Set Inferences and Sensitivity Analysis in Semiparametric Conditionally Identified Models*

Juan Carlos Escanciano†
Indiana University

Lin Zhu‡
Tsinghua University

October 22nd, 2013

Abstract

This paper provides tools for partial identification inference and sensitivity analysis in a general class of semiparametric models. The main working assumption is that the finite-dimensional parameter of interest and the possibly infinite-dimensional nuisance parameters are identified conditionally on other nuisance parameters being known. This structure arises in numerous applications and leads to relatively simple inference procedures. The paper develops uniform convergence for a set of semiparametric two-step GMM estimators, and it uses the uniformity to establish set inferences, including confidence regions for the identified set and the true parameter. Sensitivity analysis considers a domain of variation for the unidentified parameter that can be well outside its identified set, which demands inference to be established under misspecification. The paper also introduces new measures of sensitivity. Inferences are implemented with new bootstrap methods. Several example applications illustrate the wide applicability of our results.

Keywords: Partial Identification; Semiparametric models; Sensitivity analysis.

JEL classification: C14, C21, C25

*We would like to thank Stéphane Bonhomme, Yanqin Fan, Kei Hirano, Arthur Lewbel, Adam Rosen, Enrique Sentana, as well as seminar participants at many institutions for helpful comments and discussions. This is a revised version of the paper circulated under the title: "Inferences in Semiparametric Partially Identified Models: An Empirical Process Approach," dated on December 8th, 2011.

†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 ECO2012-33053.

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

1 Introduction

In many economic models, point identification of parameters of interest is often at the cost of ad-hoc assumptions. Relaxing these assumptions in general leads to a loss of point identification, and instead only a strict (non-singleton) subset of the parameter space can be identified from a large sample, a situation that has been referred to as partial identification or set identification; see Manski (2003, 2007) for textbook treatments and Tamer (2010) for a survey on recent theoretical and empirical developments. The literature on partial identification has increasingly grown over the last two decades, but much of this literature has been restricted to models involving only finite-dimensional parameters. Furthermore, there are rather few tools available for applied economists to quantify the sensitivity of inferences to critical identifying assumptions with different degrees of identification power. In this paper, we propose a new unified framework and new tools to implement inference and sensitivity analysis in a broad class of partially identified semiparametric models.

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

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

where W is a d_w-dimensional observable random vector, and $\psi(\cdot, \theta, h(\cdot), \tau)$ is a measurable moment function from $\mathbb{R}^{d_w}$ to $\mathbb{R}^{d_\psi}$, for each $\gamma := (\theta, h(\cdot), \tau) \in \Gamma := \Theta \times \mathcal{H} \times \mathcal{T}$. We are particularly concerned with situations where θ_0, the parameter of interest, is not identified by the moment restrictions (1). However, we find plausible that in many applications the parameter θ is conditionally identified in the sense that, for each fixed $\tau \in \mathcal{T}_0 \subset \mathcal{T}$ and for an identified $h_0(\cdot, \tau) \in \mathcal{H}$, there exists a unique solution to (1) in θ_0, say $\theta_0(\tau) \in \Theta$.[1] The conditional identification implies that the identified set for θ is the finite-dimensional manifold

$$\Theta_0 := \{\theta \in \Theta : \theta = \theta_0(\tau) \text{ for some } \tau \in \mathcal{T}_0\},$$

where $\mathcal{T}_0 := \{\tau \in \mathcal{T} : E[\psi(W, (\theta_0(\tau), h_0(\cdot, \tau), \tau))] = 0\}$ is possibly unknown.

This setting turns out to be general enough to be applicable to many parametric and semiparametric partially identified econometric models considered in the literature, and to many other new applications, while permitting the use of relatively standard methods of analysis. It provides a framework under which sensitivity to identification can be quantitatively assessed. Roughly speaking, the nuisance parameter τ embodies assumptions that "complete" the model, and variation of $\theta_0(\tau)$ in τ quantifies the sensitivity of the parameter of interest to these assumptions. The set $\mathcal{T}_0$ could be interpreted as the set of observationally equivalent identifying assumptions. Sensitivity analysis often considers variation of $\theta_0(\tau)$ on sets $\mathcal{T}_1$ with $\mathcal{T}_1 \nsubseteq \mathcal{T}_0$, and the theory we propose accommodates this possibility.

This paper provides the following contributions within the setting described above: (i) it establishes uniform (in τ) asymptotic results for semiparametric two-step Generalized Method of Moments

[1] We assume for simplicity a two-step setting where $h_0(\cdot, \tau)$ is identified prior and independently to the identification of θ. A leading example is when $h_0(X, \theta, \tau) = E[\rho(W, \theta, \tau)|X]$, for some $\rho(\cdot)$. Example 1 below shows how our results can be easily extended to models with an infinite number of restrictions, therefore relaxing the prior identification of h_0.

(GMM) estimators for $\theta_0(\tau)$ in (1), with non-smooth moment functions and possibly misspecified models, thereby extending previous results by Chen, Linton and Van Keilegom (2003) (CLV hereafter)[2]; (ii) it justifies approximations of limiting distributions and standard errors via a multiplier-type bootstrap; (iii) it proposes set inferences, including confidence regions for the identified set Θ_0 and the true parameter, allowing for the nuisance identified set $\mathcal{T}_0$ to be unknown and estimated; and (iv) it formalizes sensitivity analysis for $\theta_0(\tau)$ on $\tau \in \mathcal{T}_1 \nsubseteq \mathcal{T}_0$. In particular, we propose new inferences for the sensitivity set $\Theta_1 = \{\theta_0(\tau) : \tau \in \mathcal{T}_1\}$, and new measures of sensitivity, such as $\partial\theta_0(\tau)/\partial\tau$ at $\tau \in \mathcal{T}_1$.

We illustrate the previous ideas with two generic examples below. The first example investigates inferences about bounded linear functionals of a function that is known to lie within a band. This is an example that has numerous applications in economics, see Manski (2003, 2007). We show how our results can be used to extend recent results by, e.g., Chandrasekhar, Chernozhukov, Molinari and Schrimpf (2012), Kline and Santos (2013) and Pacini (2012) from best linear approximation functionals to general linear continuous functionals, including but not restricted to average partial effects, gender gap distributional effects, counterfactual distributional effects, consumer surplus, and functionals of the joint distribution of outcomes in treatment effects such as the proportion of individuals who benefited from the treatment, to mention just a few. A second example application discusses set inferences and sensitivity analysis in semiparametric models with no exclusion restrictions. This example complements and extends related ideas in Conley, Hansen and Rossi (2012), derived there for parametric linear instrumental variables (IV) models, to semiparametric double-index models, including limited dependent variable models with endogeneity and sample selection models as special cases. These two examples serve to illustrate the wide applicability of the proposed procedures. The first example leads to convex identified sets with semiparametric support functions[3], and it also shows how our results can easily accommodate an infinite number of moment restrictions, which includes conditional moment restrictions or models with a continuum of quantiles as special cases. The second example yields identified sets which are in general non-convex non-linear manifolds. A previous version of the paper, Escanciano and Zhu (2012), contains further detailed examples and an empirical application of the techniques discussed here to estimating functionals of willingness-to-pay in a contingent valuation study.

The paper is organized as follows: After this introduction and a literature review, Section 3 illustrates the wide applicability of our methods with two motivating examples. Section 4 develops uniform inferences for the possibly misspecified version of the conditionally identified model in (1). Section 5 applies the uniform results to construct confidence regions for the identified set and the true parameter. Other applications in this section include a formalization of sensitivity analysis and inferences incorporating prior knowledge on the nuisance parameters. Section 6 considers some applications motivated from the examples in Section 3, showing how our conditions can be generally verified in each of them. Finally, Section 7 concludes and discusses possible extensions. Mathematical proofs are gathered in the Appendix.

[2] We extend CLV to partially identified semiparametric models and to misspecified models. The analysis under misspecification is needed here to handle the case where $\mathcal{T}_0$ is unknown and estimated, and also for sensitivity analysis. Under misspecification the definition of $\theta_0(\tau)$ has to be modified accordingly; see Section 4.

[3] For a definition of support function of a convex set see Section 2.

2 Literature Review

Our approach to partial identification has important precedents in the literature. To the best of our knowledge, this setting was first proposed for modeling partial identification by Sargan (1959) in a context of IV models. In the classical setting of demand and supply simultaneous equations, Leamer (1981) shows that the non-convex identified set for demand and supply elasticities can be written as in (1). 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 parameterizations fitting our setting with identified $\theta_0(\tau) \equiv \theta_0$ and τ completely unidentified. He further investigated the asymptotic theory of estimators and test statistics under partial identification. More recently, Altonji, Elder and Taber (2005) study the impact of attending a Catholic school on future educational attainment. In their application $\theta_0(\tau)$ is the coefficient of an endogenous binary variable in a bivariate probit model and τ denotes the correlation of the unobservable errors in the first and second stage equations.

This earlier approach to partial identification has been recently extended to general parametric moment restrictions by Arellano, Hansen and Sentana (2012). These authors provide a test for underidentification and example applications in linear IV models, dynamic panel data, Phillips curves and asset pricing models. Conley et al. (2012) and Nevo and Rosen (2012) consider set inferences in linear IV models satisfying (1). Further example applications in the context of dynamic models include structural vector autoregressions with sign restrictions, see, e.g., Rubio-Ramirez, Waggoner and Zha (2010), non-fundamental moving average representations in Lippi and Reichlin (1993,1994) or incomplete asset pricing models in Kaido and White (2009), to mention but a few. More generally, in dynamic macroeconomic models calibration has been routinely employed, which may be interpreted as a pointwise version of our partial identification approach in partially identified settings.[4] The inferences we propose here are substantially different from those in the aforementioned papers and the setting considered is significantly more general.

The scope of applications of our methods is considerably broader than the examples mentioned above; see Manski (2003, 2007). Conditionally identified models arise naturally in situations with missing, contaminated, misclassified or censored data, see, e.g., Horowitz and Manski (1995, 2000, 2006) and references below. In these models τ may parameterize 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 conditional selection probability. Our results can be used to complement their pointwise results with uniform inferences in semiparametric versions of their model.

There is also an extensive literature using sensitivity analysis as a method to quantify the impact of relaxing strong identification assumptions on parameters of interest, see, e.g., Rosenbaum and Rubin

[4] An exception to the pointwise inference in macro models is Faust, Swanson and Wright (2004), who constructed confidence intervals for scalar parameters of interest by taking the union of confidence intervals indexed by structural parameters which are not identified.

(1983), Rosenbaum (2002), Imbens (2003), or more recently, Kline and Santos (2013). In this literature the focus is not on set inference per se, but in the modulus of continuity of $\theta_0(\tau)$, for example, how large has to be τ (e.g. a measure of selection on unobservables) to obtain a "large" deviation of $\theta_0(\tau)$ from $\theta_0(\tau_0)$, where τ_0 is a benchmark case, e.g. $\tau_0 = 0$ (no selection). Our inferences deal with the whole curve $\theta_0(\tau)$, and as such, they include sensitivity analysis as a special case. Another important difference between partial identification analysis and sensitivity analysis is that in the former $\tau \in \mathcal{T}_0$, whereas in the latter typically $\tau \in \mathcal{T}_1 \nsubseteq \mathcal{T}_0$. This implies that sensitivity analysis should account for misspecification, a result that has been overlooked in the literature. These arguments led us to consider a generic set $\mathcal{T}_1 \subset \mathcal{T}$ and to allow for misspecification in our asymptotic results.[5]

Although the literature on sensitivity analysis has been mainly confined to parametric models, there are some applications in semiparametric models as well. For instance, Scharfstein, Rotnitzky and Robins (1999) study how non-ignorable drop-out affects inferences on the mean of the outcome of interest in a semiparametric model for panel data. They show 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 point-identified, and they carry out a pointwise sensitivity analysis by varying the selection bias parameter τ over a plausible (finite) range of values. More recently, Kline and Santos (2013) study sensitivity in missing data problems by setting τ to be a Kolmogorov-Smirnov distance between the distributions of missing and observed outcomes. We show in Example 1 below how our framework can accommodate inferences on parameters such as average partial effects in a general setting that includes that of Kline and Santos (2013).

A large class of models for which our results are applicable are convex identified sets. Bontemps, Magnac and Maurin (2012) consider linear "incomplete" models satisfying (1) with an infinite-dimensional nuisance parameter τ but without 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 lead to a convex and closed identified set, and Bontemps et al. (2012) exploit the convexity to develop inference based on the support function of the identified set, which is defined as $\beta(q|\Theta_0) := \sup_{\theta \in \Theta_0} q'\theta$, for all $q \in \mathbb{S}_p := \{q \in \mathbb{R}^p : q'q = 1\}$. In their model, the support function is the expectation of a suitable function indexed by $q \in \mathbb{S}_p$. Beresteanu and Molinari (2008) propose inference based on support functions for general models with convex identified sets. As shown by these authors, inference about the support function can be also carried out in the setting of (1) where τ now denotes a direction q with $\mathcal{T}_0 = \mathbb{S}_p$ and θ_0 is the value of the support function in that direction, i.e. $\theta_0(\tau) = \beta(\tau|\Theta_0)$. Our results then extend those in Beresteanu and Molinari (2008) and Bontemps et al. (2012) by permitting the moments characterizing the support function to depend on infinite-dimensional nuisance parameters as well.[6]

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

[6] Examples within this class of models are given by, e.g., Manski (2003, 2007), Blundell, Gosling, Ichimura and Meghir

Our paper belongs to the rapidly growing literature on inferences in partially identified models.[7] When the identified set is a closed interval, Horowitz and Manski (2000) develop confidence intervals for the entire identified set, while Imbens and Manski (2004) and Stoye (2009) discuss methods for constructing confidence intervals for the true value. In a general setup, Chernozhukov, Hong and Tamer (2007) develop a unified criterion function approach for estimation and inference in partially identified models, generalizing results in M-estimation theory from point identification to partial identification. They show the consistency of their level set estimator and obtain rates of convergence. Inference is based on subsampling. For an alternative proposal in the same setting see Romano and Shaikh (2008, 2010). Moment inequalities are leading examples of this literature, see e.g. Andrews and Jia (2012), Andrews and Guggenberger (2009), Andrews and Soares (2010) and Bugni (2010), among others. Models with convex identified sets are investigated in Beresteanu and Molinari (2008). See also Beresteanu, Molchanov and Molinari (2011, 2012). Recently, Kaido (2012) has investigated the connections between the criterion function approach and the support function approach when the identified set is convex. These aforementioned papers deal with partially identified models with no infinite-dimensional nuisance parameters to be estimated.

