how close the functional inequalities come to binding is related to the measure of the contact sets, as we discussed previously. The properties of the sequence b_n are designed to ensure that the estimators $\widehat{\mathcal{T}}(\beta)$ and $\widehat{\sigma}(\beta)$ adapt asymptotically to the measure of the contact sets, captured in our case by $\sigma_F^2(\beta)$. Our regularization allows us to obtain a statistic with asymptotically pivotal properties without having to estimate the contact sets themselves in a preliminary step.

3.9.4 Replacing the regularization constant κ with a vanishing sequence

Our assumptions allow for $\sigma_F^2(\beta)$ (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 that $\sigma_F^2(\beta)$ is bounded away from zero uniformly over $(\Theta\times\mathcal{F})\setminus\overline{\Lambda}_{\Theta,\mathcal{F}}$, we can replace our regularization constant κ with a positive sequence $\kappa_n \to 0$ that vanishes asymptotically. We show this in Section A5 of the online appendix.

3.10 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 development of a general theory of how to choose these tuning parameters for future work, we can provide recommendations backed by the results of our Monte Carlo experiments in Section 4. We follow the usual path and consider covariate-specific bandwidths for each continous regressor $V_m^c \in V_c$ 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. With samples of sizes $n = 1{,}000$, $n = 2{,}000$ and $n = 3{,}000$, in our Monte Carlo experiments, our bandwidth choice was $h_n \approx 0.28\cdot\widehat{\sigma}(V_m^c)$. For reference, Silverman's so-called "rule of thumb" (Silverman (1986, p. 45)) would yield $h_n \approx 0.27\cdot\widehat{\sigma}(V_m^c)$ when $n = 1{,}000$.

Given the type of bandwidths we use, we can obtain a lower bound for the value of M in Assumptions 2 and 4. Recall that r denotes the number of continuously distributed covariates in V. Fix $\epsilon > 0$. 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 $\frac{2\epsilon+\delta}{1-2(\epsilon+\delta)} < \frac{1}{4r}$. The order of kernels we use in our experiments corresponds to this value of M.

We advocate for the choices of b_n and κ to be proportional to a measure of the scale of an *envelope* of the functions $\tau_{2F}(V_i, V_j, \beta)$ and $\tau_{1F}(W_{1i}, W_{1j}, \beta_1)$ over Θ. Denote $\overline{\beta}_2 \equiv \sup_{\beta_2 \in \Theta} \|\beta_2\|$, and

$$\begin{aligned}\overline{\tau}_{2F}(V_i, V_j) \equiv &\Big(\Big(\big|R_{2F}(V_i)Q_{2F}(V_j)\big| + \big|R_{2F}(V_j)Q_{2F}(V_i)\big|\Big) + \overline{\beta}_2\Big(\|X_{2i}\| + \|X_{2j}\|\Big)\big|Q_{2F}(V_i)Q_{2F}(V_j)\big|\Big) \\ &\cdot \phi_2(V_i)\phi_2(V_j), \\ \overline{\tau}_{1F}(W_{1i}, W_{1j}) \equiv &\Big(\big|R_{1F}(W_{1j})Q_{1F}(W_{1i})\big| + \big|R_{1F}(W_{1i})Q_{1F}(W_{1j})\big|\Big) \cdot \phi_1(W_{1i})\phi_1(W_{1j}).\end{aligned} \tag{41}$$

By construction, $|\tau_{2F}(V_i, V_j, \beta)| \leq \overline{\tau}_{2F}(V_i, V_j)$ and $|\tau_{1F}(W_{1i}, W_{1j})| \leq \overline{\tau}_{1F}(W_{1i}, W_{1j})$ $\forall\ \beta \in \Theta$, so the functions in (41) are envelope functions for $\tau_{2F}(V_i, V_j, \beta)$ and $\tau_{1F}(W_{1i}, W_{1j}, \beta_1)$ over Θ. Our proposal is to choose b_n and κ to be proportional to a measure of the scale of these envelope functions. We can use $b_n^{\tau_2} = c_b \cdot \widehat{\overline{\tau}}_{2_{(0.5)}} \cdot n^{-\alpha_b}$ for the first inequality, and $b_n^{\tau_1} = c_b \cdot \widehat{\overline{\tau}}_{1_{(0.5)}} \cdot n^{-\alpha_b}$ for the second inequality, where $\widehat{\overline{\tau}}_{2_{(0.5)}}$ and $\widehat{\overline{\tau}}_{1_{(0.5)}}$ denote the sample medians of $\widehat{\overline{\tau}}_{2F}(V_i, V_j)$ and $\widehat{\overline{\tau}}_{1F}(W_{1i}, W_{1j})$, respectively. 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 Assumption 4. From here, we construct our statistic as,

