University of Toronto
Department of Economics

Working Paper 756

Subvector inference for Varying Coefficient Models with Partial Identification

By Shengjie Hong and Yu-Chin Hsu

August 31, 2023

Subvector Inference for Varying Coefficient Models with Partial Identification

Shengjie Hong* Yu-Chin Hsu† Yuanyuan Wan‡

August 5, 2023

Abstract

This paper develops inference methods for a general class of varying coefficient models defined by a set of moment inequalities and/or equalities, where unknown functional parameters are not necessarily point-identified. We propose an inferential procedure for a subvector of the parameters and establish the asymptotic validity of the resulting confidence sets uniformly over a broad family of data generating processes. We also propose a specification test for the varying coefficient models considered in this paper. Monte Carlo studies show that the proposed methods work well in finite samples.

Keywords: Varying coefficient; Moment inequalities; Partial-identification; Multiplier-bootstrap

JEL classification: C12, C14, C15

---

*School of Economics, Central University of Finance and Economics. E-mail: hongshj@vip.163.com. 39 Xueyuan South Rd., Haidian Dist., Beijing, 100081, China.

†Institute of Economics, Academia Sinica; Department of Finance, National Central University; Department of Economics, National Chengchi University. E-mail: ychsu@econ.sinica.edu.tw. 128 Academia Rd., Section 2, Nankang, Taipei, 115, Taiwan.

‡Department of Economics, University of Toronto. E-mail: yuanyuan.wan@utoronto.ca. 232, Max Gluskin House, 150 St. George St., Toronto, Ontario, M5S 3G7, Canada. Yuanyuan Wan thanks the support from the SSHRC Insight Grant #435190500.

1 Introduction

This paper considers making inferences in a general class of varying coefficient models defined by a set of moment inequalities and/or equalities, where the unknown functional parameters are not necessarily point-identified. Since the seminal paper of Hastie and Tibshirani (1993), varying coefficient models have been widely adopted in empirical researches in economics and finance for their balance of providing both dimension reduction and flexible modeling of heterogeneous effects. See, for example, Li, Huang, Li, and Fu (2002), Ang and Liu (2004), Cai, Ren, and Yang (2015) and more recently Cai, Chen, and Fang (2018), among others.[1] Li, Huang, Li, and Fu (2002) proposed a semiparametric varying coefficient model to estimate production functions in which the elasticity of inputs varies with the intermediate production and management expenses. Ang and Liu (2004) studied how to discount cash flows with time-varying expected returns based on varying coefficient models. Cai, Ren, and Yang (2015) used varying coefficient models to estimate time-varying betas and alpha in the conditional capital asset pricing model (CAPM). And Cai, Chen, and Fang (2018) used varying coefficient models to estimate the growth effect of FDI. Motivated by empirical applications, the econometric theory of varying coefficient models has been developed and extended to a variety of modeling environments. For instance, Chen and Tsay (1993) considered the time series setting and developed varying coefficient autoregressive models. Fan and Zhang (1999), Cai, Fan, and Li (2000), and Ahmad, Leelahanon, and Li (2005) discussed efficient estimation. Fan and Zhang (2000) and Fan and Li (2004) considered the panel data setting. Cai and Xu (2008) proposed quantile regression methods for a class of smooth coefficient models. Cai, Das, Xiong, and Wu (2006) and Cai, Fang, Lin, and Su (2019) studied a class of instrumental variable regression under a functional coefficient representation for the regression function. And Su, Murtazashvili, and Ullah (2013) proposed consistent inferential procedure for testing constancy of varying coefficients.

While the theoretical development of varying coefficient models is fruitful, the existing

[1] See Cai and Hong (2009) and Cai (2010) for more references on applications of varying coefficient models.

works primarily focus on cases in which the functional parameters are point-identified. However, depending upon the empirical context, the assumptions that deliver point-identification of the models may not necessarily hold. For example, in a varying coefficient linear regression model or quantile regression model, the slope parameter is not point-identified if the outcome variable is interval-observed or censored, as is quite common in many survey data. In a varying coefficient instrumental regression model, the structural parameter may not be point-identified if the instrumental variable is imperfect (e.g. not independent with the structural error). In an oligopoly market entry model, the profit function with varying coefficients is typically not point-identified if there are multiple equilibria and the equilibrium selection mechanism is unknown to researchers. It is, therefore, useful to develop inferential procedures for varying coefficients that are robust to partial identification.

Our paper contributes to the literature of varying coefficient models by filling this gap. Following Andrews and Shi (2014), we consider a general class of varying coefficient models defined by a set of conditional moment inequalities and/or equalities as follows. For any $z \in \mathcal{Z}$,

$$
\begin{aligned}
&E_P[m_j(W, \theta_0(z))|X, Z = z] \geq 0 \;\text{ a.s. } X, \text{ for } j = 1, \ldots, p \text{ and} \\
&E_P[m_j(W, \theta_0(z)|X, Z = z] = 0 \;\text{ a.s. } X, \text{ for } j = p + 1, \ldots, k,
\end{aligned} \tag{1.1}
$$

where $m_j(\cdot, \theta)$ for $j = 1, \ldots, k$ are known real-valued moment functions, $X \in \mathcal{X} \subseteq R^{d_x}$, $Z \in \mathcal{Z} \subseteq R^{d_z}$. The varying coefficient $\theta(\cdot) : \mathcal{Z} \to \Theta \subseteq \mathbb{R}^{d_\theta}$ varies with z and takes value in a compact set Θ. The random vector W contains X, Z and possibly some other random variables $Y \in \mathcal{Y} \subseteq R^{d_y}$, so that $W = (X', Y', Z')' \in R^{d_w}$ with $d_w = d_y + d_x + d_z$. In empirical applications, Y is often the outcome variable of interest. We allow X be part of Z, to overlap with Z, or to have no common components with Z. P denotes the probability measure that generates the data, and E_P denotes the expectation under the distribution P. As in existing varying coefficient literature, we consider the cases in which researchers are interested in a subvector of the varying coefficients evaluated at a given point $z_0 \in \mathcal{Z}$, that is, a subvector of $\theta_0(z_0)$. However, we allow that, conditioning on $Z = z_0$, $\theta_0(z_0)$ is only partially identified in the sense that its identified set

$$\Theta_P(z_0) = \{\theta \in \Theta : \text{ (1.1) holds with } \theta \text{ in place of } \theta_0(z).\}$$

may contain more than one element. As we will illustrate in Section 2, where we will review examples in more detail, Model (1.1) encompasses a broad class of models and applies to many empirical contexts, including those mentioned above. Of course, it also includes the conventional point-identified varying coefficient models as special cases (where the model only contains conditional moment equalties). Under this framework, we propose a multiplier-bootstrap procedure to construct confidence sets for subvectors of the functional coefficients and show that the proposed confidence sets are asymptotically valid uniformly over a set of DGPs. We also propose a specification test that is consistent for necessary implications (to be specified later) of Model (1.1).

Our approach is built upon and extends Andrews and Shi (2014, AS hereafter). AS focus on confidence sets for the whole parameter vector of $\theta_0(z)$; however, motivated by some empirical applications of varying coefficient models, we instead focus on constructing confidence sets for a subvector of the parameters. Therefore, we consider different test statistics from those in AS. Specifically, we extend the profiling-based method of Bugni, Canay, and Shi (2017), which was initially designed for subvector inference in unconditional moment inequality models with finite-dimensional parameters, to the current setup of conditional moment inequality with functional parameters. We also propose a specification test for the necessary implications of the model, which was not considered in AS. In particular, we consider testing the Model (1.1) over a set of pre-chosen grid points $\mathcal{Z}^L \equiv \{z_1, \cdots, z_L\}$. We show that over the pre-chosen $\mathcal{Z}^L$, our test controls the size uniformly and is consistent. The proposed test is, therefore, a specification test for conditional moment inequalities with infinite-dimensional parameters, and it complements the existing work of Andrews and Shi (2013) and Bugni, Canay, and Shi (2015), where the parameters are finite-dimensional.

Our paper also contributes to the literature of conditional moment inequality models.[2] Recently, a line of work studies partially identified conditional moment models; an incomplete list includes Fan (2008), Kim (2008), Andrews and Shi (2013, 2017), Chernozhukov, Lee, and Rosen (2013), Lee, Song, and Whang (2013), Armstrong (2014, 2015, 2018), Bontemps and Magnac (2017), and Hsu and Shi (2017), among others. All these papers consider finite-dimensional parameters and hence do not accommodate varying coefficients. There are a small number of papers that allow the parameter vector to contain an infinite-dimensional component, for example, Santos (2012), Tao (2015), Hong (2017), but they consider only conditional moment equalties. Chernozhukov, Newey, and Santos (2015) allows for infinite dimensional parameters and handles both moment equation and inequality. However, the null asymptotic distribution is not established for their test statistics. Instead, the critical values for their test are obtained by building a strong approximation to the test statistic and then bootstrapping a (conservatively) relaxed form of it.

The rest of the paper is organized as follows. We discuss a few motivating examples in Section 2. In Section 3, we construct the uniformly valid confidence set and propose the model specification test. In Section 4, we use Monte Carlo simulations to illustrate the finite sample performance of the proposed methods. Section 5 concludes. For ease of exposition, we collect all the proofs in the Appendix.

2 Motivating Examples

In this section, we provide a few motivating examples of partially identified varying coefficient models, all of which are special cases of Model (1.1).

Example 2.1 (Varying Coefficient Model with Interval-Outcome). *Let Y^* be a latent dependent variable with $Y^* = X'\theta_0(Z) + \epsilon$ with $E[\epsilon|X,Z] = 0$ a.s. X and Z, where*

[2] There has been a large literature on unconditional moment inequality models under partial identification, see, for example, Andrews, Berry, and Jia (2004), Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2007), Andrews and Guggenberger (2009), Romano and Shaikh (2008, 2010), Andrews and Soares (2010), Wan (2013), Menzel (2014), Bugni, Canay, and Shi (2015, 2017), Pakes, Porter, Ho, and Ishii (2015), Andrews and Kwon (2019), and Belloni, Bugni, and Chernozhukov (2019) among others. For a more thorough review, please see Canay and Shaikh (2017) and references therein.

(X, Z) *are exogenous regressors. If* Y^* *were to be observed, then the model is the classical varying coefficient regression model, see for instance Hastie and Tibshirani (1993). When the latent variable* Y^* *is unobserved, but known to lie in the observed interval* $[Y_\ell, Y_u]$*. Then, the following moment inequalities hold for any fixed* $z \in \mathcal{Z}$*.*

$$E_P[Y_u - X'\theta_0(Z)|X, Z = z] \geq 0 \quad a.s.\ X \ and$$
$$E_P[X'\theta_0(Z) - Y_\ell|X, Z = z] \geq 0 \quad a.s.\ X.$$

Example 2.2 (Varying Coefficient Quantile Regression with Interval-Outcome). *Consider the following quantile varying coefficient model:*

$$Y^* = X'\theta_0(Z) + \epsilon, \quad q_{\epsilon|X,Z}(\tau|X, Z) = 0, \quad a.s. - (X, Z) \tag{2.1}$$

where Y^* *is a latent dependent variable and* $q_{\epsilon|X,Z}(\tau|X, Z)$ *denotes the* τ*th conditional quantile of* ϵ *on* X, Z*. If* Y^* *were observed by researchers, it is the quantile varying coefficient model analyzed by Honda (2004). If* Y^* *is not directly observef but known to lie in the observed interval* $[Y_\ell, Y_u]$*, then the following moment inequalities hold for any* $z \in \mathcal{Z}$*:*

$$E_P[\tau - 1\{Y_u \leq X'\theta_0(Z)\}|X, Z = z] \geq 0 \quad a.s.\ X \ and$$
$$E_P[1\{Y_\ell \leq X'\theta_0(Z)\} - \tau|X, Z = z] \geq 0 \quad a.s.\ X.$$

Example 2.3 (Varying Coefficient Quantile Regression with Censoring). *Consider again the quantile varying coefficient model in Equation (2.1). Suppose now* Y^* *is subject to censoring according to an observed binary variable* $D \in \{0, 1\}$*:* Y^* *is observed only when* $D = 1$*. Then, the following moment inequalities hold for any* $z \in \mathcal{Z}$*:*

$$E_P[\tau - 1\{Y^* \leq X'\theta_0(Z), D = 1\}|X, Z = z] \geq 0 \quad a.s.\ X \ and$$
$$E_P[1\{Y^* \leq X'\theta_0(Z), D = 1\} + 1\{D = 0\} - \tau|X, Z = z] \geq 0 \quad a.s.\ X.$$

