Inference in Semiparametric Conditional Moment Models with Partial Identification*

Shengjie Hong
Department of Economics
Tsinghua University, China
email: hongshj@sem.tsinghua.edu.cn

First Draft: April 2011
This Draft: October 2012

Abstract

This paper develops inference methods for conditional moment models in which the unknown parameter is possibly partially identified and may contain infinite-dimensional components. We consider testing the hypothesis that a given restriction on the parameter is satisfied by at least one element of the identification set. We propose using the sieve minimum of a Kolmogorov-Smirnov type statistic as the test statistic, derive its asymptotic distribution, and provide consistent bootstrap critical values. In this way a broad family of restrictions can be consistently tested, making the proposed procedure applicable to various types of inference. In particular, we show how to: (1) test the semiparametric model specification; (2) construct confidence sets for unknown parametric components; and (3) construct confidence sets for unknown functions at a given point. The specification test is consistent against fixed alternatives. The confidence sets have correct asymptotic coverage probability, excluding any value outside the identification set with asymptotic probability one. Our methods are robust to partial identification, and allow for the moment functions to be nonsmooth. As an illustration, we apply the proposed inference methods to study the quantile instrumental variable Engel curves for gasoline in Brazil. A Monte Carlo study demonstrates finite sample performance.

Keywords: Conditional moment equalities, identification set, model specification test, confidence set, Sieve space, ill-posedness, bootstrap.

*I thank Bruce Hansen and Jack Porter for guidance and encouragement. I am indebted to Andres Aradillas-Lopez, Demian Pouzo, Andres Santos, Xiaoxia Shi, and Ping Yu for helpful discussions. Any error is my own responsibility.
1 Introduction

In this paper, we consider inference in conditional moment models of the following form:

\[ E \left[ m (Y, \theta_0, h_0(X)) \right] | Z = 0 \]  

(1)

where the vector-valued moment function \( m (\cdot) \) is known up to a finite dimensional parameter \( \theta_0 \in \Theta \) and an unknown function \( h_0 \in \mathcal{H} : \mathcal{X} \to \mathbb{R}^H \). \((X, Y, Z)\) are the observable variables whose true probability distributions are unknown. Model (1) is a natural semi-parametric extension of the parametric conditional moment models

\[ E \left[ m (Y, X, \theta_0) \right] | Z = 0 \]  

(2)

which are well known from the work of Hansen (1982), Chamberlain (1987), Newey (1990), and others. We allow for the unknown parameter \( \alpha_0 \equiv (\theta_0, h_0) \) to be partially identified in the sense that the identification set

\[ \mathcal{A}_I \equiv \{ \alpha \in \Theta \times \mathcal{H} : E \left[ m (Y, \theta, h(X)) \right] | Z = 0 \text{ a.s. on } Z \} \]

may have more than one element.

The conditional moment restrictions in (1) have been studied extensively under point-identification assumptions. This line of work led to a rich literature, including important papers such as Newey and Powell (2003), Ai and Chen (2003), and Darolles, Fan, Florens, and Renault (2011). Newey and Powell (2003) provide consistent estimators for \( \alpha_0 \). Ai and Chen (2003) establish the \( \sqrt{n} \) asymptotic normality and efficiency of their estimators for \( \theta_0 \). Both papers utilize sieve estimation. Darolles et al. (2011) provide consistent estimators for \( \alpha_0 \) using a Tikhonov regularization approach. Recently, Chen, Chernozhukov, Lee, and Newey (2011) derive sufficient conditions for achieving local point-identification in Model (1). These regularity conditions provide new insights for understanding both semi-parametric identification and its challenges. Canay, Santos, and Shaikh (2011) study the testability of necessary conditions for point-identification in some nonparametric models with endogeneity.

A growing number of papers have drawn attention to cases where point-identification fails in special cases of model (1). For instance, Santos (2011) shows examples of partial identification in nonparametric instrumental variable (IV) regressions. As another example, Aradillas-Lopez (2010) shows examples of partial identification in the context of incomplete information entry-games. The possible lack of point-identification makes it desirable to allow for partial identification in the conditional moment models. Pioneered by Manski (1990, 2003), partial identification analysis for parametric models has

\[ ^1\text{We allow for } X, Y, Z \text{ to have common components.} \]

We extend the existing literature by developing methods for hypothesis testing under the semiparametric specification of Model (1) without assuming point-identification. First, we propose a general procedure for testing the hypothesis that a given restriction on the parameter $\alpha$ is satisfied by at least one element of the identification set $A_I$. Then we show that a broad family of restrictions can be consistently tested, making the proposed procedure applicable to various types of inference. In particular, we demonstrate how to test the model specification, construct confidence sets for the parametric component $\theta$, and construct confidence sets for the unknown function $h(\cdot)$ at a given point. Our methods are robust to partial identification, and allow for the moment function $m(\cdot)$ to be pointwise nonsmooth in $\alpha$.

Our confidence sets for $\theta$, denoted by $CS_n$, have the correct asymptotic coverage probability in the following sense: For a targeted level $1 - \rho$,

$$\lim \inf_{n \to \infty} \inf_{\theta \in \Theta_I} \Pr (\theta \in CS_n (1 - \rho)) \geq 1 - \rho \quad (3)$$

where $\Theta_I$ is the identification set for $\theta$. The infimum in (3) is carried out before the limit is taken, which implies uniformity of correct asymptotic coverage probability over $\Theta_I$.

A potential alternative way of constructing CS’s for $\theta$ is to project the CS’s for $(\theta, h)$ onto $\Theta$. However, this alternative has the following drawbacks: (i) As pointed out by Hahn and Ridder (2009), CS’s built from projection of some higher dimensional CS’s can be very conservative; (ii) To the best of our knowledge, among existing methods, Andrews and Shi (2011)’s is the only one that can be potentially used for constructing CS’s for the infinite-dimensional $(\theta, h)$ with partial identification. Yet, Andrews and Shi (2011) are primarily concerned with finite-dimensional parameters. The computational cost of using their method for infinite-dimensional parameters can be prohibitive; (iii) In general it is impossible to write out the resulting CS’s for $\theta$ in a form that is practically useful.

Conditional moment model (1) has been recognized as an important branch of semiparametric modeling in applied econometrics. It encompasses nonparametric IV $E [Y - h_0(X) | Z] = 0$, partially linear IV $E [Y - X' \theta_0 - h_0 (X_2) | Z] = 0$, single index IV $E [Y - h_0 (X' \theta_0) | Z] = 0$, nonparametric quantile IV $E [1 \{Y \leq h_0 (X)\} | Z] = \tau$, and semiparametric quantile IV as special cases. Model (1) also arises in many interaction-based economic environments with optimizing agents. For example, consider the incom-
plete information entry-game in Aradillas-Lopez (2010): It can be shown that the parameters in the payoff functions are characterized by conditional moment equalities. And the moment functions would contain the unobserved belief under Bayesian-Nash equilibrium as a nuisance parameter.


In Section 2, we regularize the parameter space \( A \) and formally describe the hypothesis test to be considered. This section serves as a road map for the following sections. In Section 3, we introduce the general testing procedure, derive the asymptotic distribution of the test statistic, and show consistency results for the hypothesis test. Bootstrap critical values are provided. We discuss the specification test as a special case of the general test at the end of this section. In Section 4, we show how to utilize the test procedure developed in Section 3 to construct confidence sets. Finite sample performance is studied by Monte Carlo simulations in Section 5. Section 6 is an empirical illustration of the proposed inference methods applied to study the quantile IV Engel curves for gasoline. Section 7 concludes. Mathematical proofs of all the theorems stated in the paper are given in Appendix A.

## 2 Inference with Partial Identification

Model (1) defines the true value of the parameter \( \alpha = (\theta, h (\cdot)) \) by the following conditional moment restriction:

\[
E [m (Y, \theta, h (X)) | Z] = 0
\]

where \( \theta \) is finite dimensional, and \( h \) is infinite dimensional.

We say that the parameter \( \alpha = (\theta, h) \) is partially identified by Model (1), if the identification set
\[ \mathcal{A}_f \equiv \{ (\theta, h) \in \Theta \times \mathcal{H} : E[m(Y, \theta, h(X)) | Z] = 0 \} \]

contains more than one element. Santos (2011) provides several examples of partially identified nonparametric IV. In our simulation study (reported in Section 5) we consider a partially identified semiparametric quantile IV, which is another example of partially identified Model (1).

In what follows, we first regularize the parameter space \( \mathcal{A} = \Theta \times \mathcal{H} \). Then we formally describe the general form of the hypothesis tests that we consider.