$$\begin{aligned}\widehat{\mathcal{T}}_2(\beta) &\equiv \frac{1}{n \cdot (n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \widehat{\tau}_2(V_i, V_j, \beta) \cdot \mathbb{1}\Big\{\widehat{\tau}_2(V_i, V_j, \beta) \geq -b_n^{\tau_2}\Big\}, \\ \widehat{\mathcal{T}}_1(\beta_1) &\equiv \frac{1}{n \cdot (n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \widehat{\tau}_1(W_{1i}, W_{1j}, \beta_1) \cdot \mathbb{1}\Big\{\widehat{\tau}_1(W_{1i}, W_{1j}, \beta_1) \geq -b_n^{\tau_1}\Big\}, \\ \widehat{\mathcal{T}}(\beta) &\equiv \widehat{\mathcal{T}}_2(\beta) + \widehat{\mathcal{T}}_1(\beta_1),\end{aligned} \tag{42}$$

We propose a regularization parameter κ of the form $\kappa = c_\kappa \cdot (\widehat{\overline{\tau}}_{2_{(0.5)}} \wedge \widehat{\overline{\tau}}_{1_{(0.5)}})$. In our Monte Carlos, we set the values of the constants c_b and c_κ small enough (in the order of 10^{-20}) that the tuning parameters b_n and κ were equivalent to zero in our computations. This was done to enhance the power of the underlying test in the construction of our confidence sets and, with this, the informative properties of our results. Overall, with the above choices, our experiments results yielded informative confidence intervals for the parameters of the model, and coverage frequencies for the true parameter value that were in line with our asymptotic results. We describe the details in Section 4.

3.11 Sharp identified set and our approach

Our approach is based on pairwise inequalities that are sharp, in the sense that they exploit all the restrictions in Assumption 1 for the functionals involved, and cannot be improved upon without additional restrictions. From here, we constructed a CS based on a statistic that detects violations to our pairwise inequalities that occur with nonzero probability over a pre-specified inference range. While the functional inequalities we rely upon are sharp given our assumptions, we do not make the claim that they are *sufficient* to fully characterize the sharp identified set for the parameters themselves. This region can be described, at a high-level, using the definition of observational equivalence for incomplete models in Chesher and Rosen (2017). Given the nature of our model, an inferential procedure based on the exact characterization of the sharp identified set of parameters would require searching over all observationally equivalent *structures* (β, λ_F, H_F) that are consistent with our assumptions. This would require us to address the issue of how to approximate the nonparametric functionals λ_F and H_F. One of the advantages of our approach (and the usual pairwise-difference methods in general) is precisely the ability to bypass this problem. We conclude this discussion by pointing back to Section 2.1.2 and Result 1, where we showed that there exist conditions under which the parameters of the model can be point-identified through our pairwise inequalities. We leave the problem of inference on the sharp identified set for the type of models studied here for future research.