The literature on general semiparametric partially identified models is more recent and rather scarce. Song, Kosorok and Fine (2009) use profile likelihood methods to propose optimal tests in semiparametric models with parameters that are not identified under the null, extending previous results by Andrews and Ploberger (1994). Chen, Tamer and Torgovitsky (2011) propose inverting profiled likelihood ratio tests to construct confidence sets for finite-dimensional parameters. These works allow for the finite-dimensional parameter to be estimated at a slower rate than the regular (parametric) one.[8] Hong (2012) considers semiparametric conditional moment models, with infinite-dimensional parameters approximated by sieves, extending previous important results by Santos (2012). Our paper differs considerably from these existing papers in the setting and objectives. We consider a two-step approach, and more importantly, our results exploit the conditional identification assumption (identified sets parameterized by τ), which leads to simple and efficient implementations in our setting. Thus, sieve nonparametric methods and our methods are complements rather than substitutes, and they can be potentially combined in extensions of our basic setting where τ is infinite-dimensional.[9]

3 Motivating Examples

The following examples illustrate the wide applicability of our methods. Henceforth, A' denotes the transpose of the matrix or vector A, $1(B)$ denotes the indicator of the event B (=1 if B occurs

(2007), Fan and Zhu (2009), Lee (2009), Chandrasekhar et al. (2012), Pacini (2012) and Kline and Santos (2013), to mention but a few.

[7] This literature is too extensive to be fully discussed here. The reader is referred to Tamer (2010) for a survey.

[8] 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).

[9] All our results, with the exception of Theorem 5.2, are directly applicable to an infinite-dimensional τ. A sieve approach, combined with our uniform results, can be used to extend Theorem 5.2. This extension is, however, beyond the scope of this paper, and it is deferred for future research.

and =0 otherwise), for a generic real-valued function $r(x)$, $r^+(x) := r(x)1(r(x) > 0)$ and $r^-(x) := r(x)1(r(x) < 0)$ denote the positive and negative parts, respectively, and for a generic random vector X, F_X, f_X and $\mathcal{S}_X$ denote the cumulative distribution function (cdf), (Lebesgue) density and support of X, respectively.

Example 1 (Bounded linear functionals of band-identified functions) In many partially identified models, the function of interest is known to lie within a band; see Manski (2003, 2007) and Chandrasekar et al. (2012) for numerous interesting examples, including regression with interval data, sample selection and quantile treatment effects. In this setting there is a real-valued function of interest, which is denoted by $\varphi(x, \alpha)$, e.g., a structural mean, distribution or quantile function conditional on $X = x$. It is known that φ is identified within a band

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

for identified lower and upper functions $l(x, \alpha)$ and $u(x, \alpha)$, respectively, which can be estimated nonparametrically. The index $\mathcal{A}$ may denote, for instance, the set of quantiles of interest and/or other parameters measuring the level of missingness in the data, as in Kline and Santos (2013). Let $L_2(F_X)$ be the Hilbert space of square-integrable measurable functions of X. Assume $\varphi(\cdot, \alpha) \in L_2(F_X)$ for each $\alpha \in \mathcal{A}$. Suppose we are interested in a linear bounded functional of $\varphi(x, \alpha)$, say $\theta_0(\alpha) = T\varphi(\cdot, \alpha)$. For instance, T might be an average incremental effect functional $T\varphi(\cdot, \alpha) = E[\varphi(1, X_2, \alpha) - \varphi(0, X_2, \alpha)]$, where $X = (X_1, X_2)$ and $X_1 \in \{0, 1\}$ is a binary variable (e.g., gender or treatment indicator), or a best linear approximation functional $T\varphi(\cdot, \alpha) = E[XX']^{-1} E[X\varphi(X, \alpha)]$ as in Chandrasekar et al. (2012), Kline and Santos (2013) or Pacini (2012), or a counterfactual distribution functional effect $T\varphi(\cdot, \alpha) = \int \varphi(x, \alpha)F^*(dx)$, where $F^*(x)$ is an identified counterfactual distribution for covariates, see, e.g., Chernozhukov, Fernández-Val and Melly (2013).[10]

In the general case, by the Riesz Representation Theorem there exists a function $r \equiv r_T \in L_2(F_X)$, called the Riesz representer of T, such that

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

and the identified set for $\theta_0(\alpha)$ is the closed interval $[l(\alpha), u(\alpha)]$, where $l(\alpha) := E[L(X, \alpha)]$, $u(\alpha) := E[U(X, \alpha)]$,

$$L(X, \alpha) := l(X, \alpha)r^+(X) + u(X, \alpha)r^-(X)$$

and

$$U(X, \alpha) := u(X, \alpha)r^+(X) + l(X, \alpha)r^-(X).$$

For example, for $T\varphi(\cdot, \alpha) = E[\varphi(1, X_2, \alpha) - \varphi(0, X_2, \alpha)]$, the Riesz representer is

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

[10]We focus on univariate functionals but our results can be trivially extended to multivariate ones by using one-dimensional projections, as with the support function of a convex set.

where $p(x_2) := P[X_1 = 1|X_2 = x_2]$ is a propensity score, assumed to satisfy $0 < p(x_2) < 1$, for all $x_2 \in \mathcal{S}_{X_2} \subset \mathbb{R}^{d_{x_2}}$. For $T\varphi(\cdot, \alpha) = E[\partial\varphi(X, \alpha)/\partial x_1]$, $r(x) = -\partial \log f_X(x)/\partial x_1$. Notice that for many examples of T the corresponding r is nonparametric, which may lead to moment functions that are non-smooth in an infinite-dimensional nuisance parameter. Our semiparametric setting, much like CLV, allows for this possibility. Interestingly enough, in this new generic example with a nonparametric representer r there is no asymptotic contribution from estimating r in the non-smooth indicators, which facilitates inference. See Section 8.2 in the Appendix for details.

We can write this general example into our setting by letting $\theta_0(\tau) = \lambda l(\alpha) + (1 - \lambda)u(\alpha)$, with $\tau = (\lambda, \alpha) \in \mathcal{T}_0 := [0, 1] \times \mathcal{A}$. That is, the set of solutions corresponding to the moment functions

$$\psi(W, \theta_0, h_0, \tau) = \theta_0 - \lambda L(W, \alpha) - (1 - \lambda)U(W, \alpha),$$

where $W = X$ and $h_0(\cdot, \tau) \equiv h_0(\cdot, \alpha) = (l(\cdot, \alpha), u(\cdot, \alpha), r(\cdot))$.

The application of our procedures to this class of examples leads to extensions of previous results by Chandrasekar et al. (2012), who deal with best linear approximation functionals. Special examples include average partial or increment effects, average structural functions, counterfactual distributions or functionals of the joint distribution of outcomes in treatment effects such as the proportion of individuals who benefited from the treatment.[11] This generic example also illustrates how our setting can be modified to accommodate an infinite number of moment restrictions and/or convex identified sets with semiparametric support functions. Section 6 below discusses in detail an application of our results to construct uniform (in τ) confidence bands for gender gap wage distributional effects.

Example 2 (Semiparametric double-index models with no exclusion restrictions) In many applications in microeconometrics exclusion restrictions, in the form of zero coefficients in an outcome equation, are imposed to achieve point identification. Economic theory is often silent about such restrictions, and, as we show below, the sensitivity of the conclusions can be assessed by fixing such coefficients to τ, rather than zero, and proceeding with a formal sensitivity analysis. Conley et al. (2012) applied a similar idea in a linear IV setting, where τ is the coefficient of the instruments in the structural equation. See also Altonji et al. (2005) for an alternative approach in a setting with no exclusion restrictions. We show how our results can be used to investigate sensitivity to exclusion restrictions in non-linear and possibly non-separable semiparametric models. Non-linear and non-separable models are representative of situations where the verification of the conditional identification assumption is non-trivial. Likewise, these are typical applications where $\mathcal{T}_0$ may be unknown and needs to be estimated.

Many semiparametric models, including models with endogenous regressors estimated by control function or limited dependent variable models with selection, lead to double-index restrictions of the form $E[Y|X] = F_0[X'\beta_0, g_0(X)]$; see, e.g., Blundell and Powell (2004), Escanciano, Jacho-Chávez and

[11] Our setting here does not provide sharp bounds for cases where φ satisfies additional restrictions to (2), such as having identified linear functionals; see, e.g., Firpo and Ridder (2010) and Fan and Zhu (2009) for results on the joint distribution of outcomes in treatment effects, and see also Heckman and Vytlacil (2007, Section 10) for a survey of partial identification in this context.

Lewbel (2012) and references therein for numerous examples. This setting includes a wide class of selection models, such as standard Heckman-type selection models, extensions of Tobit models like double hurdle models, and censored binary choice models, among others.

A researcher estimating this model would impose that one of the coefficients in β_0 is zero, say that of X_1, where $X = (X_1, X_2, X_3')'$. Suppose, however, that she is concerned about the sensitivity of inferences to the maintained exclusion restriction. Then, she can write $X'\beta_0 = \tau X_1 + X_2 + \theta_0' X_3$ (here the coefficient of X_2 is normalized to one since F_0 is unknown). In this application $h_0(\cdot, \tau) = (F_0(\cdot, \tau), g_0)$. We show below that our conditional identification assumption holds under mild conditions when we take as moments the score equations from the semiparametric least squares (SLS) estimator of Escanciano et al. (2012). In this example, the mapping $\tau \to \theta_0(\tau)$ is nonlinear, the nuisance parameter F_0 depends on both $\theta_0(\tau)$ and τ, and the set $\mathcal{T}_0$ is unknown. A robust approach to identification, as the one suggested here, does not impose that $\mathcal{T}_0$ is a singleton. Similarly, sensitivity analysis can be carried out by estimating $\theta_0(\tau)$ over a set $\mathcal{T}_1$ that includes the exclusion restriction $\tau = 0$. A particular case of this semiparametric double-index model is investigated in more detail in Section 6.

4 Uniform Inference

We first elaborate further on the model introduced in (1). Notice that, though we do not make it explicit in (1), the nuisance function $h_0(\cdot)$ may contain (θ, τ) as additional arguments (see, e.g., Example 2 above). Our estimation and inference results are developed to account for potential misspecification of the moment restriction (1), and we also discuss simplification of the results when the model is correctly specified. The asymptotic results developed below for semiparametric GMM estimation under global misspecification with non-smooth moment functions are of independent interest and complement some of the previous results given by Hall and Inoue (2003) and Ai and Chen (2007) for smooth moments and point-identified models.

We assume that for each fixed value of $(\theta, \tau) \in \Theta \times \mathcal{T}_1$, there is available a first-step nonparametric estimator $\widehat{h}(\cdot)$ for $h_0(\cdot)$ with certain convergence properties as specified in Assumption A1 and A2 below. Throughout we use the following notation. Let $|\cdot|$ denote the Euclidean norm, i.e. $|A| := (tr(A'A))^{1/2}$, where $tr(A)$ is the trace of the matrix A. Let $vec(A)$ denote the vectorization of matrix A and $\otimes$ denote the Kronecker product. For a measurable function g of W, define the norms $\|g\|_\infty = \sup_{w \in \mathcal{S}_W} |g(w)|$ and $\|g\|_r := (E[|g(W)|^r])^{1/r}$. 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. $(\theta, h, \tau) := (\theta, h(\cdot, \theta, \tau), \tau)$.

Suppose the observed data $\{W_i\}_{i=1}^n$ are an independent and identically distributed (i.i.d.) sequence of random vectors following the same distribution as W. For a measurable function f we denote the empirical expectation and empirical process by

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

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 $\|A\| = (tr(A'\Sigma A))^{1/2}$ for some fixed symmetric positive definite matrix Σ. Our theory can be easily extended to the case where Σ depends on τ and is estimated, provided the model is correctly specified. However, when the model is misspecified such extension is cumbersome; see Ai and Chen (2007, p.10) for discussion. Thus, for simplicity we assume hereafter that Σ is known. To develop asymptotic results under misspecification, we maintain the conditional identification assumption throughout, that is, for each $\tau \in \mathcal{T}_1$, we assume $\|M(\theta, h_0, \tau)\|$ is uniquely minimized at $\theta_0(\tau)$.[12] For each fixed $\tau \in \mathcal{T}_1$, we consider a two-step GMM-estimator for $\theta_0(\tau)$:

$$\widehat{\theta}(\tau) := \arg\min_{\theta \in \Theta} \left\| M_n\left(\theta, \widehat{h}, \tau\right)\right\|, \tag{3}$$

where $\widehat{h}(\cdot)$ is the first-step estimator of $h_0(\cdot)$ for each fixed pair (θ, τ). In some 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. Denote $Z_n(\tau) := \sqrt{n}(\widehat{\theta}(\tau) - \theta_0(\tau))$. Under mild regularity conditions, $Z_n(\cdot)$ belongs to the Banach space $\ell^\infty(\mathcal{T}_1)$ of uniformly bounded functions on $\mathcal{T}_1$, which is equipped with the sup-norm, $\|z\|_{\mathcal{T}_1} := \sup_{\tau \in \mathcal{T}_1} |z(\tau)|$. In this paper we consider weak convergence, denoted by $\rightsquigarrow$, of $Z_n(\cdot)$ in the metric space $\ell^\infty(\mathcal{T}_1)$ endowed with the sup-norm in the sense of J. Hoffmann-Jørgensen (see, e.g., Dudley 1999 p. 94). For simplicity of presentation, we abstract from measurability issues that may arise as a result of using the space $\ell^\infty(\mathcal{T}_1)$ endowed with the sup-norm. See, e.g., van der Vaart and Wellner (1996) for a formal treatment of lack of measurability and for definitions of Glivenko-Cantelli and P-Donsker classes.

