Set Inferences and Sensitivity Analysis in Semiparametric Partially Identified Models*

Juan Carlos Escanciano† — *Indiana University*

Lin Zhu‡ — *Tsinghua University*

This version: April 20th, 2013. First version: December 8th, 2011.

Abstract

This paper provides inference tools for partially identified semiparametric models. The main working assumption is that the finite-dimensional parameter of interest and the infinite-dimensional nuisance parameters are identified conditionally on other nuisance parameters being known. This structure arises in numerous applications and permits the use of standard tools from empirical processes theory. We develop uniform convergence for a set of two-step GMM estimators and use the uniformity to establish set inference. We allow for the identified set of nuisance parameters to be unknown and estimated, which in general requires to establish inference under local misspecification. A global misspecification analysis is useful in a formal development of sensitivity analysis. Inference is implemented with the assistance of bootstrap methods. Several examples illustrate the wide applicability of our results.

Keywords: Partial Identification; Semiparametric models; Sensitivity analysis; Willingness to pay; Empirical process theory.

JEL classification: C12, C13, C14

*We would like to thank Stéphane Bonhomme, Kei Hirano, Arthur Lewbel, Enrique Sentana, Ellie Tamer, as well as seminar participants at UCL, ESEM Malaga, Oxford, CEMFI and Georgetown for helpful discussions.

†Department of Economics, Indiana University, 105 Wylie Hall, 100 South Woodlawn Avenue, Bloomington, IN 47405–7104, USA. E-mail: jescanci@indiana.edu. Web Page: http://mypage.iu.edu/~jescanci/.

Research funded by the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908.

‡School of Economics and Management, Tsinghua University, Beijing 100084, China. E-mail: zhulin@sem.tsinghua.edu.cn.

1 Introduction

In many economic models, point identification of parameters of interest is often at the cost of ad-hoc assumptions on some nuisance parameters. Relaxing those assumptions in general leads to loss of point identification, and instead only a set of parameters can be identified from the data, a situation that has been referred to as partial identification or set identification; see Manski (2003, 2007) for textbook treatments and Tamer (2010) for a recent survey on theoretical and empirical developments. The vast majority of existing inference methods in a partially identified context deal with models with finite-dimensional parameters. In this paper, we propose a general, yet simple, framework to implement estimation and inference in partially identified *semiparametric* models with tools that are relatively standard in the econometrics literature.

We investigate inferences in a class of semiparametric models containing a vector of finite-dimensional parameters $(\theta_0, \tau_0) \in \Theta \times \mathcal{T}$, $\Theta \subset \mathbb{R}^{d_\theta}$, $\mathcal{T} \subset \mathbb{R}^{d_\tau}$, a possibly infinite-dimensional nuisance parameter $h_0 \in \mathcal{H}$, for a suitable class of functions $\mathcal{H}$, and satisfying the moment restrictions

$$E[\psi(W, \theta_0, h_0(W), \tau_0)] = 0, \tag{1}$$

where W is a d_w-dimensional observable random vector, and $\psi(\cdot, \theta, h(\cdot), \tau)$ is a measurable moment function from $\mathbb{R}^{d_w}$ to $\mathbb{R}^{d_\psi}$, for each $\gamma := (\theta, h(\cdot), \tau) \in \Gamma := \Theta \times \mathcal{H} \times \mathcal{T}$. We are particularly concerned with a situation where θ_0, the parameter of interest, is not identified by the moment restrictions (1). That is, the identified set

$$\Gamma_0 := \{\gamma \in \Gamma : E[\psi(W, \gamma)] = 0\},$$

contains at least two distinct elements with different θ components. However, we still assume the parameter θ and the infinite-dimensional nuisance parameter h are partially identified in the sense that, for each fixed $\tau \in \mathcal{T}_0 \subset \mathcal{T}$, there exists a unique solution to (1) in (θ_0, h_0), say $\theta_0(\tau) \in \Theta$ and $h_0(\cdot, \tau) \in \mathcal{H}$. The possibly unknown set $\mathcal{T}_0$ can be viewed as the projection of Γ_0 onto the nuisance parameter set $\mathcal{T}$. Hence, the identified set for θ is $\Theta_0 := \{\theta \in \Theta : \theta = \theta_0(\tau) \text{ for some } \tau \in \mathcal{T}_0\}$. That is, our partial identification assumption implies that the identified set Θ_0 is a finite-dimensional manifold.

This setting turns out to be general enough to be applicable to many parametric and semiparametric partially identified models considered in the literature, and to many other new applications, while permitting the use of relatively standard methods of analysis. Roughly speaking, the nuisance parameter τ can embody classical assumptions that "complete" the model and deliver point identification, and variation of $\theta_0(\tau)$ in τ quantifies the sensitivity of the parameter of interest to these assumptions. The set $\mathcal{T}_0$ can be interpreted as the set of competing alternative identifying assumptions that are consistent with the data and it can contain a finite or an infinite number of elements.

We consider a limited information approach.[1] For each fixed $\tau \in \mathcal{T}_0$, we assume the existence of a first-step estimator for $h_0(\cdot, \tau)$. Given this first-step estimator, we then propose a semiparametric two-step Generalized Method of Moments (GMM) estimator for $\theta_0(\tau)$ in (1). We establish the asymptotic distribution for the estimator $\widehat{\theta}(\cdot)$ as a process in τ, and use this to construct confidence regions

[1] This is just for simplicity in the notation. We show below how our results can be easily extended to a full information setting.

for Θ_0 and unknown members of Θ_0. Inferences are implemented with the assistance of bootstrap methods. Extensions of this setting are considered that allow for $\mathcal{T}_0$ to be unknown and estimated, and more generally, allow for inference on $\theta_0(\tau)$ for $\tau \in \mathcal{T}_1 \subsetneq \mathcal{T}_0$. These cases require an analysis under misspecification, where the definition of $\theta_0(\tau)$ has to be modified accordingly.

This approach to partial identification has precedents in the literature, albeit for specific, mostly parametric, models. To the best of our knowledge, this setting was first proposed for modeling partial identification by Sargan (1959) in a context of instrumental variables (IV) models. In the classical setting of demand and supply simultaneous equations, Leamer (1981) shows that the identified set can be written as in (1) for a non-convex identified set. For instance, in a demand and supply model with no exogenous variables and uncorrelated errors, the identified set for the slope parameters is a section of a hyperbola. For more general linear simultaneous equations, Phillips (1989) shows the existence of reparametrizations fitting our setting with $\theta_0(\tau) \equiv \theta_0$. He further investigated the asymptotic theory of estimators and test statistics under partial identification. This earlier approach to partial identification has been recently extended to general parametric moment restrictions by Arellano, Hansen and Sentana (2012). These authors provide a test for underidentification and example applications in linear IV models, dynamic panel data, Phillips curves and asset pricing models. Further example applications in the context of dynamic models include structural vector autoregressions with sign restrictions, see Rubio-Ramirez, Waggoner and Zha (2010). More generally, in dynamic macro models calibration has been routinely employed, which can be interpreted as a pointwise version of our partial identification approach. The inferences we consider here are different from those in the aforementioned papers, in that they apply to the identified set and unknown members of this set.

The scope of applications of our methods is substantially broader than the examples mentioned above, see Manski (2003, 2007). Parametrized identified sets that fit our setting arise naturally in models with missing, contaminated, misclassified or censored data, see e.g. Horowitz and Manski (1995, 2000, 2006) and references below. In these models τ may parametrize the counterfactual conditional probability or selection probability. For instance, Chernozhukov, Rigobon and Stoker (2010) study inferences for the parameter associated with a Tobin regressor (an endogenous, censored and selected regressor). Their model satisfies (1) for certain (non-smooth) moments, with τ denoting the selection conditional probability. Our results can be used to complement their pointwise results with uniform inferences in semiparametric versions of their model.

There is also an extensive literature providing sensitivity analysis as a method to quantify the impact of relaxing strong identification assumptions on parameters of interest, see, e.g., Rosenbaum and Rubin (1983), Imbens (2003), or more recently, Kline and Santos (2012). Typically, in this literature the focus is not on set inference, but in the modulus of continuity of $\theta_0(\tau)$, for example, how large has to be τ (e.g. a measure of selection on unobservables) to obtain a given deviation of $\theta_0(\tau)$ from $\theta_0(\tau_0)$, where τ_0 is a benchmark case, e.g. $\tau_0 = 0$ (no selection). Our inference deals with the whole curve $\theta_0(\tau)$, and therefore includes sensitivity analysis as a special case. Another important difference between set inference and sensitivity analysis is that in the former $\tau \in \mathcal{T}_0$ whereas in the latter $\tau \in \mathcal{T}_1 \subsetneq \mathcal{T}_0$; see e.g. Imbens (2003), Altonji, Elder and Taber (2005) or Conley, Hansen and Rossi (2012), to mention just a few. This implies that typical sensitivity analyses should account for misspecification, a result that has been overlooked in the literature. These arguments led us to consider a generic set $\mathcal{T}_1 \subset \mathcal{T}$ and allow for misspecification in our uniform results.[2]

Although the literature on sensitivity analysis has been mainly confined to parametric models, there are some applications in semiparametric models. For instance, Scharfstein, Rotnitzky and Robins (1999) study how non-ignorable drop-out affects inferences on the mean of the outcome of interest in a semiparametric model for panel data. They showed that their model satisfies (1) for certain moments, where θ is the mean of the outcome variable, h is a cumulative conditional hazard function, and τ is a selection bias parameter. For a fixed τ, they show that θ and h are identified, and they carried out a pointwise sensitivity analysis by varying the selection bias parameter τ over a plausible range of values. More recently, Kline and Santos (2012) show how to nonparametrically parametrized missing data problems using a Kolmogorov-Smirnov distance between the distributions of missing and observed outcomes. We show below how our setting can accommodate inferences on parameters such as average partial effects in a more general setting that includes Kline and Santos (2012) as a special case.

Closely related to our work is work by Bontemps, Magnac and Maurin (2012), who consider linear "incomplete" models satisfying (1) with an infinite-dimensional nuisance parameter τ but no nuisance parameters h. They show that many examples fall under this structure, including the case of regression models with interval dependent data, which has been a leading example investigated in the literature, see e.g. Manski and Tamer (2002). The convexity and boundedness of $\mathcal{T}$ and the linearity of the moment function ψ in the nuisance function τ leads to a convex and closed identified set, and Bontemps et al. (2012) exploit the convexity to develop inference based on the support function of the identified set. In their model, the support function is the expectation of a suitable function indexed by a finite-dimensional parameter (a direction in the unit sphere). See also Beresteanu and Molinari (2008) for general models with convex identified sets. As shown by these authors, inference about the support function can be also carried out in the setting of (1) where τ now denotes a direction and θ_0 is the value of the support function in that direction, see Section 2 for details on this formulation. Our results then extend those in Bontemps et al. (2012) and Beresteanu and Molinari (2008) by permitting the moments characterizing the support function to depend on infinite-dimensional nuisance parameters as well.

We illustrate our ideas with three examples below. The first example investigates inferences about the mean of functions of willingness to pay in contingent valuations. Semiparametric point-identification results in this setting were investigated in Lewbel, McFadden and Linton (2010). In a second generic example the parameter of interest is a bounded linear functional of a function that is known to lie within a band. This is an example that has numerous applications in economics, see Manski (2003, 2007). We show how our results can be used to extend the recent results by Chandrasekhar, Chernozhukov, Molinari and Schrimpf (2012), Kline and Santos (2012) and Pacini (2012) from best linear approximation functionals to general linear continuous functionals, including but not restricted to average partial or increment effects and counterfactual distributional effects.[3] A third example shows how to carry out

[2] Allowing for misspecification is also important in sensitivity analysis because often ad-hoc functional form assumptions on the unobserved heterogeneity, missing data or selection distributions are used.

[3] Our main tool is the Riesz' representation theorem. We show how this tool avoids making additional bound restrictions

sensitivity analysis in non-separable models with limited exogeneity (e.g. no exclusion restrictions). We illustrate the main ideas with a semiparametric binary choice model with selection. Examples 1 and 2 lead to convex identified sets with semiparametric support functions, while Example 3 leads to an identified set which is a non-convex nonlinear manifold. Example 2 is also useful in showing how our theory can accommodate an infinite number of moment restrictions, and therefore, full-information settings.

Our paper belongs to the rapidly growing literature on inferences in partially identified models. When the identified set is a closed interval, Horowitz and Manski (2000) develop confidence intervals for the entire identified set, while Imbens and Manski (2004) and Stoye (2009) discuss methods for constructing confidence intervals for the true value. In a general setup, Chernozhukov, Hong and Tamer (2007) develop a unified criterion function approach for estimation and inference in partially identified models, generalizing results in M-estimation theory from point identification to partial identification. They show the consistency of their level set estimator and obtain rates of convergence. Inference is based on subsampling. For an alternative proposal in the same setting see Romano and Shaikh (2008, 2010). Moment inequalities are leading examples of this literature, see e.g. Andrews and Jia (2012), Andrews and Guggenberger (2009), Andrews and Soares (2010) and Bugni (2010), among others. In models with convex identified sets, Beresteanu and Molinari (2008) propose general methods based on set-valued random elements and support functions, establishing limiting distributions for their test statistics. Recently, Kaido (2011) has investigated the connections between the criterion function approach and the support function approach when the identified set is convex. These aforementioned papers deal with partially identified models with no infinite-dimensional nuisance parameters to be estimated.

The literature on general semiparametric partially identified models is more recent and rather scarce. Song, Kosorok and Fine (2009) use profile likelihood methods to propose optimal tests in semiparametric models with parameters that are not identified under the null, extending previous results in Andrews and Ploberger (1994). Chen, Tamer and Torgovitsky (2010) propose inverting profiled likelihood ratio tests to construct confidence sets for finite-dimensional parameters. These works allow for the finite-dimensional parameter to be estimated at a slower rate than the regular (parametric) one. Hong (2011) considers semiparametric conditional moment models, with infinite-dimensional parameters approximated by sieves, extending previous important results by Santos (2010). In this paper we restrict attention to regular estimators and finitely parametrized infinite-dimensional nuisance parameters, so our results are less general in that respect, but in examples featuring a partial identification structure such as ours, the proposed methods lead to much simpler and more efficient implementations[4]. Thus, sieve nonparametric methods and our methods are complements rather than substitutes, and they can be potentially combined in extensions of our basic setting where τ is infinite-dimensional.

The rest of the paper is organized as follows. Section 2 illustrates the general applicability of our

on the derivatives of unidentified components of the model, cf. Manski (1989, p.348).

[4] Our results can be extended to non-regular situations using well-known results from the empirical process literature. For an interesting application of these tools in the context of partially identified models see Lee, Seo and Shin (2011).

methods with several motivating examples. Section 3 develops uniform inferences for the possibly misspecified version of the partially identified model in (1). Section 4 applies the uniform results to construct confidence regions for the identified set and the true parameter. Other applications in this section include a formalization of sensitivity analysis and inference incorporating prior knowledge on the nuisance parameters. Section 5 revisits some of the motivating examples, showing how our conditions can be generally verified in each of them. Finally, Section 6 concludes and discusses possible extensions. Mathematical proofs are relegated to the Appendix.

2 Motivating Examples

The following examples illustrate the wide applicability of our methods.

Example 1 (Willingness-to-Pay) Let U be the maximum amount an individual would be willing to pay (WTP) for a good or resource, and let the individual's response be $Y = 1(U > V)$ where V is the bid price, usually referred as the experimental design. We only observe (Y, X, V), where X is a vector of covariates, e.g. education, income, etc, and, as is standard in the literature, we assume the conditional independence between U and V, conditional on X, in short $(U \perp V)|X$. We are interested in estimating the mean of a non-decreasing function of WTP, i.e. $\theta_0 = E[r(U)]$, for example, $r(U) = U$. Example 2 below shows how to extend our bounds in this example for a non-monotonic function $r(\cdot)$. We stress that there are different example applications with the same structure as the WTP example. Such an example is current status data, where we are interested in, e.g., the mean of unemployment duration U for an individual, but we only observe (Y, X, V), where $Y = 1(U > V)$, V is a "check-up time", and X is a vector of individual's covariates. We focus on WTP in referendum contingent valuations, but the methods proposed here apply *mutatis mutandis* to all these examples sharing the same structure. See Lewbel, McFadden and Linton (2010) and Magnac and Maurin (2008) for numerous examples sharing a similar structure.

Let $G(u|x) := \Pr(U > u|X = x)$ be the conditional survival function, and denote the conditional cumulative distribution function (cdf) as $F(u|x) := 1 - G(u|x)$. Then,

$$\theta_0 = E\left[\int r(u)\, dF(u|X)\right].$$

Lewbel et al. (2010) point out that nonparametric identification of $G(u|x)$ (or equivalently $F(u|x)$) requires the support of U to be contained in the support of V, while almost all experiments only contain a small number of discrete values for V, even in large samples. They deal with this discrepancy by considering a support for V that, although finite in finite-samples, becomes dense as $n \to \infty$ and contains that of U in the limit, effectively assuming point identification at infinity. If we relax this assumption, partial identification naturally arises in this framework. We analyze the partial identified case in this paper. Our results complement those of Magnac and Maurin (2008), who investigated partial identification in a setting where $U = \theta' X + \varepsilon$, ε is unobserved and possibly correlated with X, A' denotes the transpose of A, and the parameter of interest is θ. Here we focus on the case where the WTP is an unrestricted function of X and unobservables. Notice that by the conditional independence assumption, we have $G(v|x) = E[Y|V = v, X = x]$, which can be identified only at the (finite) support of V, say $\mathcal{V} := \{v_1, v_2, \ldots, v_m\}$, with $0 \equiv v_0 < v_1 < v_2 < \cdots < v_m$. Denoting $h_{0j}(x) := G(v_j|x)$, the identified set can be characterized as $\Theta_0 := \{\theta = E\left[\int r(u)\, dF(u|X)\right]$: for some cdf F satisfying $F(v_j|x) = 1 - h_{0j}(x)$ for all $1 \le j \le m\}$.

We provide a further characterization of Θ_0. In what follows we fix x in the support of X. Denote the support of U as $\mathcal{U}$, and assume $r(\cdot)$ is differentiable. Furthermore, the lower bound $\underline{r} := \inf_{u \in \mathcal{U}} r(u) > -\infty$ and upper bound $\overline{r} := \sup_{u \in \mathcal{U}} r(u) < \infty$ are assumed to be known.[5] Then, applying integration by parts, we have

$$\int_{\mathcal{U}} r(u)\, dF(u|x) = \underline{r} + \int_{\mathcal{U}} \dot{r}(u)\, G(u|x)\, du,$$

where $\dot{r}(u) := \partial r(u)/\partial u$. Since $\dot{r}(\cdot)$ is non-negative and $G(u|x)$ is non-increasing in u for each fixed x, then for $v_{j-1} \le u \le v_j$, it holds that $\dot{r}(u)\, h_{0j}(x) \le \dot{r}(u)\, G(u|x) \le \dot{r}(u)\, h_{0j-1}(x)$, $j = 1, \ldots, m+1$, with $h_{00}(x) := 1$ and $h_{0m+1}(x) := 0$. Hence for $j = 1, \ldots, m+1$,

$$(r(v_j) - r(v_{j-1}))\, h_{0j}(x) \le \int_{v_{j-1}}^{v_j} \dot{r}(u)\, G(u|x)\, du \le (r(v_j) - r(v_{j-1}))\, h_{0j-1}(x),$$

with $r(v_0) := \underline{r}$ and $r(v_{m+1}) := \overline{r}$. Then, defining $h_0(x) := (h_{01}(x), \ldots, h_{0m}(x))'$, $\Delta r_j := r(v_j) - r(v_{j-1})$, $\Delta_{ur} := (\Delta r_2, \ldots, \Delta r_{m+1})'$ and $\Delta_{lr} := (\Delta r_1, \ldots, \Delta r_m)'$, we get

$$L_r(h_0|x) := \underline{r} + \Delta_{lr}' h_0(x) \le \int_{\mathcal{U}} r(u)\, dF(u|x) \le r(v_1) + \Delta_{ur}' h_0(x) =: U_r(h_0|x).$$

It is worth pointing out that this bound is sharp for each fixed x if the support $\mathcal{U}$ is independent of X. Thus, the sharp bound for the identified set Θ_0 is the interval $E[L_r(h_0|X)] \le \theta_0 \le E[U_r(h_0|X)]$. Our bounds extend those in Green, Jacowitz, Kahneman and McFadden (1998) to account for covariates and general functions of the WTP.

To write this example into our framework, let $\psi(w, \theta, h_0, \tau) := \theta - (1-\tau) L_r(h_0|x) - \tau U_r(h_0|x)$ with $W := (Y, X, V)$, $\tau \in \mathcal{T}_0 := [0,1]$ and where $L_r(h_0|x)$ and $U_r(h_0|x)$ are as defined above. Notice that $h_{0j}(x) = E[Y|V = v_j,\ X = x]$ can be nonparametrically estimated for each $1 \le j \le m$. Since in this example the identified set $\Theta_0 \subset \mathbb{R}$ is a closed interval, it can be characterized by the extremes of the interval $E[L_r(h_0|X)]$ and $E[U_r(h_0|X)]$, respectively. More generally, convex and closed identified sets can be characterized by the support function defined generically as

$$\beta(q|\Theta_0) := \sup_{\theta \in \Theta_0} q'\theta, \text{ for all } q \in \mathbb{S}_p := \{q \in \mathbb{R}^p : q'q = 1\}.$$

For instance, for the WTP example, for $q \in \{1, -1\}$,

$$\beta(q|\Theta_0) = E\left[\frac{1+q}{2} U_r(h_0|X) - \frac{1-q}{2} L_r(h_0|X)\right].$$

[5] In our partially identified model there is only partial information on the support of U, so in general $\mathcal{U}$ is not estimable without further information. However, we expect inference not to be very sensitive to the knowledge of $\underline{r}$ and $\overline{r}$ in many applications. To see this, note that, as we show below, the derivative of $\theta_0(\tau)$ with respect to $\underline{r}$ is $\partial\theta_0(\tau)/\partial\underline{r} = (\tau - 1)(1 - E[E[Y|V = v_1, X = x]])$. Hence, if v_1 is such that $E[Y|V = v_1, X = x] = 1$, then $\theta_0(\tau)$ does not depend on $\underline{r}$. Similar conclusions hold with $\overline{r}$. In applications for which $r(\cdot)$ has no natural bounds, we suggest to implement the bounds with $\underline{r} = r(\underline{u})$ and $\overline{r} = r(\overline{u})$, plugging in estimators of $\underline{u} = \max\{v \in \mathcal{V} : E[Y|V = v] = 1\}$ and $\overline{u} = \min\{v \in \mathcal{V} : E[Y|V = v] = 0\}$. For more general estimates accounting for covariates see Chernozhukov, Lee and Rosen (2010).

Hence, we can also write the support function characterization into our framework by letting $\tau := q \in \mathcal{T}_0 := \{-1, 1\}$, $\theta_0 := \beta(q|\Theta_0)$ and $\psi(W, \theta_0, h_0, \tau) := \theta_0 - 0.5\{(1+\tau) U_r(h_0|X) - (1-\tau) L_r(h_0|X)\}$.

More generally, a large class of semiparametric models with convex and closed identified sets Θ_0 lead to support functions satisfying

$$\beta(q|\Theta_0) = E[\phi(W, q)],$$

see Bontemps et al. (2012) and Beresteanu and Molinari (2008) for examples. Then, letting $\tau := q \in \mathcal{T}_0 := \mathbb{S}_p$, $\theta_0(\tau) := \beta(\tau|\Theta_0)$, and $\psi(W, \theta, \tau) := \theta - \phi(W, \tau)$, one can rewrite the above equation as

$$E[\psi(W, \theta_0, \tau)] = 0.$$

As shown with the WTP example, in a semiparametric context the moment function ϕ may also depend on infinite-dimensional nuisance parameters, and in this more general situation our results can be used to obtain inferences on the semiparametric support function under mild regularity conditions.