4 Monte Carlo experiments

To study the finite-sample performance of our methodology we revisit the interval-data example from Section 2.1. Our selection and outcome equations are given by,

$$\begin{aligned} Y_1^* &= \beta_{10}^1 \cdot X_1^1 + \beta_{10}^2 \cdot X_1^2 + \varepsilon_1, \\ Y_2^* &= \beta_{20}^0 + \beta_{20}^1 \cdot X_2^1 + \beta_{20}^2 \cdot X_1^1 + \varepsilon_2. \end{aligned} \tag{43}$$

Note that the regressor X_1^1 appears in both the outcome and selection equations. The true parameter values are set to 1 for all the slope coefficients: $\beta_{10}^1 = \beta_{10}^2 = \beta_{20}^1 = \beta_{20}^2 = 1$, and the intercept in the outcome equation is set to $\beta_{20}^0 = 0.5$. The latent variables $(\varepsilon_1, \varepsilon_2)$ are distributed as,

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix} \right)$$

We set $\sigma_{12} = 0.5$. The regressors X_1^1 and X_2^1 are generated as i.i.d $\mathcal{N}(0,1)$, independent of $(\varepsilon_1, \varepsilon_2)$.

We assume a setting where the econometrician observes the regressors X_1^1 and X_2^1, but only observes interval data $(\underline{X}_1^2, \overline{X}_1^2)$ for X_1^2, with $\underline{X}_1^2 \leq X_1^2 \leq \overline{X}_1^2$ w.p.1. First, we generate $\underline{X}_1^2 \sim \mathcal{N}(0,1)$. From here, we generate X_1^2 and $\overline{X}_1^2$ as follows. Let $\alpha_1 \sim U[0,1]$ and $\xi_1 \sim U[0, \Delta_0]$, with $\Delta_0 > 0$ being a constant parameter. We generate, $\overline{X}_1^2 = \underline{X}_1^2 + \xi_1$, and $X_1^2 = \alpha_1 \cdot \underline{X}_1^2 + (1 - \alpha_1) \cdot \overline{X}_1^2$. The random variables $(X_1^1, X_2^1, \underline{X}_1^2, \alpha_1, \xi_1)$ are all independent of each other, and independent of $(\varepsilon_1, \varepsilon_2)$. Following the notation in the paper, we have $V \equiv (X_1^1, X_2^1, \underline{X}_1^2, \overline{X}_1^2)$ and $W_1 \equiv (X_1^1, \underline{X}_1^2, \overline{X}_1^2)$. The parameter Δ_0 is a measure of how wide our bounds are for the unobserved regressor X_1^2, since $0 \leq \overline{X}_1^2 - \underline{X}_1^2 \leq \Delta_0$, and $E[\overline{X}_1^2 - \underline{X}_1^2] = \frac{\Delta_0}{2}$. Some of our analysis will look at various values of Δ_0.

4.1 Pairwise inequalities in our designs

The pairwise inequalities are those given in equation (9). We will refer to the first and the second inequalities in (9) as the "outcome-equation inequality" and "selection-equation inequality" respectively. That is,

$$\begin{aligned}\Big(\big(\mu_{2F}(V_i) - X_{2i}'\beta_{20}\big) - \big(\mu_{2F}(V_j) - X_{2j}'\beta_{20}\big)\Big)\mathbb{1}\Big\{X_{1Uj}'\beta_{10} \leq X_{1Li}'\beta_{10}\Big\} \leq 0 & \quad \text{outcome-equation inequality} \\ \big(\mu_{1F}(W_{1j}) - \mu_{1F}(W_{1i})\big)\cdot \mathbb{1}\Big\{X_{1Uj}'\beta_{10} \leq X_{1Li}'\beta_{10}\Big\} \leq 0 & \quad \text{selection-equation inequality}\end{aligned}$$

In our case, we have $X_{1U}'\beta_{10} = \beta_{10}^1 \cdot X_1^1 + \beta_{10}^2 \cdot \overline{X}_1^2$ and $X_{1L}'\beta_{10} = \beta_{10}^1 \cdot X_1^1 + \beta_{10}^2 \cdot \underline{X}_1^2$ as the bounds for the unobserved control function $\beta_{10}^1 \cdot X_1^1 + \beta_{10}^2 \cdot X_1^2$. As we noted previously, an intercept in either equation is differenced out in our inequalities, so our relevant parameter space will include only the slope coefficients in both equations.