4.1 Consistency

In this section we discuss the consistency of the estimator defined in (3) under the following assumptions.

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

$$\sup_{\tau \in \mathcal{T}_1} \left\{ \left\| M_n\left(\widehat{\theta}(\tau), \widehat{h}, \tau\right)\right\| - \inf_{\theta \in \Theta} \left\| M_n\left(\theta, \widehat{h}, \tau\right)\right\| \right\} = o_P(1). \tag{4}$$

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

$$\inf_{\tau \in \mathcal{T}_1} \left\{ \inf_{\theta: |\theta - \theta_0(\tau)| \geq \varepsilon} \|M(\theta, h_0, \tau)\| - \|M(\theta_0(\tau), h_0, \tau)\| \right\} \geq \eta(\varepsilon). \tag{5}$$

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

$$\sup_{\theta \in \Theta, \tau \in \mathcal{T}_1, \|h - h_0\|_{\mathcal{H}} \leq \delta_n} \|M_n(\theta, h, \tau) - M(\theta, h, \tau)\| = o_P(1). \tag{6}$$

[12] Under misspecification $\theta_0(\tau)$ is a pseudo-parameter; see White (1982).

These assumptions are uniform versions of those in CLV for consistency. Like these authors, we also allow for non-smooth moment functions $\psi(\cdot)$ as long as $M(\cdot)$ is continuous. Consistency and rates of convergence for nonparametric estimators indexed by nuisance parameters are investigated in Andrews (1995) or Escanciano, Jacho-Chávez and Lewbel (2013) for kernel estimates, or by Song (2008) for series estimators. These results can be used to verify A1(iv). Assumption A1(v) is implied by a Glivenko-Cantelli property of the class $\Psi := \{\psi(\cdot, \theta, h, \tau) : \theta \in \Theta,\ h \in \mathcal{H},\ \tau \in \mathcal{T}_1\}$, for which well known sufficient conditions are available in the literature, see, e.g., van der Vaart and Wellner (1996).

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

Theorem 4.1 *Under Assumption A1, it holds that* $\sup_{\tau\in\mathcal{T}_1} \left|\widehat{\theta}(\tau) - \theta_0(\tau)\right| = o_P(1)$.

4.2 Weak Convergence

Consider now the weak convergence of $Z_n(\cdot) = \sqrt{n}(\widehat{\theta}(\cdot) - \theta_0(\cdot))$ as a stochastic process in $\ell^\infty(\mathcal{T}_1)$ endowed with the sup-norm $\|\cdot\|_{\mathcal{T}_1}$. Given consistency, we can work, as usual, in a small or even shrinking neighborhood of $\theta_0(\cdot)$ and h_0. Define the sensitivity set $\Theta_1 := \{\theta_0(\tau) : \tau \in \mathcal{T}_1\}$. With some abuse of the notation, define a δ-expansion of the parameter sets $\Theta_1^\delta := \{\theta \in \Theta : \inf_{\theta_1\in\Theta_1} |\theta - \theta_1| \leq \delta\}$ and $\mathcal{H}^\delta := \{h \in \mathcal{H} : \|h - h_0\|_{\mathcal{H}} \leq \delta\}$, where the pseudo-metric $\|\cdot\|_{\mathcal{H}}$ is also modified according to the smaller parameter set Θ_1^δ. We first introduce the definition of pathwise functional derivative to deal with the estimation effects of $\widehat{h}$. For each $(\theta, \tau) \in \Theta_1^\delta \times \mathcal{T}_1$, we say that $M(\theta, h, \tau)$ is pathwise differentiable at $h \in \mathcal{H}^\delta$ in the direction $\left[\overline{h} - h\right]$ if $\left\{h + \lambda\left(\overline{h} - h\right) : \lambda \in [0, 1]\right\} \subset \mathcal{H}^\delta$ and

$$\lim_{\lambda\to 0} \frac{M\left(\theta, h + \lambda\left(\overline{h} - h\right), \tau\right) - M(\theta, h, \tau)}{\lambda} \text{ exists.}$$

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

Assumption A2: Suppose that $\theta_0(\tau)$ is in the interior of Θ for each $\tau \in \mathcal{T}_1$, and that $\sup_{\tau\in\mathcal{T}_1} |\widehat{\theta}(\tau) - \theta_0(\tau)| = o_P(1)$. In addition, assume:

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

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

(ii) The estimator $\widehat{\theta}(\tau)$ satisfies that uniformly in $\tau \in \mathcal{T}_1$,

$$\left(E_n\psi_\theta\left(W, \widehat{\theta}(\tau), \widehat{h}, \tau\right)\right)' \Sigma E_n\psi\left(W, \widehat{\theta}(\tau), \widehat{h}, \tau\right) = o_P\left(n^{-1/2}\right). \tag{7}$$

(iii) Smoothness in θ : (a) for each $\tau \in \mathcal{T}_1$, the map $\theta \to M(\theta, h_0, \tau)$ is twice continuously differentiable at θ_0, with first-order derivative $V_{\theta_0}(\tau)$. Furthermore, (b) suppose that $A_0(\tau) := V_{\theta_0}(\tau)'\Sigma V_{\theta_0}(\tau) + (M(\tau)'\Sigma \otimes I_{d_\theta})V_{\theta_0\theta_0}(\tau)$ is of full rank for all $\tau \in \mathcal{T}_1$, $\sup_{\tau\in\mathcal{T}_1} |A_0(\tau)| < \infty$ and $\sup_{\tau\in\mathcal{T}_1} \left|A_0^{-1}(\tau)\right| < \infty$.

(iv) Smoothness in h : (a) for each $(\theta,\tau) \in \Theta_1^{\delta} \times \mathcal{T}_1$, the pathwise derivative $V_h(\theta,h_0,\tau)[h-h_0]$ of $M(\theta,h,\tau)$ at $h=h_0$ exists in all directions $[h-h_0] \in \mathcal{H}$; and for all $(\theta,h,\tau) \in \Theta_1^{\delta_n} \times \mathcal{H}^{\delta_n} \times \mathcal{T}_1$ with a positive sequence $\delta_n \to 0$, it holds that

$$\sup_{\tau\in\mathcal{T}_1} \left\| M(\theta,h,\tau) - M(\theta,h_0,\tau) - V_h(\theta,h_0,\tau)[h-h_0] \right\| \leq c\left\|h-h_0\right\|_{\mathcal{H}}^2 \tag{8}$$

for a constant $c \geq 0$, and

$$\sup_{\tau\in\mathcal{T}_1} \left\| V_h(\theta,h_0,\tau)[h-h_0] - V_h(\theta_0,h_0,\tau)[h-h_0] \right\| \leq o(1)\left|\theta-\theta_0\right|; \tag{9}$$

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

$$\sup_{\tau\in\mathcal{T}_1} \left\| V_\theta(\theta,h,\tau) - V_\theta(\theta,h_0,\tau) - V_{\theta h}(\theta,h_0,\tau)[h-h_0] \right\| \leq c\left\|h-h_0\right\|_{\mathcal{H}}^2 \tag{10}$$

and

$$\sup_{\tau\in\mathcal{T}_1} \left\| V_{\theta h}(\theta,h_0,\tau)[h-h_0] - V_{\theta h}(\theta_0,h_0,\tau)[h-h_0] \right\| \leq o(1)\left|\theta-\theta_0\right|; \tag{11}$$

(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 $\delta_n \to 0$, (a)

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

and, (b)

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

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

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

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

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

Moreover, we set

$$s(w,\theta_0,h_0,\tau) := \begin{pmatrix} \psi(w,\theta_0,h_0,\tau) + \phi(w,\theta_0,h_0,\tau) \\ vec(\psi_\theta(w,\theta_0,h_0,\tau) + \phi_\theta(w,\theta_0,h_0,\tau)) \end{pmatrix}, \tag{12}$$

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

Note we allow for τ to be on the boundary, e.g. $\mathcal{T}_1 = \mathcal{T}$, but we rule out parameters $\theta_0(\tau)$ on the boundary of Θ to keep the theory "simple", but see, for instance, Chernozhukov et al. (2007). Assumptions A2(i) and A2(ii) are instrumental 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 typically satisfied by letting $\psi_\theta(\cdot,\theta,h_0,\tau)=\partial\psi(\cdot,\theta,h_0,\tau)/\partial\theta'$. With non-smooth moment functions, these are high-level conditions but still are satisfied by some commonly used models, for example, the quantile regression model; see Angrist, Chernozhukov and Fernández-Val (2006). More generally, the expression for ψ_θ can be obtained from the generalized information equality, see, e.g., Newey and McFadden (1994). Assumption A2(iii) requires second-order differentiability with respect to (w.r.t.) θ due to misspecification. Similarly, A2(iv) imposes conditions on the cross term $V_{\theta h}(\cdot)$ in addition to the first-order pathwise derivative $V_h(\cdot)$. Conditions on second-order derivatives are not needed in the correctly specified case, cf. CLV. Assumption A2(v) or similar versions are commonly assumed in the semiparametric literature. Ichimura and Lee (2010) imposed slightly weaker conditions on the converging rates of the initial estimator $\widehat{h}$. Escanciano et al. (2013) provide simple primitive conditions for verifying $\Pr\left(\widehat{h}\in\mathcal{H}\right)\to 1$ with general constant kernel estimators. The rate $o_P\left(n^{-1/4}\right)$ can be relaxed to $o_P(1)$ when c in (8) is zero (i.e., under linearity). Assumption A2(vi) is usually implied by the $P-$Donsker property of the function classes $\Psi:=\{\psi(\cdot,\theta,h,\tau):\theta\in\Theta_1^\delta,\ h\in\mathcal{H}^\delta,\ \tau\in\mathcal{T}_1\}$ and $\Psi_\theta:=\{\psi_\theta(\cdot,\theta,h,\tau):\theta\in\Theta_1^\delta,\ h\in\mathcal{H}^\delta,\ \tau\in\mathcal{T}_1\}$, for which primitive conditions can be easily provided using standard empirical processes tools. Assumption A2(vii) is a high-level condition and implies Assumption (2.6) in CLV for a fixed τ. 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, e.g., Section 3 in CLV and Newey (1994) for discussion. The type of analysis needed for establishing the uniformity in $\tau\in\mathcal{T}_1$ in the expansion of Assumption A2(vii) is similar to certain uniform weighted bias calculations, which have been carried out in the literature for well-known estimators; see Andrews (1995) and Escanciano et al. (2013) for kernel estimators and Song (2008) for series estimators.

Theorem 4.2 *Under Assumption A2, it holds that*

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

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

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

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

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

and s defined in (12).

Remark 4.1 *When the model is correctly specified (i.e. $M(\tau)\equiv 0$) the assumptions and the asymptotic covariance function can be substantially simplified. In this case, for all $\tau\in\mathcal{T}_1$,*

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

Hence, the covariance function simplifies to

$$\widetilde{C}(\tau_1,\tau_2)=\left[V_{\theta_0}(\tau_1)'\Sigma V_{\theta_0}(\tau_1)\right]^{-1}V_{\theta_0}(\tau_1)'\Sigma\widetilde{K}(\tau_1,\tau_2)\Sigma V_{\theta_0}(\tau)\left[V_{\theta_0}(\tau_2)'\Sigma V_{\theta_0}(\tau_2)\right]^{-1},\qquad(13)$$

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

Example 1 (cont.) Theorem 4.2 quantifies the asymptotic effect of estimating nuisance parameters (e.g. through the function ϕ when the model is correctly specified). Interestingly enough, in Example 1 with a nonparametric Riesz representer r there is no asymptotic contribution from estimating r in the non-smooth indicators, which facilitates inference. Technically, the pathwise derivatives of mappings such as $\bar{r}(\cdot) \to E[l(X,\alpha)r(X)1(\bar{r}(X) > 0)]$ at $\bar{r}(\cdot) = r(\cdot)$ are zero, under some mild regularity conditions. See Section 8.2 for details. Practically, this means that we can treat r in the indicator functions as known, and we only need to consider estimation effects arising from the linear functionals such as $\bar{r}(\cdot) \to E[l(X,\alpha)\bar{r}(X)1(r(X) > 0)]$, which are straightforward to deal with as shown in the Appendix.

4.3 Bootstrap Approximations

The set inferences 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 can be used to provide approximations of such quantiles. CLV propose to use the ordinary nonparametric bootstrap. The nonparametric bootstrap can be used in our setting as well, and conditions for its consistency can be easily given combining our uniform results above with the arguments of CLV. The nonparametric bootstrap, however, can be computationally expensive in our context, since we need to re-estimate the pair (θ, h) for each bootstrap sample and for each fixed $\tau \in \mathcal{T}_1$. Hence, for completeness we propose here an alternative bootstrap method based on the multiplier principle, which has the computational advantage that one does not need to estimate $(\theta_0(\tau), h(\cdot,\tau))$ for each bootstrap sample. In contrast to the nonparametric bootstrap, the multiplier bootstrap requires the estimation of the influence function of the semiparametric estimator $\widehat{\theta}(\cdot)$, which can be a cumbersome task in some applications, especially under misspecification. Hence, the most practically convenient bootstrap method to choose depends on the specific application at hand. Since the theory for the nonparametric bootstrap is relatively well known, we focus in the rest of this section on the multiplier bootstrap.