### 2.1 Restrictions on The Parameter Space \( \mathcal{A} \)

The parameter space \( \mathcal{A} \) takes the form of a product space: \( \Theta \times \mathcal{H} \). We put the usual compactness assumption on \( \Theta \subseteq \mathbb{R}^{d_\theta} \). Smoothness assumptions on the unknown function are often necessary for achieving consistency of inference in non/semiparametric models. For example, Horowitz (2009) shows that it is impossible to consistently test the specification of a nonparametric IV if nonsmooth functions are allowed. It is also reasonable to believe that the unknown function is smooth in many economic applications. For the conditional moment models, we impose the same smoothness assumption on \( \mathcal{H} \) as in Santos (2011): we require \( \mathcal{H} \) to be a subset of a Sobolev space. Similar assumptions have been made by Ai and Chen (2003), Newey and Powell (2003) and many other authors.

**Definition 2.1.** (i) Let \( X \in \mathcal{X} \subseteq \mathbb{R}^{d_x} \). For \( \lambda \in \mathbb{N}_{+}^{d_x} \), \( |\lambda| \equiv \sum_{i=1}^{d_x} \lambda_i \) and \( D^\lambda h(x) \equiv \partial^{|\lambda|} h(x) / \partial x_{1}^{\lambda_{1}} \cdots \partial x_{d_x}^{\lambda_{d_x}} \). Define\(^2\):

\[ \|h\|_s = \sum_{|\lambda| \leq d} \int_{\mathcal{X}} \| D^\lambda h(x) \|_E^2 dx \]

where \( h \) is assumed to be \( d \)-times differentiable.

(ii) Define the following Sobolev space:

\[ W^s(\mathcal{X}) \equiv \{ h : \mathcal{X} \rightarrow \mathbb{R}^H \text{ s.t. } h \text{ is } d \text{-times differentiable and } \|h\|_s < \infty \} \]

We make the following assumption on the parameter space \( \mathcal{A} = \Theta \times \mathcal{H} \):

**Assumption 2.1.** (i) \( \Theta \) is a compact subset of \( \mathbb{R}^{d_{\theta}} \); (ii) \( d \geq d_x + 2 \), and \( \mathcal{H} = \{ h \in W^s(\mathcal{X}) : \|h\|_s \leq B \} \) for some \( B < \infty \).

Assumption 2.1 (ii) requires that \( h \) is at least \((d_x + 2)\)-times differentiable and belongs to a bounded subset of the Sobolev space \( W^s(\mathcal{X}) \). As shown by Gallant and Nychka (1987),

\(^2\| \cdot \|_E \) is the Euclidean norm, i.e. \( \|a\|_E = \sqrt{a'a} \).
Assumption 2.1 (ii) implies that \( \mathcal{H} \) is compact under the norm \( \| \cdot \|_c \) defined as
\[
\| h \|_c \equiv \max_{|\lambda| \leq \frac{d}{2}} \sup_{x \in X} \| D^\lambda h(x) \|_E,
\]
an important property needed in our asymptotic analysis.

## 2.2 Hypothesis Test

For a given restriction on the parameter \( \alpha \), we want to test the hypothesis that at least one element of the identification set \( \mathcal{A}_I \) satisfies the restriction. This notion of testing encompasses hypothesis tests under point-identification as a special case where we test whether the true value (the unique element of \( \mathcal{A}_I \)) satisfies a given restriction.

More specifically, denote by \( L(\alpha) = a \) a restriction on \( \alpha \) that we want to test for, where \( L \) is a mapping on \( \mathcal{A} \), and \( a \) is a constant that belongs to certain Banach space. Define the restricted set \( R \) as the set of parameter values that satisfy the given restriction:
\[
R \equiv \{ \alpha \in \mathcal{A} : L(\alpha) = a \}.
\]
(4)

Then the hypothesis test takes the following form:
\[
\begin{align*}
H_0 &: \quad \mathcal{A}_I \cap R \neq \emptyset; \\
H_1 &: \quad \mathcal{A}_I \cap R = \emptyset,
\end{align*}
\]
where \( \emptyset \) denotes the empty set.

We impose the following assumption on the restriction \( L(\alpha) = a \) that we consider testing:

**Assumption 2.2.** We only consider testing restrictions of the form \( L(\alpha) = a \) where \( L : (\mathcal{A}, \| \cdot \|_c) \rightarrow (\mathcal{L}, \| \cdot \|_L) \) is a bounded linear operator.

The seemingly restrictive requirements of boundedness and linearity of Assumption 2.2 are compensated by flexibility in choosing the Banach space \((\mathcal{L}, \| \cdot \|_L)\). And Assumption 2.2 is satisfied by a broad family of restrictions. For example, it is satisfied by the restriction \( L(\alpha) = 0 \) with \( L(\cdot) \equiv 0 \), which leads to the model specification test: \( H_0 : \mathcal{A}_I \neq \emptyset \) vs \( H_1 : \mathcal{A}_I = \emptyset \). Assumption 2.2 is also satisfied by the restrictions (i) \( \theta = \bar{\theta} \) for a constant \( \bar{\theta} \in \Theta \) and (ii) \( h(x_0) = \mu \) for a given \( x_0 \in X \) and a constant \( \mu \). As it will be shown, consistent confidence sets for \( \theta \) and \( h(x_0) \) can be built from inverting the tests for restrictions (i) and (ii), respectively. We discuss the specification test in detail at the end of Section
3. And we study the asymptotic properties of the confidence sets in Section 4. For more examples of restrictions that satisfy Assumption 2.2, see Santos (2011).

In the next section, we develop a general procedure for testing a generic restriction which satisfies the assumption specified above.

3 The Test Procedure

A popular method of handling conditional moment restrictions is transforming them into a number of (in some case infinitely many) unconditional moment restrictions. This method dates back to Bierens (1982) and is adopted by recent papers such as Dominguez and Lobato (2004), Kim (2009), Santos (2011), and Andrews and Shi (2011). We also adopt this method.

With partial identification, the transformation from conditional moment restrictions into unconditional ones needs to be done carefully in order to prevent any loss of identification power. Therefore, we require a family of instrument functions \( \{g(t, \cdot) : Z \to \mathbb{R}, t \in T\} \) such that for all random variables \( V \) with \( E(|V|) < \infty \):

\[
E(V|Z) = 0 \iff E[V \cdot g(t, Z)] = 0 \text{ for all } t \in T. \tag{5}
\]

Consequently, \( E[m(Y, \theta, h(X))|Z] = 0 \iff E[m(Y, \theta, h(X)) \cdot g(t, Z)] = 0 \text{ for all } t \in T. \)

And the identification set \( A_I \) can be equivalently written as

\[
A_I = \{(\theta, h) \in \Theta \times H : E[m(Y, \theta, h(X)) \cdot g(t, Z)] = 0 \forall t \in T\}.
\]

Instrument functions that satisfy condition(5) often exist for example: \( g(t, z) = 1(z \leq t)^3 \) with \( t \in Z \) as shown by Dominguez and Lobato (2004) and \( g(t, z) = e^{t'z} \) with any \( T \) of positive Lebesgue measure if \( Z \) is bounded almost sure as shown by Bierens (1990). A detailed discussion on valid choices of instrument functions can be found in Andrews and Shi (2011).

In summary, we impose the following assumption on the instrument functions:

**Assumption 3.1.** (i) \( g \) satisfies condition (5); (ii) \( T \in \mathbb{R}^{d_t} \), and \( g \) satisfies one of the following conditions: a) \( g(t, z) = 1(z \leq t) \) for \( t \in Z \), or b) \( T \) is compact, and \( \exists M > 0 \text{ s.t. } |g(t_1, z) - g(t_2, z)| \leq M ||t_1 - t_2||_E \text{ for all } t_1, t_2 \in T, \text{ and for all } z \in Z. \)

\( 1(z \leq t) \) is the indicator function. It equals 1 if each component in \( z \) is less than or equal to the corresponding component in \( t \), and equals 0 otherwise.
3.1 Test Statistic

Define \( J_n(\alpha, t) \equiv \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) \cdot g(t, z_i) \). \( J_n(\alpha, t) \) is the sample average of the instrument function weighted moments evaluated at \((\alpha, t)\). We propose using the following test statistic for hypothesis testing:

\[
S_n = \min_{\alpha \in A \cap R} \left\{ \sup_{t \in T_n} n [J_n(\alpha, t)]' W_n(\alpha, t) [J_n(\alpha, t)] \right\}
\]

(6)

where \( A_n = \Theta \times \mathcal{H}_n \) with \( \mathcal{H}_n \uparrow \mathcal{H} \) as \( n \to \infty \), \( T_n \uparrow T \) as \( n \to \infty \), and \( W_n \) is a weighting matrix.

\( S_n \) is the minimum of a Kolmogorov-Smirnov type statistic. A Cramér-von Mises (CvM) version of \( S_n \), in which \( \sup \) is replaced by \( \int_{T_n} \ldots dQ(t) \) with some distribution function \( Q(\cdot) \), is an interesting potential alternative. We expect \( S_n \) and its CvM alternative to exhibit different power properties, but do not expect one to dominate the other. We leave the CvM alternative for future research. As the expression \( W_n(\alpha, t) \) suggests, we allow for \( W_n \) to be data dependent and/or \((\alpha, t)\)-dependent.

