Inferences in Semiparametric Partially Identified Models: An Empirical Process Approach*

Juan Carlos Escanciano†
Indiana University

Lin Zhu‡
Indiana University

First draft: December 8, 2011. This draft: May 12, 2012

Abstract

This paper provides inference tools for partially identified and possibly misspecified semiparametric econometric models. The main working assumption is that the finite-dimensional parameter of interest and the infinite-dimensional nuisance parameter are identified conditionally on other nuisance parameters being known. This structure arises in numerous applications and permits the use of standard tools from empirical process methods. Our approach can be viewed as a formalization of sensitivity analysis in semiparametric models. We develop uniform convergence for the set of partially identified estimates and use the uniformity to establish set inference and hypothesis tests of interest. A practically convenient multiplier-type bootstrap is proposed to consistently estimate asymptotic quantiles for constructing confidence regions for the identified set and for estimating standard errors. We also extend the classical overidentification restrictions test to our semiparametric partially identified setting, and use this test to construct a confidence set for the true parameter that is robust to lack of identification of the nuisance parameter. Several examples illustrate the wide applicability of our results. In particular, we use a willingness-to-pay example in contingent valuations to demonstrate that our methods can accommodate generic convex identified sets with semiparametric support functions.

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

JEL classification: C12, C13, C14

*We would like to thank Arthur Lewbel, Ellie Tamer and Kei Hirano 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.

‡Department of Economics, Indiana University, 105 Wylie Hall, 100 South Woodlawn Avenue, Bloomington, IN 47405–7104, USA. E-mail: lz2@umail.iu.edu. Web Page: http://mypage.iu.edu/~lz2/

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 will lead to loss of point identification, and instead only a set of parameters can be identified from the model, a situation that has been termed as partial identification or set identification, see Manski (2003) for a textbook treatment and Tamer (2010) for a recent survey on theoretical and empirical developments. The vast majority of existing methods 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 $(\theta_0, \tau_0) \in \Theta \times \mathcal{T}$, $\Theta \subset \mathbb{R}^{d_\theta}$, $\mathcal{T} \subset \mathbb{R}^{d_\tau}$, and 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 $h_0(\cdot, \tau) \in \mathcal{H}$. Here $\mathcal{T}_0$ can be viewed as the projection of Γ_0 onto the nuisance parameter set $\mathcal{T}$. Hence, in our setting the identified set for θ is $\Theta_0 := \{\theta \in \Theta : \theta = \theta_0(\tau) \text{ for some } \tau \in \mathcal{T}_0\}$.

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 $\mathcal{T}$ hence can be viewed as the set of competing alternative identifying assumptions. Our approach is flexible in that it allows for a bottom-up identification approach, in which one starts with a large set $\mathcal{T}_0$ and adds identifying restrictions that restrict the size of $\mathcal{T}_0$, or a top-down identification approach where $\mathcal{T}_0$ is a singleton (point identification), and one proceeds relaxing assumptions that increase the size of $\mathcal{T}_0$.

Given a first-step estimate for h_0, we propose a semiparametric estimate for θ_0 in (1) using a GMM objective function. We establish the asymptotic distribution for the estimate $\widehat{\theta}(\cdot)$ as a process in τ, and use this to construct confidence regions for Θ_0 and unknown members of Θ_0 using our uniform results. Inference, including consistent estimation of standard errors, is implemented with the assistance of a multiplier-type bootstrap. We also show that inferences when $\mathcal{T}_0$ is unknown and identified from (1)

necessarily require an analysis of local misspecification, which motivates us to develop inference for misspecified models (i.e. relaxing (1)). See Section 3 for further discussion.

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 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 (2011). These authors provide example applications in linear Instrumental Variables (IV) models, dynamic panel data, Phillips curves and asset pricing models. For further example applications in the context of dynamic parametric moment restrictions see the companion paper in Escanciano and Zhu (2011). 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, and the obtained results can be used to extend theirs to semiparametric, possibly misspecified, models with infinite-dimensional nuisance parameters or non-smooth moment functions ψ, arising in e.g. quantile restrictions.

The scope of applications of our setting is substantially broader than the examples mentioned above. 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) 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 probability; see Section 2 for details. Our results can be used to extend their pointwise results to uniform results in semiparametric versions of their model.

There is an extensive literature providing sensitivity analysis as a method to evaluate the impact of relaxing strong identification assumptions on parameters of interest, see e.g. Rosenbaum and Rubin (1983) or more recently Altonji, Elder and Taber (2005). Our results can be seen as a formalization of sensitivity analysis as an inferential method in parametric and semiparametric partially identified models. Although this literature 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.[1] 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

[1] We thank Kei Hirano for pointing out this reference to us.

sensitivity analysis by varying the selection bias parameter τ over a plausible range of values.

Closely related to our work is the work by Bontemps, Magnac and Maurin (2011), who consider linear 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. (2011) 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. Hence, 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. Our results then complement those in Bontemps et al. (2011) and Beresteanu and Molinari (2008) in that we allow for the moments characterizing the support function to depend on infinite-dimensional nuisance parameters. We illustrate this important extension below with an example to inferences about the mean of functions of willingness to pay in contingent valuations.

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 any member of the identified set. In a more 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 (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010) and Bugni (2010). 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. Chen, Tamer and Torgovitsky (2010) propose inverting profiled likelihood ratio tests to construct confidence sets for finite-dimensional parameters. Song, Kosorok and Fine (2009) also 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). These works allow for the finite-dimensional parameter to be estimated at a slower rate than the regular (parametric) ones and focus on specific aspects of inference. Hong (2011) considers semiparametric conditional moment models, with infinite-dimensional parameters approximated by sieves, extending previous re-

sults by Santos (2010). We restrict attention to regular estimators and finitely parametrized nuisance parameters, so our results are less general in that respect, but the additional structure that we impose leads to simple implementations in which potentially infinite-dimensional optimization is not required, so popular nuisance parameter estimates such as kernel or series estimates can be used.[2] Thus, sieve methods and our methods are seen as complements rather than substitutes, and they can be potentially combined as discussed below.

The rest of the paper is organized as follows. Section 2 illustrates the general applicability of our methods with several motivational examples. Section 3 develops inferences for the possibly misspecified version of the partially identified model in (1). Section 4 revisits the willingness-to-pay example introduced in Section 2 and shows that our general theory applies to this example, provided that some regularity conditions hold. Section 5 concludes and discusses possible extensions. Mathematical proofs are relegated to the Appendix.

2 Motivational Examples

The following examples illustrate the wide applicability of our methods.

Example 1 (Willingness-to-Pay and Convex Identified Sets) 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, i.e. $(U \perp V)|X$. We are interested in estimating the mean value of a non-decreasing function of WTP, i.e. $\theta_0 = E[r(U)]$, for example $r(U) = U$, see Lewbel, McFadden and Linton (2010) for other interesting choices of $r(\cdot)$. For a non-monotonic function $r(\cdot)$, our methods can still be applied using results in Kemperman (1968), although the characterization of the identified set is more involved. We stress that there are different example applications with the same structure as the WTP example. Such an example is current status data, in which 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.

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,

$$\begin{aligned}\theta_0 &= E[r(U)] \\ &= E\left[\int r(u)\, dF(u|X)\right].\end{aligned}$$

[2] 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). See also the discussion below for extensions to infinite dimensional τ.

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 values for V, even in the case of large sample sizes. 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, effectively assuming point identification in the population. If we remove this assumption partial identification naturally arises in this framework. We analyze the partial identified case in this paper. 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 $v_1 < v_2 < \cdots < v_m$. Denoting $h_{0j}(x) := G(v_j|x)$, the identified set can be characterized as $\Theta_0 := \{\theta = E\left[\int r(u)\, dF(u|X)\right] : \text{for some cdf } F \text{ satisfying } F(v_j|x) = 1 - h_{0j}(x) \text{ for all } 1 \leq j \leq m\}$.

In what follows we fix x in the support of X. Denote the support of U as $\mathcal{U}$, and assume that $covh(\mathcal{V}) \subset covh(\mathcal{U})$, where $covh$ denotes convex hull, and suppose $r(\cdot)$ 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.[3] Then, applying integration by parts, we have

$$\begin{aligned}\int_{\mathcal{U}} r(u)\, dF(u|x) &= \int_{\mathcal{U}} r(u)\, d(-G(u|x)) \\ &= \underline{r} + \int_{\mathcal{U}} \dot{r}(u)\, G(u|x)\, du,\end{aligned}$$

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 $v_{j-1} \leq u \leq v_j$, it holds that $\dot{r}(u)\, h_{0j}(x) \leq \dot{r}(u)\, G(u|x) \leq \dot{r}(u)\, h_{0j-1}(x)$ with $h_{00}(x) := 1$ and $h_{0m+1}(x) := 0$. Hence for $j = 1, \ldots, m+1$

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

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

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

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 $E[L_r(h_0|X)] \leq \theta_0 \leq E[U_r(h_0|X)]$.

[3] 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. In applications for which $r(\cdot)$ has no natural bounds, we suggest to implement the bounds with $\underline{r} = r(\underline{u})$ and $\overline{r} = r(\overline{u})$, pluging in estimators of $\underline{u} = \max\{v \in \mathcal{V} : \sup_x F(v|x) = 0\}$ and $\overline{u} = \min\{v \in \mathcal{V} : \inf_x F(v|x) = 1\}$. The asymptotic distribution theory for these estimates has been studied in Chernozhukov, Lee and Rosen (2010).