Example 2 (Bounded linear functionals of band-identified functions) The WTP example is a special case of a very general class of models leading to functions the lie within a band; see Chandrasekar et al. (2012) for numerous examples, including regression with interval data, sample selection and quantile treatment effects. In this setting the function of interest is denoted by $\varphi(x, \alpha)$, e.g., a conditional mean, distribution or quantile function conditional on $X = x$. It is known that φ is identified within a band

$$l(x, \alpha) \leq \varphi(x, \alpha) \leq u(x, \alpha) \text{ for a.s. } x \text{ and all } \alpha \in \mathcal{A}, \tag{2}$$

for identified lower and upper functions $l(x, \alpha)$ and $u(x, \alpha)$, respectively, which can be estimated nonparametrically. The index $\mathcal{A}$ may denote, for instance, the set of quantiles of interest and/or other parameters measuring the level of missingness in the data, as in Kline and Santos (2012). Let F_X denote the cdf of X and let $L_2(F_X)$ be the Hilbert space of squared integrable measurable functions of X. Asume $\varphi(\cdot, \alpha) \in L_2(F_X)$ for each $\alpha \in \mathcal{A}$. Suppose we are interested in a linear bounded functional of $\varphi(x, \alpha)$, say $\theta_0(\alpha) = L\varphi(\cdot, \alpha)$. For instance, L might be an average increment effect functional $L\varphi(\cdot, \alpha) = E[\varphi(1, X_2, \alpha) - \varphi(0, X_2, \alpha)]$, where $X = (X_1, X_2)$ and $X_1 \in \{0, 1\}$ is a binary variable, or a best linear approximation functional $L\varphi(\cdot, \alpha) = E[XX']^{-1} E[X\varphi(X, \alpha)]$ as in Chandrasekar et al. (2012), Kline and Santos (2012) or Pacini (2012). In the general case, by the Riesz Representation theorem there exists an $r \equiv r_L \in L_2(F_X)$ such that

$$\theta_0(\alpha) = E[\varphi(X, \alpha) r(X)],$$

and the identified set for $\theta_0(\alpha)$ is

$$l(\alpha) \leq \theta_0(\alpha) \leq u(\alpha) \text{ for all } \alpha \in \mathcal{A},$$

where

$$l(\alpha) = E[l(X, \alpha) r(X) 1(r(X) > 0) + u(X, \alpha) r(X) 1(r(X) \leq 0)]$$

and

$$u(\alpha) = E\left[u(X,\alpha)r(X)1(r(X) > 0) + l(X,\alpha)r(X)1(r(X) \le 0)\right].$$

Then, we can apply the arguments of Example 1 to transform this into our setting by writing $\theta_0(\tau) = \lambda l(\alpha) + (1-\lambda)u(\alpha)$, $\tau = (\lambda, \alpha) \in \mathcal{T}_0 := [0,1] \times \mathcal{A}$. Our results then extend previous results in Chandrasekar et al. (2012), who deal with best linear functionals, to other parameters of interest such as average partial or increment effects, average structural functions or counterfactual distributions. In this generic example the nuisance parameter is $h_0(\cdot,\tau) = (l(\cdot,\alpha), u(\cdot,\alpha), r(\cdot))'$.[6]

We illustrate some of the ideas discussed above with an application to inferences on the gender gap distributional effects and counterfactual analysis in wage equations. For a general treatment of partial identification in sample selection models see Manski (1989). Let Y^* be a latent wage. We only observe wages for working individuals. The selection variable is D. That is, we only observe $Y = Y^*D$, together with D and a vector of covariates X. The structural function of interest is $\varphi(X,\alpha) = P[Y^* \le \alpha|X]$, which is known to satisfy (2), see Manski (1989), with

$$l(x,\alpha) \equiv E[1(Y \le \alpha)D|X = x] \text{ and } u(x,\alpha) \equiv E[1(Y \le \alpha)D|X = x] + 1 - g_0(x), \tag{3}$$

for all $\alpha \in \mathcal{A}$, where $\mathcal{A}$ is a compact set of $\mathbb{R}$ that represents the quantiles of interest, and $g_0(x) := E[D|X = x]$. Let $X = (X_1, X_2)$, where X_1 denotes gender, $X_1 = 1$ for women; $X_1 = 0$ for men. Two functionals of interest are the gender gap distributional effects

$$L\varphi(\cdot,\alpha) = E\left[\varphi(1, X_2, \alpha) - \varphi(0, X_2, \alpha)\right]$$

and the counterfactual distribution functional effect

$$L\varphi(\cdot,\alpha) = \int \varphi(x,\alpha)F^*(dx),$$

where $F^*(x)$ is a given counterfactual distribution for covariates that is possibly unknown but identified from the data. In the latter example it is convenient to change the domain of definition of L to $L_2(F_X^*)$. The previous discussion applies to these two examples and our uniform results can be used to construct uniform confidence bands for the gender gap and counterfactual distributions. Section 5 below considers in detail the example on gender gap distributional effects.

Example 3 (Sensitivity to exclusion restrictions) In many applications in microeconometrics exclusion restrictions, in the form of a zero coefficient in an outcome equation, are imposed. Economic theory is often silent about such restrictions, and a more robust inference approach is obtained if we fixed such coefficients to τ, rather than zero, and proceed with a partial identification approach or a sensitivity analysis. Conley et al. (2012) considered this approach in a linear IV setting, where τ is the coefficient of the instruments in the structural equation. Here we extend their results to semiparametric, possibly non-separable, models.[7] Also, the type of inferences that we proposed are different from theirs.

[6] In some applications, such as in the WTP example or in counterfactual exercises, r may be known and h_0 only contains l and u.

[7] Non-separability and nonlinearity complicates to a large extend the verification of the partial identification assumption.

We illustrate the general idea with an example on a selection model with binary outcome. Suppose a latent binary variable Y^* satisfies the ordinary threshold crossing binary response model $Y^* = 1(X'\beta_0 - e \geq 0)$. The econometrician is assumed to know relatively little about selection D other than it is binary, so let D be given by the nonparametric threshold crossing model $D = 1[g_0(X) - u \geq 0]$ where u is uniformly distributed. Assume (e, u) is drawn from an unknown joint distribution function $F(e, u)$ with $(e, u) \perp X$. As before, we only observe (Y, D, X). Point-identification in this model has been investigated in Escanciano, Jacho-Chavez and Lewbel (2012), while its estimation is discussed in Escanciano, Jacho-Chavez and Lewbel (2011,2012) and, for index selection, in Klein, Shen and Vella (2011). Here we consider a partial identification approach.

Suppose for simplicity that $X = (X_1, X_2, X_3)'$. A researcher estimating this model would impose that one of the coefficients in β_0 is zero, say that of X_1. Suppose, however, that she is concerned about the sensitivity of inferences to the maintained exclusion restriction. Then, we can write $X'\beta_0 = \tau X_1 + X_2 + \theta_0 X_3$ (here the coefficient of X_2 is normalized to one since the distribution of e is unknown). It is straightforward to show that our partial identification assumption holds under mild smoothness conditions. This can be seen by taking derivatives of $E[Y|X = x]$ with respect to x, and solving for θ_0. This suggests that the score equations from the semiparametric likelihood estimator in Klein, Shen and Vella (2011) or from the semiparametric least squares estimator of Escanciano et al. (2012) can be used as the set of moment restrictions in our setting. In this example, the mapping $\theta_0(\tau)$ is nonlinear, the nuisance parameter F depends on both $\theta_0(\tau)$ and τ, and the set $\mathcal{T}_0$ is unknown. A robust approach to identification, as the one suggested here, does not impose that this set is a singleton. Similarly, sensitivity analysis can be carried out by estimating $\theta_0(\tau)$ over a set $\mathcal{T}_1$ that includes the case $\tau = 0$.

3 Uniform Inference

We first elaborate further on the model introduced in (1). Notice that though we do not make it explicit in (1), the nuisance function $h_0(\cdot)$ may contain (θ, τ) as additional arguments. To develop asymptotic results under misspecification, we maintain the partial identification assumption throughout, that is, for each $\tau \in \mathcal{T}_1$, we assume $\|E[\psi(W, \theta, h_0, \tau)]\|$ is uniquely minimized at $\theta_0(\tau)$. Our estimation and inference results are developed to account for potential misspecification of the moment restriction (1), and we also discuss simplification of the results when the model is correctly specified. The asymptotic results developed below for semiparametric GMM estimation under global misspecification with non-smooth moment functions are also of independent interest and extend some of the previous results given by Ai and Chen (2007) for smooth moments and $\mathcal{T}_1 = \{\tau_1\}$.

We assume that for each fixed value of $(\theta, \tau) \in \Theta \times \mathcal{T}_1$, there is a first-step nonparametric estimator $\widehat{h}(\cdot)$ for $h_0(\cdot)$ available with certain convergence properties as specified in Assumption A1 and A2 below. Throughout we use the following notation. Let $|\cdot|$ denote the Euclidean norm, i.e. $|A| := (tr(A'A))^{1/2}$, where $tr(A)$ is the trace of the matrix A. Let $vec(A)$ denote the vectorization of matrix A and $\otimes$ denote the Kronecker product. For a measurable function g of W, define the norms $\|g\|_\infty = \sup_{w \in \mathcal{W}} |g(w)|$ and $\|g\|_r := (E[|g(W)|^r])^{1/r}$, where $\mathcal{W}$ is the support of W. The function space $\mathcal{H}$ is endowed with a pseudo-metric $\|\cdot\|_{\mathcal{H}}$ which is a sup-norm metric with respect to the (θ, τ)-arguments and a pseudo-metric with respect to w. For example, $\|h\|_{\mathcal{H}} := \sup_{\theta,\tau} \|h(\cdot,\theta,\tau)\|_{\infty}$ or $\|h\|_{\mathcal{H}} := \sup_{\theta,\tau} \|h(\cdot,\theta,\tau)\|_{r}$. In what follows, we suppress (θ,τ) in the nuisance function h to save space, but it should be understood conformably, i.e. $(\theta,h,\tau) := (\theta,h(\cdot,\theta,\tau),\tau)$. Our framework can be viewed as an extension of Chen, Linton and Van Keilegom (2003) (CLV hereafter) in two directions: first to partially identified semiparametric models, and second to misspecified models. The latter seems to be relevant even in the point identified case, as there is no general treatment of misspecification for two-step GMM estimators with non-smooth objective functions, to the best of our knowledge.

Suppose the observed data $\{W_i\}_{i=1}^{n}$ are an independent and identically distributed (i.i.d.) sequence following the same distribution as W. Here n is the sample size. Henceforth, we use the following generic notation, for a measurable function f we denote the empirical expectation and empirical process by

$$E_n f(W) := n^{-1}\textstyle\sum_{i=1}^{n} f(W_i) \text{ and } \mathbb{G}_n f(W) := n^{-1/2}\textstyle\sum_{i=1}^{n}\{f(W_i) - E[f(W_i)]\}.$$

Then, let $M_n(\theta,h,\tau) := E_n[\psi(W,\theta,h(W),\tau)]$, $M(\theta,h,\tau) := E[\psi(W,\theta,h(W),\tau)]$, and define the weighted Euclidean norm of a matrix A as $\|A\| = (tr(A'\Xi A))^{1/2}$ for some fixed symmetric positive definite matrix Ξ. Our theory can be easily extended to the case where Ξ depends on τ and is estimated. For each fixed $\tau \in \mathcal{T}_1$, we consider a GMM-estimator for $\theta_0(\tau)$:

$$\widehat{\theta}(\tau) := \arg\min_{\theta\in\Theta} \left\| M_n\left(\theta,\widehat{h},\tau\right)\right\|, \tag{4}$$

where $\widehat{h}(\cdot)$ is the first-step estimator of $h_0(\cdot)$ for each fixed pair (θ,τ). For practical applications, it would suffice to consider an estimator that is close to the minimizer, and we will make this point clear in the following assumptions. Since $\widehat{\theta}(\cdot)$ is viewed as a process in τ, we call $Z_n(\tau) := \sqrt{n}(\widehat{\theta}(\tau) - \theta_0(\tau))$ a *Z-process*. Under mild regularity conditions, $Z_n(\cdot)$ belongs to the Banach space $\ell^{\infty}(\mathcal{T}_1)$ of uniformly bounded functions on $\mathcal{T}_1$, which is equipped with the sup-norm, $\|z\|_{\mathcal{T}_1} := \sup_{\tau\in\mathcal{T}_1}|z(\tau)|$. In this paper we consider the weak convergence, denoted by $\rightsquigarrow$, of $Z_n(\cdot)$ in the metric space $\ell^{\infty}(\mathcal{T}_1)$ endowed with the sup-norm in the sense of J. Hoffmann-Jørgensen (see, e.g., Dudley 1999 p. 94). We abstract from measurability issues that may arise as a result of using the space $\ell^{\infty}(\mathcal{T}_1)$ endowed with the sup-norm. See van der Vaart and Wellner (1996) for a formal treatment of lack of measurability.

3.1 Consistency

In this section we discuss the consistency of the estimator defined in (4) under the following assumptions.

Assumption A1: Suppose that $\theta_0(\tau) \in \Theta$ solves the minimization problem $\inf_{\theta\in\Theta}\|M(\theta,h_0,\tau)\|$ for each $\tau \in \mathcal{T}_1$, where Θ is a compact set in $\mathbb{R}^{d_\theta}$. In addition, assume
(i) The estimator $\widehat{\theta}(\tau)$ satisfies

$$\sup_{\tau\in\mathcal{T}_1}\left\{\left\|M_n\left(\widehat{\theta}(\tau),\widehat{h},\tau\right)\right\| - \inf_{\theta\in\Theta}\left\|M_n\left(\theta,\widehat{h},\tau\right)\right\|\right\} \le o_P(1). \tag{5}$$

(ii) Uniform Partial Identification: for all $\varepsilon > 0$, there exists $\eta(\varepsilon) > 0$ such that

$$\inf_{\tau\in\mathcal{T}_1}\left\{\inf_{\theta:|\theta-\theta_0(\tau)|\ge\varepsilon}\|M(\theta,h_0,\tau)\| - \|M(\theta_0(\tau),h_0,\tau)\|\right\} \ge \eta(\varepsilon). \tag{6}$$

(iii) Uniform Continuity: uniformly for all $\theta \in \Theta$ and $\tau \in \mathcal{T}_1$, $M(\theta, h, \tau)$ is continuous in h at $h = h_0$ with respect to the metric $\|\cdot\|_{\mathcal{H}}$.
(iv) $\left\|\widehat{h} - h_0\right\|_{\mathcal{H}} = o_P(1)$.
(v) Uniform convergence: for all sequences of positive numbers $\{\delta_n\} \to 0$,

$$\sup_{\theta\in\Theta,\tau\in\mathcal{T}_1,\|h-h_0\|_{\mathcal{H}}\le\delta_n} \|M_n(\theta, h, \tau) - M(\theta, h, \tau)\| = o_P(1). \tag{7}$$

These assumptions are uniform versions of those in CLV for consistency. Like these authors, we also allow for non-smooth moment functions $\psi(\cdot)$ as long as $M(\cdot)$ is continuous. Consistency and rates of convergence for nonparametric estimators indexed by nuisance parameters are investigated in Andrews (1995), Sperlich (2009) or Escanciano, Jacho-Chávez and Lewbel (2011) for kernel estimates, or by Song (2008) for series estimators. These results can be used to verify A1(iv). Assumption A1(v) is implied by a Glivenko-Cantelli property of the class $\Psi := \{\psi(\cdot, \theta, h, \tau) : \theta \in \Theta,\ h \in \mathcal{H},\ \tau \in \mathcal{T}_1\}$.

Our first result shows the uniform consistency of $\widehat{\theta}(\cdot)$.

Theorem 3.1 *Under Assumption A1, it holds that* $\sup_{\tau\in\mathcal{T}_1}\left|\widehat{\theta}(\tau) - \theta_0(\tau)\right| = o_P(1)$.

3.2 Weak Convergence

Consider now the weak convergence of $Z_n(\cdot) = \sqrt{n}(\widehat{\theta}(\cdot) - \theta_0(\cdot))$ as a stochastic process in $\ell^\infty(\mathcal{T}_1)$ endowed with the sup-norm $\|\cdot\|_{\mathcal{T}_1}$. Given consistency, we can work, as usual, in a small or even shrinking neighborhood of $\theta_0(\cdot)$ and h_0. Define $\Theta_{01} := \{\theta_0(\tau) : \tau \in \mathcal{T}_1\}$. With some abuse of the notation, define a δ-enlargement of the parameter sets $\Theta_\delta := \{\theta \in \Theta : d_H(\theta, \Theta_{01}) \le \delta\}$ and $\mathcal{H}_\delta := \{h \in \mathcal{H} : \|h - h_0\|_{\mathcal{H}} \le \delta\}$, where the pseudo-metric $\|\cdot\|_{\mathcal{H}}$ is also modified according to the smaller parameter set Θ_δ. We first introduce the definition of pathwise functional derivative to deal with the estimation effects of $\widehat{h}$. For each $(\theta, \tau) \in \Theta_\delta \times \mathcal{T}_1$, we say that $M(\theta, h, \tau)$ is pathwise differentiable at $h \in \mathcal{H}_\delta$ in the direction $\left[\overline{h} - h\right]$ if $\left\{h + \lambda\left(\overline{h} - h\right) : \lambda \in [0, 1]\right\} \subset \mathcal{H}$ and

$$\lim_{\lambda\to0} \frac{M\left(\theta, h + \lambda\left(\overline{h} - h\right), \tau\right) - M(\theta, h, \tau)}{\lambda} \text{ exists.}$$

To simplify the notation we drop the dependence on true values. For instance, $M(\tau) := M(\theta_0(\tau), h_0, \tau)$. Define $V_\theta(\theta, h, \tau) := \partial M(\theta, h, \tau)/\partial\theta'$, $V_{\theta_0}(\tau) := V_\theta(\theta_0(\tau), h_0, \tau)$, $V_h(\theta, h, \tau)\left[\overline{h} - h\right]$ is the pathwise derivative of $M(\theta, h, \tau)$ along the direction $\overline{h} - h$, $V_{\theta\theta}(\theta, h, \tau) := \partial vec(V_\theta(\theta, h, \tau))/\partial\theta'$, $V_{\theta_0\theta_0}(\tau) := V_{\theta\theta}(\theta_0(\tau), h_0, \tau)$, $V_{\theta h}(\theta, h, \tau)\left[\overline{h} - h\right]$ is the pathwise derivative of $V_\theta(\theta, h, \tau)$ along the direction $\overline{h} - h$. We will suppress τ from $\theta_0(\tau)$ and $\widehat{\theta}(\tau)$ whenever there is no ambiguity. For the weak convergence we need the following assumptions.

Assumption A2: Suppose that $\theta_0(\tau)$ is in the interior of Θ for each $\tau \in \mathcal{T}_1$, and that $\sup_{\tau\in\mathcal{T}_1}|\widehat{\theta}(\tau) - \theta_0(\tau)| = o_P(1)$. In addition, assume:
(i) There exists a ($d_\psi \times d_\theta$ matrix) function $\psi_\theta(\cdot, \theta, h, \tau)$ such that for any positive sequence $\delta_n \to 0$,

$$\sup_{\tau\in\mathcal{T}_1} \sup_{|\theta-\theta_0|\le\delta_n,\|h-h_0\|_{\mathcal{H}}\le\delta_n} |E_n\psi_\theta(Z, \theta, h, \tau) - V_\theta(\theta, h, \tau)| = o_P(1).$$

(ii) The estimator $\widehat{\theta}(\tau)$ satisfies

$$\left(E_n\psi_\theta\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)\right)'\Xi E_n\psi\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)=o_P\left(n^{-1/2}\right). \tag{8}$$

(iii) Smoothness in θ : (a) for each $\tau\in\mathcal{T}_1$, the map $\theta\to M(\theta,h_0,\tau)$ is twice continuously differentiable at θ_0, with first-order derivative $V_{\theta_0}(\tau)$. Furthermore, (b) suppose that $A_0(\tau):=V_{\theta_0}(\tau)'\Xi V_{\theta_0}(\tau)+(M(\tau)'\Xi\otimes I_{d_\theta})V_{\theta_0\theta_0}(\tau)$ is of full rank for all $\tau\in\mathcal{T}_1$, $\sup_{\tau\in\mathcal{T}_1}\|A_0(\tau)\|<\infty$ and $\sup_{\tau\in\mathcal{T}_1}\left\|A_0^{-1}(\tau)\right\|<\infty$.
(iv) Smoothness in h : (a) for each $(\theta,\tau)\in\Theta_\delta\times\mathcal{T}_1$, the pathwise derivative $V_h(\theta,h_0,\tau)[h-h_0]$ of $M(\theta,h,\tau)$ at $h=h_0$ exists in all directions $[h-h_0]\in\mathcal{H}$; and for all $(\theta,h,\tau)\in\Theta_{\delta_n}\times\mathcal{H}_{\delta_n}\times\mathcal{T}_1$ with a positive sequence $\delta_n\to0$, it holds that

$$\sup_{\tau\in\mathcal{T}_1}\|M(\theta,h,\tau)-M(\theta,h_0,\tau)-V_h(\theta,h_0,\tau)[h-h_0]\|\le c\|h-h_0\|_{\mathcal{H}}^2 \tag{9}$$

for a constant $c\ge0$, and

$$\sup_{\tau\in\mathcal{T}_1}\|V_h(\theta,h_0,\tau)[h-h_0]-V_h(\theta_0,h_0,\tau)[h-h_0]\|\le o(1)\,\delta_n; \tag{10}$$

(b) similarly, the pathwise derivative $V_{\theta h}(\theta,h_0,\tau)[h-h_0]$ of $V_\theta(\tau)$ at $h=h_0$ also exists, and

$$\sup_{\tau\in\mathcal{T}_1}\|V_\theta(\theta,h,\tau)-V_\theta(\theta,h_0,\tau)-V_{\theta h}(\theta,h_0,\tau)[h-h_0]\|\le c\|h-h_0\|_{\mathcal{H}}^2 \tag{11}$$

and

$$\sup_{\tau\in\mathcal{T}_1}\|V_{\theta h}(\theta,h_0,\tau)[h-h_0]-V_{\theta h}(\theta_0,h_0,\tau)[h-h_0]\|\le o(1)\,\delta_n; \tag{12}$$

(v) $\Pr\left(\widehat{h}\in\mathcal{H}\right)\to1$, and $\left\|\widehat{h}-h_0\right\|_{\mathcal{H}}=o_P\left(n^{-1/4}\right)$.
(vi) Stochastic Equicontinuity: for all sequences of positive numbers $\delta_n\to0$, (a)

$$\sup_{\tau\in\mathcal{T}_1}\sup_{|\theta-\theta_0|\le\delta_n,\|h-h_0\|_{\mathcal{H}}\le\delta_n}\|\mathbb{G}_n\psi(W,\theta,h,\tau)-\mathbb{G}_n\psi(W,\theta_0,h_0,\tau)\|=o_P(1).$$

and, (b)

$$\sup_{\tau\in\mathcal{T}_1}\sup_{|\theta-\theta_0|\le\delta_n,\|h-h_0\|_{\mathcal{H}}\le\delta_n}\|\mathbb{G}_n\psi_\theta(W,\theta,h,\tau)-\mathbb{G}_n\psi_\theta(W,\theta_0,h_0,\tau)\|=o_P(1).$$

(vii) $\sqrt{n}V_h(\theta_0,h_0,\tau)[\widehat{h}-h_0]$ admits an asymptotic expansion (uniformly in τ):

$$\sqrt{n}V_h(\theta_0,h_0,\tau)[\widehat{h}-h_0]=\frac{1}{\sqrt{n}}\sum_{i=1}^n\phi(W_i,\theta_0,h_0,\tau)+o_P(1),$$

and $\sqrt{n}V_{\theta h}(\theta_0,h_0,\tau)[\widehat{h}-h_0]$ also admits an asymptotic expansion (uniformly in τ):

$$\sqrt{n}V_{\theta h}(\theta_0,h_0,\tau)[\widehat{h}-h_0]=\frac{1}{\sqrt{n}}\sum_{i=1}^n\phi_\theta(W_i,\theta_0,h_0,\tau)+o_P(1).$$

Moreover, we set

$$s(w,\theta_0,h_0,\tau):=\begin{pmatrix}\psi(w,\theta_0,h_0,\tau)+\phi(w,\theta_0,h_0,\tau)\\ vec(\psi_\theta(w,\theta_0,h_0,\tau)+\phi_\theta(w,\theta_0,h_0,\tau))\end{pmatrix}, \tag{13}$$

and assume the function class $\{w\to s(w,\theta_0,h_0,\tau):\tau\in\mathcal{T}_1\}$ is Donsker.