According to the proof of Theorem 4.2, the asymptotic linear expansion for the estimator $\widehat{\theta}$ is

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

where $\Delta(\theta_0,h_0,\tau) = A_0(\tau)^{-1}[B_0(\tau)\;\; D_0(\tau)]$ with A_0, B_0, D_0 and $s(\cdot)$ defined as in Theorem 4.2. 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 Δ as a result of differentiation. This is for instance the case in models under quantile restrictions, where the influence function depends on a conditional density, in addition to the original parameters of the model. Our results below allow for this possibility under sufficient regularity conditions for the nonparametric estimators of these additional nonparametric objects. Notice that if h_0 includes θ as an argument, e.g. h_0 is profiled, then estimation of $V_{\theta_0}(\tau)$ and $V_{\theta_0\theta_0}(\tau)$ might be difficult. However, in many applications it is feasible to obtain uniformly consistent estimators. Even for the profiled case, we can still estimate $V_{\theta_0}(\tau)$ and $V_{\theta_0\theta_0}(\tau)$ in some widely used semiparametric models, for instance, mean regression and quantile regression models, see Example 2 below. 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 $\widehat{V}_\theta(\widehat{\theta}, \widehat{h}, \tau) := E_n \psi_\theta\left(W, \widehat{\theta}, \widehat{h}, \tau\right)$. Similarly, we assume we can estimate $V_{\theta_0\theta_0}(\tau)$ by $\widehat{V}_{\theta\theta}(\widehat{\theta}, \widehat{h}, \tau)$ and $M(\tau)$ by $E_n\left[\psi\left(W, \widehat{\theta}, \widehat{h}, \tau\right)\right]$. Denote the corresponding estimator of $\Delta(\theta_0, h_0, \tau)$ by $\widehat{\Delta}(\widehat{\theta}, \widehat{h}, \tau)$. We propose the following multiplier-type bootstrap to approximate the asymptotic distribution of a continuous functional of $Z_n(\cdot)$, say $\Psi(Z_n)$:

Algorithm 1: (Multiplier Bootstrap Approximation)
Step 1: Generate an i.i.d. sequence of random variables $\{u_i\}_{i=1}^n$ with mean zero, unit variance and bounded $(2+\eta)$ moments, with $\eta > 0$, which are also independent of $\{W_i\}_{i=1}^n$. Possible choices of distributions for u_i include the standard normal and Bernoulli. For example, one can use the Bernoulli distribution with $\Pr(u_i = (1-\sqrt{5})/2) = (\sqrt{5}+1)/(2\sqrt{5})$, and $\Pr(u_i = (1+\sqrt{5})/2) = (\sqrt{5}-1)/(2\sqrt{5})$, as advocated in e.g. Mammen (1993).
Step 2: For each fixed τ, compute $\widehat{Z}_n^*(\tau) = \widehat{\Delta}(\widehat{\theta}, \widehat{h}, \tau) n^{-1/2} \sum_{i=1}^n \left\{s\left(W_i, \widehat{\theta}, \widehat{h}, \tau\right) - E_n s\left(W, \widehat{\theta}, \widehat{h}, \tau\right)\right\} u_i$.
Step 3: Compute the functional of interest $\Psi(\widehat{Z}_n^*)$.
Step 4: Repeat Step 1-3 B times and approximate the distribution of $\Psi(Z_n)$ with the empirical cdf of the B bootstrap realizations of $\Psi(\widehat{Z}_n^*)$.

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

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

$$\sup_{\theta \in \Theta_1^\delta, h \in \mathcal{H}^\delta, \tau \in \mathcal{T}_1}\left|\widehat{V}_\theta(\theta, h, \tau) - V_\theta(\theta, h, \tau)\right| = o_P(1),$$
$$\sup_{\theta \in \Theta_1^\delta, h \in \mathcal{H}^\delta, \tau \in \mathcal{T}_1}\left|\widehat{V}_{\theta\theta}(\theta, h, \tau) - V_{\theta\theta}(\theta, h, \tau)\right| = o_P(1);$$

both $V_\theta(\theta, h, \tau)$ and $V_{\theta\theta}(\theta, h, \tau)$ are continuous in (θ, h) at $(\theta, h) = (\theta_0, h_0)$ uniformly in $\tau \in \mathcal{T}_1$.
(iii) The class of functions $\mathcal{S} = \{s(\cdot, \theta, h, \tau) : \theta \in \Theta_1^\delta, h \in \mathcal{H}^\delta, \tau \in \mathcal{T}_1\}$ is $P-$Donsker with uniformly bounded mean, i.e. $\sup_{(\theta,h,\tau) \in \Theta_1^\delta \times \mathcal{H}^\delta \times \mathcal{T}_1} E\left[|s(W, \theta, h, \tau)|\right] < \infty$; moreover, both $E\left[s(W, \theta, h, \tau)\right]$ and $E\left[s(W, \theta, h, \tau) s(W, \theta, h, \tau)'\right]$ are continuous in (θ, h), uniformly in τ.

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 4.2. Assumption A3(ii) is a high-level condition, which is used to show the consistency of $\widehat{\Delta}(\widehat{\theta}, \widehat{h}, \tau)$ to $\Delta(\theta_0, h_0, \tau)$ uniformly in $\tau \in \mathcal{T}$. This assumption requires verification of the uniform convergence of

$\widehat{V}_\theta$ and $\widehat{V}_{\theta\theta}$, which is relatively easy in many cases. For example, if $\psi(\cdot,\theta,h,\tau)$ is twice continuously differentiable in θ, then under some mild regularity conditions, we can establish a Glivenko-Cantelli property for $\{\partial/\partial\theta\psi(\cdot,\theta,h,\tau) : \theta \in \Theta_1^\delta, h \in \mathcal{H}, \tau \in \mathcal{T}_1\}$ and $\widehat{V}_\theta$ is the sample analog of $V_\theta(h,\tau) = E[\partial\psi(\cdot,\theta,h,\tau)/\partial\theta]$; similar primitive conditions can be imposed for $\widehat{V}_{\theta\theta}$. As for Assumption A3(iii), we directly assume a $P-$Donsker property for the function class $\mathcal{S}$ which can be verified by applying standard arguments, see van der Vaart and Wellner (1996).

We use the notion of bootstrap consistency in probability introduced in Giné and Zinn (1990). Let BL_1 denote the set of all functionals on $\ell^\infty(\mathcal{T}_1)$ with a Lipschitz norm bounded by 1, i.e. for any $f \in BL_1$, $\sup_{z\in\ell^\infty(\mathcal{T}_1)}|f(z)| \le 1$ and for $z_1, z_2 \in \ell^\infty(\mathcal{T}_1)$, $|f(z_1) - f(z_2)| \le \|z_1 - z_2\|_{\mathcal{T}_1}$. Let $E^*[\cdot]$ and $Var^*(\cdot)$ denote the expectation and variance of bootstrap statistics conditional on $\{W_i\}_{i=1}^n$, respectively.

Theorem 4.3 *Under Assumption A3, it holds that $\widehat{Z}_n^*(\cdot)$ weakly converges to $L(\cdot)$ in $\ell^\infty(\mathcal{T}_1)$ conditional on $\{W_i\}_{i=1}^n$ in probability, i.e.*

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

and

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

where $L(\cdot)$ and $C(\cdot,\cdot)$ are defined as in Theorem 4.2.

Remark 4.2 *Equation (14) and the continuous mapping theorem imply the consistency of bootstrap quantiles of continuous functionals of Z_n, whereas (15) shows the consistency of bootstrap standard errors. Similar results to (15) for the ordinary nonparametric bootstrap are not available in the literature in this generality. If the model is correctly specified, then we do not need to estimate $V_{\theta_0\theta_0}(\tau)$ and the assumptions can be simplified as shown in Section 8.1 in the Appendix. Details are omitted.*

5 Set Inferences, Sensitivity Analysis and Prior Information

5.1 Set Inferences

5.1.1 Inference on the Identified Set

We apply our previous uniform results to obtain inference on the identified set $\Theta_0 = \{\theta_0(\tau) : \tau \in \mathcal{T}_0\}$ for a correctly specified model. We allow for the possibility that $\mathcal{T}_0$ is unknown and estimated consistently by $\hat{\mathcal{T}}_0$. A candidate for $\hat{\mathcal{T}}_0$ can be obtained extending the ideas in Chernozhukov et al. (2007) to our semiparametric context with infinite-dimensional nuisance parameters.

Specifically, a natural estimate for $\mathcal{T}_0$ is the level set estimator

$$\hat{\mathcal{T}}_0 = \left\{\tau \in \mathcal{T}_1 : \left\|M_n\left(\widehat{\theta}(\tau), \widehat{h}(\cdot, \widehat{\theta}(\tau), \tau), \tau\right)\right\|^2 \le \hat{c}/n\right\}, \tag{16}$$

for a suitable positive level value $\hat{c}$. In practice we can choose $\hat{c} = \log n$, as suggested by Chernozhukov et al. (2007). The following result provides conditions for convergence of $d_H\left(\hat{\mathcal{T}}_0, \mathcal{T}_0\right)$ to zero, where d_H denotes the Hausdorff distance

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

with $d_H\left(a, B\right) := \inf_{b\in B}|a-b|$. This result is of independent interest.

Lemma 5.1 *Let Assumption A2 hold. Then,* $d_H\left(\hat{\mathcal{T}}_0, \mathcal{T}_0\right) = o_P(1)$ *and* $\Pr\left(\mathcal{T}_0 \subset \hat{\mathcal{T}}_0\right) \to 1$ *for* $\hat{\mathcal{T}}_0$ *in (16) and any* $\hat{c} \to \infty$*, such that* $\hat{c}/n \to 0$ *as* $n \to \infty$*.*

Even in applications where $\mathcal{T}_0$ is known, it is often practically convenient to consider a discrete approximation to $\mathcal{T}_0$, so that our inferences below can be easily carried out (i.e., the sup is replaced by a max over a finite set). Our results permit discrete approximations as long as the discrete set $\hat{\mathcal{T}}_0$ converges to $\mathcal{T}_0$ in the Hausdorff metric as $n \to \infty$. To include all these possibilities, we derive our inferences on Θ_0 for a generic estimator $\hat{\mathcal{T}}_0$ satisfying a consistency condition.

Our first result is a direct corollary of Theorem 4.1 and shows the consistency in the Hausdorff metric of the estimated identified set $\widehat{\Theta}_0 = \{\widehat{\theta}(\tau) : \tau \in \hat{\mathcal{T}}_0\}$. Let $\mathcal{T}_0^{\delta} := \{\tau \in \mathcal{T} : \inf_{\tau'\in\mathcal{T}_0}|\tau - \tau'| \leq \delta\}$ be a small expansion of $\mathcal{T}_0$. In situations where we can guarantee that $\Pr\left(\hat{\mathcal{T}}_0 \subset \mathcal{T}_0\right) \to 1$ there is no need to consider a $\delta-$expansion and local misspecification, so Assumption A2 above can be replaced by Assumption A2' in the Appendix. Unfortunately, it is the other inclusion what we typically obtain, cf. Lemma 5.1.

Corollary 5.1 *Let Assumption A1 above hold for* $\mathcal{T}_1 = \mathcal{T}_0^{\delta}$*, and assume* $\theta_0(\cdot)$ *is uniformly continuous on* $\mathcal{T}_1$ *and that* $d_H\left(\hat{\mathcal{T}}_0, \mathcal{T}_0\right) = o_P(1)$. *Then, it holds that* $d_H\left(\widehat{\Theta}_0, \Theta_0\right) = o_P(1)$.

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

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

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

Proposition 5.1 *Let Assumptions A2 and A3 hold with* $\mathcal{T}_1 = \mathcal{T}_0^{\delta}$*, and assume* $\Pr\left(\mathcal{T}_0 \subset \hat{\mathcal{T}}_0\right) \to 1$. *Then,*

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

The required inclusion condition $\Pr\left(\mathcal{T}_0 \subset \hat{\mathcal{T}}_0\right) \to 1$ is not very restrictive, see Lemma 5.1. It is worth stressing that estimation of $\mathcal{T}_0$ does not have an asymptotic impact on our confidence region for the identified set. This follows from the fact that, under our assumptions,

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

see Chernozhukov, Lee and Rosen (2013) for a related result in a different setting.

5.1.2 Inference on the True Parameter

In this subsection, we consider testing for the null hypothesis that some given parameter vector lies in the identified set, and we use this test to construct confidence sets for the "true" parameter, see Imbens and Manski (2004) for related 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, our proposal in that setting requires a rank condition and an order condition that are mirror images of the classical overidentification conditions required in the standard J-test. The convex case is less general, but it has the advantage that in some cases where the order condition of the general case fails, it still can be applied.

The General Case The testing hypotheses are

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

These hypotheses are of interest in their own. For instance, Altonji et al. (2005) aim to test for significance of the impact of attending a Catholic school on educational attainment in a partially identified parametric model. Their example fits our setting with a null hypothesis $0 \in \Theta_0$. A stronger sense of no-impact is $\theta_0(\tau) = 0$ for all $\tau \in \mathcal{T}_0$, which can be tested using the results of the previous section.

Recalling that $M(\theta, h_0, \tau) = E[\psi(W, \theta, h_0(\cdot, \theta, \tau), \tau)]$, 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 $\widetilde{\mathcal{T}}$ is non-empty under the null hypothesis.

A simple approach to obtain a confidence region for the true parameter consists in taking the union of confidence intervals for $\theta_0(\tau)$ constructed at each $\tau \in \mathcal{T}_0$. This union-intersection principle has been commonly used in the literature. However, this approach leads to conservative inference. A more efficient method can be based on the following test statistic:

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

where $\widehat{h}$ is a consistent estimator for $\widetilde{h}_0$. This criterion-based approach is a semiparametric extension of the method proposed by Chernozhukov et al. (2007). For a given $\widetilde{\theta}$ under the null, $T_n(\widetilde{\theta})$ can be viewed as a generalization of the classical overidentification test statistic to partially identified semiparametric models. To proceed, we need the following assumptions.