To write this example into our framework, let $\psi(w,\theta,h_0,\tau) := \theta - L_r(h_0|x) - \tau(U_r(h_0|x) - L_r(h_0|x))$ 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 identified and nonparametrically estimated for each $1 \leq j \leq m$. Since in this example the identified set $\Theta_0 \subset \mathbb{R}$ is a closed interval, it can be characterized with the extremes of the interval $E[L_r(h_0|X)]$ and $E[U_r(h_0|X)]$. More generally, convex and closed identified sets can be characterized by the support function defined as

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

where $|\cdot|$ is the Euclidean norm, i.e. $|A| := (tr(A'A))^{1/2}$, and where A' and $tr(A)$ denote the transpose and trace of the matrix A, respectively. For instance, for the WTP example

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

Hence, we can also write the support function characterization into our framework by letting $\tau := q$, $\theta_0 := \beta(q|\Theta_0)$ 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. (2011) and Beresteanu and Molinari (2008) for examples. Then, letting $\tau := q$, $\theta_0(\tau) := \beta(\tau|\Theta_0)$, and $\psi(W,\theta,\tau) := \theta - \phi(W,\tau)$, one can rewrite the above equation as

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

with $\tau \in \mathcal{T}_0 := \mathbb{S}_p$. 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 (Tobin Regressor with Nonparametric Control Function) Consider the model:

$$\begin{aligned} Y &= \widetilde{X}\beta_0 + \widetilde{U}, \\ \widetilde{X} &= Z'\gamma_0 + \widetilde{V}, \\ \widetilde{U} &= h^*\left(\widetilde{V}\right) + \varepsilon, \end{aligned}$$

where Y is a scalar dependent variable, $\widetilde{X}$ is an unobserved scalar endogenous regressor, Z is a vector of instruments, $\widetilde{V}$ is median independent of Z, ε is mean independent of $\left(\widetilde{X}, Z\right)$, and $h^*(\cdot)$ is nonparametric control function. $\widetilde{X}$ is not observed, and instead, a censored version of it is observed:

$$\begin{aligned} X &= I(R=1)\, I\left(\widetilde{X} > 0\right)\widetilde{X}, \\ R &= \begin{cases} 1 \text{ with probability } 1-\pi, \\ 0 \text{ with probability } \pi, \end{cases} \end{aligned}$$

where $I(\cdot)$ is indicator function, R is unobserved but independent of Z. This is a semiparametric version of the Tobin regressor example studied in Chernozhukov, Rigobon and Stoker (2011), in which the control function is assumed to be nonparametric. The observed regressor X is selected, censored and endogenous, which is called Tobin regressor in Chernozhukov et al. (2011). It turns out that the parameter of interest β_0 is not identified in this example, since R is not observed, which implies the selection probability π is unknown and it is impossible to distinguish the effects of selection and censoring. However, as Chernozhukov et al. (2011) showed, for each fixed value of π, the parameter of interest β_0 can be identified, for instance, from the following moment conditions (see Powell (1984)):

$$E[I\left(Z'\gamma_0 > 0\right)(\tau - I\left(X \leq Z'\gamma_0\right))Z] = 0, \tag{2}$$

$$E\left[I\left(X > 0\right)\{Y - X\beta_0 - h_0\left(X - Z'\gamma_0, \beta_0, \tau\right)\}Q\right] = 0, \tag{3}$$

where $Q := (X, Z')'$, and $\tau := 0.5(1 - \pi) + \pi$. For each β and τ, define the (profiled) function

$$h_0(v, \beta, \tau) := E[Y - X\beta | X - Z'\gamma_0(\tau) = v, X > 0]$$

where $\gamma_0(\tau)$ is obtained from equation (2). Hence, to write this example into our framework, let

$$\psi(W, \theta_0, h_0, \tau) := \begin{pmatrix} I(Z'\gamma_0 > 0)[\tau - I(X \leq Z'\gamma_0)]Z \\ I(X > 0)\{Y_i - X\beta_0 - h_0(X - Z'\gamma_0, \beta_0, \tau)\}Q \end{pmatrix},$$

with $W := (Y, X, Z')'$, $\theta_0 := (\beta_0, \gamma_0')'$, then $E[\psi(W, \theta_0, h_0, \tau)] = 0$ follows.

Chernozhukov et al. (2011) also allow π to be a conditional probability function, so τ is infinite dimensional in their general setting. In our example, τ is finite-dimensional, but the model contains an infinite-dimensional nuisance parameter h_0.

Example 3 (Semiparametric Hazard Rate with Non-ignorable Drop-Out) Scharfstein, Rotnitzky, and Robins (1999) consider lack of identification and estimation in panel data models in which some objects might drop out at a random time Q prior to the recording time T when the outcome of interest Y is observed. The drop-out time Q may be due to factors related to the outcome Y, hence it includes information for estimating features of Y. Assume that $\{W_i = (Q_i, \Delta_i, \Delta_i Y_i, X_i)\}_{i=1}^n$ are observed, where $\Delta_i = 1(Q_i \geq T)$ and Q_i is set to be T if $\Delta_i = 1$, X_i contains other observed variables for individual i, and for simplicity we assume X_i does not depend on time. The parameter of interest is $\theta_0 = E[Y]$, and the conditional hazard of Q, given (Y, X), is assumed to follow a stratified Cox proportional hazards model of the form

$$\lambda_Q(t|X, Y) = \lambda_0(t|X)\exp(\tau Y),$$

where $\lambda_Q(t|X, Y) := \lim_{\delta \to 0} \Pr[t \leq Q \leq t + \delta | X, Y, Q \geq t]/\delta$, $\lambda_0(t|X)$ is an unknown positive function and τ is an unknown parameter. Scharfstein et al. (1999) show that (θ_0, τ) cannot be jointly identified without further assumptions, but for each fixed τ, θ_0 can be identified from the following moment restriction:

$$E[\psi(W_i, \theta_0, \Lambda_0, \tau)] = 0,$$

where $\Lambda_0(t|X) := \int_0^t \lambda_0(u|X)\,du$, and

$$\psi(W_i,\theta,\Lambda_0,\tau) = \frac{\Delta_i}{\pi(X_i,Y_i)}\left(Y_i - \theta - E\left[(1-\Delta_i)\,b(X_i,Q_i;\theta)\,|X_i,Y_i\right]\right) + (1-\Delta_i)\,b(X_i,Q_i;\theta)$$

with $\pi(X_i,Y_i) := \Pr(\Delta_i = 1|X_i,Y_i) = \exp(-\Lambda_0(T|X_i)\exp(\tau Y_i))$, and $b(x,t;\theta)$ is a function chosen by the researcher, see Scharfstein et al. (1999). Furthermore, the conditional expectation in the above equation can be evaluated as

$$E\left[(1-\Delta_i)\,b(X_i,Q_i;\theta)\,|X_i,Y_i\right] = \int_0^T b(X_i,t;\theta)\exp(-\Lambda_0(t|X_i)\exp(\tau Y_i))\exp(\tau Y_i)\,d\Lambda_0(t|X_i).$$

In the moment equation, $h_0 := \Lambda_0$ is unknown, θ_0 is the parameter of interest and τ is the nuisance parameter.

Example 4 (Normalization and Partial Identification) In many partially identified econometric models in which our structure arises the common approach to inference has been to focus on an arbitrary known choice of τ, say τ_0, and carry out standard inference on $\theta_0(\tau_0)$. This approach is commonly referred as "normalization" condition. The impact of normalization on inference has been overlooked, with rather few exceptions; see Hamilton, Waggoner and Zha (2007). This stands in contrast with the extensive list of applications in which this problem arises, see e.g. single-index models, limited dependent variable models, cointegrated models, etc. We provide here a systematic study of the impact of different normalizations in partially identified semiparametric models. Normalizations have been often applied to parametrizations of the model for which $\theta_0(\tau) \equiv \theta_0(\tau_0)$ for all $\tau \in \mathcal{T}$, i.e. $\theta_0 \equiv \theta_0(\tau_0)$ is identified. Even in these situations, once we account for the partial identification nature of the model, new issues arise, such as what the limiting distribution of the estimator of θ_0 is when the normalization is not impose (see Phillips (1989)) or how to improve inference on θ_0 when all normalizations are simultaneously considered. We do not address issues on efficiency here and rather focus on the case where θ_0 is not point identified, but these will be investigated in future research.

3 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, for instance, the nuisance function h_0 in Example 2 contains both θ and τ. Although in many applications, the nuisance parameter set $\mathcal{T}_0$ is known, e.g. Example 1, there are cases in which $\mathcal{T}_0$ is unknown and needs to be consistently estimated (with an estimator denoted as $\hat{\mathcal{T}}_0$), as in Example 2. Consistency here is in the sense of Hausdorff metric $d_H\left(\hat{\mathcal{T}}_0,\mathcal{T}_0\right) = o_P(1)$, where

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

and $d_H(a, B) := \inf_{b\in B} |a - b|$. On the other hand, by definition, the nuisance parameter set $\mathcal{T}_0$ is the largest possible set such that for each $\tau \in \mathcal{T}_0$, equation (1) holds for some $\theta \in \Theta$. Hence, for $\tau \in \hat{\mathcal{T}}_0 \backslash \mathcal{T}_0$, equation (1) fails to hold for any $\theta \in \Theta$ and misspecification arises in this situation. This is a generic feature of inferences under partial identification. 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 $\mathcal{T}_0$ is known and there is no need to consider misspecification. The asymptotic results developed below for semiparametric GMM estimation under global misspecification with nonsmooth moment functions are also of independent interest.

Given consistent estimator $\hat{\mathcal{T}}_0$ of $\mathcal{T}_0$, it suffices to consider a small enlargement of $\mathcal{T}_0$, $\mathcal{T}_0^\delta := \{\tau \in \mathcal{T} : \inf_{\tau'\in\mathcal{T}_0} |\tau - \tau'| \leq \delta\}$. Then for any $\tau \in \mathcal{T}_0^\delta \backslash \mathcal{T}_0$, by definition of $\mathcal{T}_0$, we have $E\psi(W, \theta, h_0, \tau) \neq 0$ for all $\theta \in \Theta$, i.e. the model is misspecified for $\tau \in \mathcal{T}_0^{\delta_0} \backslash \mathcal{T}_0$. To develop asymptotic results under misspecification, we maintain the partial identification assumption, that is, for each $\tau \in \mathcal{T}_0^\delta$, we assume $\|E\psi(W, \theta, h_0, \tau)\|$ is uniquely minimized at $\theta_0(\tau)$.

We assume that for each fixed value of $(\theta, \tau) \in \Theta \times \mathcal{T}_0^\delta$, 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. 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 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 not general treatment of misspecification for non-smooth objective functions, to the best of our knowledge.