Example 2.4 (Entry Game with Incomplete Information). *Consider a simultaneous oligopoly entry game of complete information (Ciliberto and Tamer, 2009):*

$$\begin{aligned} Y_1 &= 1\left\{X'\beta_{1,0}(Z) + \gamma_{1,0}(Z)Y_2 - \varepsilon_1 \geq 0\right\}; \\ Y_2 &= 1\left\{X'\beta_{2,0}(Z) + \gamma_{2,0}(Z)Y_1 - \varepsilon_2 \geq 0\right\}, \end{aligned}$$

where, for $j = 1, 2$, player j's payoff from entering the market, given the other player's action Y_{-j}, is specified as $X'\beta_{j,0}(Z) + \gamma_{j,0}(Z)Y_{-j} - \varepsilon_j$. In practice, it is common to assume the magnitude of the strategic interaction $\gamma_0(Z) = (\gamma_{1,0}(Z), \gamma_{2,0}(Z))'$ depends on certain observed market characteristics Z. Assume the profit shock ε_j has a known joint distribution $G(\varepsilon_1, \varepsilon_1; \alpha)$ up to finite dimensinoal parameter α and is independent with (X, Z). Following Ciliberto and Tamer (2009), and similar to the example in AS, we can characterize the equilibrium outcome by the following conditional moment equalities and inequalities for a given market characteristics z:

$$\begin{aligned} &E\left[Pr\left(0,0|X, Z = z, \theta_0\right) - \left(1 - Y_1\right)\left(1 - Y_2\right)|X, Z = z\right] = 0; \\ &E\left[Pr\left(1,1|X, Z = z, \theta_0\right) - Y_1 Y_2|X, Z = z\right] = 0; \\ &E\left[Pr\left(0,1|X, Z = z, \theta_0\right) - \left(1 - Y_1\right) Y_2|X, Z = z\right] \geq 0; \\ &E\left[Pr\left(1,0|X, Z = z, \theta_0\right) - Y_1\left(1 - Y_2\right)|X, Z = z\right] \geq 0, \end{aligned}$$

where $\theta_0 = \left(\alpha_0, \beta_{1,0}\left(z\right)', \beta_{2,0}\left(z\right)', \gamma_0'\right)'$ and

$$\begin{aligned} &Pr\left(0,0|X, Z = z, \theta\right) = 1 - G_1\left(X'\beta_1\left(z\right); \alpha\right) - G_2\left(X'\beta_2\left(z\right); \alpha\right) + G\left(X'\beta_1\left(z\right), X'\beta_2\left(z\right); \alpha\right); \\ &Pr\left(1,1|X, Z = z, \theta\right) = G\left(X'\beta_1\left(z\right) + \gamma_1(z), X'\beta_2\left(z\right) + \gamma_2(z); \alpha\right); \\ &Pr\left(0,1|X, Z = z, \theta\right) = G_2\left(X'\beta_2\left(z\right); \alpha\right) - G\left(X'\beta_1\left(z\right) + \gamma_1(z), X'\beta_2\left(z\right); \alpha\right); \\ &Pr\left(1,0|X, Z = z, \theta\right) = G_1\left(X'\beta_1\left(z\right); \alpha\right) - G\left(X'\beta_1\left(z\right), X'\beta_2\left(z\right) + \gamma_2(z); \alpha\right), \end{aligned}$$

with G_1 and G_2 denoting marginal CDF's of ε_1 and ε_2, respectively. Typically, we are interested in (some subvector of) $\beta_0\left(z\right) = \left(\beta_{1,0}\left(z\right)', \beta_{2,0}\left(z\right)'\right)'$ and/or γ_0, which is a sub-

vector of θ_0*.*

Example 2.5 (Imperfect IV). *Consider the a version of the model entertained by Cai, Fang, Lin, and Su (2019):*

$$Y = X_1'\theta(Z) + \epsilon,$$

where X_1 *is a vector of endogenous variables and* Z *is an exogenous control variable. Let* X_2 *be a vector of instruments satisfying* $E_P[\epsilon|X_2, Z = z] = 0$ *for some* z *and a.s. in* X_2*. One possible application of this model is estimating return to education, where* Y *is the wage income,* X_1 *is the endogenous education level,* Z *can be experiences or demographic variables,* ϵ *is the unobserved talent, and the instrumental variable* X_2 *is parents' education. The varying coefficient captures the idea that return to education depends on experiences. However, the validity of parents' education as an instrumental variable may hold in some empirical contexts but not in others. For example, Kédagni and Mourifié (2020) found evidence that the independence assumption can fail to hold even conditioning on children's ability. On the other hand, it is more reasonable to assume that the children's talent is positively correlated with their parents' education conditioning on* Z*, that is,* $E[X_2\epsilon|Z = z] \geq 0$ *for all* z*, which implies*

$$\mathbb{E}[X_2(Y - X_1'\theta(Z)|Z = z] \geq 0.$$

This is a special case of our Model (1.1) in which $X = (X_1, X_2)$ *does not appears in the conditioning variables.*

Example 2.6 (Testing Local Average Treatment Effect (LATE) Assumptions). *Consider a potential outcome model with binary treatment* $D \in \{0, 1\}$ *and binary instrument* $T \in \{0, 1\}$*. Let* X_1 *and* X_0 *be two potential outcomes, and* D_0 *and* D_1 *be two potential treatments. Let* Z *be a vector of covariates (here we name variables differently from the conventional treatment effect literature to match our notation). Suppose for any* $z \in \mathcal{Z}$*, we have (i)* $(X_1, X_0, D_0, D_1) \perp T|Z = z$*, (ii)* $\mathbb{P}(D = 1|T = 1, Z = z) \neq \mathbb{P}(D = 1|T = 0, Z = z)$*, and (iii)* $D_1 \geq D_0$ *or* $D_0 \geq D_1$ *a.s., then the conditional local average* *treatment effect $E_P[X_1 - X_0|Z = z]$ is identified by the Wald estimand. Mourifié and Wan (2017, Corollary 1) formulated the testable implication of LATE identifying assumptions (i)–(iii) as a set of conditional moment inequalities:*

$$
\begin{aligned}
&E_P[c_1(Z)D(1-T) - c_0(Z)DT|Z = z, X] \leq 0, \quad a.s.\ X \\
&E_P[c_0(Z)(1-D)T - c_1(Z)(1-D)(1-T)|Z = z, X] \leq 0, \quad a.s.\ X \\
&E_P[c_1(Z) - T|Z = z] = 0; \\
&E_P[c_0(Z) - (1-T)|Z = z] = 0.
\end{aligned}
$$

It fits the Model (1.1) with $\theta_0(Z) = (c_1(Z), c_0(Z))$ be the varying coefficient, and $W = (T, Z, D, X, Z')'$. In this case, the random coefficients $c_1(z)$ and $c_0(z)$ are point-identified as the conditional probability $P(T = 1|Z = z)$ and $P(T = 0|Z = z)$, respectively. Researchers are interested in testing the model specification instead of estimation. Unlike Mourifié and Wan (2017)'s algorithm, we allow Z be either discrete or continuous.[3]

3 Confidence Set

In this section, we will propose a profiled test statistics for constructing confidence set (CS) of subvectors of $\theta_0(z_0)$, for instance the first component $\theta_{01}(z_0)$, where $z_0 \in \mathcal{Z}$ is a pre-specified value that an applied researcher is interested in. Without loss of generality, we assume that the support of X, $\mathcal{X} = [0, 1]^{d_x}$.[4] We consider a countable set of instrument

[3]Mourifié and Wan (2017)'s implementation procedure is built upon the Stata package of Chernozhukov, Kim, Lee, and Rosen (2015) and accommodates only a single continuous conditioning variable. So a continuous Z needs to be discretized. Our method, on the other hand, allows for both discrete and continuous Z.

[4]We can always normalize an observed x_{ij} to the unit interval by applying the transformation $\Phi\left(\frac{x_{ij} - \bar{x}_j}{\hat{\sigma}_{x,j}}\right)$, where Φ is the standard nomral CDF, and $(\bar{x}_j, \hat{\sigma}_{x,j})$ are sample mean and standard deviation of observations $\{x_{1j}, x_{2j}, \cdots, x_{nj}\}$, respectively.

functions that are indicator functions of hyper-cubes in $\mathcal{X}$:

$$\mathcal{G}_{\text{c-cube}} = \{g_\ell(\cdot) = 1(\cdot \in C_\ell) : \ell \equiv (x, r) \in \mathcal{L}_{\text{c-cube}}\}, \text{ where}$$
$$C_\ell = \left(\times_{j=1}^{d_x}(x_j, x_j + r]\right) \text{ and}$$
$$\mathcal{L}_{\text{c-cube}} = \left\{(x, q^{-1}) : \ q \cdot x \in \{0, 1, 2, \cdots, q-1\}^{d_x}, \text{ and } q = 1, 2, \cdots\right\}. \tag{3.1}$$

For notation simplicity, we let $C_1 = C_{(\mathbf{0},1)} = \mathcal{X}$ and $g_1 = g_{(\mathbf{0},1)} = 1$. One can also consider other instrument functions that satisfy Andrews and Shi (2013, Assumption CI). Let $f_z(z_0)$ denote the probability density function (pdf) of Z evaluated at $Z = z_0$ and assume that $f_z(z_0) \geq \delta > 0$. Let $\mu_\ell(\theta, z_0) = E_P[g_\ell(X) \cdot m(W, \theta)|Z = z_0] \cdot f_z(z_0)$. AS show that the moment conditions in (1.1) are equivalent to

$$\mu_{\ell,j}(\theta, z_0) \geq 0 \ \text{ for } j = 1, \ldots, p \text{ and} \tag{3.2}$$
$$\mu_{\ell,j}(\theta, z_0) = 0 \ \text{ for } j = p+1, \ldots, k, \text{ for all } \ell \in \mathcal{L}.$$

In this paper, unlike AS, we are interested in constructing CS for a subvector of $\theta_0(z_0)$ or a functional of $\theta_0(z_0)$ for a fixed z_0 . We focus on the case that we are interested in the first element of the vector of parameter, $\theta_{0,1}(z_0)$.[5] A valid CS, $\widehat{CS}_n$, with confidence level $1 - \alpha$ for $\theta_{0,1}(z_0)$ should satisfy that

$$\liminf_{n\to\infty} \inf_{(\theta_1, P)\in\mathcal{H}_0} Pr_P(\theta_1 \in \widehat{CS}_n) \geq 1 - \alpha. \tag{3.3}$$

where $\mathcal{H}_0$ is a collection of (θ_1, P) and will be made specific later.

To construct our test statistics, we define some notation. Let $K(\cdot)$ denote a kernel function with support on $[-1, 1]^{d_z}$ and h_n is a bandwidth. For $j = 1, \ldots, k$, define

$$\hat{\mu}_{\ell,n}(\theta, z_0) = \frac{1}{nh_n^{d_z}} \sum_{i=1}^{n} K\Big(\frac{Z_i - z_0}{h_n}\Big) g_\ell(X_i) \cdot m(W_i, \theta)$$

[5]We can extend our method to the case in which researchers are interested in $\lambda(z_0) \equiv \lambda(\theta(z_0))$ for some function $\lambda : \Theta \to \Lambda \subseteq \mathbb{R}^{d_\lambda}$, as Bugni, Canay, and Shi (2017) for unconditional moment inequalities.

which will be a consistent estimator for $\mu_\ell(\theta, z_0)$ and with undersmoothing, we have $\sqrt{nh_n^{d_z}}(\hat{\mu}_{\ell,n}(\theta, z_0) - \mu_\ell(\theta, z_0))$ will be a k-dimensional mean zero Guassian process with covariance kernel $\rho_2 \cdot Cov_P[g_{\ell^{(1)}}(X) \cdot m(W, \theta^{(1)}), g_{\ell^{(2)}}(X) \cdot m(W, \theta^{(2)}) | Z = z_0] \cdot f_z(z_0)$, where the constant $\rho_2 = \int_u K^2(u)du$. Let $\hat{\mu}_{1,n}(\theta, z_0) = \frac{1}{nh_n^{d_z}} \sum_{i=1}^n K\left(\frac{Z_i - z_0}{h_n}\right) m(W_i, \theta)$ and define

$$\widehat{\Sigma}_n(\theta, 1, z_0) = \frac{1}{nh_n^{d_z}} \sum_{i=1}^n \left(K\Big(\frac{Z_i - z_0}{h_n}\Big)(m(W_i, \theta) - \hat{\mu}_{1,n}(\theta, z_0)\right) \left(K\Big(\frac{Z_i - z_0}{h_n}\Big) m(W_i, \theta) - \hat{\mu}_{1,n}(\theta, z_0)\right)',$$

$$\widehat{\Sigma}_n(\theta, \ell, z_0) = \frac{1}{nh_n^{d_z}} \sum_{i=1}^n \left(K\Big(\frac{Z_i - z_0}{h_n}\Big) g_\ell(X_i) m(W_i, \theta) - \hat{\mu}_{\ell,n}(\theta, z_0)\right) \cdot \left(K\Big(\frac{Z_i - z_0}{h_n}\Big) g_\ell(X_i) m(W_i, \theta) - \hat{\mu}_{\ell,n}(\theta, z_0))\right)',$$

$$\widehat{\Sigma}_{\epsilon,n}(\theta, \ell, z_0) = \widehat{\Sigma}_n(\theta, \ell, z_0) + \epsilon \cdot \text{diag}\left(\widehat{\Sigma}_n(\theta, 1, z_0)\right).$$

Let $S(m, \Sigma)$ be a testing function, which can be chosen as one of the following two forms.

$$S(m, \Sigma) = \sum_{j=1}^p \left[\frac{m_j}{\sigma_j}\right]_-^2 + \sum_{j=p+1}^k \left[\frac{m_j}{\sigma_j}\right]^2, \text{ or}$$

$$S(m, \Sigma) = \max\left\{ \left[\frac{m_1}{\sigma_1}\right]_-^2, \ldots, \left[\frac{m_p}{\sigma_p}\right]_-^2, \left[\frac{m_{p+1}}{\sigma_{p+1}}\right]^2, \ldots, \left[\frac{m_k}{\sigma_k}\right]^2 \right\}$$

where $[a]_- = \min\{0, a\}$ and $\sigma_j = \sqrt{\Sigma_{jj}}$. Then for a fixed value of θ_1, we can define the following Cramér-von-Mises-type (CvM) (profiled) test statistic as

$$\widehat{TS}_n(\theta_1) \equiv \inf_{\theta \in \Theta(\theta_1)} \widehat{T}_n(\theta, z_0),$$

where $\Theta(\theta_1) \equiv \{\tilde{\theta} \in \Theta : \ \tilde{\theta}_1 = \theta_1\}$ is the possible value that the rest of parameters can take when the first parameter is fixed at θ_1, and

$$\widehat{T}_n(\theta, z_0) = \sum_{q=1}^{Q_n} \frac{1}{q^2 + 100} \sum_{\{\ell : \ r = q^{-1}\}} q^{-d_x} S(\sqrt{nh_n^{d_z}} \hat{\mu}_{\ell,n}(\theta, z_0), \hat{\Sigma}_{\epsilon,\ell,n}(\theta, z_0)).$$

with $Q_n \to \infty$ as $n \to \infty$.