We impose the following assumption on the weighting matrix \( W_n \):

**Assumption 3.2. (the weighting matrix)** (i) \( W_n(\alpha, t) \) is positive semi-definite; (ii) There is a positive definite \( W(\alpha, t) \) such that: a) \( W \leq W(\alpha, t) \leq \overline{W} \) for all \((\alpha, t) \in A \times T \) for some positive definite matrices \( \underline{W} \) and \( \overline{W} \); and (b) \( \sup_{A \times T} |W_n(\alpha, t) - W(\alpha, t)| = O(n^{-1/2}) \) in a per-element sense.

Assumption 3.2 (i) is a natural requirement for weighting matrices. Assumption 3.2 (ii) requires that the weighing matrix \( W_n(\alpha, t) \) converges at the \( 1/\sqrt{n} \) rate to a limiting matrix \( W(\alpha, t) \) that is bounded from above and bounded away from zero.

3.2 Definitions

Our main interest is the asymptotic behavior of the test statistic \( S_n \). In this subsection, we introduce several important definitions that are needed for developing the asymptotic results.

---

4An example of weighting matrices that depend both on the data and the choice of the parameter value is the continuous-updating type weighting matrices.

5For two positive semi-definite matrices \( A \) and \( B \), we say \( A \geq B \) if \( A - B \) is positive semi-definite.
Define \( \hat{\alpha}_n \in \arg \min_{\alpha \in A_n \cap R} \left\{ \sup_{t \in T} [J_n(\alpha, t)]' W_n(\alpha, t) [J_n(\alpha, t)] \right\} \).

(7)
as the minimizer of the objective function \( \sup_{t \in T} [J_n(\alpha, t)]' W_n(\alpha, t) [J_n(\alpha, t)] \).

Just like in many point-identification analyses, the asymptotic behavior of \( \hat{\alpha}_n \) provides important insight on that of \( S_n \) within the partial identification framework. In particular, we find the distance between \( \hat{\alpha}_n \) and the identification set \( A_I \) with respect to norms \( \| \cdot \|_{L^2} \) and \( \| \cdot \|_w \) (to be defined immediately) plays a crucial role in our analysis.

\( \| \cdot \|_{L^2} \) is the \( L^2 \) norm, defined as

\[
\| \alpha \|_{L^2} = \| \theta \|_E + \| h \|_{L^2} = \| \theta \|_E + \sqrt{E \left[ (h(X))^2 \right]}.
\]

And the distance between \( \hat{\alpha}_n \) and \( A_I \) with respect to \( \| \cdot \|_{L^2} \) is defined as

\[
d_{L^2}(\hat{\alpha}_n, A_I) \equiv \inf_{\alpha \in A_I} \| \hat{\alpha}_n - \alpha \|_{L^2}.
\]

The second norm \( \| \cdot \|_w \) is a pseudo norm that is weaker than \( \| \cdot \|_{L^2} \). To define \( \| \cdot \|_w \), we need to introduce the following notion of functional derivatives:

**Definition 3.1. (pathwise derivatives)** (i) For any functional \( f(\cdot) : \mathcal{H} \to \mathbb{R} \), define

\[
\frac{df(h_0)}{dh} [\Delta_h] = \frac{df(h_0 + \tau \Delta_h)}{d\tau} |_{\tau=0}
\]
as the pathwise first derivative of \( f \) at \( h_0 \) in the direction of \( \Delta_h \);

(ii) Define

\[
\frac{d^2 f(h_0)}{dh^2} [\Delta, \Delta] = \frac{d^2 f(h_0 + \tau \Delta)}{d\tau^2} |_{\tau=0}
\]
as the pathwise second derivative of \( f \) at \( h_0 \) in the direction of \( \Delta \);

(iii) For any function \( f(\cdot) : \mathcal{A} \to \mathbb{R} \), define

\[
\frac{df(\alpha_0)}{d\alpha} [\Delta] = \frac{df(\alpha_0 + \tau \Delta)}{d\tau} |_{\tau=0} = \frac{\partial f(\theta_0, h_0)}{\partial \theta'} (\Delta_\theta) + \frac{df(\theta_0, h_0)}{dh} [\Delta_h]
\]
as the pathwise first derivative of \( f \) at \( \alpha_0 = (\theta_0, h_0) \) in the direction of \( \Delta = (\Delta_\theta, \Delta_h) \).

**Definition 3.2. (pseudo norm \( \| \cdot \|_w \))** Denote by \( \rho(\alpha, t) \equiv E [m(Y, \theta, h(X)) \cdot g(t, Z)] \). When the “\( \arg \min \)” is not well defined, \( \hat{\alpha}_n \) is defined as \( Q_n(\hat{\alpha}_n) \leq \inf_{\alpha \in A_n} Q_n(\alpha) + o_p(1) \) where \( Q_n(\alpha) \equiv \sup_{t \in T} [J_n(\alpha, t)]' W_n(\alpha, t) [J_n(\alpha, t)] \).
Define

\[ \|\hat{\alpha}\|_w = \sup_{\alpha_0 \in A_I} \sup_{t \in T} \left\| \frac{d\rho(\alpha_0, t)}{d\alpha}[\hat{\alpha}] \right\|_E. \]

And the distance between \( \hat{\alpha}_n \) and \( A_I \) with respect to \( \|\cdot\|_w \) is defined as

\[ d_w(\hat{\alpha}_n, A_I) \equiv \inf_{\alpha \in A_I} \|\hat{\alpha}_n - \alpha\|_w. \]

We derive the convergence rates of \( d_{L_2}(\hat{\alpha}_n, A_I) \) and \( d_w(\hat{\alpha}_n, A_I) \), and summarize the results in Lemma 2 and Lemma 3 in Appendix A. These results are used in deriving the asymptotic distribution of our test statistic. Notably, they parallel the convergence rates results under point-identification in Chen and Pouzo (2011).

**Remark 3.1.** (i) According to Definition 3.1, in the direction of \( \Delta = (\Delta_\theta, \Delta_h) \)

\[ \frac{d\rho(\alpha_0, t)}{d\alpha}[\Delta] = \left. \frac{d\rho(\alpha_0 + \tau\Delta, t)}{d\tau} \right|_{\tau=0} = \frac{\partial\rho(\alpha_0, t)}{\partial\theta'}(\Delta_\theta) + \frac{d\rho(\alpha_0, t)}{dh}[\Delta_h]. \]

(ii) To better understand Definition 3.1, it is helpful to consider examples.

(a) For a partially linear IV where

\[ \rho(\theta, h, t) = E \left[ (Y - X_1\theta - h(X_2)) \cdot g(t, Z) \right], \]

in the direction of \( (\Delta_\theta, \Delta_h) \), its pathwise first derivative at \( \alpha_0 \) is

\[ -E \left[ (X_1\Delta_\theta + \Delta_h(X_2)) \cdot g(t, Z) \right], \tag{8} \]

and its pathwise second derivative is 0;

(b) For a partially linear quantile IV where

\[ \rho(\theta, h, t) = E \left[ \left\{ 1 \left( Y - X_1\theta - h(X_2) \leq 0 \right) - \tau \right\} \cdot g(t, Z) \right], \]

in the direction of \( (\Delta_\theta, \Delta_h) \), its pathwise first derivative at \( \alpha_0 \) is

\[ E \left[ f_{Y|X,Z} \left( X_1\theta_0 + h_0(X_2) \right) \cdot (X_1\Delta_\theta + \Delta_h(X_2)) \cdot g(t, Z) \right], \tag{9} \]

and its pathwise second derivative is

\[ E \left[ f'_{Y|X,Z} \left( X_1\theta_0 + h_0(X_2) \right) \cdot (X_1\Delta_\theta + \Delta_h(X_2))^2 \cdot g(t, Z) \right]. \tag{10} \]
(iii) (a) In the point-identified case, $\|\alpha\|_w = \max_{t \in T} \left\| \rho (\alpha_0, t) \right\|_E$. Compared with

$$
\|\alpha\|_0 = \sqrt{E \left\{ \left| dE (m (Y, \alpha_0) | Z) \right| \left\| \alpha \right\|_E^2 \right\}}
$$

which is the standard weak norm considered in the literature (see, for example, Ai and Chen (2003) and Chen and Pouzo (2011)),

$$
\|\alpha\|_w = \sup_{t \in T} \left\| dE \left( m (Y, \alpha_0) \cdot g (t, Z) \right) \left\| \alpha \right\|_E \right\|
$$

$$
\leq C \cdot \|E \left[ dE (m (Y, \alpha_0) | Z) \right] \|_E
\leq C \cdot \|\alpha\|_0.
$$

Therefore, $\|\cdot\|_w \preceq \|\cdot\|_0$.

(b) When $m (y, h (x)) = y - h (x)$, the $\|h\|_w = \max_{t \in T} \left| E \left[ h (X) g (Z, t) \right] \right|$ which coincides with Santos (2011).