Assumption A2(i) and (ii) are critical in deriving asymptotic distribution of $\sqrt{n}(\widehat{\theta}-\theta_0)$ under misspecification. When the moment function $\psi(w,\theta,h_0,\tau)$ is smooth in θ, these conditions are satisfied by letting $\psi_\theta(\cdot,\theta,h_0,\tau)=\partial\psi(\cdot,\theta,h_0,\tau)/\partial\theta'$. With non-smooth moment functions, these are high-level conditions but still are satisfied by some commonly used models, for example, the quantile regression model; see Angrist, Chernozhukov and Fernandez-Val (2006). More generally, the expression for ψ_θ can be obtained from the generalized information equality, see, e.g., Newey and McFadden (1994). Assumption A2(iii) requires second-order differentiability w.r.t. θ due to misspecification. Similarly, A2(iv) imposes conditions on the cross term $V_{\theta h}(\cdot)$ in addition to the first-order pathwise derivative $V_h(\cdot)$. Conditions on second derivatives are not needed in the correct specified case, see CLV. Assumption A2(v) or similar versions are commonly assumed in the semiparametric literature. Ichimura and Lee (2010) imposed slightly weaker conditions on the converging rates of the initial estimator $\widehat{h}$. Escanciano, Jacho-Chávez and Lewbel (2011) provide simple primitive conditions for verifying $\Pr\left(\widehat{h}\in\mathcal{H}\right)\to 1$. Assumption A2(vi) is usually implied by the Donsker property of the function classes $\Psi:=\{\psi(\cdot,\theta,h,\tau):\theta\in\Theta_\delta,\ h\in\mathcal{H}_\delta,\ \tau\in\mathcal{T}_1\}$ and $\Psi_\theta:=\{\psi_\theta(\cdot,\theta,h,\tau):\theta\in\Theta_\delta,\ h\in\mathcal{H}_\delta,\ \tau\in\mathcal{T}_1\}$, for which primitive conditions can be easily provided using standard empirical processes tools. Assumption A2(vii) is a high-level condition and implies Assumption (2.6) in CLV. Notice that verification of Assumption (2.6) in CLV usually leads to the above asymptotic linear expansion by plugging in, for example, the Bahadur representation for the nonparametric estimator $\widehat{h}-h_0$ in the Riesz representation of the linear mapping V_h. See Section 3 in CLV for discussion. The type of analysis needed for establishing the uniformity in $\tau\in\mathcal{T}_1$ in the expansion of Assumption (2.6) is similar to certain uniform weighted bias calculations, which have been carried out in the literature for well-known estimators; see Andrews (1995) and Escanciano, Jacho-Chávez and Lewbel (2011) for kernel estimators and Song (2008) for series estimators.

Theorem 3.2 *Under Assumption A2, it holds that*

$$\sqrt{n}(\widehat{\theta}-\theta_0)\rightsquigarrow L,$$

where L is a Gaussian process in $\ell^\infty(\mathcal{T}_1)$ with zero mean and covariance function

$$C(\tau_1,\tau_2)=A_0(\tau_1)^{-1}\left[B_0(\tau_1)\ \ D_0(\tau_1)\right]K(\tau_1,\tau_2)\left[B_0'(\tau_2)\ \ D_0'(\tau_2)\right]'A_0'(\tau_2)^{-1},$$

with $A_0(\tau)$ defined in A2(iii), $B_0(\tau)=V_{\theta_0}(\tau)'\Xi$, $D_0(\tau)=(M(\theta_0,h_0,\tau)'\Xi\otimes I_{d_\theta})$,

$$K(\tau_1,\tau_2):=Cov\left(s(W,\theta_0,h_0,\tau_1),s(W,\theta_0,h_0,\tau_2)\right),$$

and s is defined in (13).

Remark 3.1 *When the model is correctly specified the assumptions and the asymptotic covariance function can be substantially simplified. In this case, for all $\tau\in\mathcal{T}_1$,*

$$A_0(\tau)=V_{\theta_0}(\tau)'\Xi V_{\theta_0}(\tau)\ \text{ and }\ D_0(\tau)=0.$$

Hence, the covariance function simplifies to

$$\widetilde{C}(\tau_1,\tau_2) = \left[V_{\theta_0}(\tau_1)' \Xi V_{\theta_0}(\tau_1)\right]^{-1} V_{\theta_0}(\tau_1)' \Xi \widetilde{K}(\tau_1,\tau_2) \Xi V_{\theta_0}(\tau) \left[V_{\theta_0}(\tau_2)' \Xi V_{\theta_0}(\tau_2)\right]^{-1},$$

with $\widetilde{K}(\tau_1,\tau_2) = E\left[\widetilde{s}(W,\theta_0,h_0,\tau_1)\,\widetilde{s}'(W,\theta_0,h_0,\tau_2)\right]$, $\widetilde{s}(W,\theta_0,h_0,\tau_1) := \psi(W,\theta_0,h_0,\tau)+\phi(W,\theta_0,h_0,\tau)$. *Similarly, Assumptions A2(i-ii), A2(iv-b) and A2(vi-b) are not needed anymore. For the sake of completeness, we list the conditions needed for the correct specification case in the Appendix.*

3.3 Bootstrap Approximations

The set inference developed in this paper will involve the limit distribution of continuous functionals of $Z_n(\cdot) = \sqrt{n}(\widehat{\theta}(\cdot) - \theta_0(\cdot))$. Quantiles of these limiting distributions are generally unknown, and bootstrap methods provide natural approximations. CLV propose to use the ordinary nonparametric bootstrap. The nonparametric bootstrap can be used in our setting as well, and conditions for its consistency can be easily given combining our uniform results with the arguments of CLV. The nonparametric bootstrap, however, can be computationally expensive in our context, since we need to re-estimate the pair (θ, h) for each bootstrap sample and for each fixed $\tau \in \mathcal{T}_1$. Hence, for completeness we propose here an alternative bootstrap method based on the multiplier principle, which has the advantage that one does not need to estimate $(\theta_0(\tau), h(\cdot,\tau))$ for each bootstrap sample. In contrast to the nonparametric bootstrap, the multiplier bootstrap requires the estimation of influence functions of the semiparametric estimator $\widehat{\theta}(\cdot)$, which can be a cumbersome task in some applications, especially under misspecification. Hence, the most practically convenient bootstrap method to choose depends on the specific application at hand. Since the theory for the nonparametric bootstrap is well known, we focus in this section on the multiplier bootstrap.

According to the proof of Theorem 3.2, the asymptotic linear expansion for the estimator $\widehat{\theta}$ is

$$Z_n(\tau) = -A(\theta_0,h_0,\tau)\mathbb{G}_n s(W,\theta_0,h_0,\tau) + o_P(1),$$

where $A(\theta_0,h_0,\tau) = A_0(\tau)^{-1}\left[B_0(\tau)\ \ D_0(\tau)\right]$ with A_0, B_0, D_0 and $s(\cdot)$ defined as in Theorem 3.2. To simplify the exposition and with some abuse of notation, in what follows we include in h_0 any nonparametric object that may appear in the influence function s and the matrix A as a result of differentiation. This is for instance the case in models under quantile restrictions, in which the influence function depends on a conditional density, in addition to the original parameters of the model. Our results below allow for this possibility under sufficient regularity conditions for the nonparametric estimators of these additional nonparametric objects. Notice that if h_0 includes θ as an argument, e.g. h_0 is profiled, then estimation of $V_{\theta_0}(\tau)$ and $V_{\theta_0\theta_0}(\tau)$ might be difficult. However, in many applications it is feasible to obtain uniformly consistent estimators. Even for the profiled case, we can still estimate $V_{\theta_0}(\tau)$ and $V_{\theta_0\theta_0}(\tau)$ in some widely used semiparametric models, for instance, mean regression and quantile regression models, see Example 2 in CLV. Suppose there exists a uniformly (in (θ,h,τ)) consistent estimator $\widehat{V}_\theta(\theta,h,\tau)$ of $V_\theta(\theta,h,\tau)$, as in Assumption A2(i), then $V_{\theta_0}(\tau)$ can be estimated by $\widehat{V}_\theta(\widehat{\theta},\widehat{h},\tau) \equiv E_n\psi_\theta\left(Z,\widehat{\theta},\widehat{h},\tau\right)$. Similarly, we assume we can estimate $V_{\theta_0\theta_0}(\tau)$ by $\widehat{V}_{\theta\theta}(\widehat{\theta},\widehat{h},\tau)$ and $M(\tau)$ by $E_n\left[\psi\left(W,\widehat{\theta},\widehat{h},\tau\right)\right]$. Denote the corresponding estimator of $A(\theta_0,h_0,\tau)$ as $\widehat{A}(\widehat{\theta},\widehat{h},\tau)$. We propose the following multiplier-type bootstrap to approximate the asymptotic distribution of a continuous functional of $Z_n(\cdot)$, say $\Psi(Z_n)$:

Algorithm 1: (Multiplier-Bootstrap Approximation)

Step 1: Generate an i.i.d. sequence of random variables $\{u_i\}_{i=1}^n$ with mean zero, unit variance and bounded $2+\eta$ moments, with $\eta > 0$, which are also independent of $\{W_i\}_{i=1}^n$. Possible choices of distributions for u_i include the standard normal and Bernoulli. For example, one can use the Bernoulli distribution with $\Pr(u_i = (1-\sqrt{5})/2) = (\sqrt{5}+1)/(2\sqrt{5})$, and $\Pr(u_i = (1+\sqrt{5})/2) = (\sqrt{5}-1)/(2\sqrt{5})$, as advocated in e.g. Mammen (1993).

Step 2: For each fixed τ, compute $\widehat{Z}_n^*(\tau) = A(\widehat{\theta},\widehat{h},\tau)n^{-1/2}\sum_{i=1}^n \left\{ s\left(W_i,\widehat{\theta},\widehat{h},\tau\right) - E_n s\left(W,\widehat{\theta},\widehat{h},\tau\right)\right\} u_i$.

Step 3: Compute the functional of interest $\Psi(\widehat{Z}_n^*)$.

Step 4: Repeat Step 1-3 B times and approximate the distribution of $\Psi(Z_n)$ with the empirical cdf of the B realizations of $\Psi(\widehat{Z}_n^*)$.

Our next theorem shows that $\widehat{Z}_n^*(\cdot)$ weakly converges to the same distribution as $Z_n(\cdot)$, for which we need the following assumption.

Assumption A3: In addition to Assumption A2(ii), assume:
(i) $\sup_{\tau\in\mathcal{T}_1}\left|\widehat{\theta}(\tau) - \theta_0(\tau)\right| = o_P(1)$; $\Pr\left(\widehat{h}\in\mathcal{H}\right)\to 1$, and $\left\|\widehat{h}-h_0\right\|_{\mathcal{H}} = o_P(1)$.
(ii) There exist estimators $\widehat{V}_\theta(\theta,h,\tau)$ and $\widehat{V}_{\theta\theta}(\theta,h,\tau)$ such that

$$\sup_{\theta\in\Theta_\delta,h\in\mathcal{H}_\delta,\tau\in\mathcal{T}_1}\left|\widehat{V}_\theta(\theta,h,\tau) - V_\theta(\theta,h,\tau)\right| = o_P(1),$$
$$\sup_{\theta\in\Theta_\delta,h\in\mathcal{H}_\delta,\tau\in\mathcal{T}_1}\left|\widehat{V}_{\theta\theta}(\theta,h,\tau) - V_{\theta\theta}(\theta,h,\tau)\right| = o_P(1);$$

both $V_\theta(\theta,h,\tau)$ and $V_{\theta\theta}(\theta,h,\tau)$ are continuous in (θ,h) at $(\theta,h)=(\theta_0,h_0)$ uniformly in $\tau\in\mathcal{T}_1$.
(iii) The class of functions $\mathcal{S} := \{s(\cdot,\theta,h,\tau) : \theta\in\Theta_\delta, h\in\mathcal{H}_\delta, \tau\in\mathcal{T}_1\}$ is Donsker.

Notice that because of the possible addition of new infinite-dimensional nuisance parameters into the influence function, the parameter set $\mathcal{H}$ might be different from the one used in Theorem 3.2. Assumption A3(ii) is a high-level condition, which is used to show the consistency of $\widehat{A}(\widehat{\theta},\widehat{h},\tau)$ to $A(\theta_0,h_0,\tau)$ uniformly in $\tau\in\mathcal{T}$. This assumption requires verification of the uniform convergence of $\widehat{V}_\theta$ and $\widehat{V}_{\theta\theta}$, which is relatively easy in many cases. For example, if $\psi(\cdot,\theta,h,\tau)$ is twice differentiable in θ, then under some mild regularity conditions, we can establish a Glivenko-Cantelli property for $\{\partial/\partial\theta\psi(\cdot,\theta,h,\tau) : \theta\in\Theta_\delta, h\in\mathcal{H}, \tau\in\mathcal{T}_1\}$ and $\widehat{V}_\theta$ is the sample analog of $V_\theta(h,\tau) = E[\partial\psi(\cdot,\theta,h,\tau)/\partial\theta]$; similar primitive conditions can be imposed for $\widehat{V}_{\theta\theta}$. As for Assumption A3(iii), we directly assume a Donsker property for the function class $\mathcal{S}$ which can be verified by applying standard arguments, see van der Vaart and Wellner (1996).

We use the notion of bootstrap consistency in probability introduced in Giné and Zinn (1990). Let BL_1 denote the set of all functionals on $\ell^\infty(\mathcal{T}_1)$ with a Lipschitz norm bounded by 1, i.e. for any $f\in BL_1$, $\sup_{z\in\ell^\infty(\mathcal{T}_1)}|f(z)|\le 1$ and for $z_1,z_2\in\ell^\infty(\mathcal{T}_1)$, $|f(z_1)-f(z_2)|\le\|z_1-z_2\|_{\mathcal{T}_1}$. Let

$E^*[\cdot]$ and $Var^*(\cdot)$ denote the expectation and variance of bootstrap statistics conditional on $\{W_i\}_{i=1}^n$, respectively.

Theorem 3.3 *Under Assumption A3, it holds that $\widehat{Z}_n^*(\cdot)$ weakly converges to $L(\cdot)$ in $\ell^\infty(\mathcal{T}_1)$ conditional on $\{W_i\}_{i=1}^n$ in probability, i.e.*

$$\sup_{f\in BL_1}\left|E^*\left[f\left(\widehat{Z}_n^*\right)\right]-E\left[f(L)\right]\right|=o_P(1) \tag{14}$$

and

$$\sup_{\tau\in\mathcal{T}_1}\left|Var^*(\widehat{Z}_n^*(\tau))-C(\tau,\tau)\right|=o_P(1), \tag{15}$$

where $L(\cdot)$ and $C(\cdot,\cdot)$ are defined as in Theorem 3.2.

Remark 3.2 *The conclusion in (14) of the previous theorem, together with the continuous mapping theorem, can be applied to obtain consistency of bootstrap quantiles of continuous functionals of Z_n, whereas (15) shows the consistency of bootstrap standard errors. Similar results to (15) for the ordinary nonparametric bootstrap are not available in the literature in this generality. If the model is correctly specified, then we do not need to estimate $V_{\theta_0\theta_0}(\tau)$ and the assumptions can be simplified. Details are omitted.*

4 Set Inference, Sensitivity Analysis and Prior Information

4.1 Set Inference

4.1.1 Inference on the Identified Set

We apply our previous uniform results to obtain inference on the identified set Θ_0 for a correctly specified model. We allow for the possibility that $\mathcal{T}_0$ is unknown and estimated consistently by $\hat{\mathcal{T}}_0$. A candidate for $\hat{\mathcal{T}}_0$ can be obtained extending the ideas in Chernozhukov et al. (2007) to our semi-parametric context with infinite-dimensional nuisance parameters. In certain two-step estimators the results of Chernozhukov et al. (2007) are directly applicable. For instance, this the case for parameters θ_0 that are functions of τ and possibly h, but that otherwise do not restrict the identified set of τ. That is, the moments have the form $\psi=(\psi_1(W,\tau),\theta-\psi_2(\tau,h))$. A natural estimate for $\mathcal{T}_0$ in the general case is

$$\hat{\mathcal{T}}_0=\left\{\tau\in\mathcal{T}:\left\|M_n\left(\widehat{\theta}(\tau),\widehat{h}(\cdot,\widehat{\theta}(\tau),\tau),\tau\right)\right\|\le\hat{c}/n\right\}, \tag{16}$$

for a suitable level value $\hat{c}$. In practice we can choose $\hat{c}=\log n$. The following result provides conditions for convergence of $d_H\left(\hat{\mathcal{T}}_0,\mathcal{T}_0\right)$ to zero, where henceforth, d_H denotes the Hausdorff distance

$$d_H(A,B):=\max\left\{\sup_{a\in A}d_H(a,B),\sup_{b\in B}d_H(b,A)\right\},$$

with $d_H(a,B):=\inf_{b\in B}|a-b|$. This extension of Chernozhukov et al. (2007) is of independent interest.

Lemma 4.1 *Let Assumption A2 holds. Then,* $d_H\left(\hat{\mathcal{T}}_0, \mathcal{T}_0\right) = o_P(1)$ *and* $\Pr\left(\mathcal{T}_0 \subset \hat{\mathcal{T}}_0\right) \to 1$ *for* $\hat{\mathcal{T}}_0$ *in (16) for any* $\hat{c} \to \infty$, *such that* $\hat{c}/n \to 0$ *as* $n \to \infty$.

Even in applications where $\mathcal{T}_0$ is known, it can be practically convenient to consider a discrete approximation to $\mathcal{T}_0$, so that our inferences below can be easily carried out. Our results permit discrete approximations as long as the discrete set $\hat{\mathcal{T}}_0$ converges to $\mathcal{T}_0$ in the Hausdorff metric. To include all these possibilities, we derive our inferences on Θ_0 for a generic estimator $\hat{\mathcal{T}}_0$.

Our first result is a consistency result, which is a direct corollary of Theorem 3.1 and shows the consistency in the Hausdorff metric of the estimated identified set $\widehat{\Theta}_0 = \{\widehat{\theta}(\tau) : \tau \in \hat{\mathcal{T}}_0\}$. Let $\mathcal{T}_0^{\delta} := \{\tau \in \mathcal{T} : \inf_{\tau' \in \mathcal{T}_0} |\tau - \tau'| \leq \delta\}$ be a small enlargement of $\mathcal{T}_0$.

Corollary 4.1 *Let Assumption A1 hold for* $\mathcal{T}_1 = \mathcal{T}_0^{\delta}$, *and assume* $\theta_0(\cdot)$ *is equicontinuous on* $\mathcal{T}_1$ *and that* $d_H\left(\hat{\mathcal{T}}_0, \mathcal{T}_0\right) = o_P(1)$. *Then, it holds that* $d_H\left(\widehat{\Theta}_0, \Theta_0\right) = o_P(1)$.

We now use our previous uniform convergence results to construct confidence regions for the identified set Θ_0. We construct such confidence regions by inverting tests of the hypothesis $H_0 : \theta(\tau) = \theta_0(\tau)$ for all $\tau \in \mathcal{T}_0$, against the negation of H_0. One popular test statistic is the Kolmogorov-Smirnov (KS) $\sqrt{n}\sup_{\tau \in \hat{\mathcal{T}}_0}\left|\widehat{\theta}_n(\tau) - \theta_0(\tau)\right|$, whose asymptotic quantiles can be approximated by bootstrap. The KS test leads to the confidence region

$$CR_{1-\alpha,n} := \bigcup\nolimits_{\tau \in \hat{\mathcal{T}}_0}\left\{\theta \in \Theta : \left|\theta - \widehat{\theta}_n(\tau)\right| \leq \widehat{c}_{1-\alpha,n}/\sqrt{n}\right\},$$

where $\widehat{c}_{1-\alpha,n}$ is the bootstrap approximation of $c_{1-\alpha,n} := \inf\{c : \Pr(\sqrt{n}\sup_{\tau \in \hat{\mathcal{T}}_0}\left|\widehat{\theta}_n(\tau) - \theta_0(\tau)\right| \leq c) \geq 1-\alpha\}$. The coverage probability is established in the following proposition:

Proposition 4.1 *Let Assumptions A2 and A3 hold, and in addition assume* $\Pr\left(\mathcal{T}_0 \subset \hat{\mathcal{T}}_0\right) \to 1$. *Then*

$$\liminf_{n\to\infty} \Pr\left(\Theta_0 \subset CR_{1-\alpha,n}\right) \geq 1-\alpha.$$

The required inclusion condition $\Pr\left(\mathcal{T}_0 \subset \hat{\mathcal{T}}_0\right) \to 1$ is not very restrictive, see Lemma 4.1. It is worth stressing that estimation of $\mathcal{T}_0$ does not have an asymptotic impact on our confidence region on the identified set. This follows from the fact that

$$\sqrt{n}\sup_{\tau \in \hat{\mathcal{T}}_0}\left|\widehat{\theta}_n(\tau) - \theta_0(\tau)\right| = \sqrt{n}\sup_{\tau \in \mathcal{T}_0}\left|\widehat{\theta}_n(\tau) - \theta_0(\tau)\right| + o_P(1),$$

see Chernozhukov, Lee and Rosen (2010) for a related result.

4.1.2 Inference on the True Parameter

In this subsection, we consider testing for the null hypothesis that some given parameter vector lies in the identified set, and we use this test to construct confidence sets for the "true" parameter, see Imbens and Manski (2004) for discussions. We distinguish two cases for completeness: the general case and the case of convex identified sets. The general case has wide applicability but, as we show below, it requires a rank condition and a order condition that are similar to the typical overidentification conditions required in the standard J-test. The convex case is less general, but it has the advantage that in some cases where the order condition fails, it still can be applied.

The General Case The testing hypotheses are

$$H_0 : \widetilde{\theta} \in \Theta_0 \text{ against } H_1 : \widetilde{\theta} \notin \Theta_0.$$

These hypotheses are of interest in their own. For instance, Altonji et al. (2005) aim to test for significance of the impact of attending a Catholic school on educational attainment in a partially identified parametric model. Their example fits our setting with a null hypothesis $0 \in \Theta_0$. A stronger sense of no-impact is $\theta_0(\tau) = 0$ for all $\tau \in \mathcal{T}_1$, which can be tested using the results of the previous section.

Recalling that $M(\theta, h_0, \tau) = E[\psi(X, \theta, h_0(X, \theta, \tau), \tau)]$, then the null hypothesis is equivalent to

$$\exists\ \widetilde{\tau} \in \mathcal{T},\ s.t.\ M\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right) = 0,$$

where $\widetilde{h}_0(\cdot, \tau) := h_0\left(\cdot, \widetilde{\theta}, \tau\right)$. Define $\widetilde{\mathcal{T}} := \left\{\widetilde{\tau} \in \mathcal{T} : M\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right) = 0\right\}$, which satisfies that $\widetilde{\mathcal{T}} \subset \mathcal{T}_0$ and $\widetilde{\mathcal{T}} \neq \varnothing$ under the null hypothesis.