4.1.1 Identification power of the pairwise inequalities in our designs

Our design is consistent with the conditions of Result 1, which predicts that parameter values $\beta \neq \beta_0$ should violate at least one of our inequalities with nonzero probability. Tables 1 and 2 explore this for a few examples, and computes the probabilities of violations of our pairwise inequalities for parameter values $\beta \neq \beta_0$. In each case, we compute the probability of violations for different values of Δ_0. Smaller values of this parameter indicate tighter bounds for our unobserved control variable, and we expect that the probability of violations should be larger for smaller values of Δ_0. This intuition is confirmed by our results. While each $\beta \neq \beta_0$ analyzed violated both of our inequalities with positive probability, the probability of violations decreased with Δ_0.

Table 1 computes probabilities of violations to our inequalities when we change the value of one parameter at a time, keeping the rest fixed at their true values. Not surprisingly, these probabilities are greater for values that are farther away from the truth, but they are nonzero even for parameter values that are very close to β_0. As we expected, these probabilities decrease with Δ_0. In all cases, we see that both inequalities are violated with nonzero probability whenever whenever $\beta_1 \neq \beta_{10}$, so both the selection and outcome-equation inequalities are informative for β_{10}. The parameters of the outcome equation, β_{20}, only appear in the outcome-equation inequality, and we see that this inequality was violated with positive probability whenever $\beta_2 \neq \beta_{20}$. Table 2 computes the probability of violations for some additional parameter values $\beta \neq \beta_0$. Our findings there are in line with those in Table 1. Overall, our results are illustrative of the identification power of our inequalities, and they are consistent with the predictions of Result 1.

Table 1: Caption 1

(A) Outcome equation parameters: Probability of violating the outcome-equation inequality by parameter values $\beta_2 \neq \beta_{20}$							
	Alternative values of β_2^1 (true value is $\beta_{20}^1 = 1$) fixing all other parameters at their true values						
	$\beta_2^1 = -1$	$\beta_2^1 = 0$	$\beta_2^1 = 0.5$	$\beta_2^1 = 0.99$	$\beta_2^1 = 1.01$	$\beta_2^1 = 1.5$	$\beta_2^1 = 2$
$\Delta_0 = 0.25$	0.206	0.177	0.133	1.4E-3	1.3E-3	0.132	0.178
$\Delta_0 = 0.50$	0.195	0.167	0.123	9.1E-4	9.0E-4	0.123	0.167
$\Delta_0 = 1.00$	0.172	0.146	0.105	6.2E-4	6.1E-4	0.105	0.146
	Alternative values of β_2^2 (true value is $\beta_{20}^2 = 1$) fixing all other parameters at their true values						
	$\beta_2^2 = -1$	$\beta_2^2 = 0$	$\beta_2^2 = 0.5$	$\beta_2^2 = 0.99$	$\beta_2^2 = 1.01$	$\beta_2^2 = 1.5$	$\beta_2^2 = 2$
$\Delta_0 = 0.25$	0.335	0.300	0.217	1.1E-3	6.2E-4	0.059	0.079
$\Delta_0 = 0.50$	0.323	0.289	0.208	8.4E-4	3.7E-4	0.050	0.069
$\Delta_0 = 1.00$	0.296	0.265	0.190	6.8E-4	2.2E-4	0.036	0.053

Probabilities obtained from 50 million simulations.

(B) Selection equation parameters: Probability of violating either the outcome-equation inequality or the selection-equation inequality by parameter values $\beta_1 \neq \beta_{10}$							
	Alternative values of β_1^1 (true value is $\beta_{10}^1 = 1$) fixing all other parameters at their true values						
	$\beta_1^1 = -1$	$\beta_1^1 = 0$	$\beta_1^1 = 0.5$	$\beta_1^1 = 0.99$	$\beta_1^1 = 1.01$	$\beta_1^1 = 1.5$	$\beta_1^1 = 2$
$\Delta_0 = 0.25$	0.238	0.108	0.037	1.6E-6	1.9E-6	0.023	0.044
$\Delta_0 = 0.50$	0.225	0.093	0.027	3.6E-7	2.8E-7	0.016	0.037
$\Delta_0 = 1.00$	0.199	0.067	0.014	4.0E-8	1.0E-7	0.008	0.027
	Alternative values of β_1^2 (true value is $\beta_{10}^2 = 1$) fixing all other parameters at their true values						
	$\beta_1^2 = -1$	$\beta_1^2 = 0$	$\beta_1^2 = 0.5$	$\beta_1^2 = 0.99$	$\beta_1^2 = 1.01$	$\beta_1^2 = 1.5$	$\beta_1^2 = 2$
$\Delta_0 = 0.25$	0.262	0.125	0.044	2.5E-6	1.5E-6	0.019	0.037
$\Delta_0 = 0.50$	0.275	0.125	0.037	4.4E-7	4.8E-7	0.011	0.027
$\Delta_0 = 1.00$	0.300	0.126	0.027	1.6E-7	1.8E-7	0.004	0.014