Assumption A4: For fixed $\widetilde{\theta}$ given in the testing problem:
(i) Existence of Minorant: there exist $c > 0$ and $\delta > 0$, such that $\left\|M\left(\widetilde{\theta}, \widetilde{h}_0, \tau\right)\right\| \geq c(\inf_{\widetilde{\tau}\in\widetilde{\mathcal{T}}} |\tau - \widetilde{\tau}| \wedge \delta)$ for all $\tau \in \mathcal{T}$, where $a \wedge b := \min\{a, b\}$;
(ii) Smoothness in τ : $M(\widetilde{\theta}, \widetilde{h}_0, \tau)$ is continuously differentiable at all $\tau \in \mathcal{T}$, with derivative matrix $V_\tau(\widetilde{\theta}, \widetilde{h}_0, \tau) := \partial M(\widetilde{\theta}, \widetilde{h}_0, \tau)/\partial\tau$.
(iii) Smoothness in h : Assumption A2(iv)(a) holds with $\theta = \widetilde{\theta}$, $\mathcal{T}_1 = \mathcal{T}_0$ and condition (9) replaced by the following condition: for $h \in \mathcal{H}_{\delta_n}$ and $\eta_n = o(1)$, $|\tau_1 - \tau_2| \leq \eta_n$,

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

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

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

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

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

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

Theorem 5.2 *Let the null hypothesis H_0 and Assumption A4 hold. Then,*

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

where $G(\widetilde{\tau})$ is a Gaussian process on $\widetilde{\mathcal{T}}$ with covariance kernel $\widetilde{K}(\cdot, \cdot)$ defined in Remark (4.1).

Remark 5.1 *If rank$(V_\tau(\widetilde{\tau})) = d_\psi$ for some $\widetilde{\tau} \in \widetilde{\mathcal{T}}$, then $P\left(\widetilde{\theta}, \widetilde{\tau}\right)$ becomes zero, see e.g. Theorem 3.5 in Yanai, Takeuchi and Takane (2011), and as a result, the limiting distribution of $T_n(\widetilde{\theta})$ is degenerate. This is for instance the case if $d_\psi \leq d_\tau$, i.e. the number of unidentified nuisance parameters is greater than or equal to the number of moments.*

Remark 5.2 *Note that Theorem 5.2 does not require the conditional identification assumption of θ_0. The theorem also shows that if τ is conditionally identified, so $\widetilde{\mathcal{T}}$ is a singleton, then a pointwise optimal version of our test behaves like the standard J-test of overidentifying restrictions, with a limiting distribution equal to a chi-square with $d_\psi - d_\tau$ degrees of freedom. The pointwise optimal version corresponds, as usual, to the optimal choice of Σ at the unique minimizer, which is implemented following standard practice. In this situation, the test is straightforward to carry out and it does not require any resampling method. When the conditional identification of τ does not hold, one could still use the asymptotic critical values from chi-squared distribution for a continuous updated version of $T_n(\widetilde{\theta})$, but inference would be conservative for the general case. Nevertheless, the computational simplicity of this approach makes it a very attractive procedure. A more efficient bootstrap method, also more computational demanding, is described below.*

Similar test statistics to $T_n(\widetilde{\theta})$ have been used in the partial identification literature, albeit for different models. In a support function setting, Bontemps et al. (2012) circumvent the potential multiple-minimizer problem (i.e $\widetilde{\mathcal{T}}$ is not singleton) by obtaining a shrinkage estimator whose limit is well defined, and the test statistic evaluated at this estimator converges to standard normal distribution. Galichon and Henry (2009) consider a generalized Kolmogorov-Smirnov test statistic and the special structure of their testing problem delivers a straightforward asymptotic distribution. Santos (2012) develops 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 constructs 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.

For more efficient inference in the general case, we propose using a multiplier-type bootstrap described below. If $rank\left(V_\tau\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}\right)\right) < d_\psi$ for all $\widetilde{\tau} \in \widetilde{\mathcal{T}}$, then by the proof of Theorem 5.2, we have

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

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 $P\left(\widetilde{\theta},\tau\right)$ can be estimated by $\widehat{P}\left(\widetilde{\theta},\tau\right) := I_{d_\psi} - \widehat{V}_\tau\left(\widetilde{\theta},\widehat{h},\tau\right)\left[\widehat{V}_\tau\left(\widetilde{\theta},\widehat{h},\tau\right)' \Sigma \widehat{V}_\tau\left(\widetilde{\theta},\widehat{h},\tau\right)\right]^{-} \widehat{V}_\tau\left(\widetilde{\theta},\widehat{h},\tau\right)' \Sigma.$ We suggest the following bootstrap procedure:

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

Remark 5.3 *We follow Santos (2012) and use the penalty term* $\lambda_n \left\| M_n\left(\widetilde{\theta}, \widehat{h}, \tau\right)\right\|$ *in Step 2. This is necessary because the process* $n^{-1/2}\sum_{i=1}^{n} s\left(W_i, \widetilde{\theta}, \widehat{h}, \tau\right) u_i$ *is centered for all* $\tau \in \mathcal{T}$ *instead of just for* $\widetilde{\tau} \in \widetilde{\mathcal{T}}$*. By introducing this penalty term, it is ensured that* $R_n^*(\tau)$ *diverges to infinity for all* $\tau \notin \widetilde{\mathcal{T}}$*; hence when computing* $T_n^*(\widetilde{\theta})$*, the infimum is effectively evaluated only at a shrinking neighborhood of* $\widetilde{\mathcal{T}}$*. Notice that the test statistics are computed by taking infimum over the whole nuisance parameter set* $\mathcal{T}$*. An alternative approach can be conducted by first estimating the argmin set* $\widetilde{\mathcal{T}}$*, then computing the bootstrap statistics using this estimated set. However, this approach is computationally expensive for constructing confidence sets for the true parameter 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 5.2 *Suppose (i) Assumption A4 holds; (ii) the penalty sequence* λ_n *satisfies* $\lambda_n \to \infty$ *and* $\lambda_n = o\left(\sqrt{n}/\log\log n\right)$*; and (iii) the asymptotic distribution of* $T_n(\widetilde{\theta})$ *is continuous at its* $1-\alpha$ *quantile. Then, under the null hypothesis,*

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

Moreover, under the alternative hypothesis, we have

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

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

The Convex Case As discussed in Remark (5.1), when $rank(V_\tau(\widetilde{\tau})) = d_\psi$ (for example if $d_\psi \leq d_\tau$), the proposed test statistic $T_n\left(\widetilde{\theta}\right)$ is degenerate. To overcome this limitation, we complement our previous proposal with one in which we impose shape restrictions (convexity) on the identified set. Let $\beta(\tau|B)$ denote 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: Assume:
(i) Θ_0 is convex and compact;
(ii) There exists a moment function $\xi(\cdot)$ such that $\beta(\tau|\Theta_0) = E\left[\xi(W, g_0, \tau)\right]$, where $g_0(\cdot)$ is a vector of unknown functions which may include $h_0(\cdot)$ 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 $\tau\in\mathbb{S}^{d_\theta}$,

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

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 $P-$Donsker.
(v) Stochastic Equicontinuity in g :

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

If the moment $M(\theta,h_0,\tau)$ is continuous in θ for all τ, then the identified set Θ_0 is closed, and with Θ compact, it is also compact. Beresteanu and Molinari (2008), Beresteanu, Molchanov and Molinari (2011, 2012) and Bontemps et al. (2012), among others, have provided numerous examples where Assumption A5(ii) holds with parametric g_0. The rest of assumptions are standard and are similar to the ones previously discussed for the general case.

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

Theorem 5.3 *Let 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:

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

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

Remark 5.4 *We do not explicitly require differentiability of the support function $\beta(\tau|\Theta_0)$ w.r.t. τ which is equivalent to the non-existence of the exposed faces of the identified set Θ_0. However, the stochastic equicontinuity assumption might implicitly rule out the presence of exposed faces, see for example Proposition 9 in Bontemps et al. (2012), who, in their specific setting, derive the non-regular asymptotic distribution of the support function estimator with the presence of exposed faces. It should be noticed that non-differentiability does not necessarily lead to non-regular asymptotic distribution of the support function estimator, e.g. see section 5.3 in Bontemps et al. (2012).*

Remark 5.5 *To deal with the multiple minimizer problem caused by the presence of kinks, Bontemps et al. (2012) proposes a shrinkage type estimator to obtain a unique minimizer, instead we suggest taking the minimum over all possible minimizers as in the general case described in Algorithm 2.*

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

5.2 Sensitivity Analysis

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

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

The methods discussed here provide a methodological ground for sensitivity analysis. Combining the original ideas of Rosenbaum and Rubin (1983) and Imbens (2003), we suggest to report the set $\widehat{\Theta}_1=\{\widehat{\theta}(\tau):\tau\in\widehat{\mathcal{T}}_1\}$, where

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

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

Likewise, a quantity of great interest in sensitivity analysis is $\partial\theta_0(\tau)/\partial\tau$. This quantity exists under our smoothness conditions, and it can be estimated by implicit differentiation from the first order conditions

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

For instance, consider the parametric case for illustration with $d_\psi=d_\tau=1$ (the general semiparametric case involves more algebra when h_0 depends on θ_0 and τ, but it is conceptually the same). In this case (using our previous notation but without vectorization in $V_{\theta_0\theta_0}$)

$$\frac{\partial\theta_0(\tau)}{\partial\tau}=-\left[V_{\theta_0\theta_0}(\tau)M(\tau)+V_{\theta_0}(\tau)V'_{\theta_0}(\tau)\right]^{-1}\left[V_{\theta_0\tau}(\tau)M(\tau)+V_{\theta_0}(\tau)V_\tau(\tau)\right],\tag{18}$$

where $V_{\theta_0\tau}(\tau)$ is the cross-derivative w.r.t θ and τ and $V_\tau(\tau)$ is the derivative w.r.t τ. Equation (18) can be used to construct a uniform consistent estimator of $\partial\theta_0(\tau)/\partial\tau$. Details in the more general case are omitted to save space. We have previously discussed estimation of the elements involved in $\partial\theta_0(\tau)/\partial\tau$. Note that for a correctly specified model with $\tau\in\mathcal{T}_0$, (18) simplifies to

$$\frac{\partial\theta_0(\tau)}{\partial\tau}=-\left[V_{\theta_0}(\tau)V'_{\theta_0}(\tau)\right]^{-1}\left[V_{\theta_0}(\tau)V_\tau(\tau)\right].$$

It is worth stressing some connections between the sensitivity measure $\partial\theta_0(\tau)/\partial\tau$ and the theory of two-step estimators. Suppose that $\widehat{\tau}\in\mathcal{T}_1$ is a consistent estimator for $\tau\in\mathcal{T}_1$, and consider the asymptotic behaviour of the two-step estimator $\widehat{\theta}(\widehat{\tau})$. By a stochastic equicontinuity argument, under our assumptions,

$$\begin{aligned}\sqrt{n}(\widehat{\theta}(\widehat{\tau})-\theta_0(\tau))&=\sqrt{n}(\widehat{\theta}(\widehat{\tau})-\theta_0(\widehat{\tau}))+\sqrt{n}(\theta_0(\widehat{\tau})-\theta_0(\tau))\\&=\sqrt{n}(\widehat{\theta}(\tau)-\theta_0(\tau))+\sqrt{n}(\theta_0(\widehat{\tau})-\theta_0(\tau))+o_P(1).\end{aligned}$$

Then, under general conditions the second term in the right hand side is zero when $\partial\theta_0(\tau)/\partial\tau=0$ and $\sqrt{n}(\widehat{\tau}-\tau)=O_P(1)$. Therefore, roughly speaking θ_0 is not sensitive to τ when estimating the latter has no impact in the limiting distribution of the estimator of the former, and viceversa.

5.3 Incorporating Prior Information

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

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

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

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

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

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

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

$$\int_{\mathcal{T}_0} |\theta_0\,(\tau)|^2\, g(\tau) d\tau < \infty. \tag{19}$$

Then, simple arguments show that $\sqrt{n}(\widehat{\widehat{\theta}} - \widehat{\theta}) = o_P(1)$, provided $n/m \to 0$ as $n \to \infty$. Next corollary follows directly from Theorem 4.2 and the continuous mapping theorem, and hence its proof is omitted.

Corollary 5.2 *Under Assumption A2, (19) and $n/m \to 0$ as $n \to \infty$, it holds that*

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

where

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

and $C(\cdot,\cdot)$ is given in Theorem 4.2.

Conley et al. (2012) investigate alternative ways to incorporate prior knowledge on the exogeneity of instrumental variables and focus on confidence regions for the true parameter. Their methods can potentially be extended to our more complex setting here.

6 Example Applications

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

6.1 Gender Gap Distributional Effects

We illustrate some of the ideas discussed above with an application to inferences on the gender gap distributional effects in wage equations. Understanding the determinants and dynamics of the gender gap is one of the most prominent problems in labor economics. Here we discuss a robust inference approach to this problem. For a general treatment of partial identification in sample selection models see Manski (1989). For some interesting empirical applications in semiparametric partially identified models with selection see, for instance, Blundell et al. (2007) and Lee (2009). The latter reference deals with estimation of ATE in the context of a randomized experiment (the Job Corps training program) under selection, which has exactly the same structure as the gender gap effects functional investigated here. Hence, our results below can be used to extend the analysis in Lee (2009) to continuous covariates.

Let Y^* be a latent wage. We only observe wages for working individuals. The selection variable is D. That is, we only observe $Y = Y^*D$, together with D and a vector of covariates X. The structural function of interest is the wage conditional distribution $\varphi(X, \alpha) = P[Y^* \leq \alpha|X]$, which is known to satisfy (2), see Manski (1989), with

$$l(x, \alpha) \equiv m_0(x, \alpha) \text{ and } u(x, \alpha) \equiv m_0(x, \alpha) + 1 - g_0(x), \tag{20}$$

for all $\alpha \in \mathcal{A}$, where $m_0(x, \alpha) := E[D1(Y \leq \alpha)|X = x]$, $g_0(x) := E[D|X = x]$ and $\mathcal{A}$ is a compact set of $\mathbb{R}$ that represents the quantiles of interest. Let $X = (X_1, X_2)$, where X_1 denotes gender, $X_1 = 1$ for women; $X_1 = 0$ for men, and X_2 is a vector of individual's characteristics, such as education, experience, etc. We aim to do inference on the gender gap distributional effects functional

$$T\varphi(\cdot, \alpha) = E[\varphi(1, X_2, \alpha) - \varphi(0, X_2, \alpha)].$$

The special structure of T in this example permits a simpler representation of the identified set than the one previously derived, cf. Manski (1989), so that the following moments characterize the identified set for $T\varphi(\cdot, \alpha)$

$$\psi(W, \theta, h, \tau) = \theta - \lambda a(W, \alpha, h, 0) - (1 - \lambda)a(W, \alpha, h, 1),$$

where $W = X$, $\tau = (\lambda, \alpha) \in \mathcal{T}_0 := [0, 1] \times \mathcal{A}$, and for $d = 0, 1$ and with $h = (m, g)$,

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

The nuisance parameter vector is given here by $h_0 = (m_0, g_0)$. This parametrization is convenient because $M(\theta, h, \tau)$ is linear in h. Standard kernel estimators can be used to estimate m_0 and g_0. Denote the estimator by $\widehat{h} = (\widehat{m}, \widehat{g})$.