We introduce the following measure of ill-posedness to account for the ill-posed problem that may arise in our semiparametric specification:

**Definition 3.3. (sieve measure of ill-posedness)** The sieve measure of ill-posedness is defined as

$$
\psi_n = \sup_{\alpha \in A_n} \inf_{P_n (\alpha_0) \neq 0} \left( \frac{\inf_{\alpha_0 \in A_0} \|\alpha - P_n (\alpha_0)\|_{L^2}}{\inf_{\alpha_0 \in A_0} \|\alpha - P_n (\alpha_0)\|_w} \right),
$$

where $P_n (\alpha_0)$ is the projection of $\alpha_0$ onto $A_n$.$^7$

Under point-identification, the sieve measure of ill-posedness in Definition 3.3 is similar to the one defined in Chen and Pouzo (2011).

Finally, we introduce two families of functions, $V_{\alpha_0}^{k_n}$ and $V_{\alpha_0}^{\infty}$. As it will be shown, $V_{\alpha_0}^{\infty}$ appears in the asymptotic null distribution of $S_n$. $\{V_{k_n}^{\alpha_0}\}$ forms an increasing series of sets with $V_{\alpha_0}^{\infty}$ being the limiting set.

**Definition 3.4.** Let $H_{nR} = \{(r_\theta, r_h) \in \mathbb{R}^{d_\theta + k_n} : L \left( (r_\theta', r_h' p_{k_n}') \right) = 0\}$ in which $p_{k_n} = (p_1, ..., p_{k_n})'$ is the vector of basis functions for the sieve space $H_n$.

(i) Define the following family of functions:

$^7$More specifically, $P_n (\alpha_0) = P_n ((\theta_0, h_0)) = (\theta_0, P_n (h_0))$ where $P_n (h_0)$ is the projection of $h_0$ onto $H_n$. 


\[ V_{k_n}^{\alpha_0} = \left\{ v : T \to \mathbb{R}^{d_{m}} \text{ s.t. } v(t) = \frac{\partial \rho(\theta_0, h_0, t)}{\partial \theta} \cdot r_\theta + \frac{d \rho(\theta_0, h_0, t)}{dh} [p_k^{kn}]^t r_h, (r_\theta, r_h) \in H_{nR} \right\} \]

for each \( \alpha_0 \in A_f \), where \( \frac{d \rho(\theta_0, h_0, t)}{dh} [p_k^{kn}]^t \) is a \( d_m \times k_n \) matrix defined as

\[
\left( \frac{d \rho(\theta_0, h_0, t)}{dh} [p_k^{kn}] \right) ;
\]

(ii) Define \( V_{k_n}^{\alpha_0} \) as the closure of \( \bigcup_{n=1}^\infty V_{k_n}^{\alpha_0} \) under the supreme norm \( \| \cdot \|_\infty \) for each \( \alpha_0 \in A_0 \).

In words, \( V_{k_n}^{\alpha_0} \) consists of all possible first order deviations (in the sense of the pathwise derivatives) of the population moment \( E[m(Y, \cdot \cdot g(t, Z))] \) (without violating the given restriction \( L(\alpha) = a \)) from its value at \( \alpha_0 \). Note that \( H_{nR} \) forms a linear subspace in \( \mathbb{R}^{d_0 + d_k} \) because of the linearity of \( L(\cdot) \).

### 3.3 Asymptotic Results on \( S_n \)

Assumptions 2.1, 2.2, 3.1 and 3.2 state regularity conditions on the parameter space, the form of the restriction to be tested, the instrument functions, and the weighting matrix, respectively. In addition, we impose the following assumptions in order to obtain the asymptotic distribution of \( S_n \).

**Assumption 3.3.** (i) \( \{X_i, Y_i, Z_i\}_{i=1}^n \) is i.i.d. with \( X \in \mathcal{X} \subseteq \mathbb{R}^{d_x}, Y \in \mathcal{Y} \subseteq \mathbb{R}^{d_y}, Z \in \mathcal{Z} \subseteq \mathbb{R}^{d_z} \), and \( \mathcal{X} \) is bounded; (ii) \( \{m(\cdot, \alpha) : \alpha = (\theta, h) \in A\} \) is a uniformly bounded Donsker class; (iii) In a \( \| \cdot \|_{L_2} \)-neighborhood of \( A_f \) defined as \( B(A_f, \xi) = \{\alpha \in A : d_{L_2}(\alpha, A_f) \leq \xi\} \) for some \( \xi > 0 \), \( \rho(\alpha, t) \) is pathwise differentiable with respect to \( \alpha \); (iv) \( \exists \) positive constants \( c_1, c_2 \) and \( \delta \) s.t. \( c_1 \inf_{\alpha_0 \in A_f} \|\alpha - \alpha_0\|_w \wedge \delta \leq \max_{t \in T} \|\rho(\alpha, t)\|_E \leq c_2 \inf_{\alpha_0 \in A_f} \|\alpha - \alpha_0\|_w \) for all \( \alpha \in A_f \); (v) \( \psi_n = o \left( n^{1/4} \right) \).

**Assumption 3.4.** (i) The dimension of the sieve space \( H_n \), denoted by \( k_n \), satisfies \( k_n < \infty \), \( k_n \to \infty \), and \( \frac{k_n}{n} \to 0 \) as \( n \to \infty \); (ii) The eigenvalues of \( E[p_k^{kn}(X)p_k^{kn}(X)] \) are bounded, uniformly on \( n \), i.e. denote by \( \tau_n \) the largest eigenvalue of \( E[p_k^{kn}(X)p_k^{kn}(X)] \), then there is \( J > 0 \) s.t. \( \tau_n \leq J \) for all \( n \);

**Assumption 3.5.** There exists \( \Pi_n h \in H_n \) for each \( h \in H \) such that: (i) \( \|\Pi_n h - h\|_{L_2} = o(1) \) for all \( h \in H \); (ii) \( \delta_{w,n} \equiv \sup_{h_0 \in H_f} \|h_0 - \Pi_n h_0\|_w = o \left( n^{-1/2} \right) \) and \( \delta_{s,n} \equiv \sup_{h_0 \in H_f} \|h_0 - \Pi_n h_0\|_{L_2} = o \left( n^{-1/2} \right) \)
\[ o(n^{-1/4}); \text{(iii)} \text{ For some } \gamma_n = o(1) \text{ s.t. } \gamma_n \sqrt{n} \to \infty, \sup_{h \in \mathcal{H}_I} \|\Pi_n h\|_s \leq B - \gamma_n; \text{(iv)} \text{ There is } \Pi_n t \in T_n \text{ for each } t \in T \text{ such that } \sup_{t \in T} \|t - \Pi_n t\|_E = o(n^{-1/2}). \]

**Assumption 3.6.** (i) In a \( \|\cdot\|_{L_2} \)-neighborhood of \( A_I, \) \( \rho(\alpha, t) \) is pathwise twice differentiable with respect to \( \alpha; \) (ii) Define the pathwise sieve Hessian matrix of \( \rho(\alpha, t) \) as a \( k_n \times k_n \) matrix \( \Psi_n(\alpha, t) \) whose \( ij \)-th element is \( \frac{d^2 \rho(\alpha, t)}{dh^2} [p_i, p_j]. \) The eigenvalues of \( \Psi_n(\alpha, t) \) are uniformly bounded on \( A_I \times T; \) (iii) The eigenvalues of the \( d_\theta \times d_\theta \) Hessian matrix \( \frac{\partial^2 \rho(\alpha, t)}{\partial \theta^2} \) are uniformly bounded on \( A_I \times T. \)