Next, we approximate the distribution of $\widehat{TS}_n(\theta_1)$ to construct the critical value. We consider multiplier bootstrap. Let $\{U_i : i = 1, \ldots\}$ be a sequence of pseudo random variables with zero mean and unit variance that are independent of the sample path. The multiplier bootstrap process is

$$\Psi_n^u(\theta, \ell, z_0) = \frac{1}{\sqrt{nh_n^{d_z}}} \sum_{i=1}^{n} U_i \Big(K\Big(\frac{Z_i - z_0}{h_n}\Big) g_\ell(X_i) \cdot m(W_i, \theta) - \hat{\mu}_{\ell,n}(\theta, z_0) \Big).$$

Following AS, we define the slackness function as $\hat{\nu}_{\ell,n}(\theta, z_0) = \kappa_n^{-1} \sqrt{nh_n^{d_z}} \hat{\mu}_\ell(\theta, z_0)$, where $\kappa_n = \sqrt{\log(n)}$. The bootstrap version of simulated CvM test statistic for θ as

$$\widehat{T}_n^u(\theta, z_0) = \sum_{q=1}^{Q_n} \frac{1}{q^2 + 100} \sum_{\{\ell:\ r = q^{-1}\}} q^{-d_x} S\big(\Psi_n^u(\theta, \ell, z_0) + \hat{\nu}_{\ell,n}(\theta, z_0), \hat{\Sigma}_{\epsilon,\ell,n}(\theta, z_0)\big).$$

And for a fixed value of θ_1, the bootstrap test statistic as[6]

$$\widehat{TS}_n^u(\theta_1) \equiv \min_{\theta \in \Theta(\theta_1)} \widehat{T}_n^u(\theta, z_0).$$

For a fixed positive number η, for example, 10^{-6}, define $\widehat{C}_{\eta,n}(\theta_1, \alpha)$ as the $(1-\alpha)$-th quantile of the conditional distribution of $\widehat{TS}_n^u(\theta_1)$ given data plus η, i.e.,

$$\widehat{C}_{\eta,n}(\theta_1, \alpha) = \sup \Big\{ C \Big| P^u\big(\widehat{TS}_n^u(\theta_1) \le C\big) \le 1 - \alpha \Big\} + \eta. \tag{3.4}$$

The confidence set for $\theta_{0,1}(z_0)$ is then given as

$$\widehat{CS}_n = \{\theta_1 : \widehat{TS}_n(\theta_1) \le \widehat{C}_{\eta,n}(\theta_1, \alpha)\}. \tag{3.5}$$

[6] The statistic $\widehat{TS}_n^u(\theta_1)$ defined here is analogous to the statistic $T_n^{PR}(\lambda_0)$ of (2.13) in Bugni, Canay, and Shi (2017). As we show later, critical value based on $\widehat{TS}_n^u(\theta_1)$ would work. We might, in addition, consider an alternative bootstrap statistic $T_n^{DR}(\theta_1)$ analogous to their $T_n^{DR}(\lambda_0)$, and use $\min\{\widehat{TS}_n^{DR}(\theta_1), \widehat{TS}_n^u(\theta_1)\}$ for a potential power improvement. Please see discussions in Bugni, Canay, and Shi (2017, section 4.1) for a detailed discussion.

3.1 Asymptotics of Confidence Sets

Let $\{W_i\}_{i=1}^n$ denote a random sample of size n generated from P. Let $\mathcal{P}$ denote the set of P that we consider. Let F_z, F_x, and F_{xz} denote the marginal distributions of Z, X, and (X, Z) under P. Let f_z denote the density function of Z under P.

We now introduce the regularity conditions for establishing the asymptotic properties of the proposed CSs in (3.5). We first impose conditions on the moment functions $\{m_j(W, \theta) : \theta \in \Theta\}$ for $j = 1, \ldots, k$ to regulate their complexity.

Assumption 3.1 $\{(X_i, Y_i, Z_i)\}_{i=1}^n$ *is a random sample of i.i.d. observations.*

Assumption 3.2 Θ *is compact and convex.*

One special case of Assumption 3.2 is that Θ is a Cartesian product of d_θ closed intervals $\Theta = \Pi_{j=1}^{d_\theta}[\theta_{j\ell}, \theta_{ju}]$, in which case $\Theta(\theta_1) \equiv \{\tilde{\theta} \in \Theta : \tilde{\theta}_1 = \theta_1\}$ is independent with θ_1, and that $\Theta_{-1} \equiv \cup_{\theta_1}\Theta(\theta_1) = \Pi_{j=2}^{d_\theta}[\theta_{j\ell}, \theta_{ju}]$.

Assumption 3.3 *Assume that for fixed $\delta > 0$ and $0 < Q < \infty$ not depending on P,*

1. $\max_{j=1,\ldots,k} \left|m_j(w, \theta)\right| \leq M(w)$ *for all $w \in \mathcal{W}$, for all $\theta \in \Theta$ for some envelope function $M(w)$;*

2. $E_P[M(W)^4 | Z = z] \leq Q < \infty$ *on $\mathcal{N}_\delta(z_0)$ for all $P \in \mathcal{P}$;*

3. *the processes $\{m_j(W_{n,i}, \theta) : \theta \in \Theta, i \leq n, 1 \leq n\}$ for $j = 1, \ldots, k$ are manageable with respect to the envelope functions $\{M(W_{n,i}) : i \leq n, 1 \leq n\}$ where $\{W_{n,i} : i \leq n, 1 \leq n\}$ is a row-wise i.i.d. triangular array with $W_{n,i} \sim P_n$ for any sequence $\{P_n \in \mathcal{P}\}$.*

Assumption 3.3 implies that $\{n^{-1/2}h_n^{-d_z/2}K((Z_i - z_0)/h_n) \cdot g_\ell(X_i)m_j(W_{n,i}, \theta) : \theta \in \Theta, \ell \in \mathcal{L}, i \leq n, 1 \leq n\}$ are manageable with respect to the envelope functions $\{n^{-1/2}h_n^{-d_z/2}K((Z_i - z_0)/h_n) \cdot M(W_{n,i}) : i \leq n, 1 \leq n\}$.

Assumption 3.4 *For the same δ and Q as in Assumption 3.3, assume that*

1. $f_z(z) \geq \delta > 0$ *and is continuous on* $\mathcal{N}_\delta(z_0) \subset \mathcal{Z}$*;*

2. $f_z(z)$ *is twice continuously differentiable on* $\mathcal{N}_\delta(z_0)$*;*

3. $|f_z(z)| \leq Q$, $|f_z'(z)| \leq Q$ *and* $|f_z''(z)| \leq Q$ *on* $\mathcal{N}_\delta(z_0)$.

where $\mathcal{N}_\delta(z_0) = \mathcal{N}_\delta(z_0) \equiv \{z : \|z - z_0\| \leq \delta\}$.

Assumption 3.4 imposes some regularity conditions on the distribution of Z and assumes z_0 is in the interior of the support. When Z is discrete, the Model (1.1), the inference can be done using Andrews and Shi (2013) by conditioning on each realization of Z.

Let $m(W, \theta) = (m_1(W, \theta), \ldots, m_k(W, \theta))'$, $\mu(\theta, x, z) = E_P[m(W, \theta)|X = x, Z = z]$, and $\mu_j(\theta, x, z) = E_P[m_j(W, \theta)|X = x, Z = z]$. We next impose conditions on the conditional moment conditions.

Assumption 3.5 *For all* $x \in \mathcal{X}$, $\mu_j(\theta, x, z)$ *is twice continuously differentiable on* $\Theta \times \mathcal{N}_\delta(z_0)$ *. Also, for all* $x \in \mathcal{X}$*, for the same* δ *and* Q *as in Assumption 3.3 and for all* $j = 1, \ldots, k$,

1. $\|\partial \mu_j(\theta, x, z)/\partial \theta\| \leq Q$ *and* $\|\partial^2 \mu_j(\theta, x, z)/\partial \theta \partial \theta'\| \leq Q$ *on* $\Theta \times \mathcal{N}_\delta(z_0)$*;*

2. $|\mu_j(\theta, x, z)| \leq Q$, $|\partial \mu_j(\theta, x, z)/\partial z| \leq Q$ *and* $|\partial^2 \mu_j(\theta, x, z)/\partial z \partial z| \leq Q$ *on* $\Theta \times \mathcal{N}_\delta(z_0)$.

Assumption 3.6 *Assume that*

1. *The* $K(\cdot)$ *is a non-negative symmetric bounded kernel with a compact support in* R *(say* $[-1, 1]$*).*

2. $\int K(u)du = 1$ *and* $\int u_j K(u)du = 0$.

3. $h_n \to 0$, $nh_n \to \infty$ *and* $nh_n^{d_z+4} \to 0$ *as* $n \to \infty$.

Assumption 3.6 imposes conditions on kernel function and bandwidth. Assumption 3.6(i)-(ii) are satisfied for commonly used second-order kernels. While it rules out higher-order kernel, all of our results can be extended to higher-order kernel straightforwardly. Assumption 3.6(iii) requires undersmoothing, so the bias term is asymptotically negligible.

This is standard practice for nonparametric estimators being asymptotically normally distributed with mean zero and is also adopted in AS.

Assumption 3.7 *Assume that* $\kappa_n \to \infty$ *and* $\kappa_n^2 n^{-1} h_n^{-d_z} \to 0$.

Assumption 3.7 specifies the condition for the slackness tuning parameter κ_n, and it is satisfied if κ_n is proportional to $\log(n)$, or a power of $\log(n)$.

Assumption 3.8 *Assume that uniformly over* $P \in \mathcal{P}$ *given in Assumption 3.3, the following hold,*

$$\lim_{\delta \downarrow 0} \sup_{\|(\theta^{(1)} - \theta^{(2)})\| \leq \delta} \sup_{\ell \in \mathcal{L}} \max_{j=1\ldots,k} \left| Var\big(g_\ell(X) \cdot (m_j(W, \theta^{(1)}) - m_j(W, \theta^{(2)}))\big| Z = z_0\big) \right| \to 0.$$

Assumption 3.8 is imposed to ensure that when along a (sub)sequence of distributions such that $\widehat{\Psi}_n(\theta, \ell, z_0) = \sqrt{nh_n^{d_z}}(\hat{\mu}_{\ell,n}(\theta, z_0) - \mu_\ell(\theta, z_0))$ weakly converges to a tight Gaussian process along a (sub)sequence of distributions, the limiting process will have a continuous path in θ uniformly over $\ell \in \mathcal{L}$. Define population-level quantities:

$$\Sigma((\theta^{(1)}, \ell^{(1)}), (\theta^{(2)}, \ell^{(2)})) = \rho_2 \cdot Cov_P(g_{\ell^{(1)}}(X) \cdot m(W, \theta^{(1)}), g_{\ell^{(2)}}(X) \cdot m(W, \theta^{(2)}) | Z = z_0) \cdot f_z(z_0)$$
$$\Sigma((\theta, \ell)) = \Sigma((\theta, \ell), (\theta, \ell)),$$
$$\Sigma((\theta, 1)) = \rho_2 \cdot Cov_P(m(W, \theta), m(W, \theta) | Z = z_0) \cdot f_z(z_0),$$
$$\Sigma_\epsilon((\theta, \ell)) = \Sigma((\theta, \ell)) + \epsilon \cdot \Sigma((\theta, 1)),$$

and define

$$T_P(\theta, z_0) = \sum_{q=1}^{\infty} \frac{1}{q^2 + 100} \sum_{\{\ell:\ r = q^{-1}\}} q^{-d_x} S(\mu_\ell(\theta, z_0), \Sigma_{\epsilon,\ell}((\theta))). \tag{3.6}$$

Assumption 3.9 *Let* $\mathcal{P}_0$ *be the collection of* $P \in \mathcal{P}$ *such that* $\Theta_P(z_0)$ *is not empty. Then for all* $P \in \mathcal{P}_0$, $T_P(\theta, z_0) \geq c \min\{\delta, \inf_{\tilde{\theta} \in \Theta_P(z_0)} \|\theta - \tilde{\theta}\|^2\}$ *for some constants* $c > 0$ *and* $\delta > 0$.