Probabilities obtained from 50 million simulations.

4.2 Parameter space for our implementation

As we described in Section 2.1.2, the scale of the slope coefficients β_1 in the selection equation needs to be normalized. We did so in our implementation by fixing $\beta_1^1 = 1$. From here, the parameter space for $(\beta_{12}, \beta_{21}, \beta_{22})$ was set to be $[0,4] \times [0,4] \times [0,4]$. In our simulations, the parameter Δ_0 was set to $\Delta_0 = 1$. Recall that this parameter is a measure of the width of the bounds for our selection-equation control function. We have chosen the largest value among those analyzed in Table 1. As we described previously, we set the covariance between $(\varepsilon_1, \varepsilon_2)$ to $\sigma_{12} = 0.5$.

Table 2: Caption 2

	Probability of violating the selection-equation inequality				Probability of violating the outcome-equation inequality			
	β^a	β^b	β^c	β^d	β^a	β^b	β^c	β^d
$\Delta_0 = 0.25$	0.166	0.043	0.003	6.0E-4	0.022	0.006	1.4E-5	1.9E-6
$\Delta_0 = 0.50$	0.159	0.034	0.002	3.8E-4	0.016	0.003	3.2E-6	3.6E-7
$\Delta_0 = 1.00$	0.147	0.022	0.001	2.4E-4	0.008	7.0E-4	6.4E-7	4.0E-8

Probabilities obtained from 50 million simulations.
Description of parameter values analyzed:
β^a: $\beta_1^1 = 1.2$, $\beta_1^2 = 0.8$, $\beta_2^1 = 1.25$, $\beta_2^2 = 0.75$
β^b: $\beta_1^1 = 0.9$, $\beta_1^2 = 1.1$, $\beta_2^1 = 0.9$, $\beta_2^2 = 1.1$
β^c: $\beta_1^1 = 1.01$, $\beta_1^2 = 0.99$, $\beta_2^1 = 1.01$, $\beta_2^2 = 0.99$
β^d: $\beta_1^1 = 0.995$, $\beta_1^2 = 1.005$, $\beta_2^1 = 0.995$, $\beta_2^2 = 1.005$
True parameter values: $\beta_{10}^1 = \beta_{10}^2 = \beta_{20}^1 = \beta_{20}^2 = 1$.

4.3 Inference range, kernels and bandwidths

We have $r \equiv 4$ continuously distributed observable conditioning variables, $V \equiv (X_1^1, X_2^1, \underline{X}_1^2, \overline{X}_1^2)$. Our inference range $\mathcal{V}$ included all the observations i for which each element in V_i is between the 0.001 and the 0.999 quantiles in the sample. Thus, in a sample of size $n = 1,000$, we only eliminate observations that include the smallest or the largest observed values of V_i element-wise. The weight functions $\phi_2(V_i)$ and $\phi_1(W_{1i})$ are simply indicator functions for the event that $V_i \in \mathcal{V}$ and $W_{1i} \in \mathcal{V}$, respectively. Next, since $r = 4$, in order to comply with the restrictions in Assumption 4, and following the discussion in Section 3.10, we use a bias-reducing kernel of order $M = 10$. Our kernel is multiplicative, $K(v) = \prod_{m=1}^4 \kappa(v_m)$, with $\kappa(z) = \sum_{\ell=1}^5 c_\ell \cdot (S^2 - z^2)^{2\ell} \cdot \mathbb{1}\{|z| \le S\}$. By construction, $\kappa(z)$ is symmetric around zero, with support $[-S, S]$. In our experiments we used $S = 4$. The coefficients c_ℓ, are chosen to satisfy the conditions of a bias-reducing kernel of order $M = 10$. Note that our functionals satisfy the smoothness requirements stated in our assumptions for any value of M.