**Remark 3.2.** Given the compactness of \( A, \) Assumption 3.3 (ii) is usually satisfied. For example, it is satisfied by the linear moment functions in the partially linear IV as shown in Santos (2011). Assumption 3.3 (ii) is also satisfied if \( m(\cdot) \) is pointwise Lipschitz continuous in \( \alpha \) and continuous in \( y, \) and \( X \times Y \) is compact, which is shown in Appendix B. For quantile IV regressions where \( m \) is an indicator function, Assumption 3.3 (ii) is satisfied if we restrict the variability of the unknown function \( h(\cdot) \) such that \( \{ (x, y) : y \leq h(x) \} : h \in \mathcal{H} \} \) forms a Vapnik–Chervonenkis class of sets. This can be done, for example, by requiring the number of time for which the second order derivative of \( h(\cdot) \) switches signs (between \( \geq 0 \) and \( < 0 \)) to be bounded by a finite integer. Assumption 3.3 (iii) implies that the pseudo-metric \( \|\cdot\|_w \) is well defined in a neighborhood of the \( A_I. \) This condition trivially holds for partially linear IV as shown by formula (8), and also holds for partially linear quantile IV as long as the probability density \( f_{Y|X,Z}(\cdot) \) is well defined as shown by formula (9). Assumption 3.3 (iv) implies that the pseudo weak distance \( d_w(\alpha, A_I) \equiv \inf_{\alpha_0 \in A_I} \|\alpha - \alpha_0\|_w \) is Lipschitz continuous with respect to the population criterion. Similar assumptions are commonly imposed in the semiparametric literature, see, for example Chen and Pouzo (2011). In essence, Assumption 3.3 (iv) is analogous to the full rank and continuity of the Jacobian (near the true value) assumption made for parametric moment restrictions in the point-identified case. Assumption 3.3 (v) allows for the measure of ill-posedness \( \psi_n \) to grow to \( \infty \) but requires that the growth rate is slower than \( n^{1/4} \). Assumption 3.4 (i) requires that we use finite dimensional linear sieve space. Assumption 3.5 (iii) requires that any element of \( \mathcal{H}_I \) can be approximated well by some sieve space element that is in the interior of \( \mathcal{H}. \) It is worth noticing that Assumption 3.5 (iii) does not rule out scenarios in which some elements of \( \mathcal{H}_I \) are on the boundary – it only requires that each element of \( \mathcal{H}_I \) can be approximated well by some element of the sieve space that is in the interior of \( \mathcal{H}. \) Assumption 3.4 and Assumption 3.5 are satisfied by many commonly used sieves such as polynomials, P-splines and B-splines. Assumption 3.6 (i) is a stronger restriction than Assumption 3.3 (ii). It holds trivially for partially linear IV because the pathwise second derivative in such a model is always zero as shown in Remark 3.1 (ii). Assumption 3.6 (i) also holds for partially linear quantile IV as long as
$f_{Y|X,Z}(\cdot)$ is differentiable as shown by formula (10). Assumption 3.6 (ii) and (iii) together implies that $\frac{d^2 \rho(\alpha, t)}{d\alpha^2} |\Delta, \Delta| \leq D \cdot \|\Delta\|_{L2}^2$ for some $D > 0$ for all $(h, \Delta, t) \in \mathcal{H} \times \mathcal{H} \times T$. Assumption 3.6 (ii) and (iii) are needed to control the second order term in our asymptotic analysis. They hold trivially for partially linear IV, and hold for locally linear quantile IV as long as the first derivative of $f_{Y|X,Z}(\cdot)$ exists and is bounded according to (10). We note that for vector-valued $m(\cdot)$, Assumption 3.6 (ii) and (iii) should be viewed as restrictions imposed on each element of $\rho(\alpha, t)$. ■

Now we present the main asymptotic results.

**Theorem 3.1.** Let Assumptions 2.1-2.2, 3.1-3.6 hold. If $A_I \cap R \neq \emptyset$, we have

$$S_n \Rightarrow \inf_{\alpha \in A_I \cap R} \inf_{v \in V_\infty} \sup_{t \in T} \left\{ [G(\alpha, t) + v(t)]' W(\alpha, t) [G(\alpha, t) + v(t)] \right\}$$

where $G(\alpha, t)$ is a tight Gaussian process on $A \times T$.

**Theorem 3.2.** Let Assumptions 2.1-2.2, 3.1-3.6 hold. We have

$$n^{-1} S_n \xrightarrow{a.s.} \min_{\alpha \in A \cap R} \sup_{t \in T} \left\{ E \left[ m(Y, \theta, h(X)) w(t, Z) \right]' W(\alpha, t) E \left[ m(Y, \theta, h(X)) w(t, Z) \right] \right\}.$$

Theorem 3.1 shows that $S_n$ converges in distribution to a tight distribution under the null. It is worth noticing that the limiting distribution is nonpivotal in the sense that it depends on the underlying data generating process through $A_I$ and $V_\infty^\alpha$. Theorem 3.2 shows the asymptotic behavior of $S_n$ under a fixed alternative. Let

$$C \equiv \min_{\alpha \in A \cap R} \sup_{t \in T} \left\{ E \left[ m(Y, \theta, h(X)) w(t, Z) \right]' W(\alpha, t) E \left[ m(Y, \theta, h(X)) w(t, Z) \right] \right\}. \quad (11)$$

When $A_I \cap R = \emptyset$, for each $\alpha \in A \cap R$ there exists some $t \in T$ such that $E \left[ m(Y, \theta, h(X)) w(t, Z) \right] \neq 0$. Consequently, $C > 0$ due to the compactness of $A \cap R$. Therefore, Theorem 3.2 implies that $S_n$ goes to $\infty$ almost surely at the rate $O(n)$ when $H_0$ is false. Theorems 3.1 and 3.2 are the basis for our inference procedure.

### 3.4 Implementation and Critical Values

Denote by $D$ the asymptotic null distribution of $S_n$. According to Theorem 3.1,

$$D = \inf_{\alpha \in A_I \cap R} \inf_{v \in V_\infty} \sup_{t \in T} \left\{ [G(\alpha, t) + v(t)]' W(\alpha, t) [G(\alpha, t) + v(t)] \right\}. \quad (12)$$

For a targeted nominal size \( \rho \), using \( D \)'s \((1 - \rho)\)th quantile as the asymptotic critical value seems to be the most natural way of implementing Theorem 3.1 and 3.2. However, the \((1 - \rho)\)th quantile is infeasible because we are unfamiliar with the distribution of \( D \). We propose using critical values based on a bootstrap procedure that is specified below.

The Bootstrap Statistic

Recall the original test statistic is

\[
S_n = \min_{\alpha \in A \cap R} \left\{ \sup_{t \in T_n} \mathbb{E}_n \left[ J_n(\alpha, t)^\prime W_n(\alpha, t) [J_n(\alpha, t)] \right] \right\}
\]

where \( J_n(\alpha, t) \equiv \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) \cdot g(t, z_i) \).

For a bootstrap sample \( \{x_i^*, y_i^*, z_i^*\} \), let

\[
J_n^*(\alpha, t) \equiv \frac{1}{n} \sum_{i=1}^{n} \left[ m(y_i^*, \theta, h(x_i^*)) \cdot g(t, z_i^*) - \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) g(t, z_i) \right].
\tag{13}
\]

And let \( W_n^*(\alpha, t) \) be the weighting matrix calculated using the bootstrap sample. We define the bootstrap version of the test statistic as

\[
S_n^* = \min_{\alpha \in A \cap R} \left\{ \sup_{t \in T_n} \mathbb{E}_n \left[ J_n^*(\alpha, t)^\prime W_n^*(\alpha, t) [J_n^*(\alpha, t)] \right] + \lambda_n P_n(\alpha, t) \right\}
\tag{14}
\]

where \( P_n(\alpha, t) \equiv \left[ \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) g(t, z_i) \right]^2 \).

When setting \( \lambda_n = 0 \), our bootstrap procedure becomes the standard bootstrap procedure. However, we require that \( \lambda_n > 0 \) because the term \( J_n \) in \( S_n \) is only properly centered on \( A_I \) while its counterpart \( J_n^* \) in \( S_n^* \) is properly centered on the whole parameter space \( A \). By penalizing \( \alpha \not\in A_I \), \( P_n(\alpha, t) \) effectively ensures the minimization in \( S_n^* \) to be carried out asymptotically only on \( A_I \cap R \), just like \( S_n \). In essence, the need for \( P_n(\alpha, t) \) is due to the nonpivotality of the asymptotic null distribution. \( \lambda_n \in \mathbb{R} \) is a tuning parameter that determines how much weight to put on \( P_n(\alpha, t) \).

Assumption 3.7. \( \lambda_n \rightarrow \infty \) and \( \lambda_n = o(n / \log(\log(n))) \).

The following theorem summarizes the asymptotic properties for \( S_n^* \):

**Theorem 3.3.** Let Assumptions 2.1-2.2, 3.1-3.7 hold. If \( A_I \cap R \neq \emptyset \), then there exists a tight distribution \( D' \) such that \( S_n^* \Rightarrow D' \) and

\[
Pr(D > a) \leq Pr(D' > a)
\]
for any given $a \in \mathbb{R}$. In addition, if $A_I \cap R = \emptyset$

$$\lambda_n^{-1} S^*_n \xrightarrow{a.s.} \min_{\alpha \in A \cap R} \max_{t \in T} \left\{ E \left[ m(Y, \theta, h(X)) w(t, Z) \right] W(\alpha, t) E \left[ m(Y, \theta, h(X)) w(t, Z) \right] \right\}.$$  

Theorem 3.3. shows that, under the null, the bootstrap statistic $S^*_n$ converges in law to a tight distribution that stochastically dominates the asymptotic null distribution $D$ (in a non-strict sense). When the null hypothesis is false, both $S^*_n$ and $S_n$ go to infinity almost sure, but $S^*_n$ does so at a slower rate ($O(\lambda_n))$. Theorem 3.3 guarantees the consistency of the hypothesis test based on the bootstrap critical value described below.