Assumption 3.9 is an identification strength assumption. It is a type of polynomial minorant condition introduced by Chernozhukov, Hong, and Tamer (2007). A similar condition is also assumed in Bugni, Canay, and Shi (2017, Assumption A.3) for subvector inference in unconditional moment inequality models. This assumption excludes weakly identified models. For instance, it requires the instrumental variable and the endogenous variable have a correlation bounded away from zero.

We define $\mathcal{H}_0$ as the collection of (θ_1, P) such that $P \in \mathcal{P}$ and there exists a $\theta_{-1} \in \Theta_{-1}$ such that $(\theta_1, \theta_{-1}) \in \Theta_P(z_0)$. That is,

$$\mathcal{H}_0 \equiv \{(\theta_1, P) : P \in \mathcal{P},\ \text{exist } \theta_{-1} \in \Theta_{-1} \text{ such that } (\theta_1, \theta_{-1}) \in \Theta_P(z_0)\}.$$

Theorem 3.1 *Let the confidence level be $1-\alpha$. Suppose Assumptions 3.1-3.9 hold, then*

$$\liminf_{n\to\infty} \inf_{(\theta_1,P)\in\mathcal{H}_0} Pr(\theta_1 \in \widehat{CS}_n) \geq 1-\alpha. \tag{3.7}$$

In addition, if these exists $(\theta_1^*, P^*) \in \mathcal{H}_0$ such that the limiting distribution function under P^* of $\widehat{TS}_n(\theta_1)$ is continuous and strictly increasing at its $(1-\alpha)$-th quantile, then

$$\lim_{\eta\downarrow 0} \liminf_{n\to\infty} \inf_{(\theta_1,P)\in\mathcal{H}_0} Pr(\theta_1 \in \widehat{CS}_n) = 1-\alpha. \tag{3.8}$$

Remark 3.1 *The confidence sets in the Theorem 3.1 depends on z_0. In some applications, researchers may be interested in a joint inference on $\theta_{01}(\cdot)$ evaluated at multiple pre-specified values: $\mathcal{Z}^T = \{z_1, z_2, \cdots, z_T\}$.*[7] *The results of Theorem 3.1 can be readily extended to analyze this case because the confidence sets conditioning on different z_t, $t = 1, 2, \cdots, T$ are asymptotically independent due to kernel smoothing. To be specific, define $\widehat{C}_{\eta,n}^{joint}(\theta_1(z_1), \ldots, \theta(z_T), \alpha)$ as*

$$\widehat{C}_{\eta,n}^{joint}(\theta_1(z_1), \ldots, \theta(z_T), \alpha) = \sup\left\{C \middle| P^u\big(\max_{t=1,\ldots,T} \widehat{TS}_n^u(\theta_{z_t}) \leq C\big) \leq 1-\alpha\right\} + \eta.$$

[7] Researchers may also be interested in the confidence band for the functional parameter $\theta_{01}(\cdot)$. This is beyond the scope of this paper and we leave it for future research.

The joint confidence set for $\{\theta_{0,1}(z_t) : t = 1, \ldots, T\}$ *is then given as*

$$\widehat{CS}_n^{joint} = \{\{\theta_1(z_t) : t = 1, \ldots, T\} : \max_{t=1,\ldots,T} \widehat{TS}_n(\theta_1(z_t)) \leq \widehat{C}_{\eta,n}^{joint}(\theta_1(z_1), \ldots, \theta(z_T), \alpha)\}.$$

3.2 Specification Test

In many empirical settings (e.g., Example 2.6), an important concern is whether the model is correctly specified in the sense that there exists a function $\theta(\cdot)$ such that for all $Z = z$

$$\begin{cases} E_P[m_j(W, \theta(Z))|X, Z = z] \geq 0 \text{ for } j = 1, \ldots, p \\ E_P[m_j(W, \theta(Z))|X, Z = z] = 0 \text{ for } j = p+1, \ldots, k \end{cases} \quad \text{a.s. } X \tag{3.9}$$

To examine whether the model specification is consistent with the data, we propose to test the necessary condition that whether there exist a function $\theta(\cdot)$ that makes the condition moment conditions of (3.9) to hold at a fixed grid of points in the support of Z, say $\mathcal{Z}^T = \{z_1, z_2, \cdots, z_T\}$. Consequently, we aim to test the null hypothesis

$$\begin{aligned} H_0 : \quad & \text{There exist a } \theta(\cdot) \text{ s.t.} \\ & \begin{cases} E_P[m_j(W, \theta(z))|X, Z = z] \geq 0 \text{ for } j = 1, \ldots, p \\ E_P[m_j(W, \theta(z))|X, Z = z] = 0 \text{ for } j = p+1, \ldots, k \end{cases} \\ & \text{hold for all } z = z_1, \cdots, z_T \text{ and a.s. } X \end{aligned} \tag{3.10}$$

We define

$$\mathcal{P}_0 \equiv \{P \in \mathcal{P} : \text{Conditions (3.10) hold}\}.$$

Note that $\mathcal{P}_0$ implicitly depends on the grid points $\mathcal{Z}^T$. For testing the H_0 of $P \in \mathcal{P}_0$ agaist H_1 of $P \in \mathcal{P}/\mathcal{P}_0$, we propose the following test statistic

$$\widehat{T}_n \quad \equiv \quad \max_{t=1,\cdots,T} \left[\min_{\theta \in \Theta} \widehat{T}_n(\theta, z_t) \right],$$

and set the critical value $C_n^u(\alpha)$ as the $(1-\alpha)$-th quantile of $\max_{t=1,\cdots,T} \left[\min_{\theta \in \Theta} \widehat{T}_n^u(\theta, z_t) \right]$

plus η, and define the test be $\phi_n = 1[\hat{T}_n > C_n^u(\alpha)]$. It is easy to see that the test statistic $\widehat{T}_n$ and $C_n^u(\alpha)$ utilize respectively $\widehat{T}_n(\theta, z_t)$ and $\widehat{T}_n^u(\theta, z_t)$, both of which are used earlier for constructing CSs of (3.3). The following theorem establishes the consistency of the proposed procedure above for testing the null of (3.10).

Theorem 3.2 *Suppose Assumptions 3.1-3.9 hold, then*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0} Pr(\phi_n = 1) \leq \alpha. \tag{3.11}$$

In addition, if these exists $P^* \in \mathcal{P}$ such that the limiting distribution function under P^* of $\hat{T}_n$ is continuous and strictly increasing at its $(1-\alpha)$-th quantile, then

$$\lim_{\eta\downarrow 0} \limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0} Pr(\phi_n = 1) = \alpha. \tag{3.12}$$

Remark 3.2 *Focusing on the fixed grid $\mathcal{Z}^T = \{z_1, z_2, \cdots, z_T\}$, the proposed procedure does not guarantee to consistently test for the null of correct specification (i.e. the existence of a $\theta(\cdot)$ that makes the conditional moment conditions of (3.9) to hold for all values in $\mathcal{Z}$). Nevertheless, by rejecting the null of a correct specification upon rejection of H_0 of (3.10), one can consistently reject the null of a correct specification when it is false (i.e. when the model is misspecified). This is because H_0 of (3.10) is necessary but not sufficient condition for H_0 of (3.9). And empirical researchers can adopt the proposed testing procedure as a practical way of checking the model specification and can pick a larger number of grid points (of z) to make the testing result more credible.*

Remark 3.3 *In calculating the $(1-\alpha)$-th quantile of $\max_{t=1,\cdots,T}\left[\min_{\theta\in\Theta} \widehat{T}_n^u(\theta, z_t)\right]$. One can follow the idea of Bugni, Canay, and Shi (2015) and replace the minimization region Θ with $\Theta_P^{\log(n)}(z_t)$, an expansion of the identified set $\Theta_P(z_t)$, or with a set estimator $\widehat{\Theta}_P(z_t)$. In this paper, we do not further pursue in this direction.*

Corollary 3.1 *Fix $\mathcal{Z}^T = \{z_1, z_2, \cdots, z_T\}$. Suppose the conditions for Theorem 3.2 are satisfied for all $z \in \mathcal{Z}^T$. Let $T_P(\theta, z_t)$ be as defined in Equation (3.6) with z_t in place of*

z_0 *and* P_n *be a sequence of DGP such that*

$$c_n = \max_{t=1,\cdots,T} \inf_{\theta\in\Theta} T_{P_n}(\theta, z_t) > 0.$$

Then for any chosen $\eta < +\infty$, *we have* $\liminf_{n\to\infty} Pr(\phi_n = 1) = 1$ *if* $c_n \to c > 0$. *If* $nh^{d_z}c_n \to c > 0$, *and let* $r(c) \equiv \liminf_{n\to\infty} Pr(\phi_n = 1)$, *then we have* $\lim_{c\to+\infty} r(c) = 1$.

Remark 3.4 *The condition* $\max_{t=1,\cdots,T} \inf_{\theta\in\Theta} T_{P_n}(\theta, z_t) = c_n > 0$ *is a high level condition.* $c_n \to c \in (0,\infty)$ *can occur if a moment inequality is violated at a particular* z_t. *For example, for some* $j = 1,\cdots,p$, $E_{P_n}[m_j(W,\theta_0(Z))|X, z = z_t] \equiv h_j(X, z_t)$ *such that* $h_j(X, z_t) < -\delta < 0$ *over a subset of* $\tilde{\mathcal{X}}_{z_t}$ *with* $Pr(X \in \tilde{\mathcal{X}}_{z_t}|Z = z_t) > 0$. *It can also occur when a moment equality is violated, for example, for some* $j = p+1,\cdots,k$, $E_p[m_j(W,\theta_0(Z))|X, z = z_t] \equiv b_j(X, z_t)$ *such that* $|b_j(X, z_t)| > \delta > 0$ *over a subset of* $\tilde{\mathcal{X}}_{z_t}$ *with* $Pr(X \in \tilde{\mathcal{X}}_{z_t}|Z = z_t) > 0$.

4 Simulation

This section provides some Monte Carlo simulations to illustrate our method and demonstrate its finite sample performance. In Section 4.1, we mainly focus on the property of the proposed confidence set. In Section 4.2, we investigate the property of the proposed specification test. Throughout all the simulation exercises, we consider three sample sizes $n \in \{500, 1000, 2000, 4000\}$, bootstrap sample size $B = 1000$, number of replications $R = 1000$, input parameter of the interval class $Q_n = 10$. The instrumental functions are selected based on Equation (3.1). We use the second-order Epanechnikov kernel function and (under-smoothed) rule-of-thumb bandwidth: $h = h_{rot} \times n^c$ for $c = -0.1$.[8] Finally, we choose the infinitesimal constant $\eta = 10^{-6}$, and the slackness constant $\kappa_n = \sqrt{\log n}$.

[8]We also tried other reasonable choices of Q_n (such as 5 or 15) and c (such as -0.05 or -0.08). The results are qualitatively similar.