Regarding the bandwidth choice, we applied our procedure for generated samples of sizes $n = 1,000$, $n = 2,000$ and $n = 3,000$. We employed covariate-specific bandwidths for each element of V. We used, $h_n(V^\ell) = 0.32 \cdot \widehat{\sigma}(V^\ell)$ for $n = 1,000$, $h_n(V^\ell) = 0.28 \cdot \widehat{\sigma}(V^\ell)$ for $n = 2,000$, and $h_n(V^\ell) = 0.26 \cdot \widehat{\sigma}(V^\ell)$ for $n = 3,000$. For reference, Silverman's so-called "rule of thumb" (Silverman (1986, p. 45)) would yield $h_n(V^\ell) \approx 0.27 \cdot \widehat{\sigma}(V_m^c)$ for a sample of size $n = 1,000$. Our bandwidths are slightly larger in magnitude reflecting the fact that their rates of convergence are slower than those of their rule-of-thumb counterparts. Our choices of $b_n^{\tau_\ell}$ and κ were set small enough that they were computationally equivalent to zero in our experiments[16]. This was done in an effort

[16]Strictly speaking, we followed the prescription we described in Section 3.10 (see equations 41 and 42), with $b_n^{\tau_\ell} = 10^{-20} \cdot \widehat{\overline{\tau}}_{\ell(0.50)}$ for $\ell = 1, 2$, and $\kappa = 10^{-20} \cdot (\widehat{\overline{\tau}}_{2(0.5)} \wedge \widehat{\overline{\tau}}_{1(0.5)})$ for all three sample sizes n analyzed. The proportionality constant of 10^{-20} was small enough to make these tuning parameters computationally equivalent to zero in

to enhance the power of the underlying test in the construction of our CS and, with this, the informative properties of our results.

4.4 Results

4.4.1 Confidence intervals for individual parameters

We simulated 500 samples of sizes $n = 1,000$, $n = 2,000$ and $n = 3,000$. For each simulated sample, we constructed a CS with target coverage probability 95% using our approach. Each CS was obtained from a randomly generated grid of points. This grid was the same for all simulations, to facilitate comparisons across simulations. Table 3 includes confidence intervals (CI) for each individual parameter. These CIs are constructed as projections from the CS as follows. We found the smallest and the largest values for each parameter among the parameter values that were included in our CS with frequency at least 95% in our simulations. As our results show, the CIs become tighter as n grows, and they all include the true parameter values. The width of the CI is shortest for the slope coefficient β_1^2 in the selection equation; this is perhaps not surprising since, as we discussed above, both of our pairwise inequalities are informative about β_1^2. The widths of the CIs for the outcome equation coefficients were comparable. The width was slightly larger for β_2^2, the slope coefficient of the only regressor that appears in both equations. We do not know whether this is indicative of a general result where our procedure is more informative for coefficients of regressors that appear only in one equation, or if it is simply a feature of our experiment designs. As n grows, all of our CIs become tighter and increasingly bounded away from the boundary of the parameter space. Our results are in line with our discussion in Section 2.1.2, which illustrated that our pairwise inequalities are informative for the parameters of the model.

Table 3: Individual confidence intervals obtained as projections from confidence sets with target coverage probability 95%

	Outcome equation		Selection equation
	β_2^1	β_2^2	β_1^2
$n = 1,000$	[0.019 , 2.126]	[0.267 , 3.229]	[0.611 , 2.298]
$n = 2,000$	[0.123 , 1.881]	[0.594 , 2.566]	[0.748 , 1.801]
$n = 3,000$	[0.478 , 1.621]	[0.840 , 2.279]	[0.795 , 1.627]

Results show the smallest and largest values for each parameter among all parameter values included in our CS with frequency ≥ 95% in 500 simulations. Parameter space for $(\beta_2^1, \beta_2^2, \beta_1^2)$ is $[0,4]\times[0,4]\times[0,4]$. β_1^1 is normalized to 1. True parameter values are $\beta_{20}^1 = \beta_{20}^2 = \beta_{10}^2 = \beta_{10}^1 = 1$.

our experiments.