This paper deals with uniform estimation of $\theta_0(\cdot)$ as a device to make inference on Θ_0. Suppose the observed data $\{W_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.) sequence following the same distribution as W. Henceforth, we use the following generic notation, for a measurable function f we denote the empirical expectation and empirical process by

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

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, under suitable and well-known conditions. For each fixed $\tau \in \mathcal{T}_0^\delta$, 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. The corresponding estimated set is defined as $\widehat{\Theta}_0 := \{\widehat{\theta}(\tau) : \tau \in \widehat{\mathcal{T}}_0\}$. 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 $\ell^\infty(\mathcal{T}_0^\delta)$ of uniformly bounded functions on $\mathcal{T}_0^\delta$, which is equipped with the sup-norm, $\|z\|_{\mathcal{T}_0^\delta} := \sup_{\tau\in\mathcal{T}_0^\delta} |z(\tau)|$. In this article we consider the weak convergence, denoted by $\rightsquigarrow$, of $Z_n(\cdot)$ in the metric space $\ell^\infty(\mathcal{T}_0^\delta)$ endowed with the sup-norm in the sense of J. Hoffmann-Jørgensen (see, e.g., Dudley 1999 p. 94). In this paper we abstract from measurability issues that may arise as a result of using the space $\ell^\infty(\mathcal{T}_0^\delta)$ endowed with the sup-norm, as for all practical matters these issues are irrelevant. 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 $\tau \in \mathcal{T}_0^\delta$, where Θ is a compact set in $\mathbb{R}^{d_\theta}$. In addition, assume
(i) The estimator $\widehat{\theta}(\tau)$ satisfies

$$\sup_{\tau\in\mathcal{T}_0^\delta} \left\{ \left\| M_n\left(\widehat{\theta}(\tau), \widehat{h}, \tau\right)\right\| - \inf_{\theta\in\Theta} \left\| M_n\left(\theta, \widehat{h}, \tau\right)\right\| \right\} \leq o_P(1). \tag{5}$$

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

$$\inf_{\tau\in\mathcal{T}_0^\delta} \left\{ \inf_{\theta:|\theta-\theta_0(\tau)|\geq\varepsilon} \|M(\theta, h_0, \tau)\| - \|M(\theta_0(\tau), h_0, \tau)\| \right\} \geq \eta(\varepsilon). \tag{6}$$

(iii) Uniform Continuity: uniformly for all $\theta \in \Theta$ and $\tau \in \mathcal{T}_0^\delta$, $M(\theta, h, \tau)$ 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 $\{\delta_n\} \to 0$,

$$\sup_{\theta\in\Theta,\tau\in\mathcal{T}_0^\delta,\|h-h_0\|_{\mathcal{H}}\leq\delta_n} \|M_n(\theta, h, \tau) - M(\theta, h, \tau)\| = o_P(1). \tag{7}$$

(vi) Nuisance parameter identified set: There is an estimator $\widehat{\mathcal{T}}_0$ that is consistent for $\mathcal{T}_0$, i.e. $d_H\left(\widehat{\mathcal{T}}_0, \mathcal{T}_0\right) = o_P(1)$.

These assumptions are uniform versions of those in CLV for consistency. Like these authors, we also allow for nonsmooth 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 $\Psi := \{\psi(\cdot, \theta, h, \tau) : \theta \in \Theta,\ h \in \mathcal{H},\ \tau \in \mathcal{T}_0^\delta\}$. When $\mathcal{T}_0$ is known, we can take $\delta = 0$ in Assumption A1 and A1(vi) is obviously not required. A candidate for $\widehat{\mathcal{T}}_0$

can be obtained extending the ideas in Chernozhukov, Hong and Tamer (2007) to our semiparametric context with infinite dimensional nuisance parameters. This is beyond the scope of this paper, but it should follow from combining our results with theirs.

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

Theorem 3.1 *Under Assumption A1(i-v), for any sequence of estimators satisfying (5) it holds that* $\sup_{\tau\in\mathcal{T}_0^\delta}\left|\widehat{\theta}(\tau)-\theta_0(\tau)\right|=o_P(1)$.

As a corollary of Theorem 1, we obtain that the estimated set $\widehat{\Theta}_0=\{\widehat{\theta}(\tau):\tau\in\widehat{\mathcal{T}}_0\}$ is a consistent estimator of the identified set Θ_0 in the sense of Hausdorff metric.

Corollary 3.1 *Under Assumption A1 and additionally assuming that* $\theta_0(\cdot)$ *is equi-continuous on* $\mathcal{T}_0^\delta$, *we have* $d_H\left(\widehat{\Theta}_0,\Theta_0\right)=o_P(1)$.

If $\mathcal{T}_0$ is known, then we do not need the equi-continuity condition on $\theta_0(\cdot)$ to establish the consistency of $\widehat{\Theta}_0$.

3.2 Weak Convergence

In this section we establish the weak convergence of $Z_n(\cdot)=\sqrt{n}(\widehat{\theta}(\cdot)-\theta_0(\cdot))$ as a stochastic process in $\ell^\infty(\mathcal{T}_0^\delta)$ endowed with the sup-norm $\|\cdot\|_{\mathcal{T}_0^\delta}$. Given consistency, we can work, as usual, in a small or even shrinking neighborhood of $\theta_0(\cdot)$ and h_0. With some abuse of the notations, define a δ-enlargement of the parameter sets $\Theta_\delta:=\{\theta\in\Theta:d_H(\theta,\Theta_0)\le\delta\}$ 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 $(\theta,\tau)\in\Theta_\delta\times\mathcal{T}_0^\delta$, 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 make the presentation clear, we first introduce some notations and abbreviated versions. Let $vec(A)$ denote the vectorization of matrix A and $\otimes$ denote the Kronecker product. We drop the dependence on true values to simplify the notation. 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 $\tau\in\mathcal{T}_0^\delta$, and that $\sup_{\tau\in\mathcal{T}_0^{\delta_0}}|\widehat{\theta}(\tau)-\theta_0(\tau)|=o_P(1)$. In addition, assume:

(i) There exists a ($d_\psi \times d_\theta$ matrix) function $\psi_\theta(\cdot, \theta, h_0, \tau)$ such that for any positive sequence $\delta_n \to 0$,

$$\sup_{\tau \in \mathcal{T}_0^\delta} \sup_{|\theta-\theta_0| \le \delta_n, \|h-h_0\|_{\mathcal{H}} \le \delta_n} |E_n \psi_\theta(Z, \theta, h, \tau) - V_\theta(\theta, h, \tau)| = o_P(1).$$

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

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

(iii) Smoothness in θ : for each $\tau \in \mathcal{T}_0^\delta$, the map $\theta \to M(\theta, h_0, \tau)$ is twice continuously differentiable at θ_0, with first-order derivative $V_{\theta_0}(\tau)$ that is of full rank at θ_0 and $\sup_{\tau \in \mathcal{T}_0^\delta} \|V_{\theta_0}(\tau)\| < \infty$ and $\sup_{\tau \in \mathcal{T}_0^\delta} \|V_{\theta_0}(\tau)\|^{-1} < \infty$, and the second-order derivative $V_{\theta_0\theta_0}(\tau)$. Furthermore, suppose that $V_{\theta_0}(\tau)'\Xi V_{\theta_0}(\tau) + (M(\tau)'\Xi \otimes I_{d_\theta}) V_{\theta_0\theta_0}(\tau)$ is of full rank.
(iv) Smoothness in h : (1) for each $(\theta, \tau) \in \Theta_\delta \times \mathcal{T}_0^\delta$, 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 $(\theta, h, \tau) \in \Theta_{\delta_n} \times \mathcal{H}_{\delta_n} \times \mathcal{T}_0^\delta$ with a positive sequence $\delta_n \to 0$, it holds that

$$\sup_{\tau \in \mathcal{T}_0^\delta} \|M(\theta, h, \tau) - M(\theta, h_0, \tau) - V_h(\theta, h_0, \tau)[h - h_0]\| \le c \|h - h_0\|_{\mathcal{H}}^2 \tag{9}$$

for a constant $c > 0$, and

$$\sup_{\tau \in \mathcal{T}_0^\delta} \|V_h(\theta, h_0, \tau)[h - h_0] - V_h(\theta_0, h_0, \tau)[h - h_0]\| \le o(1)\delta_n; \tag{10}$$

(2) 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 it also holds that

$$\sup_{\tau \in \mathcal{T}_0^\delta} \|V_\theta(\theta, h, \tau) - V_\theta(\theta, h_0, \tau) - V_{\theta h}(\theta, h_0, \tau)[h - h_0]\| \le c \|h - h_0\|_{\mathcal{H}}^2 \tag{11}$$

and

$$\sup_{\tau \in \mathcal{T}_0^\delta} \|V_{\theta h}(\theta, h_0, \tau)[h - h_0] - V_{\theta h}(\theta_0, h_0, \tau)[h - h_0]\| \le o(1)\delta_n; \tag{12}$$

(v) $\Pr\left(\widehat{h} \in \mathcal{H}\right) \to 1$, 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 $\delta_n \to 0$,

$$\sup_{\tau \in \mathcal{T}_0^\delta} \sup_{|\theta-\theta_0| \le \delta_n, \|h-h_0\|_{\mathcal{H}} \le \delta_n} \|\mathbb{G}_n \psi(W, \theta, h, \tau) - \mathbb{G}_n \psi(W, \theta_0, h, \tau)\| = o_P(1).$$

and

$$\sup_{\tau \in \mathcal{T}_0^\delta} \sup_{|\theta-\theta_0| \le \delta_n, \|h-h_0\|_{\mathcal{H}} \le \delta_n} \|\mathbb{G}_n \psi_\theta(W, \theta, h, \tau) - \mathbb{G}_n \psi_\theta(W, \theta_0, h_0, \tau)\| = o_P(1).$$

(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, let

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

and assume the function class $\left\{w \to s(w, \theta_0, h_0, \tau) : \tau \in \mathcal{T}_0^\delta\right\}$ 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 nonsmooth moment functions, these are high-level conditions but still are satisfied by some commonly used models, for example, the quantile regression model; see Angrist et al. (2006). Assumption A2(iii) requires second-order differentiablility 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 $V_h(\cdot)$. 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 $\Pr\left(\widehat{h} \in \mathcal{H}\right) \to 1$, even in more complicated settings such as ours, where infinite-dimensional parameters are in turn indexed by other nuisance parameters. Assumption A2(vi) is usually implied by the Donsker property of the function class $\Psi := \{\psi(\cdot, \theta, h, \tau) : \theta \in \Theta_\delta, h \in \mathcal{H}_\delta, \tau \in \mathcal{T}_0^\delta\}$ and $\Psi_\theta := \{\psi_\theta(\cdot, \theta, h, \tau) : \theta \in \Theta_\delta, h \in \mathcal{H}_\delta, \tau \in \mathcal{T}_0^\delta\}$, for which primitive conditions can be provided using standard empirical process 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.