4.1 Finite Sample Performances of the CS

4.1.1 Linear Regression with Interval-Outcome

Let Y be a latent dependent variable with $Y^* = X'\theta_0(Z) + \epsilon$ with $E[\epsilon|X, Z] = 0$ a.s. X and Z, where (X, Z) are exogenous regressors. The latent variable Y^* is known to lie in the observed interval $[Y_\ell, Y_u]$. Then, the following moment inequalities hold for any fixed $Z = z_0 \in \mathcal{Z}$, the support of Z:

$$E_P[Y_u - X_1\theta_{10}(Z) - X_2\theta_{20}(Z)|X, Z = z_0] \geq 0 \quad \text{a.s. } X \text{ and} \tag{4.1}$$

$$E_P[X_1\theta_{10}(Z) + X_2\theta_{20}(Z) - Y_\ell|X, Z = z_0] \geq 0 \quad \text{a.s. } X. \tag{4.2}$$

We consider the following specification:

$$Y^* = X_1\theta_{10}(Z) + X_2\theta_{20}(Z) + \epsilon,$$

where $X_1 \sim N(0, 1)$, $X_2 \sim N(0, 1)$, $Z \sim U[2, 6]$, $\epsilon \sim N(0, 1)$ are all mutually independent. For some $\delta > 0$, let $Y_u = \delta(\text{Ceil}[Y^*/\delta])$ and $Y_\ell = \delta(\text{Ceil}[Y^*/\delta] - 1)$, where $\text{Ceil}[x]$ rounds x to integer toward $+\infty$. Under this construction, the bracket length $Y_u - Y_\ell$ is exactly δ and the researchers only observe this bracket around the true value of Y^*. We consider the following varying coefficients:

$$\theta_{10}(z) = (1.6 + 0.6z)e^{-0.4(z-3)^2}, \quad \theta_{20}(z) = 2(1 + \cos(z))$$

This specification of θ_{10} is taken from Cai, Fang, Lin, and Su (2019). We focus on $z_0 = 4$, which implies the true value of $\theta_{10}(z_0)$ equals to 2.68. In this model, the upper and lower bounds of the identified set for $\theta_{10}(z_0)$ would be θ_1 is $[\theta_{1,\ell b}, \theta_{1,ub}]$, where

$$\theta_{1,\ell b} = \inf_{\theta \in \Theta} \theta_1 \quad \text{s.t.} \quad E_P[Y_\ell \mid X = x, Z = z_0] \leq x^\intercal\theta \leq E_P[Y_u \mid X = x, Z = z_0], \quad x\text{–a.e.}$$

$$\theta_{1,ub} = \sup_{\theta \in \Theta} \theta_1 \quad \text{s.t.} \quad E_P[Y_\ell \mid X = x, Z = z_0] \leq x^\intercal\theta \leq E_P[Y_u \mid X = x, Z = z_0], \quad x\text{–a.e.}$$

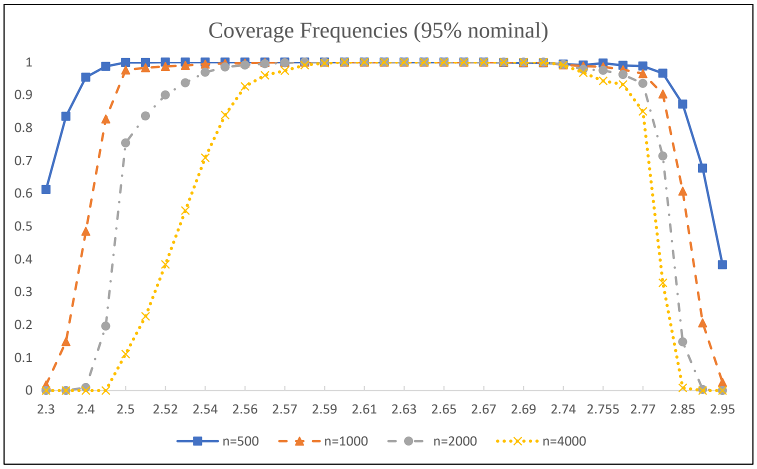

Figure 1: Coverage Frequency: Linear Model

For this linear regression with interval-observed outcome variable designs, we consider interval lengths $\delta = 0.5$, which implies the identified set to be $[2.6, 2.73]$.[9] We calculate coverage frequencies at 95% nominal levels for different values of θ_1 and plot in Figure 1.[10] We can see that the coverage frequency is no smaller than the nominal level near the identified set, which shows that our CS is asymptotically valid. It is higher than the nominal level because the asymptotic coverage probability for points in the interior of the identified set is higher (and converges to 1 as n goes to infinity). We can also see that the coverage probability declines quickly when moving away from the identified set. Also, the coverage frequency decreases quickly as the sample size increases for each given θ outside of the identified set. This shows that our CS has a good finite sample power.

[9] The "approximated identified sets" reported here are calculated by evaluating sample objective functions with a very large sample size ($n = 100,000$) and $Q_n = 10$. Therefore these sets are essentially approximations of the approximated identified region of the set of unconditional moment inequalities corresponding to $Q_n = 10$, and they should be larger than the true identified sets of the conditional moment inequalities. We also consider $= 0.1$ and $\delta = 1.0$. The results are qualitatively similar and therefore omitted to save space.

[10] When the dimension of the parameter vector is high, instead of considering a fixed grid points, one can use the EAM algorithm of Kaido, Molinari, and Stoye (2019) to select testing points to reduce computation cost.

4.1.2 Entry Game with Complete Information

Consider a simultaneous entry game of complete information with two players:

$$
\begin{aligned}
Y_1 &= 1\left\{\theta_{1,0}(Z)Y_2 - \varepsilon_1 \geq 0\right\}; \\
Y_2 &= 1\left\{\theta_{2,0}(Z)Y_1 - \varepsilon_2 \geq 0\right\},
\end{aligned}
$$

where the coefficient $\theta_{1,0}(z) = -\frac{e^z-1}{e-1}$, $\theta_{2,0}(z) = -\frac{e^{1-z}-1}{e-1}$, $Z \sim U[0,1]$, and $(\varepsilon_1, \varepsilon_2) \sim \mathcal{N}\left(\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)\right)$. In this model, the strength of the strategic interaction depends on observed variable Z. We assume that players play a pure strategy Nash equilibrium, and when there are multiple equilibria, the nature tosses a fair coin to select. Researchers observe Y_1, Y_2 and Z, but do not know the functional form of $\theta_{j,0}(z)$, $j = 1, 2$. Researchers are also agnostic about the equilibrium selection mechanism. The goal is to construct the confidence interval for $\theta_{1,0}(z_0)$ at $z_0 = 0.5$.

Let $\Phi_\rho(t_1, t_2)$ be the probability of the event $\{\varepsilon_1 \leq t_1 \quad \& \quad \varepsilon_2 \leq t_2\}$. The necessary condition of Nash equilibrium implies the following conditional moment restrictions:

$$
\begin{aligned}
E_P\left[0.5 - \Phi_\rho(\theta_{1,0}(Z), 0) - (1-Y_1)Y_2 \mid Z = z\right] &\geq 0, \\
E_P\left[0.5 - \Phi_\rho(0, \theta_{2,0}(Z)) - Y_1(1-Y_2) \mid Z = z\right] &\geq 0, \\
E_P\left[\Phi_\rho(\theta_{1,0}(Z), \theta_{2,0}(Z)) - Y_1 Y_2 \mid Z = z\right] &= 0, \\
E_P\left[\Phi_\rho(0,0) - (1-Y_1)(1-Y_2) \mid Z = z\right] &= 0.
\end{aligned}
$$

In this model, the unknown parameters are $(\theta_{1,0}(\cdot), \theta_{2,0}(\cdot), \rho)$. However, ρ is identified from the fourth moment equality. Therefore we solve ρ from the fourth equation and focus on the first three conditional moment restrictions:

$$\Phi_\rho(\theta_1, 0) \leq 0.5 - p(0, 1|z), \tag{4.3}$$

$$\Phi_\rho(0, \theta_2) \leq 0.5 - p(1, 0|z), \tag{4.4}$$

$$\Phi_\rho(\theta_1, \theta_2) = p(1, 1|z), \tag{4.5}$$

where $p(\ell, k|z) \equiv \Pr(Y_1 = \ell, Y_2 = k|Z = z)$. Note that given the joint normal distribution of epsilons, the upper and lower bound of the identified set for $\theta_{01}(z_0)$ can be analytically calculated from Equations (4.3) to (4.5). In particular, Equation (4.5) says that the joint identified set is a curve in the two-dimensional space. Equation (4.3) and Equation (4.4) provide the coordinates of the two endpoints of the curve. Based on our calculation, the identified set for $\theta_{01}(z_0)$ is $[-0.47, -0.29]$ when $\rho = 0.5$.

Crefentry-a reports the coverage frequencies under different sample sizes for different θ_1 values. We also considered other values of ρ and other significance levels but omitted the results due to the qualitative similarity. When parameter values are inside the identified set, the coverage frequency is above the nominal levels, demonstrating its validity. The coverage frequencies are closer to the nominal level when θ_1 is closer to the boundary of the identified set. When θ_1 moves away from the identified set, the coverage frequencies decline dramatically and declines faster for larger sample sizes.

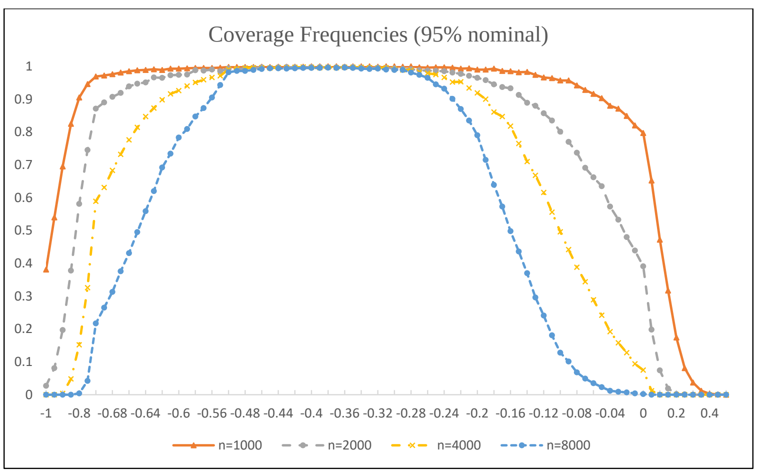

Figure 2: Coverage Frequency: Entry Game

4.2 Specification Test

In this subsection, we examine the finite sample performance of our specification test, for which we consider examples that are parallel to those that we used in the previous subsection for confidence sets. For each example, we change the DGP so that the moment inequalities are mis-specified.

4.2.1 Linear Regression with Interval-Outcome

The DGP is the same as the linear regression with an interval-outcome example, except now we also consider cases in which $\delta < 0$. In such cases, the model is mis-specified, and we should expect high rejection frequency. To implement the test, we consider five grid points in the space of z: $\{3.0, 3.5, 4.0, 4.5, 5.0\}$. We expect that the test performs better if we use finer grids.

Table 1: Rejection Frequency: Linear Regression with Interval Outcome

δ	n	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$\delta = -1.0$	$n = 500$	1.000	1.000	1.000
	$n = 1000$	1.000	1.000	1.000
	$n = 2000$	1.000	1.000	1.000
$\delta = -0.5$	$n = 500$	1.000	1.000	1.000
	$n = 1000$	1.000	1.000	1.000
	$n = 2000$	1.000	1.000	1.000
$\delta = -0.2$	$n = 500$	0.7625	0.6350	0.3950
	$n = 1000$	0.9725	0.8775	0.5125
	$n = 2000$	1.000	1.000	0.9300
$\delta = 0.0$	$n = 500$	0.1925	0.1100	0.0450
	$n = 1000$	0.1500	0.0725	0.0200
	$n = 2000$	0.1125	0.0750	0.0200
$\delta = 0.2$	$n = 500$	0.0450	0.0325	0.0050
	$n = 1000$	0.0075	0.0025	0.000
	$n = 2000$	0.000	0.000	0.000
$\delta = 0.5$	$n = 500$	0.000	0.000	0.000
	$n = 1000$	0.000	0.000	0.0000
	$n = 2000$	0.000	0.000	0.000
$\delta = 1.0$	$n = 500$	0.000	0.000	0.000
	$n = 1000$	0.000	0.000	0.000
	$n = 2000$	0.000	0.000	0.000

The following Table 1 reports the rejection frequencies under different significance level α and δ. When the model is correctly specified and has a positive interval length

($\delta > 0$), the rejection frequency is very low and close to zero. This is not surprising because the true model lies in the "interior" of the null hypothesis. When the model is correctly specified but point-identified ($\delta = 0$), we are in the knife-edge case, and the rejection frequency is close to the nominal value when the sample size is large enough. Finally, when the model is mis-specified ($\delta > 0$), our test can detect it and show good power—the rejection frequencies increases as the size of the misspecification increases.

4.2.2 Entry Game with Complete Information

Our second example corresponds to the design in Section 4.1.2. We consider the same game and use the same set of inequalities, except that change the error terms $(\varepsilon_1, \varepsilon_2) \sim \mathcal{N}\left(\begin{pmatrix} -\delta \\ -\delta \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$ and Φ is the standard normal CDF. However, the researcher incorrectly parameterizes the joint distribution as to be $(\varepsilon_1, \varepsilon_2) \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$. In this desisn, the size of δ measures the mangnitude of the mis-sepcification. For example, as $\delta \to +\infty$, the probability of the outcome $(0, 0)$ will converges to zero, but if under the mis-sepcified model, for any given value of ρ, the outcome $(0, 0)$ would occur with positive probability. We consider five z values when implement the test: $z \in \{0.2, 0.35, 0.5, 0.65, 0.8\}$. Table 2 reports the rejection frequencies when $\rho = 0$. When $\delta = 0$, the model is correctly specified, and we see the rejection frequencies are below nominal values across the broad. When $\delta > 0$, the model is mis-specified. Our test rejects the model with large frenqucies and the rejection rate increases with both sample size n and mis-specification magnitude δ.