4.4.2 Coverage of the true parameter value

In our designs, the probability that the lower and upper bounds for the unobserved control variable are equal to each other is zero and, evaluated at the true parameter value, the contact sets have measure zero. In this case, Theorem 1 predicts that our confidence set will include the true parameter value with asymptotic probability 1 (see part (ii) of equation (36)). Table 4 presents the frequency with which we would have included the true parameter value in our 95% CS. That is, the frequency with which our test-statistic, evaluated at the true parameter value, was below the $z_{0.95}$ critical value. Our results are in line with the asymptotic predictions of Theorem 1, with coverage frequencies very close to 1 in each of the sample sizes analyzed.

Table 4: Frequency with which the true parameter value was included in our confidence sets with target coverage probability 95%

$n = 1,000$	$n = 2,000$	$n = 3,000$
0.998	0.998	0.996

500 simulations

Overall, our Monte Carlo experiments suggest that the implementation of our approach with arguably "intuitive" tuning parameter choices leads to reasonably informative results consistent with our asymptotic predictions.

5 Other cases where control functions may be partially identified

This paper focused on models where control functions are partially identified because some control variables are unobserved, but (lower and upper) bounds are available for them. There are other scenarios where control functions may be partially identified. We briefly discuss two possibilities here.

5.1 Models where all regressors are observed, but the parameters of the control functions are partially identified

For illustration, let us keep our focus on our bivariate sample-selection model. Suppose now that all regressors in both equations (X_1, X_2) are observed, but that the parameters of the selection equation, β_{10} are (possibly) partially identified. Due to the discrete-choice nature of the selection equation, this can happen, for example, if X_1 consists only of discrete regressors (Khan, Ouyang, and Tamer (2021)). Suppose we maintain the exclusion restrictions of Assumption 1. This predicts that $P_F(Y_1 = 1|W_1, X_1) = P_F(Y_1 = 1|X_1'\beta_{10})$ a.s for the selection equation. What we have in this instance is a generalization of the pairwise-difference models studied by James L. Powell and coauthors, with the only caveat that the parameters are (possibly) partially identified. Leaving aside issues of sharpness for the moment, a characterization of the identified set of parameters can be given by a combination of the above exclusion restriction and the usual pairwise conditional moment condition in the model,

$$\Theta^I \equiv \Big\{(\beta_1,\beta_2)\in\Theta : P_F(Y_{1i}=1|X_{1i}) = P_F(Y_{1i}=1|X_{1i}'\beta_1), \quad \text{and}$$
$$E_F\Big[\Big(Y_{2i}-X_{2i}'\beta_2\Big)-\Big(Y_{2j}-X_{2j}'\beta_2\Big)\ \Big|\ Y_{1i}=Y_{1j}=1, X_{1i}'\beta_1 = X_{1j}'\beta_1\Big] = 0 \ \text{ a.s}\Big\}$$

In models like this, we would be able to drop the monotonicity restrictions for λ_F and H_F in Assumption 1 while still avoiding the need to estimate these functions. Other versions of this case may include the case where (X_1, X_2) are observed, but the selection equation is characterized by moment inequalities; for example, if Y_1 is unobserved, but we observe $\overline{Y}_1$ such that $\overline{Y}_1 \geq Y_1$ w.p.1, so that $Y_1 = 1$ only if $\overline{Y}_1 = 1$. In either scenario, inference can be based on methods that combine conditional moment equalities/inequalities (e.g, Andrews and Shi (2013), Chernozhukov, Lee, and Rosen (2013)), properly modifying these procedures to account for the pairwise nature of the second conditional moment.