**Calculating the Critical Value**

For a targeted nominal size $\rho$, the critical value is calculated as follows:

1. Generate $B$ bootstrap samples from the original sample $\{x_i, y_i, z_i\}_{i=1}^n$;
2. From each bootstrap sample, calculate $S^*_n,b$ according to formula (14);
3. Calculate the critical value $C_n^* (1 - \rho)$ as the $(1 - \rho)$th percentile of the set of values $\left\{ S^*_n,b \right\}_{b=1}^B$.

The null hypothesis $H_0 : A_I \neq \emptyset$ is rejected if $S_n > C_n^* (1 - \rho)$.

We conclude by showing the consistency of the hypothesis test based on the bootstrap critical value $C_n^* (1 - \rho)$:

**Theorem 3.4. (consistency of the hypothesis test)** Consider hypothesis test of $H_0 : A_I \neq \emptyset$ against $H_1 : A_I \cap R = \emptyset$. Let Assumptions 2.1-2.2, 3.1-3.7 hold. Then the test has the following properties:

$$\limsup_{n \to \infty} Pr \left( S_n > C_n^* (1 - \rho) \mid H_0 \right) \leq \rho;$$  

$$\lim_{n \to \infty} Pr \left( S_n \leq C_n^* (1 - \rho) \mid H_1 \right) = 0.$$  

Clearly, Theorem 3.4 shows that the probability of type I error is asymptotically no larger than the nominal size $\rho$, and the probability of type II error is asymptotically 0.

**Model Specification Test**

In the specification test for conditional moment models, we test whether the conditional moment restriction is satisfied by some element of the parameter space. It fits in to our
testing framework as a special case in which the restricted set \( R = \mathcal{A}. \)

Formally, in the specification test, we consider the following \( H_0 \) and \( H_1 \):

\[
H_0 : \mathcal{A}_I \neq \emptyset; \\
H_1 : \mathcal{A}_I = \emptyset,
\]

Consequently, the test statistic in this case becomes

\[
S_n = \min_{\alpha \in \mathcal{A}_n} \left\{ \sup_{t \in T_n} \left[ J_n (\alpha, t) \right]^{\prime} W_n (\alpha, t) \left[ J_n (\alpha, t) \right] \right\}.
\]

And the corresponding bootstrap statistic becomes

\[
S^*_n = \min_{\alpha \in \mathcal{A}_n} \left\{ \sup_{t \in T_n} \left[ J_n^* (\alpha, t) \right]^{\prime} W^*_n (\alpha, t) \left[ J_n^* (\alpha, t) \right] + \lambda_n P_n (\alpha, t) \right\}.
\]

Following the general procedure developed in this section, we calculate the bootstrap critical value \( C^*_n (1 - \rho) \) for a targeted size \( \rho \) and reject the null hypothesis that the model is correctly specified if \( S_n > C^*_n (1 - \rho) \).

It follows directly from Theorem 3.4 that the specification test above is consistent. This result is reported as the following corollary:

**Corollary 3.1. (consistency of the specification test)** Let Assumptions 2.1-2.2, 3.1-3.7 hold. The specification test is consistent against fixed alternative in the sense that

\[
\limsup_{n \to \infty} \Pr (\text{Type I Error}) \leq \rho; \\
\lim_{n \to \infty} \Pr (\text{Type II Error}) = 0.
\]

**4 Confidence Sets**

In the previous section, we developed a consistent procedure for hypothesis testing in conditional moment models of the form

\[
E [m (Y, \theta, h (X)) | Z] = 0
\]

where the parameter \( \alpha = (\theta, h) \) is possibly partially identified. As previously mentioned, confidence sets for \( \theta \) can be built by inverting the test for the restriction \( \theta = \bar{\theta} \). And

\footnote{Recall \( R = \{ \alpha \in \mathcal{A} : L (\alpha) = a \} \). When \( L (\cdot) \equiv 0 \) which is clearly a bounded linear operator and \( a = 0 \), \( R = \mathcal{A} \).}
confidence sets for \( h(\cdot) \) at a given point \( x_0 \) can be built by inverting the test for \( h(x_0) = \mu \).

In this section, we show exactly how. And we study the asymptotic properties of the resulting confidence sets.

Although we only consider in this paper the two types of confidence sets above, confidence sets for other features of the parameter can be constructed by similarly inverting the corresponding test. For example, we can build confidence sets for the whole \( h \) by inverting the test for \( h = \bar{h} \) where \( \bar{h} \) is any given element of \( \mathcal{H} \).

4.1 Confidence Sets for \( \theta \)

To construct confidence sets for \( \theta \), we start with considering testing the restriction \( \theta = \bar{\theta} \) for a given constant \( \bar{\theta} \in \Theta \). This test fits into our framework as a special case where \( L(\cdot) \) is a projection operator such that \( L((\theta, h)) = \theta \).

**Testing \( \theta = \bar{\theta} \)**

When testing the restriction \( \theta = \bar{\theta} \), \( \mathcal{A} \cap R = \{ (\bar{\theta}, h) : h \in \mathcal{H} \} \). In this case, the test statistic becomes

\[
S_n(\bar{\theta}) = \min_{h \in \mathcal{H}} \left\{ \sup_{t \in T_n} \left[ J_n((\bar{\theta}, h), t) \right]' W_n((\bar{\theta}, h), t) \left[ J_n((\bar{\theta}, h), t) \right] \right\}.
\]  

And the corresponding bootstrap statistic and critical value are

\[
S_n^*(\bar{\theta}) = \min_{h \in \mathcal{H}} \left\{ \sup_{t \in T_n} \left[ J_n^*(((\bar{\theta}, h), t)) \right]' W_n^*((\bar{\theta}, h), t) \left[ J_n^*((\bar{\theta}, h), t) \right] + \lambda_n P_n((\bar{\theta}, h), t) \right\}
\]  

and \( C_n^*(1 - \rho, \bar{\theta}) \), respectively.

Notice in (15) and (16) how the minimization are carried out: different values of \( h \) are searched on while \( \theta \) is fixed at \( \bar{\theta} \). Therefore both \( S_n(\bar{\theta}) \) and \( S_n^*(\bar{\theta}) \) can be viewed as a profiled minimum. We write \( S_n \), \( S_n^* \) and \( C_n^*(1 - \rho) \) as functions of \( \bar{\theta} \) to emphasize their dependence on the conjectured value \( \bar{\theta} \).

**The Confidence Sets**

We construct confidence sets for \( \theta \) by inverting the above test. For a targeted nominal level \( 1 - \rho \), the confidence set is constructed as

\[
CS_n(1 - \rho) = \{ \bar{\theta} \in \Theta : S_n(\bar{\theta}) \leq C_n^*(1 - \rho, \bar{\theta}) \}.
\]
In words, the confidence set consists of all values of $\bar{\theta}$ at which the corresponding test fails to reject the null.

Theorem 3.4 guarantees the resulting confidence sets are consistent in the following sense:

$$\inf_{\theta \in \Theta} \liminf_{n \to \infty} \Pr (\theta \in \text{CS}_n (1 - \rho)) \geq 1 - \rho.$$  \hfill (18)

We are able to show that another notion of consistency, which is stronger than (18), holds for the confidence sets without additional assumptions:

**Theorem 4.1.** Let Assumptions 2.1-2.2, 3.1-3.7 hold. Then:

$$\liminf_{n \to \infty} \inf_{\theta \in \Theta} \Pr (\theta \in \text{CS}_n (1 - \rho)) \geq 1 - \rho$$  \hfill (19)

and

$$\lim_{n \to \infty} \Pr (\theta \in \text{CS}_n (1 - \rho)) = 0 \text{ for any } \theta \notin \Theta.$$

Compared with (18), the infimum in (19) is carried out before the limit is taken. Therefore, Theorem 4.1 shows that the confidence sets have correct asymptotic coverage probability uniformly over the identification set $\Theta_I$ for a fixed data generating process (DGP). We note that our confidence sets do not necessarily achieve the uniformity over drifting sequences of data generating processes as in Andrews and Soares (2010). Theorem 4.1 also shows that the confidence sets are consistent against fixed alternatives.

### 4.2 Confidence Sets for $h(x_0)$

We construct confidence sets for $h(\cdot)$ at a given point, say $x_0$, by inverting the test for the restriction $h(x_0) = \mu$ where $\mu$ is constant in the range of $h(\cdot)$.

**Testing $h(x_0) = \mu$**

When testing the restriction $h(x_0) = \mu$, the test statistic becomes

$$S_n (\mu) = \min_{\alpha \in A_n : h(x_0) = \mu} \left\{ \sup_{t \in T_n} [J_n (\alpha, t)]' W_n (\alpha, t) [J_n (\alpha, t)] \right\}.$$  \hfill (20)

And the corresponding bootstrap statistic and critical value are

$$S_n^* (\mu) = \min_{\alpha \in A_n : h(x_0) = \mu} \left\{ \sup_{t \in T_n} [J_n^* (\alpha, t)]' W_n^* (\alpha, t) [J_n^* (\alpha, t)] + \lambda_n P_n (\alpha, t) \right\}$$  \hfill (21)

and $C_n^* (1 - \rho, \mu)$, respectively. We write $S_n, S_n^*$ and $C_n^* (1 - \rho)$ as functions of $\mu$ to emphasize their dependence on the conjectured value $\mu$.  