5 Conclusion

In this paper, we provide a consistent inference procedure for varying coefficients that are defined by a set of moment inequalities and/or equalties. The proposed procedure is based on multiplier-bootstrap and as shown, can be readily used to construct confidence sets for subvector of interest of the parameters. We establish that the resulting confidence sets are asymptotically valid uniformly over a broad family of DGPs and robust to partial identification. We also propose a specification test for a finite number of necessary

Table 2: Rejection Frequency: Entry Game

δ	n	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$\delta = 0.0$	$n = 500$	0.0075	0.0025	0.0000
	$n = 1000$	0.0200	0.0075	0.0000
	$n = 2000$	0.0200	0.0050	0.0000
$\delta = 0.2$	$n = 500$	0.6350	0.3675	0.0050
	$n = 1000$	0.9050	0.7750	0.2275
	$n = 2000$	1.0000	0.9725	0.7425
$\delta = 0.4$	$n = 500$	0.9975	0.9850	0.6800
	$n = 1000$	1.0000	1.0000	0.9925
	$n = 2000$	1.0000	1.0000	1.0000

implications of the varying coefficient models we considered.

Appendix

A Notations

We introduce more notations. Let Ω be a specified closed set of $k \times k$ covariance matrices. Recall that

$$\Sigma_P((\theta^{(1)}, \ell^{(1)}), (\theta^{(2)}, \ell^{(2)})) = \rho_2 \cdot Cov_P(g_{\ell^{(1)}}(X) \cdot m(W, \theta^{(1)}), g_{\ell^{(2)}}(X) \cdot m(W, \theta^{(2)})|Z = z_0) \cdot f_z(z_0)$$

$$\Sigma_P((\theta, \ell)) = \Sigma_P((\theta, \ell), (\theta, \ell)),$$

$$\Sigma_P((\theta, 1)) = \rho_2 \cdot Cov_P(m(W, \theta), m(W, \theta)|Z = z_0) \cdot f_z(z_0),$$

$$\Sigma_{P,\epsilon}((\theta, \ell)) = \Sigma_P((\theta, \ell)) + \epsilon \cdot \Sigma_P((\theta, 1)),$$

$$\mu_\ell(\theta, z_0) = E_P(m(W, \theta)g_\ell(X)|Z = z_0) \cdot f_z(z_0).$$

For a given pair of $(\ell^{(1)}, \ell^{(2)})$, let $\mathcal{C}(\Theta^2)$ denote the space of continuous functions $\Sigma_P((\cdot, \ell^{(1)}), (\cdot, \ell^{(2)})) : \Theta^2 \to \Omega$. For notation simplicity, we write Σ_P to denote $\Sigma_P((\theta^{(1)}, \ell^{(1)}), (\theta^{(2)}, \ell^{(2)}))$ when it causes no confusion.

For a given θ_1, define

$$\Lambda_{n,P}(\theta_1) = \{(\theta, \xi) \in \Theta(\theta_1) \times \{R^k_{\pm\infty}\}_{\ell \in \mathcal{L}} : \xi_\ell = \sqrt{nh_n^{d_z}}\mu_\ell(\theta, z_0)\},$$

$$\Lambda^*_{n,P}(\theta_1) = \{(\theta, \xi) \in \Theta(\theta_1) \times \{R^k_{\pm\infty}\}_{\ell \in \mathcal{L}} : \xi_\ell = \kappa_n^{-1}\sqrt{nh_n^{d_z}}\mu_\ell(\theta, z_0)\},$$

$$\widehat{\Lambda}^*_{n,P}(\theta_1) = \{(\theta, \xi) \in \Theta(\theta_1) \times \{R^k_{\pm\infty}\}_{\ell \in \mathcal{L}} : \xi_\ell = \kappa_n^{-1}\sqrt{nh_n^{d_z}}\hat{\mu}_\ell(\theta, z_0)\}.$$

where $\mu_\ell(\theta, z_0) = E_P[g_\ell(X) \cdot m(W, \theta)|Z = z_0] \cdot f_z(z_0)$.

For any two points (θ, ξ) and (θ', ξ') in $\Theta \times \{R^k_{\pm\infty}\}_{\ell \in \mathcal{L}}$, define the metric as

$$d((\theta, \xi), (\theta', \xi')) = \Big[\sum_{j=1}^{d_\theta}(\Phi(\theta_j) - \Phi(\theta'_j))^2 + \sum_{q=1}^{\infty}\frac{1}{q^2 + 100}\sum_{\{\ell:\ r=q^{-1}\}} q^{-d_x}\sum_{j=1}^{k}(\Phi(\xi_{j,\ell}) - \Phi(\xi'_{j,\ell}))^2\Big]^{1/2},$$

where $\Phi(\cdot)$ is the CDF of the standard normal. Then it is true that the space $(\Theta \times \{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}}, d)$ constitutes a compact metric space because that $R_{\pm\infty}$ is a compact space under metric d_R with $d_R(r, r') = |\Phi(r) - \Phi(r)|$, $r, r' \in R_{\pm\infty}$. Let $\mathcal{S}(\Theta \times \{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}})$ denote the collection of compact subsets of the metric space $(\Theta \times \{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}}, d)$. Note that this is true only when the dimension of $\Theta \times \{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}}$ is countably many infinite and this is the main reason that we have to use instrument functions $\mathcal{G}_{\text{c-cube}}$ that is countably many. Let d_H denote the Hausdorff metric associated to the metric d, i.e., for any sets $A, B \subseteq \Theta \times \{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}}$,

$$d_H(A, B) = \max\Big\{ \sup_{(\theta,\xi)\in A} \inf_{(\theta',\xi')\in B} d((\theta, \xi), (\theta', \xi')), \sup_{(\theta',\xi')\in B} \inf_{(\theta,\xi)\in A} d((\theta, \xi), (\theta', \xi'))\Big\}.$$

At last, define the metric space $(\Theta(\theta_1) \times \{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}}, d)$ and the collection of its compact subsets $\mathcal{S}(\Theta(\theta_1) \times \{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}})$ analogously.

B Lemmas

Lemma B.1 *Suppose Assumptions 3.1-3.9 hold. Let $\{(\lambda_{u_n}, P_{u_n} \in \mathcal{H}_0)\}_{n\geq 1}$ be a (sub)sequence of parameters and distributions such that for some $(\Sigma, \Lambda_{\mathcal{L}}) \in \{\mathcal{C}(\theta^2)\}_{(\ell_1,\ell_2)\in\mathcal{L}^2} \times \mathcal{S}(\Theta \times \{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}})$, (i) $\Sigma_{P_{u_n}} \to \Sigma$ uniformly and (ii) $\Lambda_{u_n,P_{u_n}}(\theta_{u_n}) \xrightarrow{H} \Lambda_{\mathcal{L}}$. Then, along the (sub)sequence,*

$$\widehat{TS}_{u_n}(\theta_{1,u_n}) \xrightarrow{d} \inf_{(\theta,\lambda_{\mathcal{L}})\in\Lambda_{\mathcal{L}}} \sum_{q=1}^{\infty} \frac{1}{q^2 + 100} \sum_{\{\ell:\ r=q^{-1}\}} q^{-d_x} S(\Psi_\Sigma(\theta, \ell) + \lambda_\ell, \Sigma_\epsilon(\theta, \ell)), \tag{B.1}$$

where $\Psi_\Sigma : \Theta \times \mathcal{L} \to R^k$ is a R^k-valued tight Guassian process with covariance kernel $\Sigma \in \mathcal{C}(\theta^2)$, and $\Sigma_\epsilon = \Sigma(\theta, \ell) + \epsilon\Sigma(\theta, 1)$.

Proof. Without loss of generality, we let $u_n = n$. Recall that

$$\widehat{TS}_n(\theta_1) \equiv \inf_{\theta\in\Theta(\theta_1)} \widehat{T}_n(\theta, z_0),$$

where $\Theta(\theta_1) \equiv \{\tilde{\theta} \in \Theta : \ \tilde{\theta}_1 = \theta_1\}$ and

$$\widehat{T}_n(\theta, z_0) = \sum_{q=1}^{\infty} \frac{1}{q^2 + 100} \sum_{\{\ell: \ r=q^{-1}\}} q^{-d_x} S(\sqrt{nh_n^{d_z}} \hat{\mu}_n(\theta, \ell, z_0), \hat{\Sigma}_{\epsilon,n}(\theta, \ell, z_0)).$$

Let $\widehat{\Psi}_n(\theta, \ell, z_0) = \sqrt{nh_n^{d_z}}(\hat{\mu}_{\ell,n}(\theta, z_0) - \mu_\ell(\theta, z_0))$. We have

$$\begin{aligned}
\widehat{TS}_n(\theta_1) &= \inf_{\theta \in \Theta(\theta_1)} \sum_{q=1}^{\infty} \frac{1}{q^2 + 100} \sum_{\{\ell: \ r=q^{-1}\}} q^{-d_x} S(\sqrt{nh_n^{d_z}} \hat{\mu}_{\ell,n}(\theta, z_0), \hat{\Sigma}_{\epsilon,n}(\theta, \ell, z_0)) \\
&= \inf_{(\theta,\xi) \in \Lambda_{n,P}(\theta_1)} \sum_{q=1}^{\infty} \frac{1}{q^2 + 100} \sum_{\{\ell: \ r=q^{-1}\}} q^{-d_x} S(\widehat{\Psi}_n(\theta, \ell, z_0) + \xi_\ell, \hat{\Sigma}_{\epsilon,n}(\theta, \ell, z_0)).
\end{aligned}$$

For a generic uniform continuous function $\gamma : \Theta \times \mathcal{L} \to \mathbb{R}^K$, define

$$\begin{aligned}
g_n(\gamma(\cdot), \Sigma(\cdot)) &\equiv \inf_{(\theta,\xi) \in \Lambda_{n,P}(\theta_1)} \sum_{q=1}^{\infty} \frac{1}{q^2 + 100} \sum_{\{\ell: \ r=q^{-1}\}} q^{-d_x} S(\gamma(\theta, \ell) + \xi_\ell, \Sigma_\epsilon(\theta, \ell)), \ \text{and} \\
g(\gamma(\cdot), \Sigma(\cdot)) &\equiv \inf_{(\theta,\xi) \in \Lambda_{\mathcal{L}}(\theta_1)} \sum_{q=1}^{\infty} \frac{1}{q^2 + 100} \sum_{\{\ell: \ r=q^{-1}\}} q^{-d_x} S(\gamma(\theta, \ell) + \xi_\ell, \Sigma_\epsilon(\theta, \ell)).
\end{aligned}$$

Let $\{\gamma_n(\cdot), \Sigma_n(\cdot)\}_{n \geq 1}$ be a sequence of functions such that

$$\lim_{n \to \infty} \sup_{\theta \in \Theta(\theta_1)} \sum_{q=1}^{\infty} \frac{1}{q^2 + 100} \sum_{\{\ell: \ r=q^{-1}\}} q^{-d_x} \|(\gamma_n(\theta, \ell), \Sigma_n(\theta, \ell)) - (\gamma_n(\theta, \ell), \Sigma(\theta, \ell))\| = 0$$

where $\| \cdot \|$ denotes the Euclidean norm, then by the same argument of Theorem 3.1 of Bugni, Canay, and Shi (2015), we can show that

$$\lim_{n \to \infty} g_n(\gamma_n(\cdot), \Sigma_n(\cdot)) = g(\gamma(\cdot), \Sigma(\cdot)).$$

Therefore, Lemma B.1 holds following the extended continuous mapping theorem (Van Der Vaart and Wellner, 1996, Theorem 1.11.1) and by observing $\Psi_n \xrightarrow{d} \Psi_\Sigma$. □

Lemma B.2 *Suppose Assumptions 3.1-3.9 hold. Let* $\{(\lambda_{u_n}, P_{u_n} \in \mathcal{H}_0)\}_{n \geq 1}$ *be a (sub)sequence of parameters and distributions such that for some* $(\Sigma, \Lambda^*_{\mathcal{L}}) \in \{\mathcal{C}(\theta^2)\}_{(\ell_1,\ell_2) \in \mathcal{L}^2} \times \mathcal{S}^2(\Theta \times$

$\{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}})$, (i) $\Sigma_{P_{u_n}} \to \Sigma$ uniformly and (ii) $\Lambda^*_{u_n,P_{u_n},\mathcal{L}}(\theta_{u_n}) \overset{H}{\to} \Lambda^*_{\mathcal{L}}$. Then, there exists a further subsequence $\{k_n\}_{n\geq 1}$ of $\{u_n\}_{n\geq 1}$,

$$\widehat{TS}^u_{k_n}(\theta_{k_n}) \overset{d}{\to} \inf_{(\theta,\lambda_{\mathcal{L}})\in\Lambda^*_{\mathcal{L}}} \sum_{q=1}^{\infty} \frac{1}{q^2+100} \sum_{\{\ell:\ r=q^{-1}\}} q^{-d_x} S(\nu_\Sigma(\theta,\ell)+\lambda_\ell, \Sigma_\epsilon(\theta,\ell)), \tag{B.2}$$

conditional on the sample path almost surely.