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

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

where $\mathcal{C}_1^\eta(\mathcal{S}_{X_2})$ is a subset of continuous, bounded (by 1) functions on the convex, bounded subset $\mathcal{S}_{X_2} \in \mathbb{R}^{d_{x_2}}$, with non-empty interior, endowed with the sup-norm $\|\cdot\|_{\mathcal{H}} = \|\cdot\|_\infty$, as defined in van der

Vaart and Wellner (1996, p.154). It is known that $\mathcal{H}$ is a $P-$Donsker class. The first order derivative w.r.t. θ and the functional derivative w.r.t. $h = (m, g)$ are given, respectively, by

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

Then Assumption A2'(ii) and (iii) are trivially satisfied. Since the moment function is linear in h we only require $\left\|\widehat{h} - h_0\right\|_{\mathcal{H}} = o_P(1)$. The latter can be checked using, for instance, the results of Escanciano et al. (2013). It is straightforward to show that $\{\psi(\cdot, \theta, h, \tau) : \theta \in \Theta, h \in \mathcal{H}, \tau \in \mathcal{T}\}$ is a $P-$Donsker class. Hence, Assumption A2'(v) holds. Likewise, conditions for the bias calculations for Assumption A2'(vi) are standard in the nonparametric literature; see, e.g., Newey (1994).

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

$$\sup_{\tau \in \widehat{\mathcal{T}}_0} \left|\sqrt{n}\left(\widehat{\theta}(\tau) - \theta_0(\tau)\right)\right| = \max_{\alpha \in \mathcal{A}_n} \max\left\{\sqrt{n}\left|\widehat{\theta}_L(\alpha) - \theta_L(\alpha)\right|, \sqrt{n}\left|\widehat{\theta}_U(\alpha) - \theta_U(\alpha)\right|\right\}, \tag{21}$$

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

$$\begin{aligned} \phi_m(W, \theta_0, h_0, \tau) := &\{D1(Y \le \alpha) - m(0, X_2, \alpha)\} \frac{1(X_1 = 0)}{1 - p(X_2)} \\ &- \{D1(Y \le \alpha) - m(1, X_2, \alpha)\} \frac{1(X_1 = 1)}{p(X_2)} \end{aligned}$$

and

$$\begin{aligned} \phi_g(W, \theta_0, h_0, \tau) := &\{D - g(0, X_2)\} \frac{\lambda 1(X_1 = 0)}{1 - p(X_2)} \\ &+ \{D - g(1, X_2)\} \frac{(1-\lambda) 1(X_1 = 1)}{p(X_2)}, \end{aligned}$$

where the additional nuisance parameter is the propensity score $p(x_2) := P[X_1 = 1 | X_2 = x_2]$, which can be easily estimated by, e.g., kernel methods. Thus, estimating $\psi + \phi$ is straightforward in this application, and the multiplier-bootstrap can be applied to approximate critical values for the limiting distribution of (21). With the bootstrap critical value $\widehat{c}_{1-\widetilde{\alpha},n}$, a uniform $(1-\widetilde{\alpha})$-confidence band for gender distributional effects (or ATE as in Lee, 2009) can be constructed as

$$\left[\widehat{\theta}_L(\alpha) - \frac{\widehat{c}_{1-\widetilde{\alpha},n}}{\sqrt{n}}, \widehat{\theta}_U(\alpha) + \frac{\widehat{c}_{1-\widetilde{\alpha},n}}{\sqrt{n}}\right], \qquad \alpha \in \mathcal{A}.$$

This confidence band accounts for the partial identification and the dependence across different quantile levels α of the estimated gender gap.

6.2 Binary Choice with Sample Selection and No Exclusions

Suppose a latent binary variable Y^* satisfies the ordinary threshold crossing binary response model $Y^* = 1(X'\beta_0 - e \geq 0)$. The econometrician is assumed to know relatively little about selection D other than it is binary, so let D be given by the nonparametric threshold crossing model $D = 1(g_0(X) - u \geq 0)$ where u is uniformly distributed in $[0,1]$. Assume (e,u) is drawn from an unknown joint distribution function $F_0(e,u)$ with $(e,u) \perp X$ and the specified uniform marginal for u. We only observe $(Y = Y^*D, D, X)$. Point-identification in this model has been investigated in Escanciano et al. (2012), while its estimation is discussed in Klein, Shen and Vella (2011) and Escanciano et al. (2012, 2013). Below we discuss sensitivity analysis and partial identification in this model.

Introducing some normalization restrictions, the model is then summarized by the equations

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

and it can be estimated from the conditional moment restrictions

$$
E[Y|X] = F_0(\tau_0 X_1 + X_2 + \theta_0' X_3, g_0(X)). \tag{22}
$$

We illustrate the general ideas with the SLS estimator. Similar ideas can be applied with likelihood methods. Denote the index $I(X, \theta_0, g_0, \tau) := (\tau X_1 + X_2 + \theta_0' X_3, g_0(X))$ and consider the moment restrictions that result as the first order conditions of the SLS estimator, i.e.

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

That is, we assume the SLS solves

$$
\theta_0(\tau) := \arg\min_{\theta \in \Theta} E\left[(Y - E[Y|I(X, \theta, g_0, \tau)])^2\right]. \tag{23}
$$

Here the nuisance parameters are $h_0 = (F_0, g_0, \partial_\theta F_0) \in \mathcal{H} \equiv \mathcal{F}^\eta \times \mathcal{C}_1^{\eta_g}(\mathcal{S}_X) \times \mathcal{F}^\eta$, where $\eta_g > d_x/2$ and $\mathcal{F}^\eta$ is the following class of functions defined in Escanciano et al. (2013). Let $\mathcal{W} := \{I(\cdot, \theta, g, \tau) : \theta \in \Theta, g \in \mathcal{C}_1^{\eta_g}(\mathcal{S}_X), \tau \in \mathcal{T}_1\}$. Let $\mathcal{F}^\eta$ be a class of measurable functions on $\mathcal{S}_X$, $q(I(x)|I)$ say, such that $I \in \mathcal{W}$ and q satisfies for a universal constant C_L and each $I_j \in \mathcal{W}$, $j = 1, 2$,

$$
\|q(I_1(\cdot)|I_1) - q(I_2(\cdot)|I_2)\|_\infty \leq C_L \|I_1 - I_2\|_\infty. \tag{24}
$$

Moreover, we assume that for each $I \in \mathcal{W}$, $q(\cdot|I) \in C_1^\eta(\mathcal{S}_I)$, for some $\eta > 1$, and that q is bounded.

Suppose we want to carry out sensitivity analysis with respect to an exclusion restriction $\tau = 0$. We apply our results with $\tau \in \mathcal{T}_1$, for a generic $\mathcal{T}_1$. Henceforth, we assume that $E[Y|I(X, \theta, g_0, \tau)]$ is twice continuously differentiable in θ. It is straightforward to show that a sufficient condition for our conditional identification assumption is that $E[Y|I(X, \theta_1, g_0, \tau)] = E[Y|I(X, \theta_2, g_0, \tau)]$ implies $\theta_1(\tau) = \theta_2(\tau)$, for all $\tau \in \mathcal{T}_1$. This condition holds, for instance, under correct specification of the double-index model and mild conditions on F_0 and the distribution of X_3. Specifically, if F_0 is strictly increasing in its first argument, and $E[X_3 X_3']$ is non-singular, then conditional identification holds for a correctly specified model.

Let $\widehat{g}$ be a kernel estimator for g_0, and for a candidate $I(X,\theta,\widehat{g},\tau)$, let $\widehat{F}(I(X,\theta,\widehat{g},\tau))$ and $\partial_\theta \widehat{F}(I(X,\theta,\widehat{g},\tau))$ be kernel estimates of $E[Y|I(X,\theta,\widehat{g},\tau)]$ and its derivative w.r.t θ, respectively. Then, consistency under our partial identification assumption holds by the uniform rate results for kernel estimates in Escanciano et al. (2013). Although we allow for misspecification, in the sense that the conditional moment (22) may not hold for $\tau \in \mathcal{T}_1$, we assume that the unconditional moment holds. This entails the solution $\theta_0(\tau)$ of (23) to be an interior solution. Ichimura and Lee (2010) also investigate conditions for single-index model estimation under misspecification. The model here is an extension of theirs to a double-index model with a nonparametric index g_0 and the uniformity aspect in τ.

We proceed to check the conditions for weak convergence in Assumption A2'. Henceforth, for simplicity of the notation we remove the dependence on true values and simply use a subscript zero, e.g. $I_0 \equiv I(X,\theta_0,g_0,\tau)$ and $\partial_\theta F_0 = \partial_\theta F_0(I_0)/\partial\theta$. Using a simplified notation, the (pathwise) derivatives are given by

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

where $\partial_g F_0(I_{0i}) := \partial E[Y|I_0 = I_{0i}(\bar{g},\tau)]/\partial\bar{g}|_{\bar{g}=g_{0i}}$, $\partial_{\theta g} F_0 := \partial^2 E[Y|I_0 = I_{0i}(\bar{g},\tau)]/\partial\theta\partial\bar{g}\big|_{\theta=\theta_0,\bar{g}=g_{0i}}$ and $I_{0i}(\bar{g},\tau) := I(X_i,\theta_0,\bar{g},\tau)$. Assumption A2'(iii) is satisfied with $c = 0$ for the corresponding terms to F and $\partial_\theta F$, and hence, we only require rates on $\|\widehat{g} - g_0\|_\infty$. Standard empirical processes arguments and the definition of $\mathcal{H}$ then imply that Assumption A2'(v) holds. Verification of Assumption A2'(vi) is long but involves standard bias calculations. Details are available from the authors upon request. In the correctly specified case where the conditional moment (22) holds there is no contribution in Assumption A2'(vi) from the pathwise derivative $V_{\partial_\theta F_0}$ and from the second component of V_g, but in the misspecified case these terms contribute to the asymptotic distribution of the estimators.

For sensitivity analysis we propose to estimate consistently

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

by plugging in consistent kernel estimates for $\partial_\theta F_0$, F_0, $\partial_{\theta\theta} F_0$, $\partial_\tau F_0$ and $\partial_{\tau\theta} F_0$ in the expressions above. Denote these kernel estimators by $\partial_\theta \widehat{F}$, $\widehat{F}$, $\partial_{\theta\theta}\widehat{F}$, $\partial_\tau \widehat{F}$ and $\partial_{\tau\theta}\widehat{F}$, respectively. Then, we estimate $\partial\theta_0(\tau)/\partial\tau$ by

$$\widehat{\frac{\partial\theta(\tau)}{\partial\tau}} = -\left(E_n\left[-\partial_\theta \widehat{F}\partial_\theta' \widehat{F} + (Y - \widehat{F})\partial_{\theta\theta}\widehat{F}\right]\right)^{-1} E_n\left[-\partial_\tau \widehat{F}\partial_\theta \widehat{F} + (Y - \widehat{F})\partial_{\tau\theta}\widehat{F}\right].$$

Of particular interest is the derivative at zero, measuring the sensitivity of the parameters of interest to the exclusion restriction. When (22) holds, the estimator of $\partial\theta_0(\tau)/\partial\tau$ can be simplified, and corresponds to the least square estimate of $-\partial_\tau \widehat{F}$ on $\partial_\theta \widehat{F}$, which can be implemented with standard regression packages.

If a partial identification approach to this problem is considered, it proceeds by first estimating $\mathcal{T}_0$, using (16). Then, one can apply our methods above to construct a confidence region for the identified set or for the true parameter. In this application there is no closed form expression for $\theta_0(\tau)$. Hence, it is convenient to consider discrete approximations of $\mathcal{T}_1$ or $\mathcal{T}_0$ to implement our inferences. Similarly, the expression for the influence function is too involved in the misspecified case, so the nonparametric bootstrap seems to be more convenient in this example than the multiplier-bootstrap. The algorithm for implementing our inferences in this example is as follows:

Algorithm 3: (Nonparametric Bootstrap Approximation)
Step 1: Choose a grid approximation $\{\tau_1, ..., \tau_m\}$ of $\mathcal{T}_1$ or $\widehat{\mathcal{T}}_0$. For each $j = 1, ..., m$: compute kernel estimates of $E[Y|V(X, \theta, \widehat{g}, \tau_j)]$, say $\widehat{F}\,(V(X, \theta, \widehat{g}, \tau_j))$. Then, compute the SLS estimator $\widehat{\theta}\,(\tau_j)$.
Step 2: Draw a bootstrap sample from the empirical distribution of the data and repeat Step 1 above with the bootstrap sample.
Step 3: Repeat Step 2 B times. Approximate the distribution $\sqrt{n}\left(\widehat{\theta}\,(\cdot) - \theta_0\,(\cdot)\right)$ by the empirical distribution of $\left\{\sqrt{n}\left(\widehat{\theta}^{*l}\,(\cdot) - \widehat{\theta}\,(\cdot)\right)\right\}_{l=1}^{B}$, where $\widehat{\theta}^{*l}$ denotes the $l-th$ iteration of Step 2.

7 Conclusions

In this paper, we develop inference procedures for a class of semiparametric partially identified models where the identified set for the Euclidean parameter of interest is a finite-dimensional manifold. Within this setting, we can employ tools from empirical process theory to derive uniform convergence results for the set estimates. Our framework allows for the presence of infinite-dimensional nuisance parameters and unknown nuisance identified sets. Our inference methods permit but do not require convex identified sets. When applied to convex identified sets, they allow for semiparametric support functions, which complements the support function approach considered in Beresteanu and Molinari (2008) and Bontemps et al. (2012). We propose bootstrap procedures to approximate the asymptotic distribution of functionals of the set estimator. Based on the uniform weak convergence results and consistency of the bootstrap, we construct a simple confidence region for the identified set and the true parameter.