Theorem 3.2 *Under Assumption A2, it holds that*

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

where L is a Gaussian process in $\ell^\infty(\mathcal{T}_0^\delta)$ 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

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

$K(\tau_1, \tau_2) := Cov\left(s(W, \theta_0, h_0, \tau_1), s(W, \theta_0, h_0, \tau_2)\right)$, *and s as defined in (13).*

Remark 3.1 *When $\mathcal{T}_0$ is known and there is no need to consider misspecification, we would consider the weak convergence in $\ell^\infty(\mathcal{T}_0)$ and the asymptotic covariance function can be simplified substantially. In this case, for all $\tau \in \mathcal{T}_0$,*

$$A_0(\tau) = V_{\theta_0}(\tau)' \Xi V_{\theta_0}(\tau), \ B_0(\tau) = V_{\theta_0}(\tau)' \Xi, \text{ and } D_0(\tau) = 0.$$

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 $\widetilde{K}(\tau_1, \tau_2) = E\left[\widetilde{s}(W, \theta_0, h_0, \tau_1) \widetilde{s}'(W, \theta_0, h_0, \tau_2)\right]$, $\widetilde{s}(W, \theta_0, h_0, \tau_1) := \psi(W, \theta_0, h_0, \tau) + \phi(W, \theta_0, h_0, \tau)$. Without misspecification, some key assumptions in A2 can be dropped. For the sake of completeness, we list the conditions needed for this case.

Assumption A2': Suppose that $\theta_0(\tau)$ is in interior of Θ for each $\tau \in \mathcal{T}_0$, and that $\sup_{\tau \in \mathcal{T}_0} |\widehat{\theta}(\tau) - \theta_0(\tau)| = o_P(1)$. In addition, assume
(i) The estimator $\widehat{\theta}(\tau)$ also satisfies

$$\sup_{\tau \in \mathcal{T}_0} \left\{ \left\| M_n\left(\widehat{\theta}(\tau), \widehat{h}, \tau\right) \right\| - \inf_{\theta \in \Theta_\delta} \left\| M_n\left(\theta, \widehat{h}, \tau\right) \right\| \right\} \leq o_P\left(n^{-1/2}\right).$$

(ii) Smoothness in θ : for each $\tau \in \mathcal{T}_0$, 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 $\sup_{\tau \in \mathcal{T}_0} \|V_{\theta_0}(\tau)\| < \infty$ and $\sup_{\tau \in \mathcal{T}_0} \|V_{\theta_0}(\tau)\|^{-1} < \infty$.
(iii) Smoothness in h : for each $(\theta, \tau) \in \Theta_\delta \times \mathcal{T}_0$, 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 $(\theta, h, \tau) \in \Theta_{\delta_n} \times \mathcal{H}_{\delta_n} \times \mathcal{T}_0$ with a positive sequence $\delta_n \to 0$, it holds that

$$\sup_{\tau \in \mathcal{T}_0} \|M(\theta, h, \tau) - M(\theta, h_0, \tau) - V_h(\theta, h_0, \tau)[h - h_0]\| \leq c \|h - h_0\|_{\mathcal{H}}^2$$

for a constant $c > 0$, and

$$\sup_{\tau \in \mathcal{T}_0} \|V_h(\theta, h_0, \tau)[h - h_0] - V_h(\theta_0, h_0, \tau)[h - h_0]\| \leq o(1) \delta_n.$$

(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 positive numbers $\delta_n \to 0$,

$$\sup_{\tau \in \mathcal{T}_0} \sup_{|\theta - \theta_0| \leq \delta_n, \|h - h_0\|_{\mathcal{H}} \leq \delta_n} \|\mathbb{G}_n \psi(W, \theta, h, \tau) - \mathbb{G}_n \psi(W, \theta_0, h, \tau)\| = o_P(1).$$

(vi) $\sqrt{n} V_h(\theta_0, h_0, \tau)[\widehat{h} - h_0]$ 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 the function class $\{s(w, \theta_0, h_0, \tau) := \psi(w, \theta_0, h_0, \tau) + \phi(w, \theta_0, h_0, \tau) : \tau \in \mathcal{T}_0\}$ is Donsker.

Notice that without misspecification, there is no need to require the existence of $\psi_\theta(\cdot)$ as in A2(i) and it suffices to impose only first-order differentiability. The proof with Assumption A2' is similar, hence omitted.

Based on the above weak convergence results, we can derive the rate of convergence for the estimated set $\widehat{\Theta}_0$.

Corollary 3.2 *Let Assumption A2 hold, and in addition assume that* $d_H\left(\widehat{\mathcal{T}}_0, \mathcal{T}_0\right) = O_P\left(n^{-1/2}\right)$ *and* $\theta_0(\cdot)$ *is Lipschitz continuous in* $\mathcal{T}_0^\delta$. *Then, it holds that* $d_H\left(\widehat{\Theta}_0, \Theta_0\right) = O_P\left(n^{-1/2}\right)$.

Again if $\mathcal{T}_0$ is known, the above corollary holds under Assumption A2' and the continuity of $\theta_0(\cdot)$ is not needed.

3.3 Bootstrap Approximation

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 a natural approximation. CLV propose to use the ordinary nonparametric bootstrap in which they need to re-estimate the pair (θ, h) for each bootstrap sample. This bootstrap procedure can be computationally expensive in our framework, since we need to do the estimation for each fixed $\tau \in \mathcal{T}_0$. Hence, we propose here an alternative bootstrap method based on the multiplier principle, which has the advantage that one only needs to estimate $(\widehat{\theta}(\tau), \widehat{h}(\cdot, \tau))$ for the original sample and hence it is computationally more attractive. We will show the consistency of this bootstrap procedure in this section. We also show the consistency of bootstrap standard errors which has not been justified in the literature even in the point-identified case with the nonparametric 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 $A(\theta_0, h_0, \tau) = A_0(\tau)^{-1}[B_0(\tau)\ D_0(\tau)]$ 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 in 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 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 Z_n :

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

The above multiplier-type bootstrap only requires estimation of (θ_0,h_0) for the original sample, hence it is computationally more efficient than the standard nonparametric bootstrap. This is especially important in our setting, where these quantities need to be estimated for each $\tau \in \mathcal{T}_0$. In contrast, the standard nonparametric bootstrap has the advantage that it does not require estimation of influence functions.

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) $\sup_{\tau\in\mathcal{T}_0^\delta}\left|\widehat{\theta}(\tau) - \theta_0(\tau)\right| = o_P(1)$; $\Pr\left(\widehat{h}\in\mathcal{H}\right)\to 1$, 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

$$\sup_{\theta\in\Theta_\delta,h\in\mathcal{H}_\delta,\tau\in\mathcal{T}_0^\delta}\left|\widehat{V}_\theta(\theta,h,\tau) - V_\theta(\theta,h,\tau)\right| = o_P(1),$$
$$\sup_{\theta\in\Theta_\delta,h\in\mathcal{H}_\delta,\tau\in\mathcal{T}_0^\delta}\left|\widehat{V}_{\theta\theta}(\theta,h,\tau) - V_{\theta\theta}(\theta,h,\tau)\right| = o_P(1);$$

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 $\tau\in\mathcal{T}_0^\delta$.
(iii) The class of functions $\mathcal{S} := \{s(\cdot,\theta,h,\tau) : \theta\in\Theta_\delta, h\in\mathcal{H}_\delta, \tau\in\mathcal{T}_0^\delta\}$ 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 some 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 $\{\partial/\partial\theta\psi(\cdot,\theta,h,\tau) : \theta\in\Theta_\delta, h\in\mathcal{H}, \tau\in\mathcal{T}_0\}$ and $\widehat{V}_\theta$ is the sample analog of $V_\theta(h,\tau) = E[\partial/\partial\theta\psi(\cdot,\theta,h,\tau)]$; 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 in probability introduced in Giné and Zinn (1990). Let BL_1 denote the set of all functionals on $\ell^\infty\left(\mathcal{T}_0^\delta\right)$ with a Lipschitz norm bounded by 1, i.e. for any $f \in BL_1$, $\sup_{z\in\ell^\infty(\mathcal{T}_0^\delta)} |f(z)| \leq 1$ and for $z_1, z_2 \in \ell^\infty\left(\mathcal{T}_0^\delta\right)$, $|f(z_1) - f(z_2)| \leq \|z_1 - z_2\|_{\mathcal{T}_0^\delta}$. 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 $\ell^\infty\left(\mathcal{T}_0^\delta\right)$ 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

$$\sup_{\tau\in\mathcal{T}_0^\delta} \left| Var^*(\widehat{Z}_n^*(\tau)) - C(\tau,\tau)\right| = o_P(1), \tag{15}$$

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. If $\mathcal{T}_0$ is known, then we do not need to estimate $V_{\theta_0\theta_0}(\tau)$ and in the second step of the bootstrap procedure, centering is not needed.*

Remark 3.3 *Uniform-in-probability versions of our results can be established under further regularity conditions. See Chapter 2.8 in van der Vaart and Wellner (1996) for general results and Romano (1989), Mikusheva (2007) and Romano and Shaikh (2010) for applications. For illustration of how these results can be obtained see our companion paper in Escanciano and Zhu (2011).*

3.4 Inference on the Identified Set

In this section, we use our previous uniform convergence results to construct confidence regions for the whole 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) test $\sqrt{n}\sup_{\tau\in\widehat{\mathcal{T}}_0} \left|\widehat{\theta}_n(\tau) - \theta_0(\tau)\right|$, whose asymptotic quantile can be approximated by the multiplier-type bootstrap in Section 3.3. The KS test leads to the confidence region

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

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

Proposition 3.1 *Let Assumptions A2 and A3 hold, and in addition assume $\Pr\left(\mathcal{T}_0 \subset \widehat{\mathcal{T}}_0\right) \to 1$. Then*

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

The required inclusion condition $\Pr\left(\mathcal{T}_0 \subset \widehat{\mathcal{T}}_0\right) \to 1$ is not very restrictive, for example, the set estimator developed in Chernozhukov, Hong and Tamer (2007) satisfies this condition.