A simple approach to obtain a confidence region for the true parameter consists in taking the union of confidence intervals for $\theta_0(\tau)$ constructed at each $\tau \in \mathcal{T}_1$, see, e.g., Chernozhukov et al. (2011). This union-intersection principle is also possible in the general class of models we consider. However, this approach leads to conservative inference. An alternative method can be based on the following test statistic:

$$T_n(\widetilde{\theta}) := \inf_{\tau \in \mathcal{T}} \left\| \sqrt{n} M_n\left(\widetilde{\theta}, \widehat{h}, \tau\right) \right\|,$$

where $\widehat{h}$ is a consistent estimator for $\widetilde{h}_0$. For given $\widetilde{\theta}$ under the null, $T_n(\widetilde{\theta})$ can be viewed as a generalization of the classical overidentification test statistic to partially identified semiparametric models. To proceed, we need the following assumptions.

Assumption A4: For fixed $\widetilde{\theta}$ given in the testing problem:
(i) Existence of Minorant: there exist $c > 0$ and $\delta > 0$, such that $\left\| M\left(\widetilde{\theta}, \widetilde{h}_0, \tau\right)\right\| \geq c(\inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} |\tau - \widetilde{\tau}| \wedge \delta)$ for all $\tau \in \mathcal{T}$, where $a \wedge b := \min\{a, b\}$;
(ii) Smoothness in τ : $M(\widetilde{\theta}, \widetilde{h}_0, \tau)$ is continuously differentiable at all $\tau \in \mathcal{T}$, with derivative matrix $V_\tau(\widetilde{\theta}, \widetilde{h}_0, \tau) := \partial M(\widetilde{\theta}, \widetilde{h}_0, \tau)/\partial\tau$.

(iii) Smoothness in h : Assumption A2(iv)(a) holds with $\theta = \widetilde{\theta}$ and condition (10) replaced by the following condition: for $h \in \mathcal{H}_{\delta_n}$, $|\tau_1 - \tau_2| \leq \eta_n$,

$$\left\| V_h\left(\widetilde{\theta}, \widetilde{h}_0, \tau_1\right)\left[h - \widetilde{h}_0\right] - V_h\left(\widetilde{\theta}, \widetilde{h}_0, \tau_2\right)\left[h - \widetilde{h}_0\right] \right\| \leq O\left(\eta_n \delta_n\right). \tag{17}$$

(iv) The estimator $\widehat{h}$ satisfies $\left\|\widehat{h} - \widetilde{h}_0\right\|_{\mathcal{H}} = o_P\left(n^{-1/4}\right)$; and uniformly in $\widetilde{\tau} \in \widetilde{\mathcal{T}}$,

$$\sqrt{n} V_h\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)[\widehat{h} - \widetilde{h}_0] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi\left(W_i, \widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right) + o_P(1);$$

and the function class $\left\{ s\left(w, \widetilde{\theta}, \widetilde{h}_0, \tau\right) := \psi\left(w, \widetilde{\theta}, \widetilde{h}_0, \tau\right) + \phi\left(w, \widetilde{\theta}, \widetilde{h}_0, \tau\right) : \tau \in \mathcal{T} \right\}$ is Donsker.
(v) $\sup_{h \in \mathcal{H}_{\delta_n}, \tau \in \mathcal{T}} \left\| \mathbb{G}_n \psi\left(W, \widetilde{\theta}, h, \tau\right) - \mathbb{G}_n \psi\left(W, \widetilde{\theta}, \widetilde{h}_0, \tau\right) \right\| = o_P(1)$.

Assumption A4(i) is used to derive the convergence rate of the minimizers of $\inf_{\tau \in \mathcal{T}} \left\| \sqrt{n} M_n\left(\widetilde{\theta}, \widehat{h}, \tau\right) \right\|$. Chernozhukov et al. (2007) imposed similar assumptions on the sample criterion function to obtain convergence rate of the set estimates. Similar conditions can also be found in semiparametric estimation with point identification, e.g. see Ai and Chen (2003). This assumption is trivially satisfied if $M\left(\widetilde{\theta}, h_0, \tau\right)$ is linear in τ, or is differentiable in τ with derivative matrix bounded away from zero near the neighborhood of the identified set $\widetilde{\mathcal{T}}$. Other assumptions are similar to those in Assumption A2.

Let I_{d_ψ} denote the $d_\psi \times d_\psi$ identity matrix, and let D^- denote the Moore–Penrose pseudoinverse of D. Define the projection matrix $P\left(\widetilde{\theta}\right) \equiv P\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right) := I_{d_\psi} - V_\tau\left(\widetilde{\tau}\right)\left[V_\tau\left(\widetilde{\tau}\right)' \Xi V_\tau\left(\widetilde{\tau}\right)\right]^{-} V_\tau\left(\widetilde{\tau}\right)' \Xi$, where $V_\tau\left(\widetilde{\tau}\right) := V_\tau(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau})$.

Theorem 4.2 *Let the null hypothesis H_0 and Assumption A4 hold, then*

$$T_n(\widetilde{\theta}) \rightsquigarrow \inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \left\| P\left(\widetilde{\theta}\right) G\left(\widetilde{\tau}\right) \right\|,$$

where $G\left(\widetilde{\tau}\right)$ is a Gaussian process on $\widetilde{\mathcal{T}}$ with covariance kernel $\widetilde{K}\left(\cdot, \cdot\right)$ defined in Remark (3.1).

Remark 4.1 *If rank$(V_\tau\left(\widetilde{\tau}\right)) = d_\psi$ for some $\widetilde{\tau} \in \widetilde{\mathcal{T}}$, then $P\left(\widetilde{\theta}\right)$ becomes zero, see e.g. Theorem 3.5 in Yanai, Takeuchi and Takane (2011), and as a result, the limiting distribution of $T_n(\widetilde{\theta})$ is degenerate. This is for instance the case if $d_\psi \leq d_\tau$, i.e. the number of nuisance parameters is greater than or equal to the number of moments.*

Similar test statistics have been used in the partial identification literature, albeit for different models. In a support function setting, Bontemps et al. (2012) circumvent the potential multiple-minimizer problem (i.e $\widetilde{\mathcal{T}}$ is not singleton) by obtaining a shrinkage estimator whose limit is well defined, and the test statistic evaluated at this estimator converges to standard normal distribution. Galichon and Henry (2009) considered a generalized Kolmogorov-Smirnov test statistic and the special structure of their testing problem delivered a straightforward asymptotic distribution. Santos (2010) developed a testing procedure for the hypothesis that at least one element of the identified set in a nonparametric instrumental variables model satisfies a conjectured restriction, for which he constructed a similar test statistic as $T_n(\widetilde{\theta})$ but with an infinite-dimensional minimizing function space. In his setting, the limiting distribution of $T_n(\widetilde{\theta})$ involves two minimizations in two infinite-dimensional spaces. The result of Theorem 4.2 directly shows that under identification of $\widetilde{\tau}$ our test behaves like the standard J-test of overidentifying restrictions, with a limiting distribution equal to a chi-square with $d_\psi - d_\tau$ degrees of freedom. One could use the asymptotic critical values from chi-squared distribution, but inference would be conservative for the general case. For more efficient inference, we propose using a multiplier-type bootstrap described below. If $rank\left(V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)\right) < d_\psi$ for all $\widetilde{\tau} \in \widetilde{\mathcal{T}}$, then by the proof of Theorem 4.2, we have

$$T_n(\widetilde{\theta}) = \inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \left\| P\left(\widetilde{\theta}\right) \frac{1}{\sqrt{n}} \sum_{i=1}^n s\left(W_i, \widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right) \right\| + o_P(1).$$

Suppose there is a uniformly (in τ and h) consistent estimator $\widehat{V}_\tau\left(\widetilde{\theta}, h, \tau\right)$ of $V_\tau\left(\widetilde{\theta}, h, \tau\right)$, then $P\left(\widetilde{\theta}, \widetilde{h}_0, \tau\right)$ can be estimated by $\widehat{P}\left(\widetilde{\theta}\right) := I_{d_\psi} - \widehat{V}_\tau\left(\widetilde{\theta}, \widehat{h}, \tau\right)\left[\widehat{V}_\tau\left(\widetilde{\theta}, \widehat{h}, \tau\right)' \Xi \widehat{V}_\tau\left(\widetilde{\theta}, \widehat{h}, \tau\right)\right]^{-} \widehat{V}_\tau\left(\widetilde{\theta}, \widehat{h}, \tau\right)' \Xi$. We suggest the following bootstrap procedure:

Algorithm 2:
Step 1: Generate $\{u_i\}_{i=1}^n$ as in Step 1 of Algorithm 1.
Step 2: For each τ, compute $R_n^*(\tau) = \left\| \widehat{P}\left(\widetilde{\theta}\right) n^{-1/2} \sum_{i=1}^n s\left(W_i, \widetilde{\theta}, \widehat{h}, \tau\right) u_i \right\| + \lambda_n \left\| M_n\left(\widetilde{\theta}, \widehat{h}, \tau\right) \right\|$, where $\{\lambda_n\}$ is a positive sequence diverging to infinity at an appropriate rate, so that $\lambda_n = o\left(\sqrt{n}/\log\log n\right)$, e.g. $\lambda_n = \log n$ suffices.
Step 3: Compute $T_n^*(\widetilde{\theta}) = \inf_{\tau \in \mathcal{T}} R_n^*(\tau)$.
Step 4: Repeat Step 1-3 B times and compute the $(1-\alpha)$ empirical bootstrap critical value $c_{n,1-\alpha}^*(\widetilde{\theta})$.

Remark 4.2 *We follow ideas in Santos (2010) and use the penalty term $\lambda_n \left\| M_n\left(\widetilde{\theta}, \widehat{h}, \tau\right) \right\|$ in Step 2. This is necessary because the process $n^{-1/2} \sum_{i=1}^n s\left(W_i, \widetilde{\theta}, \widehat{h}, \tau\right) u_i$ is centered for all $\tau \in \mathcal{T}$ instead of just for $\widetilde{\tau} \in \widetilde{\mathcal{T}}$. By introducing this penalty term, it is ensured that $R_n^*(\tau)$ diverges to infinity for all $\tau \notin \widetilde{\mathcal{T}}$; hence when computing $T_n^*(\widetilde{\theta})$, the infimum is effectively evaluated only at a shrinking neighborhood of $\widetilde{\mathcal{T}}$. Notice that the test statistics are computed by taking infimum over the whole nuisance parameter set $\mathcal{T}$. An alternative approach can be conducted by first estimating the argmin set $\widetilde{\mathcal{T}}$, then computing the bootstrap statistics using this estimated set. However, this approach is computationally expensive for constructing confidence sets for the true parameter value as we need to estimate $\widetilde{\mathcal{T}}$ for each candidate $\widetilde{\theta}$.*

Next proposition shows that the bootstrap test statistic $T_n^*(\widetilde{\theta})$ converges to the same limit as $T_n(\widetilde{\theta})$:

Proposition 4.2 *Suppose (i) Assumption A4 holds; (ii) the penalty sequence* λ_n *satisfies* $\lambda_n \to \infty$ *and* $\lambda_n = o\left(\sqrt{n}/\log\log n\right)$; *and (iii) the asymptotic distribution of* $T_n(\widetilde{\theta})$ *is continuous at its* $1-\alpha$ *quantile. Then, under the null hypothesis,*

$$\lim_{n\to\infty} \Pr\left(T_n(\widetilde{\theta}) \leq c^*_{n,1-\alpha}(\widetilde{\theta})\right) = 1-\alpha.$$

Moreover, under the alternative hypothesis, we have

$$\lim_{n\to\infty} \Pr\left(T_n(\widetilde{\theta}) > c^*_{n,1-\alpha}(\widetilde{\theta})\right) = 1.$$

With the consistency of the testing procedure, we can construct confidence region for the true parameter by collecting all $\widetilde{\theta}$ that cannot be rejected by our test, i.e. define the confidence region as $CS_{n,1-\alpha} := \left\{\widetilde{\theta} \in \Theta : T_n(\widetilde{\theta}) \leq c^*_{n,1-\alpha}(\widetilde{\theta})\right\}$. Then, it follows that $\lim_{n\to\infty} \Pr\left(\theta_0 \in CS_{n,1-\alpha}\right) = 1-\alpha$ for the true parameter θ_0 driving the underlying data generating process.

The Convex Case As discussed in Remark (4.1), when $rank(V_\tau(\widetilde{\tau})) = d_\psi$ (for example if $d_\psi \leq d_\tau$), the proposed test statistic $T_n\left(\widetilde{\theta}\right)$ is degenerate. To overcome this problem, we complement our previous proposal with one in which we impose shape restrictions (convexity) on the identified set. Recall from Example 1 that $\beta(\tau|B)$ denotes the support function of a convex set B. The testing problem is the same as in the general case. Let $\widetilde{\mathcal{S}} := \arg\min_{\tau\in\mathbb{S}^{d_\theta}}\{\beta(\tau|\Theta_0) - \tau'\widetilde{\theta}\}$ be the set of minimizers, which does not need to be a singleton. We need the following conditions:

Assumption A5: For fixed $\widetilde{\theta}$ given in the testing problem:
(i) Θ_0 is convex and compact;
(ii) There exists a moment function $\xi(\cdot)$ such that $\beta(\tau|\Theta_0) = E\left[\xi(W, g_0, \tau)\right]$, where $g_0(\cdot)$ is a vector of unknown functions which includes $h_0(\cdot)$ and possibly other nuisance parameters. Suppose g_0 admits a uniformly consistent estimator $\widehat{g}(\cdot)$ such that $\|\widehat{g} - g_0\|_{\mathcal{G}} = o_P\left(n^{-1/4}\right)$ where the pseudo-metric $\|\cdot\|_{\mathcal{G}}$ is similarly defined as $\|\cdot\|_{\mathcal{H}}$.
(iii) Smoothness in g : the pathwise derivative $V_g^c(g_0,\tau)[g-g_0]$ of $E\left[\xi(W,g,\tau)\right]$ at $g = g_0$ exists in all directions $[g-g_0] \in \mathcal{G}$; and for all $(g,\tau) \in \mathcal{G}_{\delta_n} \times \mathbb{S}^{d_\theta}$ with a positive sequence $\delta_n \to 0$, it holds that

$$\sup_{\tau\in\mathbb{S}^{d_\theta}} \left\|E\left[\xi(W,g,\tau)\right] - E\left[\xi(W,g_0,\tau)\right] - V_g^c(g_0,\tau)[g-g_0]\right\| \leq c\|g-g_0\|^2_{\mathcal{G}}$$

(iv) Uniformly in $\tau \in \mathbb{S}^{d_\theta}$,

$$\sqrt{n}V_g^c(g_0,\tau)[\widehat{g}-g_0] = \frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^n \zeta(W_i,g_0,\tau) + o_P(1);$$

and the function class $\{s^c(w,g_0,\tau) := \xi(w,g_0,\tau) + \zeta(w,g_0,\tau) : \tau \in \mathbb{S}^{d_\theta}\}$ is Donsker.
(v) Stochastic Equicontinuity in g :

$$\sup_{g\in\mathcal{G}_{\delta_n},\tau\in\mathbb{S}^{d_\theta}} \|\mathbb{G}_n\xi(W,g,\tau) - \mathbb{G}_n\xi(W,g_0,\tau)\| = o_P(1).$$

The test statistic we propose is

$$T_n^c(\widetilde{\theta}) := \inf_{\tau \in \mathbb{S}^{d_\theta}} \sqrt{n} E_n \left[\xi \left(W, \widehat{g}, \tau \right) - \tau' \widetilde{\theta} \right].$$

The test consists in rejecting H_0 for small values of $T_n^c(\widetilde{\theta})$. The asymptotic distribution of $T_n^c(\widetilde{\theta})$ is described below.

Theorem 4.3 *Let the null hypothesis H_0 and Assumption A5 hold. Then*

$$\begin{cases} T_n^c(\widetilde{\theta}) \xrightarrow{P} +\infty, & \text{if } \widetilde{\theta} \in int\left(\Theta_0\right) \\ T_n^c(\widetilde{\theta}) \rightsquigarrow \inf_{\tau \in \widetilde{\mathcal{S}}} \{ G^c\left(\tau\right) \}, & \text{if } \widetilde{\theta} \in \partial \Theta_0 \\ T_n^c(\widetilde{\theta}) \xrightarrow{P} -\infty, & \text{if } \widetilde{\theta} \notin \Theta_0 \end{cases}$$

where $G^c\left(\tau\right)$ is a Gaussian process on $\widetilde{\mathcal{S}}$ with covariance kernel $K^c\left(\cdot,\cdot\right)$ defined as:

$$K^c\left(\tau_1, \tau_2\right) := Cov\left(s^c\left(w, g_0, \tau_1\right), s^c\left(w, g_0, \tau_2\right) \right),$$

where s^c is defined in A5(iv).

Remark 4.3 *We do not explicitly require differentiability of the support function $\beta\left(\tau | \Theta_0\right)$ w.r.t. τ which is equivalent to the nonexistence of the exposed faces of the identified set Θ_0. However, the stochastic equicontinuity assumption might implicitly rule out the presence of exposed faces, see for example Proposition 9 in Bontemps et al. (2012), who, in their specific setting, derive the non-regular asymptotic distribution of the support function estimator with the presence of exposed faces. It should be noticed that non-differentiability does not necessarily lead to non-regular asymptotic distribution of the support function estimator, e.g. see section 5.3 in Bontemps et al. (2012).*

Remark 4.4 *To deal with the multiple minimizer problem caused by the presence of kinks, Bontemps et al. (2012) proposes a shrinkage type estimator to obtain a unique minimizer, instead we take the minimum over all possible minimizers.*

To implement the test we can follow the similar bootstrap procedure introduced for the general case to obtain the critical value. That is, we can use the same multiplier bootstrap method and introduce the slowly diverging sequence as an additional penalizing term. However, if we a priori know the uniqueness of the minimizer, then the limit distribution of $T_n^c\left(\widetilde{\theta}\right)$ in the least favorable case boils down to a normal distribution. In this case, or simpler case with known minimizer, there is no need to do bootstrap and instead we can just plug in the estimator of the minimizer τ_n to consistently estimate the asymptotic variance.

4.2 Sensitivity Analysis

The previous uniform results can be also applied to carry out formal sensitivity analysis. Sensitivity analysis differs from set inference in the choice of the set $\mathcal{T}_1$, and also in the set of estimands considered. Typically, in applications of sensitivity analysis τ_0 is identified, i.e. $\mathcal{T}_0 = \{\tau_0\}$, but its identification is viewed as very fragile, e.g. dependent on ad-hoc functional form assumptions, so it is not considered credible. To illustrate the main ideas, consider the specific setting considered in Imbens (2003). The aim is to investigate the sensitivity of the average treatment effect (ATE) parameter to the exogeneity assumption (unconfoundedness) in program evaluation. Rosenbaum and Rubin (1983) pioneered this sensitivity analysis by assuming that the exogeneity condition holds only after conditioning on an unobservable covariate. In a parametric setting, they investigate the sensitivity of the ATE estimator $\widehat{\theta}(\tau)$ to the coefficients τ of the unobservable covariate over an arbitrary range $\mathcal{T}_1$. Imbens (2003) suggests to interpret the sensitivity analysis in terms of partial R^2 of the unobserved covariate, in comparison with partial R^2 of observed covariates. Here we contribute to this literature by providing a formal analysis of sensitivity analysis that accounts for estimators and model uncertainty.

In the context of program evaluation the set of moments is given, for instance, by score equations resulting from the parametric specification of the potential outcomes and selection equations, see Imbens (2003) for details. Using our notation, the moments are $E[\psi(W, \theta_0, h_0(W), \tau_0)] = 0$, where θ_0 is the ATE, $h_0(W) \equiv h_0$ is here finite-dimensional and includes nuisance parameters such as the coefficients of observable covariates X, and τ_0 denotes the coefficients of the unobservable covariate in the outcome and selection equations, respectively. Following Imbens (2003), let $\hat{R}^2_{Y,par}(\tau)$ and $\hat{R}^2_{D,par}(\tau)$ be proportion of the variation in the outcome and treatment, respectively, that is explained by the unobserved covariate. Imbens (2003) reports the pairs $\left(\hat{R}^2_{Y,par}(\tau), \hat{R}^2_{D,par}(\tau)\right)$ where $\left|\widehat{\theta}_n(\tau) - \widehat{\theta}_n(0)\right| > r$, where r is some pre-specified threshold ($\tau = 0$ is the benchmark of exogeneity, and $r = \$1000$ in Imbens' application). He then compares these pairs with pairs of partial R^2 based on relevant observed covariates. The sensitivity is judged on the basis of this comparison. There are two important limitations of this approach. First, it only uses limited information on the parameter of interest, i.e. whether or not $\left|\widehat{\theta}_n(\tau) - \widehat{\theta}_n(0)\right| > r$ for a fixed r. The conclusions may sensitive to the choice of r. Second, it does not account for estimators and model uncertainty. Necessarily $E[\psi(W, \theta_0, h_0(W), 0)] \neq 0$, which suggests that standard errors of estimates need to account for misspecification.

Our methods provide a formal approach to sensitivity analysis. Combining the ideas of Rosenbaum and Rubin (1983) and Imbens (2003) we suggest to report the set $\widehat{\Theta}_1 = \{\widehat{\theta}(\tau) : \tau \in \widehat{\mathcal{T}}_1\}$, where

$$\widehat{\mathcal{T}}_1 = \left\{\tau \in \mathcal{T} : \hat{R}^2_{Y,par}(\tau) \leq \hat{r}_Y, \hat{R}^2_{D,par}(\tau) \leq \hat{r}_D\right\},$$

and $(\hat{r}_Y, \hat{r}_D)$ are, for instance, the partial R^2 of some observed covariates on the outcome and treatment, respectively. See also Altonji et al. (2005) for a related choice of $\widehat{\mathcal{T}}_1$. Inference on the set $\Theta_1 = \{\theta_0(\tau) : \tau \in \mathcal{T}_1\}$, where $\mathcal{T}_1$ is the limit set of $\widehat{\mathcal{T}}_1$, can be carried out using our uniform results, including a confidence region for Θ_1, a consistent estimator for the diameter of the set Θ_1, i.e. $d = \sup_{\tau,\tau' \in \mathcal{T}_1} |\theta_0(\tau) - \theta_0(\tau')|$, or a test for the hypothesis $0 \in \Theta_0$. Likewise, a quantity of interest in sensitivity analysis is $\partial\theta_0(\tau)/\partial\tau$. This quantity can be estimate from the first order conditions

$$(E[\psi_\theta(Z,\theta_0(\tau),h_0,\tau)])' \Xi E[\psi(W,\theta_0(\tau),h_0,\tau)] = 0,$$

by an application of the chain rule. For instance, considering a simplified case where the model is correctly specified, h_0 does not depend on θ_0 and τ, and $d_\psi = d_\theta$, we have (following our notation)

$$\frac{\partial\theta_0(\tau)}{\partial\tau} = V_\theta^{-1}(\tau)V_\tau(\tau),$$

which can be used for uniform consistent estimation of $\partial\theta_0(\tau)/\partial\tau$. Details in the more general case are omitted for the sake of space.