We have emphasized and formalized the differences between partial identification and sensitivity analysis. The latter is relatively underdeveloped, and this paper is a first attempt to fill this gap in the literature. We have formalized and extended important ideas in Rosenbaum and Rubin (1983) and Imbens (2003) to account for estimation and model uncertainty in sensitivity analysis. Additionally, we have suggested new measures of sensitivity such as the derivative $\partial\theta_0(\tau)/\partial\tau$ or the diameter $d = \sup_{\tau,\tau'\in\mathcal{T}_1} |\theta_0(\tau) - \theta_0(\tau')|$, and ways to estimate them.

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

An interesting 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 Honoré and Tamer (2006), dynamic discrete decision processes in Magnac and Thesmar (2002), nonparametric instrumental variable models in Santos (2012) or models with discrete outcomes in Chesher (2010). Although most of our theoretical results are directly applicable to a nonparametric $\mathcal{T}$, feasible versions of the proposed inferences require an approximation of $\mathcal{T}$ by sieves (e.g., for computing the supremum in $\mathcal{T}$). These feasible inferences under this more general setting can be justified combining the results of this paper with those of Santos (2012). This extension is beyond the scope of this paper and it is deferred for future research. On the contrary, the extension to conditional moment restrictions is trivial, as we can transform these models into an infinite number of unconditional moments, whose index is part of τ.[13] This illustrates the versatility of our results.

8 Appendix

8.1 Assumptions Under Correct Specification

The following condition replaces Assumption A2 under correct specification.

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

$$\sup_{\tau \in \mathcal{T}_1} \left\{ \left\| M_n \left(\widehat{\theta}(\tau), \widehat{h}, \tau \right) \right\| - \inf_{\theta \in \Theta_1^\delta} \left\| M_n \left(\theta, \widehat{h}, \tau \right) \right\| \right\} = o_P \left(n^{-1/2} \right).$$

(ii) Smoothness in θ : for each $\tau \in \mathcal{T}_1$, the map $\theta \to M(\theta, h_0, \tau)$ is continuously differentiable at $\theta_0(\tau)$, with derivative $V_{\theta_0}(\tau)$ that is of full rank and $\sup_{\tau \in \mathcal{T}_1} |V_{\theta_0}(\tau)| < \infty$ and $\sup_{\tau \in \mathcal{T}_1} \left| V_{\theta_0}^{-1}(\tau) \right| < \infty$.
(iii) Smoothness in h : for each $(\theta, \tau) \in \Theta_1^\delta \times \mathcal{T}_1$, the pathwise derivative $V_h(\theta, h_0, \tau)[h - h_0]$ of $M(\theta, h, \tau)$ at $h = h_0$ exists in all directions $[h - h_0] \in \mathcal{H}$; and for all $(\theta, h, \tau) \in \Theta_1^{\delta_n} \times \mathcal{H}^{\delta_n} \times \mathcal{T}_1$ with a positive sequence $\delta_n \to 0$, it holds that

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

for a constant $c \ge 0$, and

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

(iv) $\Pr\left(\widehat{h} \in \mathcal{H} \right) \to 1$, and $\left\| \widehat{h} - h_0 \right\|_{\mathcal{H}} = o_P\left(n^{-1/4} \right)$.

[13] That is, $E[\xi(W, \theta_0, h_0(W), \tau_1)|X] = 0$ for all τ_1 is equivalent to $E[\psi(W, \theta_0, h_0(W), \tau)] = 0$ for all τ, where $\psi(W, \theta_0, h_0(W), \tau) = \xi(W, \theta_0, h_0(W), \tau_1) 1(X \le \tau_2)$ and $\tau = (\tau_1, \tau_2)$.

(v) Stochastic Equicontinuity: for all sequences of numbers $\delta_n \to 0$,

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

(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 $\{s(w, \theta_0, h_0, \tau) := \psi(w, \theta_0, h_0, \tau) + \phi(w, \theta_0, h_0, \tau) : \tau \in \mathcal{T}_1\}$ is $P-$Donsker.

8.2 Pathwise derivatives for Example 1

In Example 1 the moment is given by

$$M(\theta_0, h_0, \tau) = \theta_0 - \lambda E[L(X, \alpha)] - (1 - \lambda) E[U(X, \alpha)].$$

We provide here the pathwise derivatives with respect to the components of $h_0(\cdot, \alpha) = (l(\cdot, \alpha), u(\cdot, \alpha), r(\cdot))$. Using our notation and standard arguments we obtain

$$V_l(\theta, h_0, \tau)\left[\bar{l} - l\right] = -\lambda E\left[(\bar{l}(X, \alpha) - l(X, \alpha)) r^+(X)\right] - (1 - \lambda) E[(\bar{l}(X, \alpha) - l(X, \alpha)) r^-(X)]$$

and

$$V_u(\theta, h_0, \tau)[\bar{u} - u] = -\lambda E\left[(\bar{u}(X, \alpha) - u(X, \alpha)) r^-(X)\right] - (1 - \lambda) E[(\bar{u}(X, \alpha) - u(X, \alpha)) r^+(X)].$$

In order to compute the pathwise derivative with respect to r, we first show that the pathwise derivatives of mappings such as $\bar{r}(\cdot) \to E[l(X, \alpha) r(X) 1(\bar{r}(X) > 0)]$ at $\bar{r}(\cdot) = r(\cdot)$ are zero, under the following conditions:

Assumption E1: (i) Assume the conditional means $m_l(r, \alpha) := E[l(X, \alpha) | r(X) = r]$ and $m_u(r, \alpha) := E[u(X, \alpha) | r(X) = r]$ are continuous and finite in a neighborhood of $r = 0$, uniformly in $\alpha \in \mathcal{A}$; (ii) for each direction $\delta(X)$ such that $r_{\rho,\delta}(X) := r(X) + \rho\delta(X)$ belongs to the parameter space for r, it holds that $|\delta(X)| < C$, for a positive constant C; (iii) the density of $r(X)$ is continuous and finite in a neighborhood of $r = 0$.

Define $a_l(r, \alpha) := m_l(r, \alpha) r$. Under Assumption E1, we can show that for all $\rho > 0$

$$\begin{aligned} &\left| \frac{E[a_l^+(r, \alpha) 1(r_{\rho,\delta}(X) > 0) - 1(r_{0,\delta}(X) > 0)]}{\rho} \right| \\ &\le \frac{E[a_l^+(r, \alpha) 1(-\rho C \le r(X) \le \rho C)]}{\rho} \\ &= \frac{1}{\rho} \left(\int_{-\rho C}^{\rho C} a_l^+(r, \alpha) f_{r(X)}(r) dr \right), \end{aligned}$$

but the right hand side converges to zero as $\rho \to 0$ under Assumption E1 by the fundamental theorem of calculus. A similar proof follows for the term involving $a_l^-(r,\alpha)$, and hence, for $a_l(r,\alpha)$. Similar arguments show that the pathwise derivative of $\bar{r}(\cdot) \to E[u(X,\alpha)r(X)1(\bar{r}(X) > 0)]$ at $\bar{r}(\cdot) = r(\cdot)$ is zero. Hence, the pathwise derivative with respect to r is given by

$$\begin{aligned} V_r(\theta, h_0, \tau)[\bar{r} - r] &= -\lambda E[l(X,\alpha)(\bar{r}(X) - r(X))1(r(X) > 0)] \\ &- \lambda E[u(X,\alpha)(\bar{r}(X) - r(X))1(r(X) < 0)] \\ &- (1-\lambda)E[u(X,\alpha)(\bar{r}(X) - r(X))1(r(X) > 0)] \\ &- (1-\lambda)E[l(X,\alpha)(\bar{r}(X) - r(X))1(r(X) < 0)]. \end{aligned}$$

8.3 Mathematical Proofs

PROOF OF THEOREM 4.1: The proof closely follows that of Theorem 1 in CLV. We would like to show that for any $\varepsilon > 0$, $\Pr\left(\sup_{\tau\in\mathcal{T}_1}\left|\widehat{\theta}(\tau) - \theta_0(\tau)\right| > \varepsilon\right) \to 0$ as $n \to \infty$. For each $\varepsilon > 0$, the event $\left\{\sup_{\tau\in\mathcal{T}_1}\left|\widehat{\theta}(\tau) - \theta_0(\tau)\right| > \varepsilon\right\}$ implies that $\exists \tau' \in \mathcal{T}_1$ s.t. $\left|\widehat{\theta}(\tau') - \theta_0(\tau')\right| \geq \varepsilon$, 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,

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

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

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

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

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

Hence,

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

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

Proof of Theorem 4.2: (1) We first show that $\left|\widehat{\theta}(\tau)-\theta_0(\tau)\right|=O_P\left(n^{-1/2}\right)$ uniformly in $\tau\in\mathcal{T}_1$. By stochastic equicontinuity in Assumption A2(vi), it follows that uniformly in $\tau\in\mathcal{T}_1$

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

and

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

Then by noticing that

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

$E\psi_\theta\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)=O_P(1)$ and $E\psi\left(W,\widehat{\theta}(\tau),\widehat{h},\tau\right)=O_P(1)$, with some algebra we obtain uniformly in $\tau\in\mathcal{T}_1$,

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

By Assumption A2 (iii)(iv)(v), using the notations introduced above Assumption A2, we have uniformly in $\tau\in\mathcal{T}_1$

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

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

Combining the three equations above, and noticing that $E\psi_\theta\left(W,\theta_0\left(\tau\right),h_0,\tau\right)'\Sigma E\psi\left(W,\theta_0\left(\tau\right),h_0,\tau\right)=0$, we obtain, uniformly in $\tau\in\mathcal{T}_1$,

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

where $A_0\left(\tau\right)=V_{\theta_0}\left(\tau\right)'\Sigma V_{\theta_0}\left(\tau\right)+\left(M\left(\tau\right)'\Sigma\otimes I_{d_\theta}\right)V_{\theta_0\theta_0}\left(\tau\right)$. By Assumption A2(iii), $\left|A_0\left(\tau\right)\right|$ is uniformly bounded both from above and from below, hence, we obtain the uniform convergence rate

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

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

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

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

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

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

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

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

It follows that

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

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

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

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 Gaussian 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 Theorem 4.3: (1) By Assumption A3(iii), the function class $\{s(w,\theta,h,\tau)\,u : \theta \in \Theta_1^{\delta}, h \in \mathcal{H}^{\delta}, \tau \in \mathcal{T}_1\}$ is $P-$Donsker. Hence by stochastic equicontinuity and Assumption A3(i), it follows that uniformly in $\tau \in \mathcal{T}_1$

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

Assumption A3(ii) implies

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

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

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

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

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

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

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

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

By law of iterated expectations,

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

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

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

(2) For notation ease, we use the following abbreviations: $\widehat{s}_i(\tau) = s\left(W_i, \widehat{\theta}, \widehat{h}, \tau\right)$, $s_{0i}(\tau) = s(W_i, \theta_0, h_0, \tau)$. The conditional mean of $\widehat{Z}_n^*(\tau)$ is 0, hence we have

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

where the second equality is due to the independence of $\{u_i\}$ and the last equality is due to uniform law of large numbers for the class of function $\mathcal{S}$. By Assumption A3 and Lemma 2.10.14 in van der Vaart and Wellner (1996), the function class $\mathcal{S}^2 := \left\{s(\cdot,\theta,h,\tau)\,s'(\cdot,\theta,h,\tau) : (\theta,h,\tau) \in \Theta_1^\delta \times \mathcal{H}^\delta \times \mathcal{T}_1\right\}$ is Glivenko-Cantelli class, hence by uniform law of large numbers and continuity of $Es(W,\theta,h,\tau)$ and $Es(W,\theta,h,\tau)\,s(W,\theta,h,\tau)'$ in (θ,h), we have uniformly in $\tau \in \mathcal{T}_1$

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n\left(\widehat{s}_i(\tau) - E\widehat{s}_i(\tau)\right)\left(\widehat{s}_i(\tau) - E\widehat{s}_i(\tau)\right)'\\&= E\left[\left(s_{0i}(\tau) - Es_{0i}(\tau)\right)\left(s_{0i}(\tau) - Es_{0i}(\tau)\right)'\right] + o_P(1).\end{aligned}$$

Hence, it follows that uniformly in $\tau \in \mathcal{T}_1$

$$Var^*\left(\widehat{Z}_n^*(\tau)\right) = \Delta(\theta_0,h_0,\tau)E\left[\left(s_{0i}(\tau) - Es_{0i}(\tau)\right)\left(s_{0i}(\tau) - Es_{0i}(\tau)\right)'\right]\Delta(\theta_0,h_0,\tau)' + o_P(1).$$

■

Proof of Lemma 5.1:

We apply Theorem 3.1 in Chernozhukov et al. (2007). Assumption A2 and Theorem 4.2 imply that their Condition C.1 holds with

$$Q_n(\tau) = \left\|M_n\left(\widehat{\theta}(\tau), \widehat{h}(\cdot,\widehat{\theta}(\tau),\tau),\tau\right)\right\|^2,$$

Q the limit of Q_n and $a_n = n$, $b_n = \sqrt{n}$. Similarly, our Theorem 4.2 implies that

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

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

Proof of Corollary 5.1:
Define the set $\widetilde{\Theta}_0 := \left\{\theta_0(\tau) : \tau \in \hat{\mathcal{T}}_0\right\}$, then by triangle inequality for the Hausdorff distance,

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

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

The following lemma is used in some of the subsequent proofs.