3.5 Inference on True Parameter Value

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 by typical arguments, 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, has the drawback that requires a rank condition and a order condition that are similar to the typical overidentification conditions required in the J-test. The convex case is less general, but it has the advantage that in some cases where the order condition fails, still it can be applied.

3.5.1 The General Case

The testing hypotheses are

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

These hypothesis are of interest in their own. For instance, in Altonji et al. (2005) the aim is to test for significance of the impact of attending a Catholic school on educational attainment. Their example fits our setting with a null hypothesis $0 \in \Theta_0$. Only the alternative is conclusive. A stronger sense of no-impact is $\theta_0(\tau) = 0$ for all $\tau \in \mathcal{T}_0$, which has been addressed in the previous section.

Recall 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 popular approach has been to take the union of confidence intervals for $\theta_0(\tau)$ constructed at each $\tau \in \mathcal{T}_0$ to obtain a confidence region for the true parameter, see e.g. Chernozhukov et al. (2011). This union-intersection approach is also possible in the general class of models we consider. However, this approach leads to conservative inference. An alternative, more efficient, approach 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)(1) holds with $\theta = \widetilde{\theta}$ and condition (10) replaced by the following condition: for $h \in \mathcal{H}_{\delta_n}$, $|\tau_1 - \tau_2| \leq \eta_n$,

$$\left\| V_h\left(\widetilde{\theta}, \widetilde{h}_0, \tau_1\right)\left[h - \widetilde{h}_0\right] - V_h\left(\widetilde{\theta}, \widetilde{h}_0, \tau_2\right)\left[h - \widetilde{h}_0\right] \right\| \leq O\left(\eta_n \delta_n\right). \tag{16}$$

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

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

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

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

Theorem 3.4 *Let the null hypothesis H_0 and Assumption A4 hold, then*

$$T_n(\widetilde{\theta}) \rightsquigarrow \inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \left\| [I_{d_\psi} - P\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)] G(\widetilde{\tau}) \right\|,$$

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

Remark 3.4 *If $rank(V_\tau(\widetilde{\tau})) = d_\psi$ for some $\widetilde{\tau} \in \widetilde{\mathcal{T}}$, then $P\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}\right)$ becomes identity, 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 the number of moments.*

Similar test statistics have been used in the partial identification literature. In a support function setting, Bontemps et al. (2011) 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 IV 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. In our setting, we are able to express the limiting distribution as just one minimization. The result of Theorem 3.4 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-squared 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 the multiplier-type bootstrap which is computationally attractive by the arguments mentioned above. 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 3.4, we have

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

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 $\widehat{P}\left(\widetilde{\theta},\widehat{h},\tau\right) := \widehat{V}_\tau\left(\widetilde{\theta},\widehat{h},\tau\right)\left[\widehat{V}_\tau\left(\widetilde{\theta},\widehat{h},\tau\right)'\Xi\widehat{V}_\tau\left(\widetilde{\theta},\widehat{h},\tau\right)\right]^- \widehat{V}_\tau\left(\widetilde{\theta},\widehat{h},\tau\right)'\Xi$. We suggest the following bootstrap procedures to approximate the critical value of $T_n(\widetilde{\theta})$, which first step is the same as before:

Step 2: For each τ, compute $R_n^*(\tau) = \left\| \left[I - \widehat{P}\left(\widetilde{\theta},\widehat{h},\tau\right)\right] \frac{1}{\sqrt{n}} \sum_{i=1}^n s\left(W_i,\widetilde{\theta},\widehat{h},\tau\right) u_i \right\| + \lambda_n \left\| M_n\left(\widetilde{\theta},\widehat{h},\tau\right)\right\|$, 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 $c_{n,1-\alpha}^*(\widetilde{\theta})$ of $\left\{T_{n,(b)}^*(\widetilde{\theta})\right\}_{b=1}^B$.

Remark 3.5 *We follow 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* $\mathcal{T}$, *which generally leads to conservative inference. A more efficient 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 3.2 *Suppose (i) Assumption A4 holds; (ii) the penalty sequence λ_n satisfies $\lambda_n \to \infty$ and $\lambda_n = o\left(\sqrt{n}/\log\log n\right)$; (ii) 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.

3.5.2 The Convex Case

As discussed in Remark (3.4), when $rank(V_\tau(\widetilde{\tau})) = d_\psi$ (for example with $d_\psi \leq d_\tau$), 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 $\beta(\tau|\Theta_0) = E[\xi(W, g_0, \tau)]$, 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 $E[\xi(W, g, \tau)]$ 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

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

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

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

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

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

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

Theorem 3.5 *Let the null hypothesis H_0 and Assumption A5 hold, then*

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

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

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

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

Remark 3.6 *We do not explicitly require differentiability of the support function $\beta(\tau|\Theta_0)$ 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 BMM, 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 nondifferentiability does not necessarily lead to non-regular asymptotic distribution of the support function estimator, e.g. see section 5.3 in BMM.*

Remark 3.7 *To deal with the multiple minimizer problem caused by the presence of kinks, BMM proposes a shrinkage type estimator to obtain a unique minimizer, instead we employ the idea in Galichon and Henry (2009) and take the minimum over all possible minimizers.*

To implement the test we can follow the similar bootstrap procedure introduced for 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, as for example in our empirical application below, then the limit distribution of $T_n^c\left(\widetilde{\theta}\right)$ 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 The WTP Example Revisited

In this section, we apply our previous results to the WTP example considered in Section 2, and we discuss primitive conditions to verify the high-level assumptions needed for the asymptotic results. Notice that the nuisance paramter set $\mathcal{T}_0 = [0,1]$ is known, hence there is no need to consider misspecification and it suffices to verify conditions in Assumption A2'. Recall that

$$\begin{aligned} U_r(h_0|x) &= r(v_1) + \Delta_{ur}' h_0(x), \\ L_r(h_0|x) &= \underline{r} + \Delta_{lr}' h_0(x). \end{aligned}$$

Then,

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

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

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

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, see the empirical application below. Suppose that $h_{0j}(\cdot) \in \mathcal{H} = \mathcal{C}_1^{\eta_h}(\mathcal{X})$ be set of continuous functions on the bounded subset $\mathcal{X} \in \mathbb{R}^{d_x}$ 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 P-Donsker class. Escanciano, Jacho-Chávez and Lewbel (2011) discussed primitive conditions for Assumption A2'(iv), for example, if $d_x = 2$, they showed that it suffices to use a third-time continuously differentiable kernel function $k(\cdot)$, and the bandwidth can be chosen as $b = an^{-1/7}$. However, since $M(\theta, h, \tau)$ is linear in h, the rate condition for $\widehat{h}$ can be eliminated, see Remark 2(iii) in CLV.

We now verify Assumption A2'(v) and A2'(vi) for this example. Notice that

$$\begin{aligned} &|\psi(w, \theta, h, \tau) - \psi(w, \theta_0, h_0, \tau)| \\ &\leq \left|0.5(1+\tau)\Delta_{ur}'\left(h(x) - h_0(x)\right)\right| + \left|0.5(1-\tau)\Delta_{lr}'\left(h(x) - h_0(x)\right)\right| \\ &\leq c\left|h(x) - h_0(x)\right|. \end{aligned}$$

Then it is easy to see 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. Suppose X has a twice continuously differentiable conditional density function

$f_j(x) := f_{X|V=v_j}(x)$, and $0 < \inf_{x\in\mathcal{X}} f_j(x) \leq \sup_{x\in\mathcal{X}} f_j(x) < \infty$ for each $j = 1, \ldots, m$. The Bahadur representation for $\widehat{h}_j - h_{0j}$ is,

$$\widehat{h}_j(x) - h_{0j}(x) = E[\widehat{m}_j(x)]^{-1} \frac{1}{n}\sum_{i=1}^n I(V_i = v_j) K_b\left(\frac{X_i - x}{b}\right)(Y_i - h_{0j}(x)) + R_{jn}(x),$$

where

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

and $R_{jn}(x)$ satisfies

$$\sup_{x\in\mathcal{X}} |R_{jn}(x)| = O_P\left(\left(\frac{\log n}{nh^{d_x}}\right)^{3/4}\right).$$

Then, under standard assumptions on the bandwidth and using a third-order kernel, we obtain the desired expansion

$$n^{1/2} E\left[\widehat{h}_j(X) - h_{0j}(X)\right] = \frac{1}{\sqrt{n}p_j}\sum_{i=1}^n I(V_i = v_j)\varepsilon_{ji} + o_P(1),$$

where $p_j := E[I(V_i = v_j)]$ and $\varepsilon_{ji} := Y_i - h_{0j}(X_i)$. Let $B_{jn} := n^{-1/2}p_j^{-1}\sum_{i=1}^n I(V_i = v_j)\varepsilon_{ji}$, then for $j \neq j'$, B_{jn} and $B_{j'n}$ are independent due to the indicator term. Hence, Assumption A2'(vi) holds.