5.2 Models where some control variables are unobserved and only lower or upper bounds are available

Our approach assumes that we have lower and upper bounds for the unobserved control variables. In some cases, we may have only lower or upper bounds but not both. For illustration, consider our bivariate sample selection example with interval data from Section 2.1. Once again, suppose X_1 is unobserved, and our observable data only provides a lower bound for the selection-equation control function, where $X_{1L}'\beta_{10} \leq X_1'\beta_{10}$ w.p.1, with X_{1L} observable. We now have $W_1 \equiv (X_{1L}, X_2)$. If we maintain all the conditions in Assumption 1, our model implies[17] $E_F[Y_2|W_1, Y_2 = 1] \leq X_2'\beta_{20} + \lambda_F(X_{1L}'\beta_{10})$, and $P_F[Y_1 = 1|W_1] \geq H_F(X_{1L}'\beta_{10})$ w.p.1. From here, a characterization of the identified set of parameters (β_1, β_2) can be given by,

$$\Theta^I \equiv \Big\{(\beta_1,\beta_2)\in\Theta : E_F[Y_{2i}|W_{1i}, Y_{2i}=1] \leq X_{2i}'\beta_2 + \lambda_F(X_{1Li}'\beta_1), \quad \text{and}$$
$$P_F[Y_{1i}=1|W_{1i}] \geq H_F(X_{1Li}'\beta_1) \ \text{ a.s}\Big\}$$

Our proposal in this case would not involve pairwise comparisons, but rather to apply nonparametric methods to estimate the functionals $\lambda_F(\cdot)$ and $H_F(\cdot)$ nonparametrically, and then use these estimators to conduct inference on β_0 based on the conditional moment inequalities given above. To this end, we can try to adapt existing nonparametric approaches to estimating control func-

[17] If we only observed an upper bound, where $X_{1U}'\beta_{10} \geq X_1'\beta_{10}$ w.p.1, the signs of our inequalities would be reversed, and we would have, $E_F[Y_2|W_1, Y_2 = 1] \geq X_2'\beta_{20} + \lambda_F(X_{1U}'\beta_{10})$, and $P_F[Y_1 = 1|W_1] \leq H_F(X_{1U}'\beta_{10})$ w.p.1.

tions, such as the series estimators proposed in Das, Newey, and Vella (2003). This would likely require additional assumptions on the nature and properties of these functionals, which we were able to bypass through our pairwise comparison approach when we have lower and upper bounds for the unobserved control variables. We leave the formal development and study of such an approach to future work.

6 Concluding remarks

Control function methods to estimate semiparametric models 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 some instances, control variables are unobservable, making pairwise matching impossible. However, in some of these 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. Using a bivariate sample selection model as a working example, we illustrated that, if the control functions enter the econometric model monotonically, 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, resulting on pairwise conditional functional inequalities. Using this result, we proposed an inferential procedure that constructs a confidence set for the parameters of the model using a statistic design to test whether a parameter value satisfies the pairwise inequalities almost surely over a pre-specified inference range. Our proposed statistic has pivotal properties and adapts asymptotically to the measure of the contact sets, avoiding the need to estimate them in a previous step. Our Monte Carlo results showed that our procedure has finite-sample properties aligned with its asymptotic predictions. We are hopeful that our exposition illustrates how our approach can be extended beyond our bivariate sample selection example to other semiparametric models. As we pointed out, it can also provide a novel way to do inference in models of strategic interaction with multiple solutions.

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

Blundell, R. and J. L. Powell (2004). Endogeneity in semiparametric binary response models. *Review of Economic Studies 71*, 655–679.

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.

Chesher, A. and A. Rosen (2017). Generalized instrumental variable models. *Econometrica 85*(3), 959–989.

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

Das, M., W. Newey, and F. Vella (2003). Nonparametric estimation of sample selection models. *Review of Economic Studies 70*(3), 33–58.

Dominguez, M. and I. Lobato (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica 72*(5), 1601–1615.

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.

Khan, S., F. Ouyang, and E. Tamer (2021). Inference on semiparametric multinomial response models. *Quantitative Economics 12*, 743–777.

Khan, S. and E. Tamer (2009). Inference on endogenously censored regression models using conditional moment inequalities. *Journal of Econometrics 152*(2), 104–119.

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

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., 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.