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

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

Proof: See proof of Lemma 1 in Chernozhukov, Lee and Rosen (2013). ■

Proof of Proposition 5.1:
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

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

where the second equality follows from the assumption $\Pr\left(\mathcal{T}_0 \subset \hat{\mathcal{T}}_0\right) \to 1$. Since $\sqrt{n}\left|\widehat{\theta}_n(\cdot) - \theta_0(\cdot)\right|$ is stochastic equicontinuous by Theorem 4.2, then according to Lemma A1, $\sqrt{n}\sup_{\tau\in\mathcal{T}_0}\left|\widehat{\theta}_n(\tau) - \theta_0(\tau)\right|$ and $\sqrt{n}\sup_{\tau\in\hat{\mathcal{T}}_0}\left|\widehat{\theta}_n(\tau) - \theta_0(\tau)\right|$ have the same asymptotic distribution, hence

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

■

Proof of Theorem 5.2: We first show the convergence rate of the minimizers 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. $d_H\left(\widetilde{\tau}_n,\widetilde{\mathcal{T}}\right) = O_P\left(n^{-1/2}\right)$, which takes two steps of proofs. At the first step, we show $d_H\left(\widetilde{\tau}_n,\widetilde{\mathcal{T}}\right) \le O_P\left(n^{-1/4}\right)$. 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}-expansion 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\{\left\|M\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}_n\right)\right\| - \left\|M(\widetilde{\theta},\widehat{h},\widetilde{\tau}_n)\right\|\right\} + \left\{\left\|M(\widetilde{\theta},\widehat{h},\widetilde{\tau}_n)\right\| - \left\|M_n(\widetilde{\theta},\widehat{h},\widetilde{\tau}_n)\right\|\right\}\\
&+ \left\{\left\|M_n(\widetilde{\theta},\widehat{h},\widetilde{\tau}_n)\right\| - \left\|M_n(\widetilde{\theta},\widehat{h},\widetilde{\tau})\right\|\right\} + \left\{\left\|M_n(\widetilde{\theta},\widehat{h},\widetilde{\tau})\right\| - \left\|M(\widetilde{\theta},\widehat{h},\widetilde{\tau})\right\|\right\}\\
&+ \left\{\left\|M(\widetilde{\theta},\widehat{h},\widetilde{\tau})\right\| - \left\|M(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau})\right\|\right\}\\
&\le \left\|M\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}_n\right) - M(\widetilde{\theta},\widehat{h},\widetilde{\tau}_n)\right\| + \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\| + \left\|M(\widetilde{\theta},\widehat{h},\widetilde{\tau}) - M(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau})\right\|\\
&= \left\|V_h\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}_n\right)[\widehat{h}-\widetilde{h}_0] + o_P\left(n^{-1/2}\right)\right\| + \left\|V_h\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}\right)[\widehat{h}-\widetilde{h}_0] + o_P\left(n^{-1/2}\right)\right\|\\
&+ \left\|M_n(\widetilde{\theta},h_0,\widetilde{\tau}_n) - M(\widetilde{\theta},h_0,\widetilde{\tau}_n) + o_P\left(n^{-1/2}\right)\right\| + \left\|M_n(\widetilde{\theta},h_0,\widetilde{\tau}) - M(\widetilde{\theta},h_0,\widetilde{\tau}) + o_P\left(n^{-1/2}\right)\right\|\\
&\le \left\|V_h\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}_n\right)[\widehat{h}-\widetilde{h}_0] - V_h\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}\right)[\widehat{h}-\widetilde{h}_0]\right\| + o_P\left(n^{-1/2}\right)\\
&+ 2\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, $\Pr\left(\left\|M\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}_n\right)\right\| < \Delta_{n1}\right) \to 1$, which implies $\widetilde{\tau}_n \in \widetilde{\mathcal{T}}_{\eta_{n1}}$, i.e. $d_H\left(\widetilde{\tau}_n,\widetilde{\mathcal{T}}\right) \le \eta_{n1} = O_P\left(n^{-1/4}\right)$. 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 slowly increasing positive sequence $\{l_n\}$ such that $l_n = o\left(\sqrt{n}\right)$ 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}-expansion 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 $\Pr\left(\left\|M\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right)\right\| \le \Delta_{n2}\right) \to 1$ which implies $\Pr\left(d_H\left(\widetilde{\tau}_n, \widetilde{\mathcal{T}}\right) \le \eta_{n2}\right) \to 1$. For fixed $\widetilde{\tau}_n \in \widetilde{\mathcal{T}}_{\eta_{n1}}$, there exists a $\widetilde{\tau} \in \widetilde{\mathcal{T}}$, such that $|\widetilde{\tau}_n - \widetilde{\tau}| \le O_P\left(n^{-1/4}\right)$. From the first step, we obtain

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

Then by Assumption A4(iii)(iv), it follows that

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

which implies $\Pr\left(\left\|M\left(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}_n\right)\right\| < \Delta_{n2}\right) \to 1$. 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 $B_{\delta_n}(\widetilde{\tau}) := \{\tau \in \mathcal{T} : |\tau - \widetilde{\tau}| \le \delta_n\}$ for all $\widetilde{\tau} \in \widetilde{\mathcal{T}}$. Then

$$\begin{aligned}&\inf_{\tau \in \mathcal{T}} \left\|\sqrt{n} M_n(\widetilde{\theta}, \widehat{h}, \tau)\right\| \\ &= \inf_{\widetilde{\tau} \in \widetilde{\mathcal{T}}} \inf_{\tau \in B_{\delta_n}(\widetilde{\tau})} \left\|\sqrt{n} M_n(\widetilde{\theta}, \widehat{h}, \tau)\right\| + o_P(1)\end{aligned} \tag{25}$$

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

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

hence continuing with (25) and $M(\widetilde{\theta}, \widetilde{h}_0, \widetilde{\tau}) = 0$,

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

By Assumption A4(ii)(iii), we have

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

Then, together with Assumption A4(iv),

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

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 $[V_\tau\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}\right)'\Sigma V_\tau\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}\right)]^-$ be the pseudoinverse of $V_\tau\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}\right)'\Sigma V_\tau\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}\right)$ if it is not invertible, which equals $[V_\tau\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}\right)'\Sigma V_\tau\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}\right)]^{-1}$ if it is invertible, then we can obtain (denote $V_\tau\left(\widetilde{\tau}\right):=V_\tau\left(\widetilde{\theta},\widetilde{h}_0,\widetilde{\tau}\right)$):

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

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

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

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),

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

which leads to our conclusion that

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

■

Proof of Proposition 5.2: (1) Under the null, Lemma A.17 in Santos (2012) indicates

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

Then the rest of the proof follows directly from the proof of Theorem 4.3.

(2) Under the alternative, $T_n\left(\widetilde{\theta}\right)$ diverges at the rate $n^{1/2}$, while $T_n^*\left(\widetilde{\theta}\right)$ diverges at the slower rate $\lambda_n = o\left(\sqrt{n}/\log\log n\right)$, hence the conclusion follows. ■.

Proof of Theorem 5.3:

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

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

Notice that by Assumption A5(iii)

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

hence,

$$\begin{aligned}
&\sqrt{n}\{E_n\xi\left(W,\widehat{g},\tau\right) - \tau'\widetilde{\theta}\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\xi\left(W_i,g_0,\tau\right) - E\xi\left(W_i,g_0,\tau\right) + \zeta\left(W_i,g_0,\tau\right)\} \qquad (26) \\
&+ \sqrt{n}\{E\xi\left(W,g_0,\tau\right) - \tau'\widetilde{\theta}\} + o_P\left(1\right).
\end{aligned}$$

(1) For any fixed $\widetilde{\theta}\in int\left(\Theta_0\right)$, there exists $\delta>0$ such that the closed ball centered at $\widetilde{\theta}$ with radius δ, $B_\delta\left(\widetilde{\theta}\right)\subset\Theta_0$. Then for any $\tau\in\mathbb{S}^{d_\theta}$, by the definition of support function, it follows that

$$\begin{aligned}
\beta\left(\tau|\Theta_0\right) &\geq \sup_{\theta\in B_\delta\left(\widetilde{\theta}\right)} \tau'\theta \\
&\geq \tau'\left(\widetilde{\theta}+\delta\tau\right) \\
&= \tau'\widetilde{\theta}+\delta
\end{aligned}$$

where the second inequality is because $\widetilde{\theta}+\delta\tau \in B_\delta\left(\widetilde{\theta}\right)$. Hence for any $\tau \in \mathbb{S}^{d_\theta}$, we have $E\xi\left(W, g_0, \tau\right) - \tau'\widetilde{\theta} \geq \delta$. Notice that the choice of δ does not depend on τ, and the first term is $O_P\left(1\right)$, thus the second term dominates and we have

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

(2) If $\widetilde{\theta} \notin \Theta_0$, then by similar arguments as in (1) and the Hyperplane Separation Theorem, there exists some $\delta > 0$ and $\widetilde{\tau} \in \mathbb{S}^{d_\theta}$, such that $E\xi\left(W, g_0, \widetilde{\tau}\right) - \widetilde{\tau}'\widetilde{\theta} \leq -\delta$. Hence we have $\inf_{\tau \in \mathbb{S}^{d_\theta}}\{E\xi\left(W, g_0, \tau\right) - \tau'\widetilde{\theta}\} < -\delta$, thus

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

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

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

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

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

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

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

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

We next show that with probability approaching one,

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

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

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

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

Lastly, by Lemma A1, 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 (26),

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

■

References

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

[2] Ai, C. and X. Chen (2007): "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables,"*Journal of Econometrics*, 141, 5–43.

[3] 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.

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

[5] 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.

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

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

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

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

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

[11] Beresteanu, A., I. Molchanov and F. Molinari (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79, 1785-1821.

[12] ———— (2012): "Partial Identification Using Random Set Theory", *Journal of Econometrics*, 166, 17-32.

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

[14] Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica*, 75, 323-363.

[15] Blundell, R. W., and J. L. Powell (2004): "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71, 655–679.

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

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

[18] Chandrasekhar, A., V. Chernozhukov, F. Molinari and P. Schrimpf (2012): "Inference for best linear approximations to set identified functions," CeMMAP working papers CWP43/12.

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

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

[21] Chernozhukov, V., I. Fernández-Val and B. Melly (2013): "Inference on Counterfactual Distributions," forthcoming *Econometrica*.

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

[23] Chernozhukov, V., S. Lee and A. M. Rosen (2013): "Intersection bounds: estimation and inference," *Econometrica*, 81, 667–737.

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

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

[26] Conley, T. G., C. B. Hansen and P. E. Rossi (2012): "Plausibly Exogenous," *Review of Economics and Statistics*, 94, 260-272.

[27] Dudley, R.M. (1999): *Uniform Central Limit Theorems*, Cambridge Studies in Advanced Mathematics, 63, Cambridge University Press, Cambridge, UK.

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

[29] ———— (2013): "Uniform Convergence of Weighted Sums of Non- and Semi-parametric Residuals for Estimation and Testing," forthcoming *Journal of Econometrics*.

[30] Escanciano, J. C. and L. Zhu (2012): "Inferences in Semiparametric Partially Identified Models: An Empirical Process Approach," Working paper.

[31] Fan, Y. and D. Zhu (2009): "Partial Identification and Confidence Sets for Functionals of the Joint Distribution of Potential Outcomes," Working paper.

[32] Faust, J., E. T. Swanson and J. H. Wright (2004): "Identifying VARS based on high frequency futures data," *Journal of Monetary Economics*, 51, 1107-1131.

[33] Firpo, S. and G. Ridder (2010): "Bounds on Functionals of the Distribution of Treatment Effects," Working paper.

[34] 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.

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

[36] Hall, A., Inoue, A., (2003): "The large sample behavior of the generalized method of moments estimator in misspecified models," *Journal of Econometrics,* 114, 361–394.

[37] Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica,* 50,1029-1054.

[38] Heckman, J and E. Vytlacil (2007): "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Economic Estimators to Evaluate Social Programs and to Forecast Their Effects in New Environments." In Handbook of Econometrics, vol. 6, edited by J. Heckman and E. Leamer, Amsterdam: Elsevier.

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

[40] Hong, S. (2012), "Inference in Semiparametric Conditional Moment Model with Partial Identification," Working Paper.

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

[42] ——— (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, 77-84.

[43] ——— (2006): "Identification and Estimation of Statistical Functionals Using Incomplete Data," *Journal of Econometrics*, 132, 445-459.

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

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

[46] Imbens, G. W and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica* 72, 1845-1857.

[47] Kaido, H. (2012): "A Dual Approach to Inference for Partially Identified Econometric Models," Working paper.

[48] Kaido, H. and H. White (2009): "Inference on Risk Neutral Measures for Incomplete Markets," *Journal of Financial Econometrics*, 7, 199-246.

[49] Klein, R., C. Shen, and F. Vella (2011): "Semiparametric Selection Models with Binary Outcomes," Working Paper.

[50] Kline, P. and A. Santos (2013): "Sensitivity to Missing Data Assumptions: Theory and an Evaluation of the U.S. Wage Structure" *Quantitative Economics*, 4, 231–267.

[51] 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.

[52] Lee, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76, 1071-1102.

[53] 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.

[54] Lippi, M., and L. Reichlin (1993), "The dynamic effects of aggregate demand and supply disturbances: comment," *The American Economic Review*, 83(3), 644–652.

[55] ——— (1994), "VAR analysis, nonfundamental representations, Blaschke matrices," *Journal of Econometrics*, 73, 307–325.

[56] Magnac, T., and D. Thesmar (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 20, 801-816.

[57] Mammen, E. (1993): "Bootstrap and wild bootstrap for high dimensional linear models," *The Annals of Statistics*, 21, 255–285.

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

[59] ——— (2003): *Partial Identification of Probability Distributions* (Berlin: Springer-Verlag).

[60] ——— (2007): *Identification for Prediction and Decision*. Cambridge, MA: Harvard Univ. Press.

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

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

[63] Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349-1382.

[64] Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, Vol. 4, 2113-2245.

[65] Pacini, D.H. (2012): "Least Square Linear Prediction with Two-Sample Data", Working Paper.

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

[67] 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.

[68] ——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169-211.

[69] Rosenbaum, P. R. (2002): *Observational Studies*, 2nd Edition, Springer-Verlag, New York.

[70] 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.

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

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

[73] 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.

[74] 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.

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

[76] 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.

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

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

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

[80] White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.

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