As discussed in Section 3.3, 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 analogues, 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

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

From above,

$$\begin{aligned}
s(W_i, \theta_0, h_0, \tau) &= \psi(W_i, \theta_0, h_0, \tau) + \phi(W_i, \theta_0, h_0, \tau) \\
&= \theta_0(\tau) - \tau\left[r(v_1) + \Delta'_{ur}(h_0(X_i) + l_0(W_i))\right] \\
&\quad - (1-\tau)\left[\underline{r} + \Delta'_{lr}(h_0(X_i) + l_0(W_i))\right],
\end{aligned}$$

where $l_0(W_i) := \left(p_1^{-1} I(V_i = v_1)(Y_i - h_{01}(X_i)), \ldots, p_J^{-1} I(V_i = v_J)(Y_i - h_{0J}(X_i))\right)'$. Hence,

$$\begin{aligned}\sqrt{n}\left(\widehat{\theta}_L - \theta_L\right) &= \frac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n s(W_i, \theta_0, h_0, 0) + o_P(1) \\ &= \frac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n \left\{\theta_L - \underline{r} - \Delta'_{lr}(h_0(X_i) + l_0(W_i))\right\} + o_P(1),\end{aligned}$$

and similarly

$$\sqrt{n}\left(\widehat{\theta}_U - \theta_U\right) = \frac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n \left\{\theta_U - r(v_1) - \Delta'_{ur}(h_0(X_i) + l_0(W_i))\right\} + o_P(1).$$

Notice that

$$\begin{aligned}&\sqrt{n}\begin{pmatrix}\widehat{\theta}_L - \theta_L \\ \widehat{\theta}_U - \theta_U\end{pmatrix} \\ &= \frac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n \left\{\begin{pmatrix}\theta_L - \underline{r} \\ \theta_U - r(v_1)\end{pmatrix} - \begin{pmatrix}\Delta'_{lr} \\ \Delta'_{ur}\end{pmatrix}(h_0(X_i) + l_0(W_i))\right\} + o_P(1) \\ &\rightsquigarrow N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\Delta'_{lr} \\ \Delta'_{ur}\end{pmatrix} Var(h_0(X_i) + l_0(W_i))(\Delta_{lr}\ \Delta_{ur})\right).\end{aligned}$$

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 bootstrap method. A conservative confidence region for the identified set can be constructed as in Section 3.4.

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 3.5.2. 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,

$$\zeta(W, h_0, \tau) = \left(\frac{1+\tau}{2}\Delta'_{ur} - \frac{1-\tau}{2}\Delta'_{lr}\right)l_0(W).$$

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. Notice that from above derivation,

$$\sqrt{n}\left(\widehat{\theta}_L - \theta_L\right) \rightsquigarrow N\left(0, \Sigma_L\right),$$
$$\sqrt{n}\left(\widehat{\theta}_U - \theta_U\right) \rightsquigarrow N\left(0, \Sigma_U\right),$$

where $\Sigma_L = \Delta_{lr}' Var\left(h_0\left(X_i\right) + l_0\left(W_i\right)\right)\Delta_{lr}$ and $\Sigma_U = \Delta_{ur}' Var\left(h_0\left(X_i\right) + l_0\left(W_i\right)\right)\Delta_{ur}$. The covariance matrix can be consistently estimated by

$$E_n\left[\left(\widehat{h}\left(X_i\right) + \widehat{l}\left(W_i\right) - \overline{\widehat{h}}\left(X_i\right) - \overline{\widehat{l}}\left(W_i\right)\right)\left(\widehat{h}\left(X_i\right) + \widehat{l}\left(W_i\right) - \overline{\widehat{h}}\left(X_i\right) - \overline{\widehat{l}}\left(W_i\right)\right)'\right]$$

where $\overline{\widehat{h}}\left(X_i\right) = E_n\left[\widehat{h}\left(X_i\right)\right]$ and $\overline{\widehat{l}}\left(W_i\right) = E_n\left[\widehat{l}\left(W_i\right)\right]$. 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.

5 Empirical Application

In this section, we re-examine an application to a contingent valuation study conducted by Hanemann et al. (1991) for protecting wetland habitats and wildlife in California's San Joaquin Valley. This application has been also considered in Hanneman et al. (1991), An (2000) and Lewbel et al. (2010), among others.[4] For each respondent, a bid value was assigned randomly first and then a second bid value was also provided which is either higher or lower than the first depending on the acceptance or rejection of the first bid. The covariates include a set of continuous random variables: age, number of years resident in California, education and income; and a set of binary indicators: sex, race and membership in an environmental organization. The sample size is 518. See An (2000) for detailed description and summary statistics. As showed in Lewbel et al. (2011), there are evidence of framing effects which make the responses to the second bids less reliable. Hence, in this empirical application we focus on the data on first bids, which have 9 bid values: $\mathcal{V} = \{55, 65, 75, 80, 125, 140, 170, 210, 250\}$. As we discussed in Example 1, we need limit values of the WTP's support: for the lower bound, a natural choice is $\underline{r} \equiv U_{\min} = 0$ by noticing that there are still some respondents rejecting the bid of 55; for the upper bound, we set $\bar{r} \equiv U_{\max} = 140$, since in the data no respondents are willing to

[4] We would like to thank Arthur Lewbel for providing the data set.

pay more than this bidding value. We also try different upper bounds taken from $\{170, 210, 250\}$, and there are no qualitative differences.

The empirical literature has recognized the partial identification aspect of contingent valuations; see e.g. An (2000, p. 488). To address this problem, the standard approach has been to use some interpolation procedure, effectively assuming identification of $G(v|x)$ for all values of v in its support, see Ann (2000, p.495) and references therein. Choosing a particular interpolation method entails indirectly choosing a particular $\tau \in [0,1]$ in our setting. There is, however, no justification to focus inference on a particular member of the identified set $\theta_0(\tau)$. The same applies to other "normalization" restrictions used in the applied literature. By considering the whole interval $[0,1]$, we are dealing with the partial identification aspect, and therefore, we provide inference for the whole identified set rather than an ad-hoc choice in this set.

Due to limited sample size, nonparametric estimation is not implementable when conditioning on those binary indicators (only quite few observations for each conditioning value), hence, we instead consider a single index structure:

$$E[Y|V = v_j, X] = E\left[Y|V = v_j, X'\gamma_0\right], \text{ for all } j = 1, \ldots, J.$$

To estimate γ_0 we consider Klein and Spady's (1993) estimator with a Gaussian kernel for the link function. The asymptotic theory requires little modification relative to the fully nonparametric case discussed in the previous section, by noticing that

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\theta}(\tau) - \theta_0(\tau)\right) \\
&= -\alpha(\tau)' \cdot \sqrt{n}\left(E_n\left[\widehat{h}\left(X'\widehat{\gamma}\right)\right] - E\left[h_0\left(X'\gamma_0\right)\right]\right) \\
&= -\alpha(\tau)' \cdot \sqrt{n}\left(E_n\left[h\left(X'\gamma_0\right)\right] - E\left[h_0\left(X'\gamma_0\right)\right]\right) \\
&- \alpha(\tau)' \cdot \sqrt{n}\left(E\left[\widehat{h}\left(X'\widehat{\gamma}\right)\right] - E\left[h_0\left(X'\gamma_0\right)\right]\right) + o_P(1),
\end{aligned}$$

where the second equality is due to stochastic equicontinuity assumption, and for each j,

$$\begin{aligned}
&\sqrt{n}\left(E\left[\widehat{h}_j\left(X'\widehat{\gamma}\right)\right] - E\left[h_{0j}\left(X'\gamma_0\right)\right]\right) \\
&= \sqrt{n}E\left[\widehat{h}_j\left(X'\widehat{\gamma}\right) - h_{0j}\left(X'\widehat{\gamma}\right)\right] + \sqrt{n}E\left[h_{0j}\left(X'\widehat{\gamma}\right) - h_{0j}\left(X'\gamma_0\right)\right].
\end{aligned}$$

With the same algebra as in Section 4, we have

$$\begin{aligned}
&\sqrt{n}E\left[\widehat{h}_j\left(X'\widehat{\gamma}\right) - h_{0j}\left(X'\widehat{\gamma}\right)\right] \\
&= \frac{1}{\sqrt{n}p_j}\textstyle\sum_{i=1}^{n} I(V_i = v_j)\left(Y_i - h_{0j}\left(X'\widehat{\gamma}\right)\right) + o_P(1) \\
&= \frac{1}{\sqrt{n}p_j}\textstyle\sum_{i=1}^{n} I(V_i = v_j)\varepsilon_{ij} + p_j^{-1}\sqrt{n}E\left[I(V_i = v_j)\left(h_{0j}\left(X'\widehat{\gamma}\right) - h_{0j}\left(X'\gamma_0\right)\right)\right] + o_P(1)
\end{aligned}$$

where $\varepsilon_{ij} = Y_i - h_{0j}\left(X'\gamma_0\right)$, and the second equality follows from stochastic equicontinuity assumption.

Hence, we have the same asymptotic representation as in Section 4,

$$\begin{aligned}&\sqrt{n}\left(E\left[\widehat{h}_j\left(X'\widehat{\gamma}\right)\right]-E\left[h_{0j}\left(X'\gamma_0\right)\right]\right)\\&=\frac{1}{\sqrt{n}p_j}\sum_{i=1}^{n}I\left(V_i=v_j\right)\left(Y_i-h_{0j}\left(X'\widehat{\gamma}\right)\right)+o_P\left(1\right),\end{aligned}$$

and the results derived in Section 4 can be directly applied here, including the consistent estimator of the covariance matrix.

Following the above estimation strategy, we obtain the estimated bounds for the mean willingness-to-pay of $[85.2, 117.4]$ by using the data on first bids; if we use the data on second bids, then the estimated bounds are quite different and given by $[217.2, 234.5]$, consistent with previous findings in the literature. The point estimates in Lewbel et al. (2010) vary from 99 to 113 for the first bids and from 143 to 715 for the second bids, corresponding to different specifications and assumptions on the model. Our results for the first bids are consistent with these point estimates, but for the second bids, they suggest more precise inferences. As mentioned earlier, the results on the second bid have to be used with caution, as they might be affected by framing or anchoring effects.

Following the procedure in Section 4, we obtain the confidence interval for the true parameter value: $[84.0, 117.7]$, and the confidence region for the whole identified set: $[83.8, 118.7]$. The above results are obtained using the rule-of-thumb bandwidth and Gaussian kernel function. Different bandwidths and different kernel functions are considered, and the resulting estimated quantities are similar to the ones reported here. Also, to check sensitivity with respect to the semiparametric single-index assumption on h_0, we have considered a fully nonparametric approach with four continuous covariates. In this case, we obtain the estimated bounds for the mean willingness-to-pay of $[87.3, 117.9]$ by using the data on first bids; if we use the data on second bids, then the estimated bounds are $[217.2, 234.5]$. Hence, the results are qualitatively the same as in the semiparametric case. Similar conclusions hold for the confidence interval for the true parameter value and the confidence region for the whole identified set.