Proof. First, by (ii) of Lemma B.5, we have

$$\sup_{(\theta,\ell))\in(\Theta(\theta_1),\mathcal{L})} \|\widehat{\Sigma}_n((\theta,\ell)) - \Sigma_P((\theta,\ell))\| \overset{p}{\to} 0,$$

and this is sufficient to show that

$$\sup_{(\theta,\ell))\in(\Theta(\theta_1),\mathcal{L})} \|\widehat{\Sigma}_{\epsilon,n}((\theta,\ell)) - \Sigma_{\epsilon,P}((\theta,\ell))\| \overset{p}{\to} 0.$$

Second, note that

$$\kappa_n^{-1}\sqrt{nh_n^{d_z}}\hat{\mu}_\ell(\theta,z_0) = \kappa_n^{-1}\widehat{\Psi}_n(\theta,\ell,z_0) + \kappa_n^{-1}\sqrt{nh_n^{d_z}}\mu_\ell(\theta,z_0)$$

and by (i) of Lemma B.5 and the fact that $\kappa_n^{-1}\to 0$, we have $d_H(\Lambda^*_{n,P}(\theta_1), \widehat{\Lambda}^*_{n,P}(\theta_1)) \overset{p}{\to} 0$. Then given that $d_H(\Lambda^*_{n,P}(\theta_1), \Lambda^*_{\mathcal{L}}) \to 0$, we have $d_H(\Lambda^*_{n,P}(\theta_1), \Lambda^*_{\mathcal{L}}) \overset{p}{\to} 0$.

Therefore, there exists a subsequence $\{k_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that (a) $\widehat{\Psi}_{k_n}(\cdot) \Rightarrow \Psi_\Sigma$ conditional on sample path almost surely, (b) $\sup_{(\theta,\ell))\in(\Theta(\theta_1),\mathcal{L})} \|\widehat{\Sigma}_n((\theta,\ell)) - \Sigma_P((\theta,\ell))\| \overset{a.s.}{\to} 0$ and (c) $d_H(\Lambda^*_{n,P}(\theta_1), \Lambda^*_{\mathcal{L}}) \overset{a.s.}{\to} 0$. Then by the same proof of Lemma B.1 and by conditional on the sample path, we have

$$\widehat{TS}^u_{k_n}(\theta_{k_n}) \overset{d}{\to} \inf_{(\theta,\lambda_{\mathcal{L}})\in\Lambda^*_{\mathcal{L}}} \sum_{q=1}^{\infty} \frac{1}{q^2+100} \sum_{\{\ell:\ r=q^{-1}\}} q^{-d_x} S(\nu_\Sigma(\theta,\ell)+\lambda_\ell, \Sigma_\epsilon(\theta,\ell)),$$

conditional on the sample path almost surely.□

Lemma B.3 *Let $\{(\theta_{1,u_n}, P_{u_n} \in \mathcal{H}_0)\}_{n\geq 1}$ be a (sub)sequence of parameters and distributions such that for some $(\Sigma, \Lambda_{\mathcal{L}}, \Lambda^*_{\mathcal{L}}) \in \{\mathcal{C}(\theta^2)\}_{(\ell_1,\ell_2)\in\mathcal{L}^2} \times \mathcal{S}^2(\Theta\times\{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}})$, (i) $\Sigma_{P_{u_n}} \to \Sigma$*

uniformly, (ii) $\Lambda_{u_n,P_{u_n},\mathcal{L}}(\theta_{1,u_n}) \xrightarrow{H} \Lambda_{\mathcal{L}}$ *and (iii)* $\Lambda^*_{u_n,P_{u_n},\mathcal{L}}(\theta_{1,u_n}) \xrightarrow{H} \Lambda^*_{\mathcal{L}}$*. Suppose Assumptions 3.1-3.9 hold. Then we have that for all* $(\theta,\xi^*) \in \Lambda^*_{\mathcal{L}}$ *such that* $\xi^*(\ell) \in R^p_{+\infty}(-\infty,\infty] \times R^{k-p}$ *for all* $\ell \in \mathcal{L}$ *with*

$$\sum_{q=1}^{\infty} \frac{1}{q^2+100} \sum_{\{\ell:\ r=q^{-1}\}} q^{-d_x} S(\xi^*(\ell), \Sigma_\epsilon(\theta,\ell)) < \infty,$$

there exists ξ *such that* $(\theta,\xi) \in \Lambda_{\mathcal{L}}$ *and* $\xi_j(\ell) \geq \xi^*_j(\ell)$ *for* $j \leq p$ *and* $\xi_j(\ell) \geq \xi^*_j(\ell)$ *for* $p < j \leq k$ *for all* $\ell \in \mathcal{L}$*.*

Proof. We apply the proof of Lemma S.3.8 of Bugni, Canay, and Shi (2017) to show our case. Without loss of generality, let $u_n = n$. If $(\theta,\xi^*) \in \Lambda^*_{\mathcal{L}}$, there exists a sequence $\{\theta_n\}$ such that $\theta_n \in \Theta(\theta_{1,n})$ with $\theta_n \to \theta$, and $\kappa_n^{-1}\sqrt{nh_n^{d_z}}\mu_\ell(\theta_n,z_0) \to \xi^*(\ell)$ for all $\ell \in \mathcal{L}$. Similar to (S.16) of Bugni, Canay, and Shi (2017), there exists a sequence of $\tilde{\theta}_n \in \Theta_{P_n}(\theta_{1,n},z_0)$ such that $\|\theta_n - \tilde{\theta}_n\| \leq O(\kappa_n/\sqrt{nh_n^{d_x}})$. Define $\hat{\theta}_n = (1-\kappa_n^{-1})\tilde{\theta}_n + \kappa_n^{-1}\theta_n$. By the same arguments of (S.17) and (S.18), we have

$$\sqrt{nh_n^{d_z}}\mu_\ell(\hat{\theta}_n,z_0) = \kappa_n^{-1}\sqrt{nh_n^{d_z}}\mu_\ell(\theta_n,z_0) + \epsilon_{1,n}(\ell) + \epsilon_{2,n}(\ell)$$

where $\epsilon_{1,n}(\ell) = (\nabla_\theta\mu_\ell(\theta_n^{**},z_0) - \nabla_\theta\mu_\ell(\theta_n^*,z_0))\sqrt{nh_n^{d_z}}(\hat{\theta}_n - \theta_n)$ with θ_n^* and θ_n^{**} both being between $\hat{\theta}_n$ and θ_n, and $\epsilon_{2,n}(\ell) = (1-\kappa_n^{-1})\sqrt{nh_n^{d_z}}\mu_\ell(\tilde{\theta}_n,z_0)$. Note that $\tilde{\theta}_n \in \Theta_{P_n}(\theta_{1,n},z_0)$ and $\kappa_n^{-1} \to 0$, so it follows that $\epsilon_{2,n,j}(\ell) \geq 0$ for $j \leq p$ and $\epsilon_{2,n,j}(\ell) = 0$ for $j > p$ for all ℓ. Note that $\nabla_\theta\mu_\ell(\theta,z_0) = E[g_\ell(X)\mu_\ell(\theta,X,Z)|Z=z_0]$ and by Assumption 3.5 1., it is true that $\|\nabla_\theta\mu_\ell(\theta_n^{**},z_0) - \nabla_\theta\mu_\ell(\theta_n^*,z_0)\| \leq cQ\|\theta_n^{**} - \theta_n^*\|$ for some positive constant c not depending on ℓ. Therefore, we have $\|\nabla_\theta\mu_\ell(\theta_n^{**},z_0) - \nabla_\theta\mu_\ell(\theta_n^*,z_0)\| = o(1)$ uniformly over ℓ. By the fact that $\sqrt{nh_n^{d_x}}\|\hat{\theta}_n - \tilde{\theta}_n\| = O(1)$, we have uniformly over ℓ,

$$\|\epsilon_{1,n}(\ell)\| \leq \|(\nabla_\theta\mu_\ell(\theta_n^{**},z_0) - \nabla_\theta\mu_\ell(\theta_n^*,z_0))\|\sqrt{nh_n^{d_x}}\|\hat{\theta}_n - \tilde{\theta}_n\| = o(1).$$

Given that the space $(\Theta \times \{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}}, d)$ constitutes a compact metric space, it is true that there exists a subsequence $\{u_n\}$ of $\{n\}$ such that $\sqrt{u_nh_{u_n}^{d_z}}\mu_\ell(\hat{\theta}_{u_n},z_0)$ and $\kappa_{u_n}^{-1}\sqrt{u_nh_{u_n}^{d_z}}\mu_\ell(\theta_{u_n},z_0)$ converge for all ℓ. To be specific, $\{R^k_{\pm\infty}, d_k\}$ where for any two points $\delta_1, \delta_2 \in R^k_{\pm\infty}$,

$d_k(\theta_1,\theta_2) = (\sum_{j=1}^{k}(\Phi(\theta_{1,j}) - \Phi(\theta_{2,j})^2)^{1/2}$ is a compact set. Note that because $\mathcal{L}$ is countable, we can order $\ell = 1, 2, \ldots$ with those ℓ's with smaller q being ordered first. For $\ell = 1$, then there exists a subsequence $\{a_{1,n}\}$ of $\{n\}$ so that

$$\xi_j(1) = \lim_{n\to\infty} \sqrt{a_{1,n}h_{a_{1,n}}^{d_z}}\mu_\ell(\hat{\theta}_{a_{1,n}}, z_0) \geq \lim_{n\to\infty} \kappa_{a_{1,n}}^{-1}\sqrt{a_{1,n}h_{a_{1,n}}^{d_z}}\mu_\ell(\theta_{a_{1,n}}, z_0) = \xi_j^*(1) \text{ for } j \leq p,$$
$$\xi_j(1) = \lim_{n\to\infty} \sqrt{a_{1,n}h_{a_{1,n}}^{d_z}}\mu_\ell(\hat{\theta}_{a_{1,n}}, z_0) = \lim_{n\to\infty} \kappa_{a_{1,n}}^{-1}\sqrt{a_{1,n}h_{a_{1,n}}^{d_z}}\mu_\ell(\theta_{a_{1,n}}, z_0) = \xi_j^*(1) \text{ for } j \leq p.$$

Similarly, for $\ell = 2$, there exists a subsequence $\{a_{2,n}\}$ of $\{a_{1,n}\}$ so that

$$\xi_j(2) = \lim_{n\to\infty} \sqrt{a_{2,n}h_{a_{2,n}}^{d_z}}\mu_\ell(\hat{\theta}_{a_{2,n}}, z_0) \geq \lim_{n\to\infty} \kappa_{a_{2,n}}^{-1}\sqrt{a_{2,n}h_{a_{2,n}}^{d_z}}\mu_\ell(\theta_{a_{2,n}}, z_0) = \xi_j^*(2) \text{ for } j \leq p,$$
$$\xi_j(2) = \lim_{n\to\infty} \sqrt{a_{2,n}h_{a_{2,n}}^{d_z}}\mu_\ell(\hat{\theta}_{a_{2,n}}, z_0) = \lim_{n\to\infty} \kappa_{a_{2,n}}^{-1}\sqrt{a_{2,n}h_{a_{2,n}}^{d_z}}\mu_\ell(\theta_{a_{2,n}}, z_0) = \xi_j^*(2) \text{ for } j \leq p.$$

Then we keep doing this for $\ell = 3, 4, \ldots$ and set $\{u_n\} = \{a_{n,n}\}$. This completes the proof.□

Lemma B.4 *Suppose Assumptions 3.1-3.9 hold. For any (sub)sequence* $\{(\theta_{u_n}, P_{u_n} \in \mathcal{H}_0)\}_{n\geq 1}$*, there exists a further subsequence* $\{k_n\}_{n\geq 1}$ *of* $\{u_n\}_{n\geq 1}$ *such that (i)* $\Sigma_{P_{k_n}} \to \Sigma$ *uniformly, (ii)* $\Lambda_{k_n,P_{k_n},\mathcal{L}}(\theta_{k_n}) \xrightarrow{H} \Lambda_{\mathcal{L}}$ *and (iii)* $\Lambda^*_{k_n,P_{k_n},\mathcal{L}}(\theta_{k_n}) \xrightarrow{H} \Lambda^*_{\mathcal{L}}$ *for some* $(\Sigma, \Lambda_{\mathcal{L}}, \Lambda^*_{\mathcal{L}}) \in \{\mathcal{C}(\theta^2)\}_{(\ell_1,\ell_2)\in\mathcal{L}^2} \times \mathcal{S}^2(\Theta \times \{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}})$.