4.3 Incorporating Prior Information

In applications, experts of the subject area may have some prior information on plausible values of the unidentified parameter τ. Since different practitioners may have different priors, a set identification approach is convenient. Suppose a researcher has a prior density $g(\tau)$ over the set of identified parameters $\mathcal{T}_0$, which is assumed to be known for simplicity. Then, the aim is to do inference on the parameter

$$\theta_0(g) = \int_{\mathcal{T}_0} \theta_0(\tau)g(\tau)d\tau,$$

provided the integral is well defined. The natural estimator for $\theta_0(g)$ is

$$\widehat{\theta} \equiv \widehat{\theta}(g) = \int_{\mathcal{T}_0} \widehat{\theta}(\tau)\, g(\tau)d\tau.$$

If $\widehat{\theta}$ cannot be computed numerically, we suggest the Monte Carlo approximation

$$\widetilde{\theta} \equiv \widetilde{\theta}(g) = \frac{1}{m}\sum_{j=1}^{m}\widehat{\theta}(\tau_j),$$

where $\{\tau_j\}_{j=1}^m$ are randomly drawn from g, and are independent of the original sample. Assume that

$$\int_{\mathcal{T}_0} |\theta_0(\tau)|^2 g(\tau)d\tau < \infty. \tag{18}$$

Then, simple arguments show that $\sqrt{n}(\widetilde{\theta} - \widehat{\theta}) = o_P(1)$, provided $n/m \to 0$ as $n \to \infty$. Then, the next corollary follows directly from Theorem 3.2 and the continuous mapping theorem, and hence its proof is omitted.

Corollary 4.2 *Under Assumption A2, (18) and $n/m \to 0$ as $n \to \infty$, it holds that*

$$\sqrt{n}(\widetilde{\theta}(g) - \theta_0(g)) \to_d N(0, \Sigma_g),$$

where

$$\Sigma_g = \int_{\mathcal{T}_0\times\mathcal{T}_0} C(\tau_1,\tau_2)\, g(\tau_1)g(\tau_2)d\tau_1 d\tau_2,$$

and $C(\cdot,\cdot)$ is given in Theorem 3.2.

Conley et al. (2012) investigate alternative ways to incorporate prior knowledge on the exogeneity of instrumental variables and focus on confidence regions for the true parameter. Their methods can potentially be extended to our more complex setting here. Likewise, one could think of constructing confidence regions by inverting the optimal tests of Andrews and Ploberger (1994) or Song, Kosorok and Fine (2009), which use an average power criteria according to g. All these extensions are beyond the scope of this paper and deserve further consideration.

5 Examples Revisited

In this section we revisit some of the motivating examples and discuss how our general conditions can be verified in each example. To avoid redundancy with existing literature, we refer the reader to references for complete sets of primitive conditions and results. We also describe implementation of our uniform results in these examples.

5.1 Willingness-To-Pay

In this example the nuisance parameter set $\mathcal{T}_0 = [0,1]$ is known, hence there is no need to consider misspecification and it suffices to verify that Assumption A2' in the Appendix holds for weak convergence of the identified set. We proceed directly with this assumption, as the consistency conditions in Assumption 1 are straightforward to check in this example. The moment function is given here by

$$\begin{aligned} M(\theta, h, \tau) = \theta - \tau \left(r(v_1) + \Delta_{ur}' E\left[h(X) \right] \right) \\ - (1-\tau) \left(\underline{r} + \Delta_{lr}' E\left[h(X) \right] \right). \end{aligned}$$

Hence, the first order derivative w.r.t. θ and the functional derivative w.r.t. h are, respectively,

$$\begin{aligned} V_{\theta_0}(\tau) &= 1, \\ V_h(\theta, h, \tau)\left[\overline{h} - h\right] &= \alpha(\tau)' E\left[\overline{h}(X) - h(X)\right], \end{aligned}$$

where $\alpha(\tau) = -\{\tau \Delta_{ur} + (1-\tau)\Delta_{lr}\}$. Then Assumption A2'(ii) and (iii) are trivially satisfied. The nuisance function $h_0(x)$ with components $h_{0j}(x) = E[Y|V = v_j, X = x]$ can be estimated by the kernel method, i.e.

$$\widehat{h}_j(x) := \frac{\sum_{i=1}^n Y_i I(V_i = v_j) K_b((X_i - x)/b)}{\sum_{i=1}^n I(V_i = v_j) K_b((X_i - x)/b)},$$

where $K_b((X_i - x)/b) := b^{-d_x} \prod_{l=1}^{d_x} k((X_{il} - x_l)/b)$ for some univariate kernel function $k(\cdot)$ and b a bandwidth parameter. Semiparametric or parametric specifications of h_0 can be also entertained. Suppose that $h_{0j}(\cdot) \in \mathcal{H} = \mathcal{C}_1^{\eta_h}(\mathcal{X})$, where $\mathcal{C}_1^{\eta_h}(\mathcal{X})$ is a subset of continuous functions on the convex, bounded subset $\mathcal{X} \in \mathbb{R}^{d_x}$, with non-empty interior, endowed with the sup-norm $\|\cdot\|_{\mathcal{H}} = \|\cdot\|_\infty$, as defined in van der Vaart and Wellner (1996, p.154) with $\eta_h > d_x/2$. It is known that $\mathcal{H}$ is a Donsker class. Escanciano et al. (2011) discussed primitive conditions for Assumption A2'(iv). $M(\theta, h, \tau)$ is linear in h, and therefore, the rate condition for $\widehat{h}$ can be eliminated, see Remark 2(iii) in CLV.

It is straightforward to show that $\{\psi(\cdot,\theta,h,\tau):\theta\in\Theta,h\in\mathcal{H},\tau\in\mathcal{T}\}$ is a Donsker class. Hence, Assumption A2'(v) holds. Similarly, well-known conditions, see, e.g., Newey (1994), show that Assumption A2'(vi) holds under standard conditions on the bandwidth and the kernel, with

$$\begin{aligned}\phi(W_i,\theta_0,h_0,\tau) &= \left(\frac{\alpha(\tau)' I(V_i=v_1)\varepsilon_{i1}}{p_1},...,\frac{\alpha(\tau)' I(V_i=v_m)\varepsilon_{im}}{p_m}\right)\\ &=: \alpha(\tau)' l_0(W_i),\end{aligned}$$

where $p_j := E[I(V_i=v_j)]$, $\varepsilon_{ji} := Y_i - h_{0j}(X_i)$ and $l_0(W_i)$ is implicitly defined.

As discussed above, we might introduce additional nuisance parameters when computing the linear expansion for the functional derivative $V_h(\theta_0,h_0,\tau)\left[\widehat{h}-h_0\right]$. Here in this example, the additional nuisance parameter is $p=(p_1,\ldots,p_m)'$, which can be easily estimated by the sample analog, and conditions in Assumption A3 are also easy to verify. Hence, for this example our conditions boil down to the same conditions as in the point-identified case, which are well established in the literature, see e.g. Newey (1994).

For inference, we first consider the confidence region for the identified set. Denote $\widehat{\theta}_L := E_n\left[L_r\left(\widehat{h}|X\right)\right]$ and $\widehat{\theta}_U := E_n\left[U_r\left(\widehat{h}|X\right)\right]$. Then, $\widehat{\theta}(\tau)=(1-\tau)\widehat{\theta}_L+\tau\widehat{\theta}_U$, and

$$\sup_{\tau\in[0,1]}\left|\sqrt{n}\left(\widehat{\theta}(\tau)-\theta_0(\tau)\right)\right| = \max\left\{\sqrt{n}\left|\widehat{\theta}_L-\theta_L\right|,\sqrt{n}\left|\widehat{\theta}_U-\theta_U\right|\right\}.$$

From the results above,

$$\sqrt{n}\begin{pmatrix}\widehat{\theta}_L-\theta_L\\ \widehat{\theta}_U-\theta_U\end{pmatrix} \rightsquigarrow N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\Delta'_{lr}\\ \Delta'_{ur}\end{pmatrix} Var\left(h_0(X_i)+l_0(W_i)\right)\left(\Delta_{lr}\ \Delta_{ur}\right)\right).$$

Hence, the critical value of the limit distribution of $\max\left\{\sqrt{n}\left|\widehat{\theta}_L-\theta_L\right|,\sqrt{n}\left|\widehat{\theta}_U-\theta_U\right|\right\}$ can be easily simulated and there is no need to use a bootstrap method.

For construction of confidence region for the true parameter value, notice that $d_\psi=d_\tau=1$ and the identified set is a closed interval, hence we can use the results developed for the convex case in Section 4.1. With dimension equal to 1, we have $\mathbb{S}^1=\{-1,1\}$ and according to Example 1,

$$\beta(\tau|\Theta_0) = E\left[\frac{1+\tau}{2}U_r(g_0|X)-\frac{1-\tau}{2}L_r(g_0|X)\right]$$

with $g_0=h_0$. The test statistic reduces to

$$\begin{aligned}T_n^c(\widetilde{\theta}) &:= \inf_{\tau\in\{1,-1\}}\sqrt{n}E_n\left[\xi\left(W,\widehat{h},\tau\right)-\tau'\widetilde{\theta}\right]\\ &= \min\left\{\sqrt{n}E_n\left[\xi\left(W,\widehat{h},1\right)-\widetilde{\theta}\right],\sqrt{n}E_n\left[\xi\left(W,\widehat{h},-1\right)+\widetilde{\theta}\right]\right\}\\ &= \min\left\{\sqrt{n}\left\{E_n\left[U_r\left(\widehat{h}|X\right)\right]-\widetilde{\theta}\right\},\ -\sqrt{n}\left\{E_n\left[L_r\left(\widehat{h}|X\right)\right]-\widetilde{\theta}\right\}\right\},\end{aligned}$$

where $\xi(W,h_0,\tau)=0.5(1+\tau)U_r(h_0|X)-0.5(1-\tau)L_r(h_0|X)$. From previous arguments, we have

$$V_g^c(h_0,\tau)[\widehat{h}-h_0] = \left(\frac{1+\tau}{2}\Delta'_{ur}-\frac{1-\tau}{2}\Delta'_{lr}\right)\frac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n l_0(W_i).$$

Hence,

$$\zeta\left(W,h_0,\tau\right)=\left(\frac{1+\tau}{2}\Delta_{ur}^{\prime}-\frac{1-\tau}{2}\Delta_{lr}^{\prime}\right)l_0\left(W\right).$$

Notice that

$$T_n^c(\widetilde{\theta})=\begin{cases}\sqrt{n}\left(\widehat{\theta}_U-\widetilde{\theta}\right), \text{ if } \widetilde{\theta}\geq(\widehat{\theta}_L+\widehat{\theta}_U)/2\\ -\sqrt{n}\left(\widehat{\theta}_L-\widetilde{\theta}\right), \text{ if } \widetilde{\theta}<(\widehat{\theta}_L+\widehat{\theta}_U)/2.\end{cases}$$

Assume $\theta_U-\theta_L>0$, so that $\Pr\left(\widehat{\theta}_U>\widehat{\theta}_L\right)\to 1$, and define $c_{\alpha L}$ and $c_{\alpha U}$ such that

$$\lim_{n\to\infty}\Pr\left(\sqrt{n}\left\{\widehat{\theta}_L-\theta_L\right\}>c_{\alpha L}\right)=\alpha \text{ and } \lim_{n\to\infty}\Pr\left(\sqrt{n}\left\{\widehat{\theta}_U-\theta_U\right\}>c_{\alpha U}\right)=\alpha,$$

respectively. Denote the estimated critical values as $\widehat{c}_{\alpha L}$ and $\widehat{c}_{\alpha U}$, the confidence interval can be constructed as

$$\left[\widehat{\theta}_L-\frac{\widehat{c}_{\alpha L}}{\sqrt{n}},\ \widehat{\theta}_U+\frac{\widehat{c}_{\alpha U}}{\sqrt{n}}\right],$$

and the test rejects the null at the $\alpha\%$ nominal level if $\widetilde{\theta}$ is outside this interval.

As expected, in this example our uniform results lead to well-known procedures for constructing confidence regions for the identified set and the true parameter. Imbens and Manski (2004) and Stoye (2009) have shown how to modify the confidence region for the true parameter to obtain confidence regions with better coverage properties, and similar modifications can be obtained for our semiparametric model here. However, constructing general versions of their confidence regions for our more general setting seems complicated and is deferred for future research. An earlier version of this paper contains an empirical application of the WTP example, re-examining the contingent valuation study conducted by Hanemann et al. (1991) for protecting wetland habitats and wildlife in California's San Joaquin Valley; see Escanciano and Zhu (2011).

5.2 Gender Gap Distributional Effects

Understanding the determinants and dynamics of the gender gap has been one of the most prominent problems in labor economics. Example 2 above discusses a robust inference approach to this problem. It is not difficult to show that the gender gap functional $L\varphi(\cdot,\alpha)=E\left[\varphi(1,X_2,\alpha)-\varphi(0,X_2,\alpha)\right]$ has the Riesz's representer

$$r(x):=\frac{x_1-p(x_2)}{p(x_2)(1-p(x_2))},$$

where $p(x_2):=P[X_1=1|X_2=x_2]$. We assume that $0<p(x_2)<1$ for all x_2 in $\mathcal{X}_{X_2}\subset\mathbb{R}^{d_{x_2}}$. Manski (1989) discussed bounds for $\varphi(1,x_2,\alpha)-\varphi(0,x_2,\alpha)$ and $\partial\varphi(x,\alpha)/\partial x$ for a fixed α and x. When the interest is in linear functionals of these quantities, the Riesz representation is a convenient universal tool that permits to work exclusively with the natural bounds for $\varphi(x,\alpha)$, without imposing bounds on $\partial\varphi(x,\alpha)/\partial x$. Of course, for specific functionals and applications direct bounds on $L\varphi(\cdot,\alpha)$ might be available based on those for φ, as is certainly the case for the gender gap functional, see Manski (1989) for the direct bounds on $\varphi(1,x_2,\alpha)-\varphi(0,x_2,\alpha)$.

For the verification of our assumptions in Example 2 it is convenient to consider a different parametrization and representation of the moments as

$$M(\theta, h, \tau) = E\left[\theta - \lambda a(W, \alpha, h, 0) - (1-\lambda) a(W, \alpha, h, 1)\right],$$

where, for $d = 0, 1$ and with $h = (m, g)$,

$$a(W, \alpha, h, d) := m(1, X_2, \alpha) - m(0, X_2, \alpha) + (-1)^{d+1}(1 - g(d, X_2)).$$

The nuisance parameters is given here by $h_0 = (m_0, g_0)$, where $m_0(d, X_2, \alpha) := E\left[D1(Y \leq \alpha) | X_1 = d, X_2\right]$. This parametrization is convenient because M is linear in h. Standard kernel estimators can be used to estimate m_0 and g_0. Denote the estimate by $\widehat{h} = (\widehat{m}, \widehat{g})$.

Checking the consistency assumption is straightforward in this example, so we focus on the more involved Assumption A2' (note the set $\mathcal{T}_0$ is known). In this application we take

$$\mathcal{H} = \left\{(m, g) : m(d, \cdot), g(d, \cdot) \in \mathcal{C}_1^{\eta}\left(\mathcal{X}_{X_2}\right) \text{ for each } d = 1, 0, \text{ where } \eta > d_{x_2}/2\right\}.$$

The first order derivative w.r.t. θ and the functional derivative w.r.t. $h = (m, g)$ are given, respectively, by

$$\begin{aligned} V_{\theta_0}(\tau) &= 1, \\ V_m(\theta, h, \tau)\left[\overline{m} - m\right] &= E\left[\overline{m}(0, X_2, \alpha) - m(0, X_2, \alpha) - \overline{m}(1, X_2, \alpha) + m(1, X_2, \alpha)\right], \\ V_g(\theta, h, \tau)\left[\overline{g} - g\right] &= -\lambda E\left[\left(\overline{g}(0, X_2) - g(0, X_2)\right)\right] + (1-\lambda) E\left[\left(\overline{g}(1, X_2) - g(1, X_2)\right)\right]. \end{aligned}$$

Then Assumption A2'(ii) and (iii) are trivially satisfied. Since the moment function is linear in h we only require $\left\|\widehat{h} - h_0\right\|_{\mathcal{H}} = o_P(1)$. Using the linearity and the definition of $\mathcal{H}$ it follows that Assumption A2'(v) holds by standard empirical processes arguments. Likewise, conditions for the uniform consistency of $\widehat{h}$ and the bias calculations for Assumption A2'(vi) are standard in the nonparametric literature.

Denote $\widehat{\theta}_L(\alpha) := E_n\left[a(W, \alpha, \widehat{h}, 1)\right]$ and $\widehat{\theta}_U(\alpha) := E_n\left[a(W, \alpha, \widehat{h}, 0)\right]$. Define the sets $\mathcal{A}_n := \{Y_j : Y_j \in \mathcal{A},\ j = 1, ..., n\}$ and $\widehat{\mathcal{T}}_0 := [0, 1] \times \mathcal{A}_n$. Then, by the arguments in the previous example

$$\sup_{\tau \in \widehat{\mathcal{T}}_0}\left|\sqrt{n}\left(\widehat{\theta}(\tau) - \theta_0(\tau)\right)\right| = \max_{\alpha \in \mathcal{A}_n} \max\left\{\sqrt{n}\left|\widehat{\theta}_L(\alpha) - \theta_L(\alpha)\right|, \sqrt{n}\left|\widehat{\theta}_U(\alpha) - \theta_U(\alpha)\right|\right\},$$

where $\theta_L(\alpha) := E\left[a(W, \alpha, h, 1)\right]$ and $\theta_U(\alpha) := E\left[a(W, \alpha, h, 0)\right]$. Here, we use the multiplier bootstrap implementation to approximate critical values, which is computationally more attractive than the nonparametric bootstrap. To that end, we need the expression for the influence function ϕ in Assumption A2'(vi), which is straightforward to obtain.

5.3 Sensitivity in Binary Choice with Sample Selection and No Exclusions

The model is given by the following equations

$$\begin{aligned} Y &= 1\left(\tau_0 X_1 + X_2 + \theta_0' X_3 - e \geq 0\right) D, \\ D &= 1\left[g_0(X) - u \geq 0\right], \end{aligned}$$

where (e, u) are drawn from an unknown joint distribution function $F_0(e, u)$, with $(e, u) \perp X$ and u is distributed as $U[0, 1]$. Hence, we have the conditional moment restrictions

$$E[Y|X] = F_0(\tau X_1 + X_2 + \theta_0' X_3, g_0(X)), \tag{19}$$

that can be used to estimate θ_0. We illustrate the general ideas with the semiparametric least squares (SLS) estimator. Similar ideas can be applied with likelihood methods. Denote the index $V(X, \theta_0, g_0, \tau) := (\tau X_1 + X_2 + \theta_0' X_3, g_0(X))$ and consider the unconditional moments restriction that result as the firsts order conditions of the SLS estimator, i.e.

$$\psi(W, \theta_0, h_0(W), \tau_0) = (Y - F_0(V(X, \theta_0, g_0, \tau)))\partial_\theta F_0(V(X, \theta_0, g_0, \tau))/\partial\theta.$$

Here the nuisance parameters are $h_0 = (F_0, g_0, \partial_\theta F_0) \in \mathcal{H} \equiv \mathcal{F}^\eta \times \mathcal{C}_1^{\eta_g}(\mathcal{X}_X) \times \mathcal{F}^\eta$, where $\eta_g > d_x/2$ and $\mathcal{F}^\eta$ is the following class of functions, see Escanciano et al. (2012). Let $\mathcal{W} := \{V(X, \theta, g, \tau) : \theta \in \Theta, g \in \mathcal{C}_1^{\eta_g}(\mathcal{X}_X), \tau \in \mathcal{T}_1\}$. Let $\mathcal{F}^\eta$ be a class of measurable functions on $\mathcal{X}_X$, $q(V(x)|V)$ say, such that $V \in \mathcal{W}$ and q satisfies for a universal constant C_L and each $V_j \in \mathcal{W}$, $j = 1, 2$,

$$\|q(V_1(\cdot)|V_1) - q(V_2(\cdot)|V_2)\|_\infty \leq C_L \|V_1 - V_2\|_\infty. \tag{20}$$

Moreover, we assume that for each $V \in \mathcal{W}$, $q(\cdot|V) \in C^\eta(\mathcal{X}_V)$, for some $\eta > 1$, and that q is bounded.

Suppose we want to carry out sensitivity analysis with respect to an exclusion restriction $\tau = 0$. We apply our results with $\tau \in \mathcal{T}_1$, for a generic $\mathcal{T}_1$, accounting for misspecification. Primitive conditions for many of our assumptions to hold are provided in Escanciano et al. (2011). Ichimura and Lee (2010) also investigate conditions for single-index model estimation under misspecification that are relevant in this application. The model here is an extension of theirs to double-index model with a nonparametric index g_0 and the uniformity aspect in τ.

Let $\widehat{g}$ be a kernel estimator for g_0, and for a candidate $V(X, \theta, \widehat{g}, \tau)$, let $\widehat{F}(V(X, \theta, \widehat{g}, \tau))$ and $\partial_\theta \widehat{F}(V(X, \theta, \widehat{g}, \tau))$ be kernel estimates of $E[Y|V(X, \theta, \widehat{g}, \tau)]$ and its derivative w.r.t θ, respectively. Then, consistency under our partial identification assumption holds by the uniform rate results for kernel estimates in Escanciano et al. (2011). Although we allow for misspecification, in the sense that the conditional moment (19) may not hold on $\tau \in \mathcal{T}_1$, we assume for simplicity that the unconditional moment holds. That is, we assume the SLS estimator is consistent for a minimizer of the least squares criteria for each $\tau \in \mathcal{T}_1$. Hence, we proceed to check the conditions for weak convergence in Assumption A2'.

We assume that $E[Y|V(X, \theta, g_0, \tau)]$ is twice continuously differentiable in θ. Henceforth, for simplicity of the notation we remove the dependence on true values and simply use a subscript zero, e.g. $V_0 \equiv V(X, \theta_0, g_0, \tau)$. Using an obvious short notation, the pathwise derivatives are given by

$$\begin{aligned}
V_{\theta_0}(\tau) &= E[-\partial_\theta F_0 \partial_\theta F_0 + (Y - F_0)\partial_{\theta\theta} F_0], \\
V_F(\theta_0, h_0, \tau)[\overline{F} - F] &= -E\left[(\overline{F} - F_0)\partial_\theta F_0\right], \\
V_g(\theta_0, h_0, \tau)[\overline{g} - g] &= -E\left[(\partial_g F_0 \partial_\theta F_0 + (Y - F_0)\partial_{\theta g} F_0)(\overline{g} - g_0)\right], \\
V_{\partial_\theta F_0}(\theta_0, h_0, \tau)[\overline{\partial_\theta F_0} - \partial_\theta F_0] &= E\left[(Y - F_0)(\overline{\partial_\theta F_0} - \partial_\theta F_0)\right],
\end{aligned}$$

where $\partial_g F_0(V_{0i}) := \partial F(V(X_i, \theta, \bar{g}, \tau))|V_0)/\partial \bar{g}|_{\bar{g}=g_{0i}}$ and $\partial_{\theta g} F_0 := \partial^2 F(V(X_i, \theta, \bar{g}, \tau))|V_0)/\partial\theta\partial\bar{g}\big|_{\bar{g}=g_{0i}}$. Assumption A2'(ii) is here assumed. Assumption A2'(iii) is satisfied with $c = 0$ for the corresponding terms to F and $\partial_\theta F$, and hence, we only require rates on $\|\widehat{g} - g_0\|_\infty$. Standard empirical processes arguments and the definition of $\mathcal{H}$ then imply that Assumption A2'(v) holds. Verification of Assumption A2'(vi) is long but involves standard bias calculations. Details are available from the authors upon request. In the correct specified case where the conditional moment (19) holds there is no contribution in Assumption A2'(vi) from the pathwise derivative $V_{\partial_\theta F_0}$ and from the second component of V_g, but in the misspecified case these terms contribute to the asymptotic distribution of the estimators.