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 can be characterized as the image of a mapping indexed by unidentified nuisance parameters. Under this setting, we can employ tools from empirical process theory to derive uniform convergence results for the set estimates. Our inference methods do not require shape restrictions on the identified set, and hence have potential application in models with nonconvex 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. (2011). To approximate the asymptotic distribution of the set estimates, we propose a multiplier-type bootstrap procedure which is easy to implement and is computationally attractive. 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. 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 semiparamtric 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 models in which the finite-dimensional parameter of interest is identified, i.e. $\theta_0(\tau)$ does not vary with τ, where τ can be the typical nuisance indexing parameter for indicator or exponential functions.

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), 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 often 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.

Though not emphasized in this paper, it would be important to study efficient estimation under partial identification. Kaido and Santos (2011) apply results on semiparametric efficiency theory for infinite-dimensional parameters to define efficient estimation in convex identified sets driven by moment inequalities. In their setting, the sample analogue (i.e. the functional applied to the empirical distribution function) turns out to be semiparametrically efficient. However, in our general setting we do not expect that to be the case in many applications. This is certainly a topic that deserves further investigation.

7 Appendix: Mathematical Proofs

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

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

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

PROOF OF THEOREM 3.1: The proof closely follows that of Theorem 1 in CLV. We wish to show that for any $\varepsilon > 0$ $\Pr\left(\sup_{\tau\in\mathcal{T}_0^\delta}\left|\widehat{\theta}(\tau)-\theta_0(\tau)\right|>\varepsilon\right)\to 0$ as $n\to\infty$. Let $\varepsilon>0$ be given, the event $\left\{\sup_{\tau\in\mathcal{T}_0^\delta}\left|\widehat{\theta}(\tau)-\theta_0(\tau)\right|>\varepsilon\right\}$ implies that $\exists\tau'\in\mathcal{T}_0^\delta$ s.t. $\left|\widehat{\theta}(\tau')-\theta_0(\tau')\right|>\varepsilon$, then by Assumption A1(ii), there exists a $\eta(\varepsilon)>0$ such that $\left\|M(\widehat{\theta}(\tau'),h_0,\tau')\right\|-\|M(\theta_0(\tau'),h_0,\tau')\|\geq\eta(\varepsilon)$. Thus,

$$\Pr\left(\sup_{\tau\in\mathcal{T}_0^\delta}\left|\widehat{\theta}(\tau)-\theta_0(\tau)\right|>\varepsilon\right)\leq\Pr\left(\sup_{\tau\in\mathcal{T}_0^\delta}\{\left\|M(\widehat{\theta}(\tau),h_0,\tau)\right\|-\|M(\theta_0(\tau),h_0,\tau)\|\}\geq\eta(\varepsilon)\right).$$

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

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}_0^\delta}\{\left\|M(\widehat{\theta}(\tau),h_0,\tau)\right\|-\|M(\theta_0(\tau),h_0,\tau)\|\}\\
&\leq\sup_{\tau\in\mathcal{T}_0^\delta}\left\|M(\widehat{\theta}(\tau),h_0,\tau)-M(\widehat{\theta}(\tau),\widehat{h},\tau)\right\|\\
&+\sup_{\tau\in\mathcal{T}_0^\delta}\left\|M_n(\widehat{\theta}(\tau),\widehat{h},\tau)-M(\widehat{\theta}(\tau),\widehat{h},\tau)\right\|\\
&+\sup_{\tau\in\mathcal{T}_0^\delta}\{\left\|M_n(\widehat{\theta}(\tau),\widehat{h},\tau)\right\|-\|M(\theta_0(\tau),h_0,\tau)\|\}.
\end{aligned}$$

By Assumption A1(v), $\sup_{\tau\in\mathcal{T}_0^\delta}\left\|M_n(\widehat{\theta}(\tau),\widehat{h},\tau)-M(\widehat{\theta}(\tau),\widehat{h},\tau)\right\|=o_P(1)$; by Assumption A1(iii) and A1(iv), it holds that $\sup_{\tau\in\mathcal{T}_0^\delta}\left\|M(\widehat{\theta}(\tau),h_0,\tau)-M(\widehat{\theta}(\tau),\widehat{h},\tau)\right\|=o_P(1)$; it remains to show the last term $\sup_{\tau\in\mathcal{T}_0^\delta}\{\left\|M_n(\widehat{\theta}(\tau),\widehat{h},\tau)\right\|-\|M(\theta_0(\tau),h_0,\tau)\|\}=o_P(1)$. Assumption A1(i) implies that uniformly in $\tau\in\mathcal{T}_0$, $\left\|M_n\left(\widehat{\theta}(\tau),\widehat{h},\tau\right)\right\|\leq\inf_{\theta\in\Theta}\left\|M_n\left(\theta,\widehat{h},\tau\right)\right\|+o_P(1)$. By triangle inequality and Assumption A1(v), uniformly in (θ,τ),

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

where $o_P(1)$ is uniformly in (θ,τ). Hence,

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}_0}\left\{\inf_{\theta\in\Theta}\left\|M_n\left(\theta,\widehat{h},\tau\right)\right\|-\|M(\theta_0(\tau),h_0,\tau)\|\right\}\\
&\leq\sup_{\tau\in\mathcal{T}_0}\inf_{\theta\in\Theta}o_P(1)+\sup_{\tau\in\mathcal{T}_0}\left\{\inf_{\theta\in\Theta}\|M\left(\theta,h_0,\tau\right)\|-\|M(\theta_0(\tau),h_0,\tau)\|\right\}\\
&=o_P(1),
\end{aligned}$$

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

PROOF OF COROLLARY 3.1:

By definition, $d_H\left(\widetilde{\Theta}_0,\Theta_0\right) = \max\left\{\sup_{\tau\in\widehat{\mathcal{T}}_0} d_H(\widehat{\theta}(\tau),\Theta_0), \sup_{\tau\in\widehat{\mathcal{T}}_0} d_H(\theta_0(\tau),\widetilde{\Theta}_0)\right\}$. Define the set $\widetilde{\Theta}_0 := \left\{\theta_0(\tau) : \tau\in\widehat{\mathcal{T}}_0\right\}$, 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 $d_H\left(\widehat{\mathcal{T}}_0,\mathcal{T}_0\right) = o_P(1)$, it follows $\Pr\left(\widehat{\mathcal{T}}_0\subset\mathcal{T}_0^{\delta}\right)\to 1$. Notice that for each $\tau\in\widehat{\mathcal{T}}_0\subset\mathcal{T}_0^{\delta}$, $d_H(\widehat{\theta}(\tau),\widetilde{\Theta}_0) \le \left|\widehat{\theta}(\tau)-\theta_0(\tau)\right|$, hence $\sup_{\tau\in\widehat{\mathcal{T}}_0} d_H(\widehat{\theta}(\tau),\Theta_0) \le \sup_{\tau\in\widehat{\mathcal{T}}_0}\left|\widehat{\theta}-\theta_0\right|$, 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 $d_H\left(\widehat{\mathcal{T}}_0,\mathcal{T}_0\right) = o_P(1)$. Thus, $d_H\left(\widetilde{\Theta}_0,\Theta_0\right) = o_P(1)$ holds.

■

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 $\tau\in\mathcal{T}_0^{\delta}$. By stochastic equicontinuity, uniformly in $\tau\in\mathcal{T}_0^{\delta_0}$

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

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

then by noticing that $E_n\psi_\theta\left(W,\theta_0(\tau),h_0,\tau\right) - E\psi_\theta\left(W,\theta_0(\tau),h_0,\tau\right) = O_P\left(n^{-1/2}\right)$, $E_n\psi\left(W,\theta_0(\tau),h_0,\tau\right) - E\psi\left(W,\theta_0(\tau),h_0,\tau\right) = O_P\left(n^{-1/2}\right)$ 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 $\tau\in\mathcal{T}_0^{\delta_0}$)

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

By Assumption A2 (iii), (iv) and (v), we have uniformly in $\tau\in\mathcal{T}_0^{\delta_0}$

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

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

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 $\tau\in\mathcal{T}_0^{\delta_0}$

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

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

$$\sup_{\tau\in\mathcal{T}_0^\delta}\left|\widehat{\theta}\left(\tau\right)-\theta_0\left(\tau\right)\right|=O_P\left(n^{-1/2}\right).$$

(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

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

It follows that

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

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

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

By Assumption A2 (vii), it follows that

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

where L is a Gassian Process 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 $K(\tau_1,\tau_2) = E\left[\left(s(W,\theta_0,h_0,\tau_1)-Es(W,\theta_0,h_0,\tau_1)\right)\left(s'(W,\theta_0,h_0,\tau_2)-Es'(W,\theta_0,h_0,\tau_1)\right)\right]$ 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 COROLLARY 3.2: Similar to the proof for Corollary 1, let $\widetilde{\Theta}_0 := \left\{\theta_0(\tau) : \tau \in \widehat{\mathcal{T}}_0\right\}$, then by triangular inequality,

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

Since $d_H\left(\widehat{\mathcal{T}}_0, \mathcal{T}_0\right) = o_P(1)$, it follows $\Pr\left(\widehat{\mathcal{T}}_0 \subset \mathcal{T}_0^\delta\right) \to 1$. Notice that for each $\tau \in \widehat{\mathcal{T}}_0 \subset \mathcal{T}_0^\delta$, $d_H(\widehat{\theta}(\tau), \widehat{\Theta}_0) \leq \left|\widehat{\theta}(\tau) - \theta_0(\tau)\right|$, hence $\sup_{\tau\in\widehat{\mathcal{T}}_0} d_H(\widehat{\theta}(\tau), \Theta_0) \leq \sup_{\tau\in\widehat{\mathcal{T}}_0}\left|\widehat{\theta} - \theta_0\right|$, then Theorem 1 implies $d_H\left(\widehat{\Theta}_0, \widetilde{\Theta}_0\right) = O_P\left(n^{-1/2}\right)$. Lipschitz condition of $\theta_0(\cdot)$ and $d_H\left(\widehat{\mathcal{T}}_0, \mathcal{T}_0\right) = O_P\left(n^{-1/2}\right)$ imply $d_H\left(\widetilde{\Theta}_0, \Theta_0\right) = O_P\left(n^{-1/2}\right)$. Thus, $d_H\left(\widehat{\Theta}_0, \Theta_0\right) = O_P\left(n^{-1/2}\right)$ holds.