19
The Confidence Sets

For a targeted nominal level $1 - \rho$, the confidence set from inverting the above test is constructed as

$$\text{CS}_n (1 - \rho) = \{ l : S_n (l) \leq C^*_n (1 - \rho, \mu) \}.$$  

Define $\mathcal{H}_{I,x_0} \equiv \{ \mu : h (x_0) = \mu \text{ for some } h \in \mathcal{H}_I \}$ as the identification set for $h (x_0)$. The following theorem shows the consistency of the confidence sets:

**Theorem 4.2.** Let Assumptions 2.1-2.2, 3.1-3.7 hold. Then the confidence sets are consistent in the sense that

$$\lim \inf_{n \to \infty} \inf_{\mu \in \mathcal{H}_{I,x_0}} \Pr (\mu \in \text{CS}_n (1 - \rho)) \geq 1 - \rho$$  

(22)

and $\lim_{n \to \infty} \Pr (\mu \in \text{CS}_n (1 - \rho)) = 0$ for any $\mu \notin \mathcal{H}_{I,x_0}$. 

Similar to the result in Theorem 4.1, this notion of CS consistency implies that the correctness of asymptotic coverage probability holds uniformly over the identification set $\mathcal{H}_{I,x_0}$.

5 Monte Carlo Simulations

In this section we study the finite sample performance of the proposed inference procedures using Monte Carlo simulations. We consider the following partially linear quantile IV:

$$E \left[ 1 \{ Y \leq X_1 \theta + h (X_2) \} | Z \right] = 0.75.$$  

(23)

We simulate the data from the following DGP:

$$Y = \sin (\pi X_2) + U,$$

$$U = \frac{1}{10} \left[ E (X_2^2 | Z) - X_2^2 \right] - 0.75 + \varepsilon.$$  

(24)

where

$Z \sim \text{Uniform } [0.5, 1],$

$X_1 \sim \text{Uniform } [0, 1] + Z,$

$X_2 \sim Z \cdot \text{Uniform } [-1, 1],$

$\varepsilon \sim \text{Uniform } [0, 1],$  

(25)
Identification

For any \( \lambda \in \mathbb{R} \), \((0, \sin(\pi x) + \lambda x)\) satisfies the conditional moment equality (23) because

\[
E \left[ 1 \{ Y \leq \sin(\pi X) + \lambda X \} \mid Z \right] \\
= Pr \left( U - \lambda X \leq 0 \mid Z \right) \\
= Pr \left( \varepsilon \leq 0.75 + \lambda X - \frac{1}{10} \left[ E \left( X^2 \mid Z \right) - X^2 \right] \mid Z \right) \\
= E \left[ \Pr \left( \varepsilon \leq 0.75 + \lambda X - \frac{1}{10} \left[ E \left( X^2 \mid Z \right) - X^2 \right] \mid X, Z \right) \mid Z \right] \\
= E \left[ 0.75 + \lambda X - \frac{1}{10} \left[ E \left( X^2 \mid Z \right) - X^2 \right] \mid Z \right] \\
= 0.75 + \lambda E \left( X \mid Z \right) \\
= 0.75.
\] (26)

Therefore, under the DGP specified in ((50)) and (25), the identification set \( A_I \) contains \((0, \sin(\pi x))\) and \((0, \sin(\pi x) + \lambda x)\) for \( \lambda > 0 \), and the parameter \((\theta_0, h_0)\) in model (23) is partially identified.

Equation (26) also shows clearly that zero is an element of \( \theta \)'s identification set \( \Theta_I \). A complete characterization of \( \Theta_I \) has not been established analytically. But through computation, we find that specific none-zero values, such as 0.3, 0.5, and 1, do not belong to \( \Theta_I \).

1 $- \rho$ Confidence Sets for $\theta$

For the instrument functions, we select the family of indicator functions

\[ g(t, z) = 1 \{ t \leq z \}, \ t \in T = Z \left( = [0.5, 1] \right), \]

and set \( T_n \) to be the grid \( \{0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1\} \). The sieve space we use is a B-Splines of order 3 with knot \( \{-1, -1, -1, 0, 1, 1, 1\} \), which implies \( k_n = 4 \). This sieve is able to provide a good approximation of \( \sin(\pi x) + \lambda x \). The critical values are based on 100 bootstrap evaluations. In our study, we performed 100 Monte Carlo repetitions of 500 observations \( (n = 500) \).
For different values of $\theta$ (0, 0.3, 0.5, and 1), Table 1 reports the simulated probability of $\theta$ being covered by the CS’s as a function of the targeted nominal level $1 - \rho$ and the tuning parameter $\lambda_n$. We have the following observations: (i) When moving the value of $\theta$ away from the identification set $\Theta_I$, from zero (a value in $\Theta_I$) to values such as 0.3, 0.5, or 1 (values outside $\Theta_I$), we see a significant decrease in the coverage probability for all three choices of $\lambda_n$. This is an indicator that the hypothesis test, by inverting which our CS’s are built, holds good power against fixed alternatives in finite samples; (ii) The coverage probability is somewhat sensitive to the choice of $\lambda_n$. Overall, choosing $\lambda_n = 63.0$, which equals approximately $n^{2/3}$, gives good size control; (iii) The choice $\lambda_n = 0$, which corresponds to the standard bootstrap procedure, is not warranted by our theory. And, as shown in Table 1, setting $\lambda_n = 0$ leads to severe size distortion.

\section{Empirical Illustration}

In this section, I apply the proposed inference methods to study the quantile IV Engel curves for gasoline in Brazil using the Pesquisa de Orçamentos Familiares 2002-2003 (POF) data. The same data set was used to study the mean IV Engel curves for gasoline by Santos (2011). A nonparametric specification of the quantile IV Engel curves takes the following form:

$$E \left[ 1 \{ Y_i \leq h (X_i) \} | Z_i \right] = \tau, \quad 0 < \tau < 1$$

(27)

where $Y_i$ is the share of total nondurable expenditures of household $i$ spent on gasoline, $X_i$ is the log of total nondurable expenditures of household $i$, and instrumental variable $Z_i$ is the log of total household income as in Blundell, Chen, and Kristensen (2007), Chen and Pouzo (2009), and Santos (2011). The data set contains $n = 4994$ observations. For the instrument function $g(t, z)$, I pick $g(t, z) = \phi \left( \frac{t_1 - z}{t_2} \right)$ with grid $\{-0.8, -0.4, 0, 0.4, 0.8\} \times$
for \((t_1, t_2)\) as in Santos (2011). I set \(\lambda_n = 292.168\) which corresponds to \(n^{2/3}\).

I consider the 50th quantile, and consequently set \(\tau = 0.5\). A log linear specification (i.e., linear with respect to the log of total nondurable expenditure \(X\)) is commonly used to parametrize Engel curves in the literature, including papers such as Leser (1963), Prais and Houthakker (1955), and Working (1943). Quantile Engel curves under the log linear specification take the form \(E\{Y_i \leq \alpha + X_i\beta\} | Z_i\} = \tau\). The log linear specification can be tested through the hypothesis test:

\[
H_0 : \mathcal{H}_I \cap R_L \neq \emptyset; \\
H_1 : \mathcal{H}_I \cap R_L = \emptyset,
\]

where \(\mathcal{H}_I\) denotes the identification set of \(h\) in Model (27), and \(R_L\) is defined as

\[
R_L = \{h \in \mathcal{H} : h(x) = \alpha + x\beta \text{ for some } (\alpha, \beta) \in \mathbb{R}^2\}.
\]

The calculated test statistic yields a p-value of 0.357. Therefore, the log linear specification is not rejected at a significance level of either 0.1 or 0.05. Interestingly, a linear specification (i.e., linear with respect to the total nondurable expenditure, \(e^X\)) is rejected at any significance level (p-value = 0).

After failing to reject the log linear specification, I calculate the 95% confidence region for the intercept \(\alpha\) and the slope \(\beta\) by inverting test of \((\alpha, \beta) = (\tilde{\alpha}, \tilde{\beta})\) as discussed in Section 4. The resulting confidence region for \((\alpha, \beta)^9\) is depicted in Figure 1 (a). Projecting this confidence region onto the \(\alpha\) axis and the \(\beta\) axis yields a confidence region of [0.0660, 0.2461] for \(\alpha\) and a confidence region of [−0.0153, 0.0028] for \(\beta\), respectively. Figure 2 (a) shows the pointwise 95% confidence bounds for the 50th quantile Engel curve under the log linear specification, obtained by plotting the maximum and minimum of \(\alpha + x\beta\) over the confidence region for \((\alpha, \beta)\) for each \(x \in [7, 13]\).