In this application there is no closed form expression for $\theta_0(\tau)$. Hence, it is convenient to consider a discrete approximation of $\mathcal{T}_1$ to implement our inferences. Similarly, the expression for the influence functions is too involved in the misspecified case, so the nonparametric bootstrap seems to be more convenient in this setting. The algorithm for implementing our inference is as follows:

Algorithm 3: (Nonparametric Bootstrap Approximation)
Step 1: Choose a grid approximation $\{\tau_1, ..., \tau_m\}$ of $\mathcal{T}_1$. For $j = 1, ..., m$: compute kernel estimates of $E[Y|V(X, \theta, \widehat{g}, \tau_j)]$, say $\widehat{F}(V(X, \theta, \widehat{g}, \tau_j))$. Then, compute the SLS estimator $\widehat{\theta}(\tau_j)$.
Step 2: Draw a bootstrap sample from the empirical distribution of the data and repeat Step 1 above with the bootstrap sample.
Step 3: Repeat Step 2 B times. Approximate the distribution $\sqrt{n}\left(\widehat{\theta}(\cdot) - \theta_0(\cdot)\right)$ by the empirical distribution of $\left\{\sqrt{n}\left(\widehat{\theta}^{*l}(\cdot) - \widehat{\theta}(\cdot)\right)\right\}_{l=1}^{B}$, where $\widehat{\theta}^{*l}$ denotes the $l-th$ iteration of Step 2.

6 Conclusions

In this paper, we develop estimation and inference procedures for a class of semiparametric partially identified models, in which the identified set for the Euclidean parameter of interest is parametrized. Under this setting, we can employ tools from empirical process theory to derive uniform convergence results for the set estimates. Our inference methods allow but do not require convex identified sets. Our framework also allows for the presence of infinite-dimensional nuisance parameters. The proposed methods can be also applied to convex identified sets characterized by semiparametric support functions, which complements the support function approach considered in Bontemps et al. (2012) and Beresteanu and Molinari (2008). We propose bootstrap procedures to approximate the asymptotic distribution of functionals of the set estimator. Based on the uniform weak convergence results and consistency of the bootstrap, we construct a simple confidence region for the identified set. Finally, by extending the classical overidentification restrictions test to our semiparametric partially identified setting, we are able to construct a confidence set for the true parameter value that is robust to partial identification of the nuisance parameter. We derive the asymptotic distribution of the test statistic, which turns out to be simpler than related test statistics in Santos (2010). When there are more moments than unidentified parameters, we can employ a multiplier-type bootstrap to obtain the critical value of the test. Otherwise, the resulting test is degenerate and we show how shape restrictions on the identified set can provide alternative non-degenerate inference. Our confidence interval attains the nominal coverage probability asymptotically under some mild conditions.

The methods developed in this paper can be potentially useful in other contexts. For example, we can extend the parametric quantile process studied in Angrist et. al. (2006) to other semiparametric quantile models with infinite-dimensional parameters, such as partially linear quantile regressions (here τ is the quantile level). The results derived here can also be used to develop consistent specification tests in generic semiparametric partially identified models.

One possible extension of our approach is to allow for an infinite-dimensional unidentified parameter τ. There are a number of applications in which this can be useful, see, e.g., dynamic binary panel data models in Honore and Tamer (2006), dynamic discrete decision processes in Magnac and Thesmar (2002), nonparametric instrumental variable models in Santos (2010) or models with discrete outcomes in Chesher (2010). Our current results on confidence regions for the true parameter are only applicable to these cases under further semi/parametric assumptions (e.g. conditional moment restrictions), and although most of our theoretical results are directly applicable to a nonparametric $\mathcal{T}$, feasible versions of the proposed inferences require an approximation of $\mathcal{T}$ by sieves. These feasible inferences under this more general setting can be justified combining the results of this paper with those of Santos (2010). This extension is beyond the scope of this paper and it is deferred for future research. On the contrary, the extension to conditional moment restrictions is trivial, as we can transform these models into an infinite number of unconditional moments, whose index is part of τ. This illustrates the versatility of our results.

7 Appendix

7.1 Further Assumptions

The following condition replaces Assumption A2 under correct specification.

Assumption A2': Suppose that $\theta_0(\tau)$ is in interior of Θ for each $\tau \in \mathcal{T}_1$, and that $\sup_{\tau\in\mathcal{T}_1} |\widehat{\theta}(\tau) - \theta_0(\tau)| = o_P(1)$. In addition, assume
(i) The estimator $\widehat{\theta}(\tau)$ also satisfies

$$\sup_{\tau\in\mathcal{T}_1}\left\{\left\|M_n\left(\widehat{\theta}(\tau), \widehat{h}, \tau\right)\right\| - \inf_{\theta\in\Theta_\delta}\left\|M_n\left(\theta, \widehat{h}, \tau\right)\right\|\right\} \leq o_P\left(n^{-1/2}\right).$$

(ii) Smoothness in θ : for each $\tau \in \mathcal{T}_1$, the map $\theta \to M(\theta, h_0, \tau)$ is continuously differentiable at $\theta_0(\tau)$, with derivative $V_{\theta_0}(\tau)$ that is of full rank and $\sup_{\tau\in\mathcal{T}_1} \|V_{\theta_0}(\tau)\| < \infty$ and $\sup_{\tau\in\mathcal{T}_1}\left\|V_{\theta_0}^{-1}(\tau)\right\| < \infty$.
(iii) Smoothness in h : for each $(\theta, \tau) \in \Theta_\delta \times \mathcal{T}_1$, the pathwise derivative $V_h(\theta, h_0, \tau)[h - h_0]$ of $M(\theta, h, \tau)$ at $h = h_0$ exists in all directions $[h - h_0] \in \mathcal{H}$; and for all $(\theta, h, \tau) \in \Theta_{\delta_n} \times \mathcal{H}_{\delta_n} \times \mathcal{T}_1$ with a positive sequence $\delta_n \to 0$, it holds that

$$\sup_{\tau\in\mathcal{T}_1}\|M(\theta, h, \tau) - M(\theta, h_0, \tau) - V_h(\theta, h_0, \tau)[h - h_0]\| \leq c\,\|h - h_0\|^2_{\mathcal{H}}$$

for a constant $c \geq 0$, and

$$\sup_{\tau\in\mathcal{T}_1} \left\| V_h\left(\theta,h_0,\tau\right)\left[h-h_0\right] - V_h\left(\theta_0,h_0,\tau\right)\left[h-h_0\right]\right\| \leq o\left(1\right)\delta_n.$$

(iv) $\Pr\left(\widehat{h}\in\mathcal{H}\right)\to 1$, and $\left\|\widehat{h}-h_0\right\|_{\mathcal{H}} = o_P\left(n^{-1/4}\right)$.
(v) Stochastic Equicontinuity: for all sequences of numbers $\delta_n \to 0$,

$$\sup_{\tau\in\mathcal{T}_1}\ \sup_{|\theta-\theta_0|\leq\delta_n, \|h-h_0\|_{\mathcal{H}}\leq\delta_n} \left\|\mathbb{G}_n\psi\left(W,\theta,h,\tau\right) - \mathbb{G}_n\psi\left(W,\theta_0,h_0,\tau\right)\right\| = o_P\left(1\right).$$

(vi) $\sqrt{n}V_h\left(\theta_0,h_0,\tau\right)[\widehat{h}-h_0]$ admits an asymptotic expansion (uniformly in τ):

$$\sqrt{n}V_h\left(\theta_0,h_0,\tau\right)[\widehat{h}-h_0] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi\left(W_i,\theta_0,h_0,\tau\right) + o_P\left(1\right),$$

and the function class $\left\{s\left(w,\theta_0,h_0,\tau\right) := \psi\left(w,\theta_0,h_0,\tau\right) + \phi\left(w,\theta_0,h_0,\tau\right) : \tau\in\mathcal{T}_1\right\}$ is Donsker.

7.2 Mathematical Proofs

Proof of Theorem 3.1: The proof closely follows that of Theorem 1 in CLV. We wish to show that for any $\varepsilon > 0$ $\Pr\left(\sup_{\tau\in\mathcal{T}_1}\left|\widehat{\theta}\left(\tau\right)-\theta_0\left(\tau\right)\right| > \varepsilon\right) \to 0$ as $n\to\infty$. Let $\varepsilon > 0$ be given, the event $\left\{\sup_{\tau\in\mathcal{T}_1}\left|\widehat{\theta}\left(\tau\right)-\theta_0\left(\tau\right)\right| > \varepsilon\right\}$ implies that $\exists\tau'\in\mathcal{T}_1$ s.t. $\left|\widehat{\theta}\left(\tau'\right)-\theta_0\left(\tau'\right)\right| \geq \varepsilon$, then by Assumption A1(ii), there exists a $\eta(\varepsilon) > 0$ such that $\left\|M(\widehat{\theta}\left(\tau'\right),h_0,\tau')\right\| - \left\|M(\theta_0\left(\tau'\right),h_0,\tau')\right\| \geq \eta\left(\varepsilon\right)$. Thus,

$$\Pr\left(\sup_{\tau\in\mathcal{T}_1}\left|\widehat{\theta}\left(\tau\right)-\theta_0\left(\tau\right)\right| > \varepsilon\right) \leq \Pr\left(\sup_{\tau\in\mathcal{T}_1}\left\{\left\|M(\widehat{\theta}(\tau),h_0,\tau)\right\| - \left\|M(\theta_0\left(\tau\right),h_0,\tau)\right\|\right\} \geq \eta\left(\varepsilon\right)\right).$$

We shall prove that the right-hand side probability tends to zero as $n\to\infty$. To that end, note that by the triangle inequality, it holds

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}_1}\left\{\left\|M(\widehat{\theta}(\tau),h_0,\tau)\right\| - \left\|M(\theta_0\left(\tau\right),h_0,\tau)\right\|\right\} \\
&\leq \sup_{\tau\in\mathcal{T}_1}\left\|M(\widehat{\theta}(\tau),h_0,\tau) - M(\widehat{\theta}(\tau),\widehat{h},\tau)\right\| \\
&+ \sup_{\tau\in\mathcal{T}_1}\left\|M_n(\widehat{\theta}(\tau),\widehat{h},\tau) - M(\widehat{\theta}(\tau),\widehat{h},\tau)\right\| \\
&+ \sup_{\tau\in\mathcal{T}_1}\left\{\left\|M_n(\widehat{\theta}(\tau),\widehat{h},\tau)\right\| - \left\|M(\theta_0\left(\tau\right),h_0,\tau)\right\|\right\}.
\end{aligned}$$

By Assumption A1(v), $\sup_{\tau\in\mathcal{T}_1}\left\|M_n(\widehat{\theta}(\tau),\widehat{h},\tau) - M(\widehat{\theta}(\tau),\widehat{h},\tau)\right\| = o_P\left(1\right)$; by Assumption A1(iii) and A1(iv), it holds that $\sup_{\tau\in\mathcal{T}_1}\left\|M(\widehat{\theta}(\tau),h_0,\tau) - M(\widehat{\theta}(\tau),\widehat{h},\tau)\right\| = o_P\left(1\right)$; it remains to show the last term $\sup_{\tau\in\mathcal{T}_1}\{\left\|M_n(\widehat{\theta}(\tau),\widehat{h},\tau)\right\| - \left\|M(\theta_0\left(\tau\right),h_0,\tau)\right\|\} = o_P\left(1\right)$. Assumption A1(i) implies that uniformly in $\tau\in\mathcal{T}_1$, $\left\|M_n\left(\widehat{\theta}\left(\tau\right),\widehat{h},\tau\right)\right\| \leq \inf_{\theta\in\Theta}\left\|M_n\left(\theta,\widehat{h},\tau\right)\right\| + o_P\left(1\right)$. By triangle inequality and Assumption A1(v), uniformly in $\left(\theta,\tau\right)$,

$$\begin{aligned}
\left\|M_n\left(\theta,\widehat{h},\tau\right)\right\| &\leq \left\|M_n\left(\theta,\widehat{h},\tau\right) - M\left(\theta,\widehat{h},\tau\right)\right\| + \left\|M\left(\theta,\widehat{h},\tau\right) - M\left(\theta,h_0,\tau\right)\right\| \\
&+ \left\|M\left(\theta,h_0,\tau\right)\right\| \\
&\leq o_P\left(1\right) + \left\|M\left(\theta,h_0,\tau\right)\right\|.
\end{aligned}$$

Hence,

$$\sup_{\tau\in\mathcal{T}_1}\left\{\inf_{\theta\in\Theta}\left\|M_n\left(\theta,\widehat{h},\tau\right)\right\|-\|M(\theta_0(\tau),h_0,\tau)\|\right\}$$
$$\le o_P(1)+\sup_{\tau\in\mathcal{T}_1}\left\{\inf_{\theta\in\Theta}\|M(\theta,h_0,\tau)\|-\|M(\theta_0(\tau),h_0,\tau)\|\right\}$$
$$=o_P(1),$$

where the last equality follows from $\inf_{\theta\in\Theta}\|M(\theta,h_0,\tau)\|=\|M(\theta_0(\tau),h_0,\tau)\|$ and the definition of $\theta_0(\tau)$. Then, the result follows. ■

Proof of Theorem 3.2: (1) First show that $\left|\widehat{\theta}(\tau)-\theta_0(\tau)\right|=O_P\left(n^{-1/2}\right)$ uniformly in $\tau\in\mathcal{T}_1$. By stochastic equicontinuity, uniformly in $\tau\in\mathcal{T}_1$

$$E_n\psi_\theta\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)$$
$$=E_n\psi_\theta(W,\theta_0(\tau),h_0,\tau)+E\psi_\theta\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)-E\psi_\theta(W,\theta_0(\tau),h_0,\tau)+o_P\left(n^{-1/2}\right),$$

$$E_n\psi\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)$$
$$=E_n\psi(W,\theta_0(\tau),h_0,\tau)+E\psi\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)-E\psi(W,\theta_0(\tau),h_0,\tau)+o_P\left(n^{-1/2}\right),$$

then by noticing that $E_n\psi_\theta(W,\theta_0(\tau),h_0,\tau)-E\psi_\theta(W,\theta_0(\tau),h_0,\tau)=O_P\left(n^{-1/2}\right)$, $E_n\psi(W,\theta_0(\tau),h_0,\tau)-E\psi(W,\theta_0(\tau),h_0,\tau)=O_P\left(n^{-1/2}\right)$ and $E\psi_\theta\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)=O_P(1)$, $E\psi\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)=O_P(1)$, some algebra gives (uniformly in $\tau\in\mathcal{T}_1$)

$$o_P\left(n^{-1/2}\right)=\left(E_n\psi_\theta\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)\right)'\Xi E_n\psi\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)$$
$$=E\psi_\theta\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)'\Xi E\psi\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)+O_P\left(n^{-1/2}\right)+o_P\left(n^{-1/2}\right).$$

By Assumption A2 (iii), (iv) and (v), we have uniformly in $\tau\in\mathcal{T}_1$

$$E\psi_\theta\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)=E\psi_\theta\left(W,\widehat{\theta}(\tau),h_0,\tau\right)+V_{\theta h}\left(\widehat{\theta}(\tau),h_0,\tau\right)\left[\widehat{h}-h_0\right]+o_P\left(n^{-1/2}\right)$$
$$=E\psi_\theta(W,\theta_0(\tau),h_0,\tau)+V_{\theta\theta}(\theta_0(\tau),h_0,\tau)(\widehat{\theta}(\tau)-\theta_0(\tau))+O_P\left(\left|\widehat{\theta}-\theta_0\right|^2\right)$$
$$+V_{\theta h}(\theta_0(\tau),h_0,\tau)\left[\widehat{h}-h_0\right]+o_P\left(\left|\widehat{\theta}-\theta_0\right|\right)+o_P\left(n^{-1/2}\right),$$

$$E\psi\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)=E\psi\left(W,\widehat{\theta}(\tau),h_0,\tau\right)+V_h\left(\widehat{\theta}(\tau),h_0,\tau\right)\left[\widehat{h}-h_0\right]o_P\left(n^{-1/2}\right)$$
$$=E\psi(W,\theta_0(\tau),h_0,\tau)+V_\theta(\theta_0(\tau),h_0,\tau)(\widehat{\theta}(\tau)-\theta_0(\tau))+O_P\left(\left|\widehat{\theta}-\theta_0\right|^2\right)$$
$$+V_h(\theta_0(\tau),h_0,\tau)\left[\widehat{h}-h_0\right]+o_P\left(\left|\widehat{\theta}-\theta_0\right|\right)+o_P\left(n^{-1/2}\right).$$

Combine three display above and notice that $E\psi_\theta\left(W,\theta_0\left(\tau\right),h_0,\tau\right)'\,\Xi E\psi\left(W,\theta_0\left(\tau\right),h_0,\tau\right)=0$, we have uniformly in $\tau\in\mathcal{T}_1$

$$o_P\left(n^{-1/2}\right)=\left(A_0+B_0\right)\left(\widehat{\theta}\left(\tau\right)-\theta_0\left(\tau\right)\right)+o_P\left(\left|\widehat{\theta}-\theta_0\right|\right)+O_P\left(n^{-1/2}\right),$$

where $A_0=V_\theta\left(\theta_0\left(\tau\right),h_0,\tau\right)'\,\Xi V_\theta\left(\theta_0\left(\tau\right),h_0,\tau\right)$, $B_0=\left(M\left(\theta_0\left(\tau\right),h_0,\tau\right)'\,\Xi\otimes I_{d_\theta}\right)V_{\theta\theta}\left(\theta_0\left(\tau\right),h_0,\tau\right)$. By assumption that A_0+B_0 is of full rank, we obtain the uniform convergence rate

$$\sup_{\tau\in\mathcal{T}_1}\left|\widehat{\theta}\left(\tau\right)-\theta_0\left(\tau\right)\right|=O_P\left(n^{-1/2}\right).$$

(2) We next derive the asymptotic distribution of $\sqrt{n}\left(\widehat{\theta}\left(\tau\right)-\theta_0\left(\tau\right)\right)$.

By Assumption A2 (vi), we have uniformly in $\tau\in\mathcal{T}$

$$\begin{aligned}\sqrt{n}E_n\psi(W,\widehat{\theta},\widehat{h},\tau)&=\mathbb{G}_n\psi(W,\widehat{\theta},\widehat{h},\tau)+\sqrt{n}E[\psi(W,\widehat{\theta},\widehat{h},\tau)]\\&=\mathbb{G}_n\psi(W,\theta_0,h_0,\tau)+\sqrt{n}E[\psi(W,\widehat{\theta},\widehat{h},\tau)]+o_P(1)\\&=\mathbb{G}_n\psi(W,\theta_0,h_0,\tau)+\sqrt{n}\{E[\psi(W,\widehat{\theta},\widehat{h},\tau)-\psi(W,\theta_0,h_0,\tau)]\}\\&+\sqrt{n}M\left(\theta_0,h_0,\tau\right)+o_P(1)\\&=\mathbb{G}_n\psi(W,\theta_0,h_0,\tau)+\sqrt{n}E_n\phi(W,\theta_0,h_0,\tau)\\&+V_{\theta_0}\left(\tau\right)\sqrt{n}(\widehat{\theta}\left(\tau\right)-\theta_0\left(\tau\right))+\sqrt{n}M\left(\theta_0,h_0,\tau\right)+o_P(1),\end{aligned}$$

where the last equality follows from Assumption A2 (iii), (iv) and (vii), and the $\sqrt{n}$-consistency of $\widehat{\theta}\left(\tau\right)-\theta_0\left(\tau\right)$. We can show by similar arguments that

$$\begin{aligned}&\left(\sqrt{n}E_n\psi_\theta(W,\widehat{\theta},\widehat{h},\tau)\right)'\,\Xi M\left(\theta_0,h_0,\tau\right)\\&=\left(\sqrt{n}E[\psi_\theta(W,\widehat{\theta},\widehat{h},\tau)]\right)'\,\Xi M\left(\theta_0,h_0,\tau\right)\\&+\left(\mathbb{G}_n\psi_\theta(W,\theta_0,h_0,\tau)\right)'\,\Xi M\left(\theta_0,h_0,\tau\right)+o_P\left(1\right)\\&=\left(\sqrt{n}E_n\phi_\theta(W,\theta_0,h_0,\tau)\right)'\,\Xi M\left(\theta_0,h_0,\tau\right)\\&+\left(M\left(\theta_0,h_0,\tau\right)'\,\Xi\otimes I_{d_\theta}\right)V'_{\theta_0\theta_0}\left(\tau\right)\sqrt{n}(\widehat{\theta}\left(\tau\right)-\theta_0\left(\tau\right))\\&+\left(\mathbb{G}_n\psi_\theta(W,\theta_0,h_0,\tau)\right)'\,\Xi M\left(\theta_0,h_0,\tau\right)+o_P\left(1\right).\end{aligned}$$

Hence, by definition of $\widehat{\theta}(\tau)$, we have

$$
\begin{aligned}
o_P(1) &= E_n\psi_\theta(W,\widehat{\theta},\widehat{h},\tau)'\Xi\left(\sqrt{n}E_n\psi(W,\widehat{\theta},\widehat{h},\tau)\right) \\
&= E_n\psi_\theta(W,\widehat{\theta},\widehat{h},\tau)'\Xi\mathbb{G}_n\psi(W,\theta_0,h_0,\tau) + E_n\psi_\theta(W,\widehat{\theta},\widehat{h},\tau)'\Xi\mathbb{G}_n\phi(W,\theta_0,h_0,\tau) \\
&+ E_n\psi_\theta(W,\widehat{\theta},\widehat{h},\tau)'\Xi V_{\theta_0}(\tau)\sqrt{n}(\widehat{\theta}(\tau)-\theta_0(\tau)) \\
&+ \left(\sqrt{n}E_n\psi_\theta(W,\widehat{\theta},\widehat{h},\tau)\right)'\Xi M(\theta_0,h_0,\tau) + o_P(1) \\
&= V_{\theta_0}(\tau)'\Xi\mathbb{G}_n\psi(W,\theta_0,h_0,\tau) + V_{\theta_0}(\tau)'\Xi\mathbb{G}_n\phi(W,\theta_0,h_0,\tau) \\
&+ V_{\theta_0}(\tau)'\Xi V_{\theta_0}(\tau)\sqrt{n}(\widehat{\theta}(\tau)-\theta_0(\tau)) + o_P(1) \\
&+ \left(\sqrt{n}E_n\phi_\theta(W,\theta_0,h_0,\tau)\right)'\Xi M(\theta_0,h_0,\tau) \\
&+ \left(M(\theta_0,h_0,\tau)'\Xi\otimes I_{d_\theta}\right)V_{\theta_0\theta_0}'(\tau)\sqrt{n}(\widehat{\theta}(\tau)-\theta_0(\tau)) \\
&+ (\mathbb{G}_n\psi_\theta(W,\theta_0,h_0,\tau))'\Xi M(\theta_0,h_0,\tau) + o_P(1).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\left[V_{\theta_0}(\tau)'\Xi V_{\theta_0}(\tau) + \left(M(\theta_0,h_0,\tau)'\Xi\otimes I_{d_\theta}\right)V_{\theta_0\theta_0}(\tau)\right]\sqrt{n}(\widehat{\theta}(\tau)-\theta_0(\tau)) \\
&= -V_{\theta_0}(\tau)'\Xi\left[\mathbb{G}_n\{\psi(W,\theta_0,h_0,\tau)+\phi(W,\theta_0,h_0,\tau)\}\right] \\
&- (M(\theta_0,h_0,\tau)'\Xi\otimes I_{d_\theta})\mathbb{G}_n\{\psi_\theta^{vec}(W,\theta_0,h_0,\tau)+\phi_\theta^{vec}(W,\theta_0,h_0,\tau)\} + o_P(1)
\end{aligned}
$$

where $\psi_\theta^{vec}(W,\theta_0,h_0,\tau) := vec(\psi_\theta(W,\theta_0,h_0,\tau))$ and $\phi_\theta^{vec}(W,\theta_0,h_0,\tau) := vec(\phi_\theta(W,\theta_0,h_0,\tau))$. Let

$$
\begin{aligned}
A_0(\tau) &= V_{\theta_0}(\tau)'\Xi V_{\theta_0}(\tau) + \left(M(\theta_0,h_0,\tau)'\Xi\otimes I_{d_\theta}\right)V_{\theta_0\theta_0}(\tau), \\
B_0(\tau) &= V_{\theta_0}(\tau)'\Xi, \\
D_0(\tau) &= (M(\theta_0,h_0,\tau)'\Xi\otimes I_{d_\theta}),
\end{aligned}
$$

then

$$
\begin{aligned}
&\sqrt{n}(\widehat{\theta}(\tau)-\theta_0(\tau)) \\
&= -A_0(\tau)^{-1}[B_0(\tau)\ \ D_0(\tau)]\mathbb{G}_n\begin{pmatrix}\psi(W,\theta_0,h_0,\tau)+\phi(W,\theta_0,h_0,\tau) \\ \psi_\theta^{vec}(W,\theta_0,h_0,\tau)+\phi_\theta^{vec}(W,\theta_0,h_0,\tau)\end{pmatrix} + o_P(1).
\end{aligned}
$$