■

PROOF OF THEOREM 3.3: (1) By Assumption A3(iii), the function class is $\{s(w, \theta, h, \tau)u : \theta \in \Theta_\delta, h \in \mathcal{H}_\delta, \tau \in \mathcal{T}_0^\delta\}$ is also Donsker. Hence by stochastic equicontinuity and Assumption A3(i), it follows that uniformly in $\tau \in \mathcal{T}_0$

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

Assumption A3(ii) implies

$$\sup_{\tau\in\mathcal{T}_0}\left|\widehat{A}(\widehat{\theta}, \widehat{h}, \tau) - A(\theta_0, h_0, \tau)\right| = o_P(1).$$

Define $Z_{0n}^*(\tau) := A(\theta_0, h_0, \tau) n^{-1/2}\sum_{i=1}^n\left\{s(W_i, \theta_0, h_0, \tau) - E s(W_i, \theta_0, h_0, \tau)\right\}u_i$, 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 $\mathbb{G}_n\left\{\left(s(W, \theta_0, h_0, \tau) - E s(W, \theta_0, h_0, \tau)\right)u\right\}$ 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 $\ell^\infty\left(\mathcal{T}_0^\delta\right)$ with a Lipschitz norm bounded by 1, i.e. for any $f \in BL_1$, $|f(z)|_{z\in\ell^\infty\left(\mathcal{T}_0^\delta\right)} \leq 1$ and $z_1, z_2 \in \ell^\infty\left(\mathcal{T}_0^\delta\right)$, $|f(z_1) - f(z_2)| \leq \|z_1 - z_2\|_{\mathcal{T}_0^\delta}$, then by Theorem 1.12.4 in van der Vaart and Wellner (1996), it is equivalent to show

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

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

$$\begin{aligned}
&\sup_{f \in BL_1} \left| E\left[f\left(\widehat{Z}_n^*\right) - f\left(Z_{0n}^*\right) | \left\{W_i\right\}_{i=1}^n \right] \right| \\
&\leq \eta \Pr\left(\left\| \widehat{Z}_n^* - Z_{0n}^* \right\|_{\mathcal{T}_0^\delta} \leq \eta | \left\{W_i\right\}_{i=1}^n \right) + 2 \Pr\left(\left\| \widehat{Z}_n^* - Z_{0n}^* \right\|_{\mathcal{T}_0^\delta} > \eta | \left\{W_i\right\}_{i=1}^n \right).
\end{aligned}$$

By law of iterated expectations,

$$\begin{aligned}
&E\left[\Pr\left(\left\| \widehat{Z}_n^* - Z_{0n}^* \right\|_{\mathcal{T}_0^\delta} > \eta | \left\{W_i\right\}_{i=1}^n \right) \right] \\
&= \Pr\left(\left\| \widehat{Z}_n^* - Z_{0n}^* \right\|_{\mathcal{T}_0^\delta} > \eta \right) \to 0,
\end{aligned}$$

hence $\Pr\left(\left\| \widehat{Z}_n^* - Z_{0n}^* \right\|_{\mathcal{T}_0^\delta} > \eta | \left\{W_i\right\}_{i=1}^n \right) \to 0$ in probability. Since η is arbitrary, we have

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

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 $\tau \in \mathcal{T}_0$,

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

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 Proposition 3.1:
By definition of $CR_{1-\alpha,n}$ and Θ_0, let $\widetilde{CR}_{1-\alpha,n} := \bigcup_{\tau \in \mathcal{T}_0} \left\{ \theta \in \Theta : \left| \theta - \widehat{\theta}_n(\tau) \right| \leq \widehat{c}_{1-\alpha,n}/\sqrt{n} \right\}$, we have

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

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

$$\begin{aligned}&\liminf \Pr\left(\sqrt{n}\sup_{\tau\in\mathcal{T}_0}\left|\widehat{\theta}_n(\tau) - \theta_0(\tau)\right| \le \widehat{c}_{1-\alpha,n}\right)\\ &= 1-\alpha.\end{aligned}$$

■

PROOF OF THEOREM 3.4: 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 $\left\|M_n(\widetilde{\theta},\widehat{h},\widetilde{\tau}_n)\right\| \le \left\|M_n(\widetilde{\theta},\widehat{h},\widetilde{\tau})\right\|$,

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

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. $\inf_{\widetilde{\tau}\in\widetilde{\mathcal{T}}}|\widetilde{\tau}_n - \widetilde{\tau}| \le \eta_{n1} = O_P\left(n^{-1/4}\right)$. For fixed $\widetilde{\tau}_n \in \widetilde{\mathcal{T}}_{\eta_{n1}}$, there exists a $\widetilde{\tau}\in\widetilde{\mathcal{T}}$, such that $|\widetilde{\tau}_n - \widetilde{\tau}| \le O_P\left(n^{-1/4}\right)$. 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 $\sqrt{n}\sup_{\tau,h}\left\|M_n\left(\widetilde{\theta},h,\tau\right) - M\left(\widetilde{\theta},h,\tau\right)\right\| \le l_n$ with probability approaching 1. Define $\widetilde{\mathcal{T}}_{\eta_{n2}} := \{\tau\in\mathcal{T} : \exists\widetilde{\tau}\in\widetilde{\mathcal{T}} \text{ s.t. } |\tau-\widetilde{\tau}|\le\eta_{n2}\}$, i.e. the η_{n2}-enlargement of $\widetilde{\mathcal{T}}$, and $\Delta_{n2} := \inf_{\tau\in\mathcal{T}\setminus\widetilde{\mathcal{T}}_{\eta_{n2}}}\left\|M\left(\widetilde{\theta},\widetilde{h}_0,\tau\right)\right\| \ge c\eta_{n2}$ by Assumption A4(i). Hence it suffices to show $\Pr\left(M\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}_n\right) \le \Delta_{n2}\right) \to 1$ which implies $\Pr\left(\inf_{\widetilde{\tau}\in\widetilde{\mathcal{T}}}|\widetilde{\tau}_n - \widetilde{\tau}| \le \eta_{n2}\right) \to 1$. According to Assumption A4(ii),

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

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)

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

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 $B(\widetilde{\tau},\delta_n) := \{\tau \in \mathcal{T} : |\tau - \widetilde{\tau}| \leq \delta_n\}$ for all $\widetilde{\tau} \in \widetilde{\mathcal{T}}$. Then

$$\begin{aligned}
&\inf_{\tau\in\mathcal{T}}\left\|\sqrt{n}M_n(\widetilde{\theta},\widehat{h},\tau)\right\| \\
&= \inf_{\widetilde{\tau}\in\widetilde{\mathcal{T}}}\inf_{\tau\in B(\widetilde{\tau},\delta_n)}\left\|\sqrt{n}M_n(\widetilde{\theta},\widehat{h},\tau)\right\| + o_P(1)
\end{aligned} \tag{17}$$

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 (17),

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

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),

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

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

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

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 $V_\tau\left(\widetilde{\tau}\right):=V_\tau\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}\right)$):

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

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

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

where $G\left(\widetilde{\tau}\right)$ 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),

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

which leads to our conclusion that

$$T_n\rightsquigarrow\inf_{\widetilde{\tau}\in\widetilde{\mathcal{T}}}\left\|[I-P\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}\right)]G\left(\widetilde{\tau}\right)\right\|.$$

■

PROOF OF PROPOSITION 3.2: (1) Under the null, Lemma 17 in Santos (2010) indicates

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

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 3.5:

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,

$$\begin{aligned}
&\sqrt{n}\{E_n\xi(W, \widehat{g}, \tau) - \tau'\widetilde{\theta}\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\xi(W_i, g_0, \tau) - E\xi(W_i, g_0, \tau) + \zeta(W_i, g_0, \tau)\} \\
&+ \sqrt{n}\{E\xi(W, g_0, \tau) - \tau'\widetilde{\theta}\} + o_P(1).
\end{aligned} \tag{18}$$

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

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

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

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

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

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

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 $d_H\left(\widetilde{\mathcal{S}}, \widetilde{\mathcal{S}}_{n,b_n}\right) = o_P(1)$. Notice that $b_n\sqrt{n} \to \infty$, we have

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

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

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

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

We next show that with probability approaching one,

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

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

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

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 $E\xi(W, g_0, \tau) - \tau'\widetilde{\theta} \ge 0$ for all τ, hence $\inf_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}} T_n^0(\tau) \ge \inf_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}} \sqrt{n}G_n(\tau)$; by definition of $\widetilde{\mathcal{S}}_{n,b_n}$, for any $\tau \in \mathbb{S}^{d_\theta}\setminus\widetilde{\mathcal{S}}_{n,b_n}$, $T_n^0(\tau) > b_n\sqrt{n}$, hence $\inf_{\tau\in\mathbb{S}^{d_\theta}\setminus\widetilde{\mathcal{S}}_{n,b_n}} T_n^0(\tau) \ge b_n\sqrt{n} > \inf_{\tau\in\widetilde{\mathcal{S}}_{n,b_n}} \sqrt{n}G_n(\tau)$ 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 (18),

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

■

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 (2008): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure", *Econometrica*, forthcoming.

[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 (2011): "Underidentification?", Working paper.

[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 (2009), "Set Identified Linear Models," *Econometrica*, forthcoming.

[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] 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.

[22] Escanciano, J. C. and L. Zhu (2011): "Inferences in Partially Identified Structural Models," working paper.

[23] 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.

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

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

[26] 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.

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

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

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

[30] ——— (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, 77-84.

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

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

[33] Kaido, H. and A. Santos (2011): "Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities," Working paper.

[34] Kemperman, J. H. B. (1968): "The General Moment Problem, A Geometric Approach," *The Annals of Mathematical Statistics*, 39, 93-122.

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

[36] 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.

[37] 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.

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

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

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

[41] ———— (2003): *Partial Identification of Probability Distributions* (Berlin: Springer-Verlag).

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

[43] Mikusheva, A. (2007): "Uniform inference in autoregressive models," *Econometrica*, 75 1411–1452.

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

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

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

[47] Romano, J. P. (1989): "Do bootstrap confidence procedures behave well uniformly in P?," *The Canadian Journal of Statistics*, 17 75–80.

[48] 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.

[49] ——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169-211.

[50] ——— (2010): "On the Uniform Asymptotic Validity of Subsampling and the Bootstrap", Working paper.

[51] 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.

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

[53] 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.

[54] 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.

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

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

[57] 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.

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

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

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

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