Proof. We apply the proof of Lemma D.7 of Bugni, Canay, and Shi (2015) to show our case. For $\ell = 1$, by the same arguments of Lemma D.7 of Bugni, Canay, and Shi (2015), we can show that there exists a subsequence $\{a_{1,n}\}$ of $\{n\}$ such that

$$\Sigma_{P_{a_{1,n}}}((\cdot,\ell_1),(\cdot,\ell_2)) \to \Sigma((\cdot,\ell_1),(\cdot,\ell_2)) \text{ uniformly for } \ell_1,\ell_2 \in \{1\},$$
$$\Lambda_{a_{1,n},P_{a_{1,n}},\ell}(\theta_{a_{1,n}}) \xrightarrow{H} \Lambda_\ell,$$
$$\Lambda^*_{a_{1,n},P_{a_{1,n}},\ell}(\theta_{a_{1,n}}) \xrightarrow{H} \Lambda^*_\ell,$$

for some $(\Sigma, \Lambda_{\mathcal{L}}, \Lambda^*_{\mathcal{L}}) \in \{\mathcal{C}(\theta^2)\}_{(\ell_1,\ell_2)\in\mathcal{L}^2} \times \mathcal{S}^2(\Theta \times \{R^k_{\pm\infty}\}_{\ell\in\mathcal{L}})$. For $\ell = 2$, we can show that there exists a subsequence $\{a_{2,n}\}$ of $\{a_{1,n}\}$ such that

$$\Sigma_{P_{a_{1,n}}}((\cdot,\ell_1),(\cdot,\ell_2)) \to \Sigma((\cdot,\ell_1),(\cdot,\ell_2)) \text{ uniformly for } \ell_1,\ell_2 \in \{1,2\},$$
$$\Lambda_{a_{2,n},P_{a_{2,n}},\ell}(\theta_{a_{2,n}}) \xrightarrow{H} \Lambda_\ell,$$
$$\Lambda^*_{a_{2,n},P_{a_{2,n}},\ell}(\theta_{a_{2,n}}) \xrightarrow{H} \Lambda^*_\ell.$$

Then we keep doing this for $\ell = 3, 4, \ldots$ and set $\{k_n\} = \{a_{n,n}\}$. This completes the proof. □

Lemma B.5 *Suppose Assumptions 3.1-3.9 hold. Let $\{P_{u_n} \in \mathcal{P}\}_{n\geq 1}$ be a (sub)sequence of distributions such that for some $\Sigma \in \{\mathcal{C}(\Theta^2)\}_{(\ell_1,\ell_2)\in\mathcal{L}^2}$, $\Sigma_{P_{u_n}} \to \Sigma$ uniformly. Then, the following statements hold:*

(i) $\widehat{\Psi}_{u_n}(\cdot) \Rightarrow \Psi_\Sigma$, where Ψ_Σ is a tight zero-mean Guassian process with covariance kernel Σ. In addition, for any fixed $\epsilon > 0$, there exists a $\delta > 0$ such that

$$Pr\Big(\sup_{\|\theta^{(1)}-\theta^{(2)}\|\leq\delta}\ \sup_{\ell\in\mathcal{L}} \|\Psi_\Sigma(\theta^{(1)},\ell) - \Psi_\Sigma(\theta^{(2)},\ell)\| \leq \epsilon\Big) = 1.$$

(ii) We have

$$\sup_{(\theta^{(1)},\ell^{(1)}),(\theta^{(2)},\ell^{(2)})\in(\Theta(\theta_1),\mathcal{L})} \|\widehat{\Sigma}_n((\theta^{(1)},\ell^{(1)}),(\theta^{(2)},\ell^{(2)})) - \Sigma_P((\theta^{(1)},\ell^{(1)}),(\theta^{(2)},\ell^{(2)}))\| \xrightarrow{p} 0, \text{ where}$$
$$\widehat{\Sigma}_n((\theta^{(1)},\ell^{(1)}),(\theta^{(2)},\ell^{(2)})) = \frac{1}{nh_n^{d_z}}\sum_{i=1}^n \Big(K\Big(\frac{Z_i - z_0}{h_n}\Big)g_{\ell^{(1)}}(X_i)m(W_i,\theta^{(1)}) - \hat{\mu}_{\ell^{(1)},n}(\theta^{(1)},z_0)\Big) \cdot \Big(K\Big(\frac{Z_i - z_0}{h_n}\Big)g_{\ell^{(2)}}(X_i)m(W_i,\theta^{(2)}) - \hat{\mu}_{\ell^{(2)},n}(\theta^{(2)},z_0)\Big)'.$$

(iii) We have $\Psi^u_n(\cdot) \Rightarrow \Psi_\Sigma$ conditional on sample path with probability 1.

Proof. Parts (i) and (ii) are the same as those of Lemma AN3 of Andrews and Shi (2014). Given part (ii), the proof of part (iii) follows from the same argument of Theorem 4.1 of Hsu (2016). □

C Proof of Theorems

Proof of Theorem 3.1. Given Lemma B.1-Lemma B.5 above, the proof to Theorem 3.1 follows the same arguments of Equation (C.5) of Bugni, Canay, and Shi (2017), and we omit the details for brevity.

Proof of Theorem 3.2. The proof of Theorem 3.2 follows analogously from those in Theorem 3.1. In particular, the limiting distribution of $\min_{\theta\in\Theta}\widehat{T}_n(\theta, z_t)$ can be obtained in a similar way as in Lemma B.1. For a set of pre-chosen grid points $\{z_1, \ldots, z_T\}$, $\min_{\theta\in\Theta\widehat{T}_n(\theta,z_t)}$ are mutually asymptotically independent, so their asymptotic joint distribution is the product of their asymptotic marginal distributions. Finally, the max operator is a continuous function, so the limiting distribution of $\widehat{T}_n$ follows by continuous mapping theorem. The validity of multipier bootstrap holds as shown in Lemma B.5.

The results in Corollary 3.1 hold because (i) the critical value $C_n^u(\alpha)$ is stochastically bounded, and (ii) $\frac{\widehat{T}_n}{nh^{d_z}} - c_n \xrightarrow{p} 0$.

References

Ahmad, I., S. Leelahanon, and Q. Li (2005): "Efficient estimation of a semiparametric partially linear varying coefficient model," *The Annals of Statistics*, 33(1), 258–283.

Andrews, D. W., S. Berry, and P. Jia (2004): "Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location," *manuscript, Yale University*.

Andrews, D. W., and P. Guggenberger (2009): "Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities," *Econometric Theory*, 25(3), 669–709.

Andrews, D. W., and S. Kwon (2019): "Inference in moment inequality models that is robust to spurious precision under model misspecification," *Cowles Foundation Discussion Paper*.

Andrews, D. W., and X. Shi (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81(2), 609–666.

——— (2014): "Nonparametric inference based on conditional moment inequalities," *Journal of Econometrics*, 179(1), 31–45.

——— (2017): "Inference based on many conditional moment inequalities," *Journal of econometrics*, 196(2), 275–287.

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

Ang, A., and J. Liu (2004): "How to discount cashflows with time-varying expected returns," *The Journal of Finance*, 59(6), 2745–2783.

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

——— (2015): "Asymptotically exact inference in conditional moment inequality models," *Journal of Econometrics*, 186(1), 51–65.

——— (2018): "On the choice of test statistic for conditional moment inequalities," *Journal of Econometrics*, 203(2), 241–255.

Belloni, A., F. A. Bugni, and V. Chernozhukov (2019): "Subvector inference in PI models with many moment inequalities," Discussion paper.

Bontemps, C., and T. Magnac (2017): "Set identification, moment restrictions, and inference," *Annual Review of Economics*, 9, 103–129.

Bugni, F. A., I. A. Canay, and X. Shi (2015): "Specification tests for partially identified models defined by moment inequalities," *Journal of Econometrics*, 185(1), 259–282.

——— (2017): "Inference for subvectors and other functions of partially identified parameters in moment inequality models," *Quantitative Economics*, 8(1), 1–38.

Cai, Z. (2010): "Functional Coefficient Models for Economic and Financial Data," *Handbook of Functional Data Analysis, eds. F. Ferraty and Y. Romain*, pp. 166–186, Oxford University Press.

Cai, Z., L. Chen, and Y. Fang (2018): "A semiparametric quantile panel data model with an application to estimating the growth effect of FDI," *Journal of Econometrics*, 206(2), 531–553.

Cai, Z., M. Das, H. Xiong, and X. Wu (2006): "Functional coefficient instrumental variables models," *Journal of Econometrics*, 133(1), 207–241.

Cai, Z., J. Fan, and R. Li (2000): "Efficient estimation and inferences for varying-coefficient models," *Journal of the American Statistical Association*, 95(451), 888–902.

Cai, Z., Y. Fang, M. Lin, and J. Su (2019): "Inferences for a partially varying coefficient model with endogenous regressors," *Journal of Business & Economic Statistics*, 37(1), 158–170.

CAI, Z., AND Y. HONG (2009): "Some recent developments in nonparametric finance," in *Nonparametric Econometric Methods*, pp. 379–432. Emerald Group Publishing Limited.

CAI, Z., Y. REN, AND B. YANG (2015): "A semiparametric conditional capital asset pricing model," *Journal of Banking & Finance*, 61, 117–126.

CAI, Z., AND X. XU (2008): "Nonparametric quantile estimations for dynamic smooth coefficient models," *Journal of the American Statistical Association*, 103(484), 1595–1608.

CANAY, I. A., AND A. M. SHAIKH (2017): "Practical and theoretical advances in inference for partially identified models," *Advances in Economics and Econometrics*, 2, 271–306.

CHEN, R., AND R. S. TSAY (1993): "Functional-coefficient autoregressive models," *Journal of the American Statistical Association*, 88(421), 298–308.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and confidence regions for parameter sets in econometric models 1," *Econometrica*, 75(5), 1243–1284.

CHERNOZHUKOV, V., W. KIM, S. LEE, AND A. M. ROSEN (2015): "Implementing intersection bounds in Stata," *The Stata Journal*, 15(1), 21–44.

CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): "Intersection bounds: Estimation and inference," *Econometrica*, 81(2), 667–737.

CHERNOZHUKOV, V., W. K. NEWEY, AND A. SANTOS (2015): "Constrained conditional moment restriction models," *arXiv preprint arXiv:1509.06311*.

CILIBERTO, F., AND E. TAMER (2009): "Market structure and multiple equilibria in airline markets," *Econometrica*, 77(6), 1791–1828.

FAN, J., AND R. LI (2004): "New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis," *Journal of the American Statistical Association*, 99(467), 710–723.

Fan, J., and J.-T. Zhang (2000): "Two-step estimation of functional linear models with applications to longitudinal data," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2), 303–322.

Fan, J., and W. Zhang (1999): "Statistical estimation in varying coefficient models," *The annals of Statistics*, 27(5), 1491–1518.

Hastie, T., and R. Tibshirani (1993): "Varying-coefficient models," *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4), 757–779.

Honda, T. (2004): "Quantile regression in varying coefficient models," *Journal of Statistical Planning and Inference*, 121(1), 113–125.

Hong, S. (2017): "Inference in semiparametric conditional moment models with partial identification," *Journal of econometrics*, 196(1), 156–179.

Hsu, Y.-C. (2016): "Multiplier bootstrap for empirical processes," Discussion paper, Institute of Economics, Academia Sinica, Taipei, Taiwan.

Hsu, Y.-C., and X. Shi (2017): "Model-selection tests for conditional moment restriction models," *The Econometrics Journal*, 20(1), 52–85.

Imbens, G. W., and C. F. Manski (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72(6), 1845–1857.

Kaido, H., F. Molinari, and J. Stoye (2019): "Confidence intervals for projections of partially identified parameters," *Econometrica*, 87(4), 1397–1432.

Kédagni, D., and I. Mourifié (2020): "Generalized instrumental inequalities: testing the instrumental variable independence assumption," *Biometrika*, 107(3), 661–675.

Kim, K. i. (2008): "Set estimation and inference with models characterized by conditional moment inequalities," *Working Paper*.

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

Li, Q., C. J. Huang, D. Li, and T.-T. Fu (2002): "Semiparametric smooth coefficient models," *Journal of Business & Economic Statistics*, 20(3), 412–422.

Menzel, K. (2014): "Consistent estimation with many moment inequalities," *Journal of Econometrics*, 182(2), 329–350.

Mourifié, I., and Y. Wan (2017): "Testing local average treatment effect assumptions," *Review of Economics and Statistics*, 99(2), 305–313.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment inequalities and their application," *Econometrica*, 83(1), 315–334.

Romano, J. P., and A. M. Shaikh (2008): "Inference for identifiable parameters in partially identified econometric models," *Journal of Statistical Planning and Inference*, 138(9), 2786–2807.

——— (2010): "Inference for the identified set in partially identified econometric models," *Econometrica*, 78(1), 169–211.

Santos, A. (2012): "Inference in nonparametric instrumental variables with partial identification," *Econometrica*, 80(1), 213–275.

Su, L., I. Murtazashvili, and A. Ullah (2013): "Local linear GMM estimation of functional coefficient IV models with an application to estimating the rate of return to schooling," *Journal of Business & Economic Statistics*, 31(2), 184–207.

Tao, J. (2015): "On the Asymptotic Theory for Semiparametric GMM Models with Partial Identification," Discussion paper.

Van Der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Wan, Y. (2013): "An integration-based approach to moment inequality models," *Manuscript. University of Toronto.*