By Assumption A2 (vii), it follows that

$$\sqrt{n}(\widehat{\theta}-\theta_0) \rightsquigarrow L,$$

where L is a Gaussian Process with zero mean and covariance function

$$C(\tau_1,\tau_2) = A_0(\tau_1)^{-1}[B_0(\tau_1)\ \ D_0(\tau_1)]K(\tau_1,\tau_2)\left[B_0'(\tau_2)\ \ D_0'(\tau_2)\right]'A_0'(\tau_2)^{-1},$$

with $K(\tau_1,\tau_2) = E\left[(s(W,\theta_0,h_0,\tau_1) - Es(W,\theta_0,h_0,\tau_1))(s'(W,\theta_0,h_0,\tau_2) - Es'(W,\theta_0,h_0,\tau_1))\right]$ where

$$s(W,\theta_0,h_0,\tau) = \begin{pmatrix}\psi(W,\theta_0,h_0,\tau)+\phi(W,\theta_0,h_0,\tau) \\ \psi_\theta^{vec}(W,\theta_0,h_0,\tau)+\phi_\theta^{vec}(W,\theta_0,h_0,\tau)\end{pmatrix}.$$

■

Proof of Theorem 3.3: (1) By Assumption A3(iii), the function class is $\{s(w,\theta,h,\tau)u : \theta \in \Theta_\delta, h \in \mathcal{H}_\delta, \tau \in \mathcal{T}_1\}$ is also Donsker. Hence by stochastic equicontinuity and Assumption A3(i), it follows that uniformly in $\tau \in \mathcal{T}_1$

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{s\left(W_i,\widehat{\theta},\widehat{h},\tau\right)-E_n s\left(W_i,\widehat{\theta},\widehat{h},\tau\right)\right\}u_i \\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{s\left(W_i,\widehat{\theta},\widehat{h},\tau\right)-Es\left(W_i,\widehat{\theta},\widehat{h},\tau\right)\right\}u_i \\
&-\left\{E_n s\left(W_i,\widehat{\theta},\widehat{h},\tau\right)-Es\left(W_i,\widehat{\theta},\widehat{h},\tau\right)\right\}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_i+o_P(1) \\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{s\left(W_i,\theta_0,h_0,\tau\right)-Es\left(W_i,\theta_0,h_0,\tau\right)\right\}u_i+o_P(1).
\end{aligned}$$

Assumption A3(ii) implies

$$\sup_{\tau\in\mathcal{T}_1}\left|\widehat{A}(\widehat{\theta},\widehat{h},\tau)-A(\theta_0,h_0,\tau)\right|=o_P(1).$$

Define $Z_{0n}^*(\tau) := A(\theta_0,h_0,\tau)\,n^{-1/2}\sum_{i=1}^{n}\{s(W_i,\theta_0,h_0,\tau)-Es(W_i,\theta_0,h_0,\tau)\}u_i$, then by Slutsky's Lemma, it follows that

$$\widehat{Z}_n^*(\tau)=Z_{0n}^*(\tau)+o_P(1).$$

Theorem 2.9.6 in van der Vaart and Wellner (1996) implies that $\mathbb{G}_n\{(s(W,\theta_0,h_0,\tau)-Es(W,\theta_0,h_0,\tau))u\}$ weakly converges to the same limit process as $\mathbb{G}_n s(W,\theta_0,h_0,\tau)$ conditioning on $\{W_i\}_{i=1}^n$ almost surely, hence, $Z_{0n}^*(\tau)$ weakly converges to the same limit as $Z_n(\tau)=\sqrt{n}\left(\widehat{\theta}(\tau)-\theta_0(\tau)\right)$ conditional on $\{W_i\}_{i=1}^n$ in probability. Hence it suffices to show $\widehat{Z}_n^*(\tau)$ and $Z_{0n}^*(\tau)$ weakly converge to the same limit conditional on $\{W_i\}_{i=1}^n$ with probability approaching 1.

Let BL_1 denote the set of all real functionals on $\ell^\infty(\mathcal{T}_1)$ with a Lipschitz norm bounded by 1, i.e. for any $f\in BL_1$, $|f(z)|_{z\in\ell^\infty(\mathcal{T}_1)}\le 1$ and $z_1,z_2\in\ell^\infty(\mathcal{T}_1)$, $|f(z_1)-f(z_2)|\le\|z_1-z_2\|_{\mathcal{T}_1}$, then by Theorem 1.12.4 in van der Vaart and Wellner (1996), it is equivalent to show

$$\sup_{f\in BL_1}\left|E\left[f\left(\widehat{Z}_n^*\right)-f\left(Z_{0n}^*\right)|\{W_i\}_{i=1}^n\right]\right|\to 0.$$

By definition, for any $\eta>0$,

$$\begin{aligned}
&\sup_{f\in BL_1}\left|E\left[f\left(\widehat{Z}_n^*\right)-f\left(Z_{0n}^*\right)|\{W_i\}_{i=1}^n\right]\right| \\
&\le\eta\Pr\left(\left\|\widehat{Z}_n^*-Z_{0n}^*\right\|_{\mathcal{T}_1}\le\eta|\{W_i\}_{i=1}^n\right)+2\Pr\left(\left\|\widehat{Z}_n^*-Z_{0n}^*\right\|_{\mathcal{T}_1}>\eta|\{W_i\}_{i=1}^n\right).
\end{aligned}$$

By law of iterated expectations,

$$\begin{aligned}
&E\left[\Pr\left(\left\|\widehat{Z}_n^*-Z_{0n}^*\right\|_{\mathcal{T}_1}>\eta|\{W_i\}_{i=1}^n\right)\right] \\
&=\Pr\left(\left\|\widehat{Z}_n^*-Z_{0n}^*\right\|_{\mathcal{T}_1}>\eta\right)\to 0,
\end{aligned}$$

hence $\Pr\left(\left\|\widehat{Z}_n^* - Z_{0n}^*\right\|_{\mathcal{T}_1} > \eta | \{W_i\}_{i=1}^n\right) \to 0$ in probability. Since η is arbitrary, we have

$$\sup_{f\in BL_1}\left|E\left[f\left(\widehat{Z}_n^*\right) - f\left(Z_{0n}^*\right) | \{W_i\}_{i=1}^n\right]\right| = o_P(1).$$

We now prove (2). The conditional mean of $\widehat{Z}_n^*(\tau)$ is 0, hence by independence of $\{u_i\}_{i=1}^n$, uniformly in $\tau \in \mathcal{T}_1$,

$$\begin{aligned} Var^*\left(\widehat{Z}_n^*(\tau)\right) &= E\left[\widehat{Z}_n^*(\tau)^2 | \{W_i\}_{i=1}^n\right] \\ &= \widehat{A}(\widehat{\theta},\widehat{h},\tau)\{\frac{1}{n}\sum_{i=1}^n \overline{s}\left(W_i,\widehat{\theta},\widehat{h},\tau\right)\overline{s}\left(W_i,\widehat{\theta},\widehat{h},\tau\right)'\}\widehat{A}(\widehat{\theta},\widehat{h},\tau)' + o_P(1) \\ &= A(\theta_0,h_0,\tau)E\left[\overline{s}(W_i,\theta_0,h_0,\tau)\overline{s}(W_i,\theta_0,h_0,\tau)'\right]A(\theta_0,h_0,\tau)' + o_P(1), \end{aligned}$$

where $\overline{s}(W_i,\theta,h,\tau) := s(W_i,\theta,h,\tau) - Es(W_i,\theta,h,\tau)$, the two equalities are due to uniform law of large number and continuity of $s(\cdot,\theta,h,\tau)$. ■

Proof of Lemma 4.1:
We apply Theorem 3.1 in Chernozhukov et al. (2007). Assumption A2 and Theorem 3.2 imply that their Condition C.1 holds with

$$Q_n(\tau) = \left\|M_n\left(\widehat{\theta}(\tau), \widehat{h}(\cdot,\widehat{\theta}(\tau),\tau),\tau\right)\right\|,$$

Q the limit of Q_n and $a_n = b_n = n$. Similarly, our Theorem 3.2 implies that

$$\sup_{\tau\in\mathcal{T}_0} nQ_n(\tau) = O_P(1).$$

Then any $\hat{c}$ satisfying the conditions of the Lemma also satisfy the conditions required in Theorem 3.1 in Chernozhukov et al. (2007), which completes the proof. ■

Proof of Corollary 4.1:
By definition, $d_H\left(\widetilde{\Theta}_0,\Theta_0\right) = \max\left\{\sup_{\tau\in\hat{\mathcal{T}}_0} d_H(\widehat{\theta}(\tau),\Theta_0), \sup_{\tau\in\hat{\mathcal{T}}_0} d_H(\theta_0(\tau),\widetilde{\Theta}_0)\right\}$. Define the set $\widetilde{\Theta}_0 := \left\{\theta_0(\tau) : \tau \in \hat{\mathcal{T}}_0\right\}$, then by triangular inequality,

$$d_H\left(\widehat{\Theta}_0,\Theta_0\right) \le d_H\left(\widehat{\Theta}_0,\widetilde{\Theta}_0\right) + d_H\left(\widetilde{\Theta}_0,\Theta_0\right).$$

Since $d_H\left(\hat{\mathcal{T}}_0,\mathcal{T}_1\right) = o_P(1)$, it follows $\Pr\left(\hat{\mathcal{T}}_0 \subset \mathcal{T}_1\right) \to 1$. Notice that for each $\tau \in \hat{\mathcal{T}}_0 \subset \mathcal{T}_1$, $d_H(\widehat{\theta}(\tau),\widetilde{\Theta}_0) \le \left|\widehat{\theta}(\tau) - \theta_0(\tau)\right|$, hence $\sup_{\tau\in\hat{\mathcal{T}}_0} d_H(\widehat{\theta}(\tau),\Theta_0) \le \sup_{\tau\in\hat{\mathcal{T}}_0}\left|\widehat{\theta} - \theta_0\right|$, then Theorem 1 implies $d_H\left(\widehat{\Theta}_0,\widetilde{\Theta}_0\right) = o_P(1)$. Equicontinuity assumption of $\theta_0(\cdot)$ implies $d_H\left(\widetilde{\Theta}_0,\Theta_0\right) = o_P(1)$ as $d_H\left(\hat{\mathcal{T}}_0,\mathcal{T}_1\right) = o_P(1)$. Thus, $d_H\left(\widetilde{\Theta}_0,\Theta_0\right) = o_P(1)$ holds. ■

Lemma A1: Assume $d_H\left(\hat{\mathcal{T}}_0,\mathcal{T}_1\right) = o_P(1)$, and $Z_n(\tau)$ is stochastic equicontinuous for all $\tau \in \mathcal{T}_1$, then

$$\left|\sup_{\tau\in\hat{\mathcal{T}}_0} Z_n(\tau) - \sup_{\tau\in\mathcal{T}_1} Z_n(\tau)\right| = o_P(1).$$

Proof: See proof of Lemma 1 in Chernozhukov, Lee and Rosen (2011). ■

Proof of Proposition 4.1:
By definition of $CR_{1-\alpha,n}$ and Θ_0, let $\widetilde{CR}_{1-\alpha,n} := \bigcup_{\tau\in\mathcal{T}_1}\left\{\theta\in\Theta : \left|\theta-\widehat{\theta}_n(\tau)\right| \le \widehat{c}_{1-\alpha,n}/\sqrt{n}\right\}$, we have

$$
\begin{aligned}
&\Pr\left(\Theta_0 \subset CR_{1-\alpha,n}\right)\\
&= \Pr\left(\{\Theta_0 \subset CR_{1-\alpha,n}\} \cap \{\mathcal{T}_1 \subset \hat{\mathcal{T}}_0\}\right) + \Pr\left(\{\Theta_0 \subset CR_{1-\alpha,n}\} \cap \{\mathcal{T}_1 \nsubseteq \hat{\mathcal{T}}_0\}\right)\\
&\ge \Pr\left(\{\Theta_0 \subset CR_{1-\alpha,n}\} \cap \{\mathcal{T}_1 \subset \hat{\mathcal{T}}_0\}\right)\\
&\ge \Pr\left(\{\Theta_0 \subset \widetilde{CR}_{1-\alpha,n}\} \cap \{\mathcal{T}_1 \subset \hat{\mathcal{T}}_0\}\right)\\
&= \Pr\left(\left\{\forall\tau\in\mathcal{T}_1,\ \exists\tau'\in\hat{\mathcal{T}}_0,\ \left|\theta_0(\tau)-\widehat{\theta}_n(\tau')\right| \le \widehat{c}_{1-\alpha,n}/\sqrt{n}\right\}\right) + o_P(1)\\
&\ge \Pr\left(\left\{\forall\tau\in\mathcal{T}_1,\ \left|\theta_0(\tau)-\widehat{\theta}_n(\tau)\right| \le \widehat{c}_{1-\alpha,n}/\sqrt{n}\right\}\right) + o_P(1)\\
&= \Pr\left(\sqrt{n}\sup_{\tau\in\mathcal{T}_1}\left|\widehat{\theta}_n(\tau)-\theta_0(\tau)\right| \le \widehat{c}_{1-\alpha,n}\right) + o_P(1),
\end{aligned}
$$

where the second equality follows from the assumption $\Pr\left(\mathcal{T}_1 \subset \hat{\mathcal{T}}_0\right) \to 1$. Since $\sqrt{n}\left|\widehat{\theta}_n(\cdot)-\theta_0(\cdot)\right|$ is stochastic equicontinuous by Theorem 3.2, then according to Lemma A1, $\sqrt{n}\sup_{\tau\in\mathcal{T}_1}\left|\widehat{\theta}_n(\tau)-\theta_0(\tau)\right|$ and $\sqrt{n}\sup_{\tau\in\hat{\mathcal{T}}_0}\left|\widehat{\theta}_n(\tau)-\theta_0(\tau)\right|$ have the same asymptotic distribution, hence

$$
\begin{aligned}
&\liminf \Pr\left(\sqrt{n}\sup_{\tau\in\mathcal{T}_1}\left|\widehat{\theta}_n(\tau)-\theta_0(\tau)\right| \le \widehat{c}_{1-\alpha,n}\right)\\
&= 1-\alpha.
\end{aligned}
$$

■

Proof of Theorem 4.2: We first show the convergence rate of the minimizer in (i), and then prove the theorem in (ii).

(i) Let $\widetilde{\tau}_n \in \arg\min_{\tau\in\mathcal{T}}\left\|M_n\left(\widetilde{\theta},\widehat{h},\tau\right)\right\|$ be any fixed minimizer, then we first show that the Hausdorff distance between $\{\widetilde{\tau}_n\}$ and $\widetilde{\mathcal{T}}$ is of order $O_P\left(n^{-1/2}\right)$, i.e. $\inf_{\widetilde{\tau}\in\widetilde{\mathcal{T}}}|\widetilde{\tau}_n-\widetilde{\tau}| = O_P\left(n^{-1/2}\right)$, which takes two steps of proofs. At the first step, we show $\inf_{\widetilde{\tau}\in\widetilde{\mathcal{T}}}|\widetilde{\tau}_n-\widetilde{\tau}| \le O_P\left(n^{-1/4}\right)$. For this purpose, let $\eta_{n1} := O_P\left(n^{-1/4}\right)$ be a positive sequence, and define $\widetilde{\mathcal{T}}_{\eta_{n1}} := \{\tau\in\mathcal{T} : \exists\widetilde{\tau}\in\widetilde{\mathcal{T}} \text{ s.t. } |\tau-\widetilde{\tau}| \le \eta_{n1}\}$ which is the η_{n1}-enlargement of $\widetilde{\mathcal{T}}$, then $\Delta_{n1} := \inf_{\tau\in\mathcal{T}\setminus\widetilde{\mathcal{T}}_{\eta_{n1}}}\left\|M\left(\widetilde{\theta},\widetilde{h}_0,\tau\right)\right\| \ge c\eta_{n1}$ by Assumption A4(i). For any $\widetilde{\tau} \in \widetilde{\mathcal{T}}$, by Assumption A4(iii)(iv)(v) and the fact that $\left\|M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n)\right\| \leq \left\|M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau})\right\|$,

$$
\begin{aligned}
\left\|M\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right)\right\| &= \{\left\|M\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right)\right\| - \left\|M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n)\right\|\} + \{\left\|M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n)\right\| - \left\|M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n)\right\|\} \\
&+ \{\left\|M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n)\right\| - \left\|M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau})\right\|\} + \{\left\|M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau})\right\| - \left\|M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau})\right\|\} \\
&+ \{\left\|M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau})\right\| - \left\|M(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau})\right\|\} \\
&\leq \left\|V_h\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right)[\widehat{h} - \widetilde{h}_0]\right\| + \left\|V_h\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)[\widehat{h} - \widetilde{h}_0]\right\| + O_P\left(n^{-1/2}\right) \\
&= o_P\left(n^{-1/4}\right).
\end{aligned}
$$

Hence, $\Pr\left(M\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right) < \Delta_{n1}\right) \to 1$, which implies $\widetilde{\tau}_n \in \widetilde{\mathcal{T}}_{\eta_{n1}}$, i.e. $\inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} |\widetilde{\tau}_n - \widetilde{\tau}| \leq \eta_{n1} = O_P\left(n^{-1/4}\right)$. For fixed $\widetilde{\tau}_n \in \widetilde{\mathcal{T}}_{\eta_{n1}}$, there exists a $\widetilde{\tau} \in \widetilde{\mathcal{T}}$, such that $|\widetilde{\tau}_n - \widetilde{\tau}| \leq O_P\left(n^{-1/4}\right)$. Then at the second step, we will make similar arguments but with different auxiliary converging sequences.

Notice that $\sup_{\tau,h}\left\|M_n\left(\widetilde{\theta}, h, \tau\right) - M\left(\widetilde{\theta}, h, \tau\right)\right\| = O_P\left(n^{-1/2}\right)$, let $\eta_{n2} := 2c^{-1}l_n n^{-1/2}$ for some positive sequence $\{l_n\}$ such that $l_n = o(\sqrt{n})$ and $\sqrt{n}\sup_{\tau,h}\left\|M_n\left(\widetilde{\theta}, h, \tau\right) - M\left(\widetilde{\theta}, h, \tau\right)\right\| \leq l_n$ with probability approaching 1. Define $\widetilde{\mathcal{T}}_{\eta_{n2}} := \{\tau \in \mathcal{T} : \exists \widetilde{\tau} \in \widetilde{\mathcal{T}} \text{ s.t. } |\tau - \widetilde{\tau}| \leq \eta_{n2}\}$, i.e. the η_{n2}-enlargement of $\widetilde{\mathcal{T}}$, and $\Delta_{n2} := \inf_{\tau \in \mathcal{T} \setminus \widetilde{\mathcal{T}}_{\eta_{n2}}}\left\|M\left(\widetilde{\theta}, \widetilde{h}_0, \tau\right)\right\| \geq c\eta_{n2}$ by Assumption A4(i). Hence it suffices to show $\Pr\left(M\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right) \leq \Delta_{n2}\right) \to 1$ which implies $\Pr\left(\inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} |\widetilde{\tau}_n - \widetilde{\tau}| \leq \eta_{n2}\right) \to 1$. According to Assumption A4(ii),

$$
\begin{aligned}
M\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right) &= (M\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right) - M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n)) + M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n) \\
&= -V_h\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right)(\widehat{h} - \widetilde{h}_0) + M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n) + o_P\left(n^{-1/2}\right),
\end{aligned}
$$

and by definition of $\widetilde{\tau}_n$, $\left\|M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n)\right\| \leq \left\|M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau})\right\|$, and Assumption A4(v)

$$
\begin{aligned}
&\left\|M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n) - V_h\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right)(\widehat{h} - \widetilde{h}_0)\right\| \\
&\leq \left\|M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n) - M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n)\right\| + (\left\|M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n)\right\| - \left\|M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau})\right\|) \\
&+ \left\|M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}) - M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau})\right\| + \left\|M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}) - V_h\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right)(\widehat{h} - \widetilde{h}_0)\right\| \\
&\leq \left|M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n) - M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}_n)\right| + \left|M_n(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}) - M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau})\right| + M(\widetilde{\theta}, \widehat{h}, \widetilde{\tau}) \\
&\leq 2l_n n^{-1/2} + \left\|V_h\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)(\widehat{h} - \widetilde{h}_0) - V_h\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right)(\widehat{h} - \widetilde{h}_0)\right\| + o_P\left(n^{-1/2}\right).
\end{aligned}
$$

Hence, by Assumption A4(iii)

$$
\begin{aligned}
\left\|M\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right)\right\| &\leq \left\|V_h\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)(\widehat{h} - \widetilde{h}_0) - V_h\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right)(\widehat{h} - \widetilde{h}_0)\right\| \\
&+ 2l_n n^{-1/2} + o_P\left(n^{-1/2}\right) \\
&\leq 2l_n n^{-1/2} + o_P\left(n^{-1/2}\right),
\end{aligned}
$$

which implies $\Pr\left(M\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right) < \Delta_{n2}\right) \to 1$. Since the convergence of l_n to ∞ can be arbitrarily slow, $\{l_n\}$ is essentially $O_P(1)$ sequence, we obtain $\inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} |\widetilde{\tau}_n - \widetilde{\tau}| = O_P\left(n^{-1/2}\right)$.

(ii) Let $\delta_n = O\left(n^{-1/2}\right)$ be a positive sequence, and define the neighborhood $B\left(\widetilde{\tau}, \delta_n\right) := \{\tau \in \mathcal{T} : |\tau - \widetilde{\tau}| \leq \delta_n\}$ for all $\widetilde{\tau} \in \widetilde{\mathcal{T}}$. Then

$$\begin{aligned} &\inf_{\tau \in \mathcal{T}} \left\| \sqrt{n} M_n(\widetilde{\theta}, \widehat{h}, \tau) \right\| \\ &= \inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \inf_{\tau \in B(\widetilde{\tau}, \delta_n)} \left\| \sqrt{n} M_n(\widetilde{\theta}, \widehat{h}, \tau) \right\| + o_P(1) \end{aligned} \tag{21}$$

By stochastic equicontinuity, $\left\| \widehat{h} - \widetilde{h}_0 \right\|_{\mathcal{H}} = o_P\left(n^{-1/4}\right)$ and $|\tau - \widetilde{\tau}| \leq \delta_n$, it holds that

$$\begin{aligned} &\sqrt{n} M_n(\widetilde{\theta}, \widehat{h}, \tau) \\ &= \sqrt{n} M_n(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}) + \sqrt{n} M(\widetilde{\theta}, \widehat{h}, \tau) + o_P(1), \end{aligned}$$

hence continuing with (21),

$$\begin{aligned} &\inf_{\tau \in \mathcal{T}} \left\| \sqrt{n} M_n\left(\widetilde{\theta}, \widehat{h}, \tau\right) \right\| \\ &= \inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \inf_{\tau \in B(\widetilde{\tau}, \delta_n)} \left\| \sqrt{n} M_n(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}) + \sqrt{n} M(\widetilde{\theta}, \widehat{h}, \tau) \right\| + o_P(1). \end{aligned}$$

By Assumption A4(ii) and (iii), and $M(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}) = 0$, we have

$$\begin{aligned} &M(\widetilde{\theta}, \widehat{h}, \tau) \\ &= V_h\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)[\widehat{h} - \widetilde{h}_0] + V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)(\tau - \widetilde{\tau}) + o_P\left(n^{-1/2}\right). \end{aligned}$$

Then, together with Assumption A4(iv),

$$\begin{aligned} &\inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \inf_{\tau \in B(\widetilde{\tau}, \delta_n)} \left\| \sqrt{n} M_n(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}) + \sqrt{n} M(\widetilde{\theta}, \widehat{h}, \tau) \right\| \\ &= \inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \inf_{\tau \in B(\widetilde{\tau}, \delta_n)} \left\| \begin{array}{c} \sqrt{n} M_n(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}) + \sqrt{n} V_h\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)[\widehat{h} - \widetilde{h}_0] \\ + V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)\sqrt{n}(\tau - \widetilde{\tau}) \end{array} \right\| + o_P(1) \\ &= \inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \inf_{\tau \in B(\widetilde{\tau}, \delta_n)} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n s\left(W_i, \widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right) + V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)\sqrt{n}(\tau - \widetilde{\tau}) \right\| + o_P(1). \end{aligned}$$

Note that $|\tau - \widetilde{\tau}| \leq \delta_n$ and $\delta_n = O\left(n^{-1/2}\right)$, then

$$\begin{aligned} &\inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \inf_{\tau \in B(\widetilde{\tau}, \delta_n)} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n s\left(W_i, \widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right) + V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)\sqrt{n}(\tau - \widetilde{\tau}) \right\| \\ &= \inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \inf_{\gamma \in \mathbb{R}^{d_\tau}} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n s\left(W_i, \widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right) + V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)\gamma \right\|. \end{aligned}$$

Denote $S_n\left(\widetilde{\tau}\right) := \frac{1}{\sqrt{n}} \sum_{i=1}^n s\left(W_i, \widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)$, then $\inf_{\gamma \in \mathbb{R}^{d_\tau}} \left\| S_n\left(\widetilde{\tau}\right) + V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)\gamma \right\|$ is the classical weighted least square problem and has a closed-form solution. Let $[V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)' \Xi V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)]^-$ be the pseudoinverse of $V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)' \Xi V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)$ if it is not invertible, which equals $[V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)' \Xi V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)]^{-1}$

if it is invertible, then we can obtain (denote $V_\tau(\widetilde{\tau}) := V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)$):

$$
\begin{aligned}
&\inf_{\gamma \in \mathbb{R}^{d_\tau}} \|S_n(\widetilde{\tau}) + V_\tau(\widetilde{\tau})\gamma\| \\
&= \left\|(I - V_\tau(\widetilde{\tau})[V_\tau(\widetilde{\tau})' \Xi V_\tau(\widetilde{\tau})]^- V_\tau(\widetilde{\tau})' \Xi) S_n(\widetilde{\tau})\right\| \\
&= \left\| P\left(\widetilde{\theta}\right) \frac{1}{\sqrt{n}} \sum_{i=1}^n s\left(W_i, \widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right) \right\|,
\end{aligned}
$$

where $P\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right) := I - V_\tau(\widetilde{\tau})[V_\tau(\widetilde{\tau})' \Xi V_\tau(\widetilde{\tau})]^{-1} V_\tau(\widetilde{\tau})' \Xi$. Since

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n s\left(W_i, \widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right) \rightsquigarrow G(\widetilde{\tau}) \text{ in } \widetilde{\mathcal{T}},
$$

where $G(\widetilde{\tau})$ is a Gaussian process on $\widetilde{\mathcal{T}}$, then if $rank(V_\tau\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)) < d_\psi$, by Theorem 1.11.1 in van der Vaart and Wellner (1996),

$$
\begin{aligned}
&\inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \left\| P\left(\widetilde{\theta}\right) \frac{1}{\sqrt{n}} \sum_{i=1}^n s\left(W_i, \widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right) \right\| \\
&\rightsquigarrow \inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \left\| P\left(\widetilde{\theta}\right) G(\widetilde{\tau}) \right\|,
\end{aligned}
$$

which leads to our conclusion that

$$
T_n \rightsquigarrow \inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \left\| P\left(\widetilde{\theta}\right) G(\widetilde{\tau}) \right\|.
$$

■

Proof of Proposition 4.2: (1) Under the null, Lemma 17 in Santos (2010) indicates

$$
\inf_{\tau \in \mathcal{T}} R_n^*(\tau) = \inf_{\tau \in \widetilde{\mathcal{T}}} \left\| \widehat{P}\left(\widetilde{\theta}\right) \frac{1}{\sqrt{n}} \sum_{i=1}^n s\left(W_i, \widetilde{\theta}, \widehat{h}, \tau\right) u_i \right\| + o_{P^*}(1).
$$

Then the rest of the proof follows directly from proofs of Theorem 3.3.

(2) Under the alternative, $T_n = O_P\left(n^{1/2}\right)$, while $T_n^* = O_P(\lambda_n) = o_P\left(n^{1/2}\right)$, hence the conclusion follows. ■.

Proof of Theorem 4.3:

By stochastic equicontinuity in Assumption A5(vii), we have

$$
\begin{aligned}
&\sqrt{n}\{E_n \xi(W, \widehat{g}, \tau) - \tau' \widetilde{\theta}\} \\
&= \sqrt{n}\{E_n \xi(W, g_0, \tau) - E\xi(W, g_0, \tau)\} + \sqrt{n}\{E\xi(W, \widehat{g}, \tau) - E\xi(W, g_0, \tau)\} \\
&+ \sqrt{n}\{E\xi(W, g_0, \tau) - \tau' \widetilde{\theta}\} + o_P(1).
\end{aligned}
$$

Notice that by Assumption ()

$$
\begin{aligned}
&\sqrt{n}\{E\xi(W, \widehat{g}, \tau) - E\xi(W, g_0, \tau)\} \\
&= \sqrt{n} V_g^c(g_0, \tau)[\widehat{g} - g_0] + o_P(1),
\end{aligned}
$$

hence,

$$\begin{aligned}
&\sqrt{n}\{E_n\xi\left(W,\widehat{g},\tau\right)-\tau'\widetilde{\theta}\} \\
&=\frac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n}\left\{\xi\left(W_i,g_0,\tau\right)-E\xi\left(W_i,g_0,\tau\right)+\zeta\left(W_i,g_0,\tau\right)\right\} \\
&+\sqrt{n}\{E\xi\left(W,g_0,\tau\right)-\tau'\widetilde{\theta}\}+o_P\left(1\right).
\end{aligned} \tag{22}$$

(1) For any fixed $\widetilde{\theta}\in int\left(\Theta_0\right)$, it holds that $\inf_{\tau\in\mathbb{S}^{d_\theta}}\{E\xi\left(W,g_0,\tau\right)-\tau'\widetilde{\theta}\}>0$; notice that the first term is $O_P\left(1\right)$, hence the second term dominates and we have

$$\inf_{\tau\in\mathbb{S}^{d_\theta}}\sqrt{n}\{E_n\xi\left(W,\widehat{g},\tau\right)-\tau'\widetilde{\theta}\}\xrightarrow{P}+\infty.$$

(2) If $\widetilde{\theta}\notin\Theta_0$, then $\inf_{\tau\in\mathbb{S}^{d_\theta}}\{E\xi\left(W,g_0,\tau\right)-\tau'\widetilde{\theta}\}<0$, hence

$$\inf_{\tau\in\mathbb{S}^{d_\theta}}\sqrt{n}\{E_n\xi\left(W,\widehat{g},\tau\right)-\tau'\widetilde{\theta}\}\xrightarrow{P}-\infty.$$

(3) If $\widetilde{\theta}$ is on the boundary of Θ_0, i.e. $\widetilde{\theta}\in\partial\Theta_0$, notice $\widetilde{\mathcal{S}}=\left\{\tau\in\mathbb{S}^{d_\theta}:E\xi\left(W,g_0,\tau\right)=\tau'\widetilde{\theta}\right\}$, then $\widetilde{\mathcal{S}}\neq\varnothing$. For this case, we modify the proof of Theorem 1 in Galichon and Henry (2009). Let $G_n\left(\tau\right):=\frac{1}{n}\sum_{i=1}^{n}\left\{\xi\left(W_i,g_0,\tau\right)-E\xi\left(W_i,g_0,\tau\right)+\zeta\left(W_i,g_0,\tau\right)\right\}$, define

$$\begin{aligned}
\widetilde{\mathcal{S}}&=\left\{\tau\in\mathbb{S}^{d_\theta}:E\xi\left(W,g_0,\tau\right)-\tau'\widetilde{\theta}\leq b\right\}, \\
\widetilde{\mathcal{S}}_{n,b}&=\left\{\tau\in\mathbb{S}^{d_\theta}:G_n\left(\tau\right)+E\xi\left(W,g_0,\tau\right)-\tau'\widetilde{\theta}\leq b\right\},
\end{aligned}$$

suppose there exists a positive sequence $\{b_n\}$ satisfying $b_n\ln\ln n+b_n^{-1}\sqrt{\ln\ln n/n}\to 0$, we first show that $d_H\left(\widetilde{\mathcal{S}},\widetilde{\mathcal{S}}_{n,b_n}\right)=o_P\left(1\right)$. Notice that $b_n\sqrt{n}\to\infty$, we have

$$\begin{aligned}
\Pr\left(\widetilde{\mathcal{S}}\subset\widetilde{\mathcal{S}}_{n,b_n}\right)&=\Pr\left(\sup_{\tau\in\widetilde{\mathcal{S}}}\sqrt{n}G_n\left(\tau\right)\leq b_n\sqrt{n}\right) \\
&\to 1,
\end{aligned}$$

it suffice to show $\sup_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}d\left(\tau,\widetilde{\mathcal{S}}\right)=o_P\left(1\right)$. For any $\varepsilon>0$, denote the ε-enlargement of $\widetilde{\mathcal{S}}$ as $\widetilde{\mathcal{S}}^{\varepsilon}$, then by definition of $\widetilde{\mathcal{S}}$, there exists $\eta\left(\varepsilon\right)>0$ such that $\inf_{\tau\in\mathbb{S}^{d_\theta}\backslash\widetilde{\mathcal{S}}^{\varepsilon}}\left\{E\xi\left(W,g_0,\tau\right)-\tau'\widetilde{\theta}\right\}>\eta\left(\varepsilon\right)$; and by definition of $\widetilde{\mathcal{S}}_{n,b_n}$, we have

$$\begin{aligned}
&\sup_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}\left\{E\xi\left(W,g_0,\tau\right)-\tau'\widetilde{\theta}\right\} \\
&\leq\sup_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}\left\{b_n-G_n\left(\tau\right)\right\} \\
&=b_n-\inf_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}G_n\left(\tau\right)=o_P\left(1\right).
\end{aligned}$$

Thus $\widetilde{\mathcal{S}}_{n,b_n}\cap\left(\mathbb{S}^{d_\theta}\backslash\widetilde{\mathcal{S}}^{\varepsilon}\right)$ is empty with probability approaching 1, implying $\widetilde{\mathcal{S}}_{n,b_n}\subset\widetilde{\mathcal{S}}^{\varepsilon}$. Since ε is arbitrary, we have $\sup_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}d\left(\tau,\widetilde{\mathcal{S}}\right)=o_P\left(1\right)$.

We next show that with probability approaching one,

$$\inf_{\tau\in\widetilde{\mathcal{S}}}\sqrt{n}G_n(\tau)\geq\inf_{\tau\in\mathbb{S}^{d_\theta}}T_n^0(\tau)\geq\inf_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}\sqrt{n}G_n(\tau),$$

where $T_n^0(\tau):=\sqrt{n}G_n(\tau)+\sqrt{n}\{E\xi(W,g_0,\tau)-\tau'\widetilde{\theta}\}$. It follows that

$$\begin{aligned}\inf_{\tau\in\widetilde{\mathcal{S}}}\sqrt{n}G_n(\tau)&=_{(i)}\inf_{\tau\in\widetilde{\mathcal{S}}}T_n^0(\tau)\\&\geq_{(ii)}\inf_{\tau\in\mathbb{S}^{d_\theta}}T_n^0(\tau)\\&=\min\left\{\inf_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}T_n^0(\tau),\inf_{\tau\in\mathbb{S}^{d_\theta}\backslash\widetilde{\mathcal{S}}_{n,b_n}}T_n^0(\tau)\right\}\\&\geq_{(iii)}\min\left\{\inf_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}\sqrt{n}G_n(\tau),\inf_{\tau\in\mathbb{S}^{d_\theta}\backslash\widetilde{\mathcal{S}}_{n,b_n}}T_n^0(\tau)\right\}\\&\geq_{(iv)}\inf_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}\sqrt{n}G_n(\tau),\end{aligned}$$

where (i) is due to the definition of $\widetilde{\mathcal{S}}$; (ii) follows from the fact that $\widetilde{\mathcal{S}}\subset\mathbb{S}^{d_\theta}$; (iii) is due to the fact $E\xi(W,g_0,\tau)-\tau'\widetilde{\theta}\geq 0$ for all τ, hence $\inf_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}T_n^0(\tau)\geq\inf_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}\sqrt{n}G_n(\tau)$; by definition of $\widetilde{\mathcal{S}}_{n,b_n}$, for any $\tau\in\mathbb{S}^{d_\theta}\backslash\widetilde{\mathcal{S}}_{n,b_n}$, $T_n^0(\tau)>b_n\sqrt{n}$, hence $\inf_{\tau\in\mathbb{S}^{d_\theta}\backslash\widetilde{\mathcal{S}}_{n,b_n}}T_n^0(\tau)\geq b_n\sqrt{n}>\inf_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}\sqrt{n}G_n(\tau)$ with probability approaching 1.

Lastly, by Lemma A1 (multiply by -1), we have

$$\left|\inf_{\tau\in\widetilde{\mathcal{S}}}\sqrt{n}G_n(\tau)-\inf_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}}\sqrt{n}G_n(\tau)\right|=o_P(1),$$

which implies that $\inf_{\tau\in\mathbb{S}^{d_\theta}}T_n^0(\tau)=\inf_{\tau\in\widetilde{\mathcal{S}}}\sqrt{n}G_n(\tau)+o_P(1)$. Hence by (22),

$$\begin{aligned}T_n^c(\widetilde{\theta})&=\inf_{\tau\in\widetilde{\mathcal{S}}}\sqrt{n}G_n(\tau)+o_P(1)\\&\rightsquigarrow\inf_{\tau\in\widetilde{\mathcal{S}}}\{G^\dagger(\tau)\}.\end{aligned}$$

■

References

[1] Ai, C. and X. Chen (2003): "Efficient Estimation of Conditional Moment Restrictions Models Containing Unknown Functions," *Econometrica,* 71, 1795–1843.

[2] Altonji, J. G., T. E. Elder, and C. R. Taber, (2005): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 13, 151-84.

[3] An, M.Y. (2000): "A semiparametric distribution for willingness to pay and statistical inference with dichotomous choice CV data," *The American Journal of Agricultural Economics,* 82, 487–500.

[4] Andrews, D. W. K. (1994): "Empirical process method in econometrics," *Handbook of Econometrics*, vol. IV, North Holland.

[5] Andrews, D. W. K. (1995): "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11, 560–596.

[6] Andrews, D. W. K., and P. Guggenberger (2009): "Validity of Subsampling and 'Plug-In Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669-709.

[7] Andrews, D. W. K. and P. Jia (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure", *Econometrica*, 80, 2805–2826.

[8] Andrews, D.W. K., and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica*, 62, 1383-1414.

[9] Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119-157.

[10] Angrist, J., V. Chernozhukov, and I. Fernandez-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure", *Econometrica,* 74, 539-563.

[11] Arellano, M., L. P. Hansen and E. Sentana (2012): "Underidentification?", *Journal of Econometrics,* 170, 256-280.

[12] Beresteanu, A. and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models", *Econometrica*, 76, 763-814.

[13] Bontemps, C., T. Magnac and E. Maurin (2012), "Set Identified Linear Models," *Econometrica*, 80, 1129–1155.

[14] Bugni, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set", *Econometrica*, 78, 735-753.

[15] Chen, X., O. Linton and I. Van Keilegom (2003): "Estimation of Semiparametric Models When the Criterion Function is Not Smooth", *Econometrica*, 71, 1591-1608.

[16] Chen, X., E. Tamer and A. Torgovitsky (2011): "Sensitivity Analysis in Partially Identified Semiparametric Likelihood Models", Working paper.

[17] Chernozhukov, V., Hong, H. and Tamer, E. (2007): "Inference on Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284.

[18] Chernozhukov, V., S. Lee and A. M. Rosen (2011): "Intersection bounds: estimation and inference," Working paper.

[19] Chernozhukov, V., R. Rigobon and T. M. Stoker (2010): "Set identification and sensitivity analysis with Tobin regressors," *Quantitative Economics*, 1, 255-277.

[20] Chesher, A. (2010): "Instrumental Variable Models for Discrete Outcomes," *Econometrica*, 78, 575-601.

[21] Conley, T. G., Hansen, C, B. and Rossi, P. E. (2012): "Plausibly Exogenous," *Review of Economics and Statistics*, 94, 260-72.

[22] Escanciano, J. C., D. T. Jacho-Chávez and A. Lewbel (2011): "Uniform Convergence of Weighted Sums of Non- and Semi-parametric Residuals for Estimation and Testing," Working paper.

[23] Escanciano, J. C., D. T. Jacho-Chávez and A. Lewbel (2012): "Identiciation and Estimation of Semiparametric Two Step Models," Working paper.

[24] Escanciano, J. C. and L. Zhu (2011): "Inferences in Semiparametric Partially Identified Models: An Empirical Process Approach," Working paper.

[25] Galichon, A., and M. Henry (2009): "A test of non-identifying restrictions and confidence regions for partially identified parameters," *Journal of Econometrics*, 152, 186–196.

[26] Green D., Jacowitz K. E., Kahneman D. and McFadden, D. (1998): "Referendum contingent valuation, anchoring, and willingness to pay for public goods", *Resource and Energy Economics*, 20, 85–116.

[27] Giné, E. and J. Zinn (1990): "Bootstrapping General Empirical Measures," *The Annals of Probability*, 18, 851-869.

[28] Hamilton, J.D., D.F. Waggoner and T. Zha, (2007), "Normalization in Econometrics," *Econometric Reviews,* 26, 221-252.

[29] Hanemann, W.M., J. Loomis and B. Kanninen (1991): "Statistical efficiency of doublebounded dichotomous choice contingent valuation," *The American Journal of Agricultural Economics,* 73, 1255–1263.

[30] Honoré, B. E. and E. Tamer (2006), "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611-629.

[31] Hong, S. (2011), "Inference in Semiparametric Conditional Moment Model with Partial Identification," Working Paper.

[32] Horowitz, J. L. and C. F. Manski (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63, 281-302.

[33] ———— (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, 77-84.

[34] ———— (2006): "Identification and Estimation of Statistical Functionals Using Incomplete Data," *Journal of Econometrics*, 132, 445-459.

[35] Ichimura, H. and S. Lee (2010): "Characterization of the Asymptotic Distribution of Semiparametric M-Estimators," *Journal of Econometrics*, 159, 252–266.

[36] Imbens, G. W. (2003): "Sensitivity to Exogeneity Assumptions in Program Evaluation," *The American Economic Review,* 93, 126–132.

[37] Kaido, H. (2011): "A Dual Approach to Inference for Partially Identified Econometric Models," Working paper.

[38] Klein, R., C. Shen, and F. Vella (2011): "Semiparametric Selection Models with Binary Outcomes," Unpublished manuscript.

[39] Klein, R. and R. Spady (1993): "An Efficient Semiparametric Estimator for Discrete Choice Models," *Econometrica*, 61, 387-421.

[40] Kline, P. and Santos, A. (2012): "Sensitivity to Missing Data Assumptions: Theory and an Evaluation of the U.S. Wage Structure" *Quantitative Economics*, forthcoming.

[41] Leamer, E. E. (1981): "Is It a Demand Curve, or Is It a Supply Curve? Partial Identification Through Inequality Constraints," *The Review of Economics and Statistics*, 3, 319-327.

[42] Lee, S., M. H. Seo and Y. Shin (2011): "Testing for Threshold Effects in Regression Models", *Journal of the American Statistical Association*, 106, 220-231.

[43] Lewbel, A., O. Linton and D. McFadden (2011): "Estimating Features of a Distribution from Binomial Data," *Journal of Econometrics*, 162, 170-188.

[44] Magnac, T. and E. Maurin (2008): "Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data," *Review of Economic Studies,* 75, 835–864.

[45] Manski, C. F. (1989), "Anatomy of the Selection Problem," *Journal of Human Resources*, Vol. 24, No. 3, 1989, pp. 343-360.

[46] Manski, C. F. (1990), "Nonparametric Bounds on Treatment Effects," *The American Economic Review*, 80, Papers and Proceedings, 319–323.

[47] ———— (2003): *Partial Identification of Probability Distributions* (Berlin: Springer-Verlag).

[48] ———— (2007): *Identification for Prediction and Decision.* Cambridge, MA: Harvard Univ. Press.

[49] Manski, C. F. and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546.

[50] Nevo, A. and Rosen, A. (2012): "Identification with Imperfect Instruments,", *Review of Economics and Statistics*, 94, 659-671.

[51] Newey, W. (1994): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233-253.

[52] Pacini, D.H. (2012): "Least Square Linear Prediction with Two-Sample Data", unpublished manuscript, University of Bristol.

[53] Phillips, P. C. B. (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5, 181-240.

[54] Powell, J. L. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics,* 25, 303–325.

[55] Romano, J. P., and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, Special Issue in Honor of T. W. Anderson.

[56] ——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169-211.

[57] ——— (2010): "On the Uniform Asymptotic Validity of Subsampling and the Bootstrap", Working paper.

[58] Rosenbaum, P.R. and D.B. Rubin (1983): "Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome," *Journal of the Royal Statistical Society Series B,* 45, 212–218.

[59] Rubio-Ramirez, J., Waggoner, D. and Zha, T. (2010): "Structural Vector Autoregressions: The Theory for Identification and Algorithms for Inference", *Review of Economic Studies*, 77, 665-696.

[60] Santos, A. (2012): "Inference in Nonparametric Instrumental Variables with Partial Identification," *Econometrica*, 80, 213-275.

[61] Sargan, J. D. (1959): "The Estimation of Relationships with Autocorrelated Residuals by the Use of Instrumental Variables," *Journal of the Royal Statistical Society,* Series B, 21, 91-105.

[62] Scharfstein, D. O., A. Rotnitzky and J. M. Robins (1999): "Adjusting for Non-ignorable Drop-Out Using Semiparametric Nonresponse Models," *Journal of the American Statistical Association*, 94, 1096-1120.

[63] Sperlich, S. (2009): "A Note on Non-parametric Estimation with Predicted Variables," *The Econometrics Journal*, 12, 382–395.

[64] Song, K. (2008): "Uniform Convergence of Series Estimators over Function Spaces," *Econometric Theory*, 24, 1463–1499.

[65] Song, R., M. R. Kosorok and J. P. Fine (2009): "On Asymptotically Optimal Tests Under Loss of Identifiability in Semiparametric Models," *The Annals of Statistics*, 37, 1100-1120.

[66] Stoye, J. (2009): "More on Confidence Regions for Partially Identified Parameters," *Econometrica*, 77, 1299-1315.

[67] Tamer, E. (2010): "Partial identification in econometrics," *Annual Review of Economics*, 2, 167-195.

[68] van der Vaart, A. W. and J. A. Wellner (1996): *Weak convergence and empirical processes*, Springer-Verlag, New York

[69] Yanai H., K. Takeuchi and Y. Takane (2011): *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition,* Springer, New York