---

9I parametrize the linear functions on the range of \(X\), \([7.468, 12.441]\), by a B-Spline of order 2 with knots \([7.468, 7.468, 12.441, 12.441]\). I obtain the the joint confidence region for the two coefficients of the B-Spline through a grid search with grid \([0.012 : 0.005 : 0.212] \times [-0.024 : 0.005 : 0.176]\), after confirming that the confidence region is in the interior of \([0.012, 0.212] \times [-0.024, 0.176]\) through an initial grid search with grid \([-3 : 0.1 : 3] \times [-3 : 0.1 : 3]\). Then I obtain the confidence regions for \((\alpha, \beta)\) as depicted in Figure 1.1 (a) through linear transformation of the confidence region for the B-Spline coefficients.
Figure 1 95% Confidence Region for $(\alpha, \beta)$
I also examine the 25th quantile ($\tau = 0.25$) and the 75th quantile ($\tau = 0.75$). For $\tau = 0.25$, a log linear specification for the corresponding Engel curve yields a p-value of 0.2467, and therefore is not rejected by the proposed test. Figure 1 (b) shows the 95% confidence region for $(\alpha, \beta)$. The projected confidence regions for $\alpha$ and $\beta$ are $[-0.0111, 0.1090]$ and $[-0.0052, 0.0068]$, respectively. Yet, a linear specification is rejected at any significance level that is larger than 0.001 (p-value = 0.001). Figure 2 (b) shows the pointwise 95% confidence bounds for the 25th quantile Engel curve under the log linear specification.

For $\tau = 0.75$, a log linear specification for the corresponding Engel curve yields a p-value of 0.26. Figure 1 (c) shows the 95% confidence region for $(\alpha, \beta)$. The projected confidence regions for $\alpha$ and $\beta$ are $[0.2601, 0.5928]$ and $[-0.0444, -0.0113]$, respectively. Yet, again, a linear specification is rejected at any significance level (p-value = 0). Figure 2 (c) shows the pointwise 95% confidence bounds for the 75th quantile Engel curve under the log linear specification.
To check the robustness of the empirical results, I conduct inference on the median Engel curve ($\tau = 0.5$) with either the tuning parameter $\lambda_n$ or the instrument function, or both, being altered to $\lambda_n = 70.668$, which corresponds to $n^{1/2}$, and $g_{\text{alt}}(t, z) = 1 \{t \leq z\}$ with grid $\{3.8, 4.3, 3.8, ..., 12.8\}$. A summary of results from different combination of choices of $\lambda_n$ and instrument function is reported as Table 2 below.

![Figure 3 Pointwise 95% Confidence Bounds Under Log Quadratic Specification](image)

<table>
<thead>
<tr>
<th>$\lambda_n$</th>
<th>Log Linear</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_n = n^{2/3}, g$</td>
<td>0.357</td>
<td>0.357</td>
</tr>
<tr>
<td>$\lambda_n = n^{1/2}, g$</td>
<td>0.293</td>
<td>0.293</td>
</tr>
<tr>
<td>$\lambda_n = n^{2/3}, g_{\text{alt}}$</td>
<td>0.593</td>
<td>0.593</td>
</tr>
<tr>
<td>$\lambda_n = n^{1/2}, g_{\text{alt}}$</td>
<td>0.250</td>
<td>0.250</td>
</tr>
</tbody>
</table>

**Table 2 P-Value and 95% Confidence Regions from Different $\lambda_n$ and Instrument Function**

The second column of Table 2 lists the p-value for the log linear specification, associated with different choices of $\lambda_n$ and instrument function. Although the p-values differ, the log linear specification is not rejected regardless of the choices of $\lambda_n$ and instrument function. To the contrary, a linear specification yields a p-value of zero, and therefore is rejected, under all four different choices of $\lambda_n$ and instrument function, as shown in the last column.

The third and fourth columns of Table 2 show the corresponding confidence regions for the intercept and slope, respectively. To get a better sense of scale and magnitude of these confidence regions, I report in the fifth column the confidences region for the median share of gasoline expenditures at the sample average of the log total expenditures $\bar{X} = 9.9151$. As shown in the fifth column, all four different choices of $\lambda_n$ and instrument function yield similar confidence regions for the median share at $\bar{X}$. 

26
For comparison purposes, I also quote in the sixth column the confidence regions for the mean share of gasoline expenditures at $\bar{X}$ from Table 2 of Santos (2011). Comparing the sixth column with the fifth column shows a noticeable shift to the left when moving from the confidence regions for the mean to the confidence regions for the median. This observation suggests that the distribution of share of gasoline expenditures at the average total expenditure level in Brazil is likely to have a mean larger than the median, and consequently is likely to be skewed to the right.

The log quadratic specification yields slightly wider confidence regions for the median share at $\bar{X}$ as shown in Table 3.

<table>
<thead>
<tr>
<th>$\lambda_0 = n^{2/3}, g$</th>
<th>$\lambda_0 = n^{1/2}, g$</th>
<th>$\lambda_0 = n^{2/3}, g_{ult}$</th>
<th>$\lambda_0 = n^{1/2}, g_{ult}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0.0917, 0.0994]</td>
<td>[0.0917, 0.0994]</td>
<td>[0.0941, 0.0971]</td>
<td>[0.0941, 0.0971]</td>
</tr>
<tr>
<td>[0.0900, 0.1054]</td>
<td>[0.0900, 0.1054]</td>
<td>[0.0804, 0.1154]</td>
<td>[0.0804, 0.1154]</td>
</tr>
</tbody>
</table>

Table 3 Confidence Regions for $h(\bar{X})$ under Two Different Specifications

7 Conclusions

In this paper, we combine ideas and techniques from semiparametric modeling and partial identification analysis to develop general methods for inference in the conditional moment model $E[ m(Y, \theta, h_0(X)) | Z] = 0$. Without assuming point-identification, we propose a consistent procedure for testing the hypothesis that a given restriction on the parameter $\alpha = (\theta, h)$ is satisfied by at least one element of the identification set. Based on the proposed testing procedure, we show how to consistently test the conditional moment restriction specification, construct confidence sets for $\theta$ and for the unknown function $h(\cdot)$ at a given point. Our methods are robust to partial identification. They extend the inference methods developed in Santos (2011) for nonparametric IV to more general conditional moment models with nonlinear and even nonsmooth moment functions.
Appendix A. Mathematical Proofs

In Appendix A, we prove Theorems 3.1-3.3, 4.1-4.2. We start with listing the lemmas that we will use in the proofs. Proofs of these lemmas are given in Appendix B.

The following lemmas are used in the proofs of the theorems:

Lemma 1. Let Assumption 3.1(ii) and 3.3(ii) hold. Then,

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (m(y_i, \theta, h(x_i)) \cdot g(t, z_i) - E[m(Y, \theta, h(X)) \cdot g(t, Z)]) \xrightarrow{\mathbb{L}} G(\alpha, t)
\]

where \(G(\alpha, t)\) is a tight Gaussian process on \(A \times T\).

Lemma 2. (convergence rate in \(\|\cdot\|_w\)) Let Assumption 2.1, 2.2, 3.1, 3.2, and 3.3(i)-(iv) hold. For any \(\hat{\alpha}_n \in \arg \min_{\alpha \in A_n \cap R} \left\{ \sup_{t \in T} [J_n(\alpha, t)]' W_n(\alpha, t) [J_n(\alpha, t)] \right\}\) as in (7)

\[
d_w(\hat{\alpha}_n, A_I \cap R) = O_p \left( \max \left\{ \delta_{w,n}, n^{-1/2} \right\} \right).
\]

Lemma 3. (convergence rate in \(\|\cdot\|_{L_2}\)) Let Assumption 2.1, 2.2, 3.1-3.3 hold. Then

\[
d_{L_2}(\hat{\alpha}_n, A_I \cap R) = O_p \left( \delta_{s,n} + \psi_n \cdot \max \left\{ \delta_{w,n}, n^{-1/2} \right\} \right).
\]

Lemma 4. Let \(A, B\) be sets of two Hilbert spaces, respectively, and \(F_n^b \subseteq F_{n+1}^b \subseteq L^\infty(A)\) with \(F_n^b\) being the closure of \(\cup F_n^b\) under \(\|\cdot\|_\infty\) for each \(b \in B\). If \(g \in L^\infty(A \times B)\), and there exists a sequence \(\{g_n\} \in L^\infty(A \times B)\) s.t. \(\|g_n - g\|_\infty = o(1)\), then:

\[
\inf_{b \in B} \inf_{f \in F_n^b} \sup_{a \in A} |g_n(a, b) - f(a)| \longrightarrow \inf_{b \in B} \inf_{f \in F_n^\infty} \sup_{a \in A} |g(a, b) - f(a)|.
\]

Lemma 5. Let \(A_n, B\) be sets with norms \(\|\cdot\|_a, \|\cdot\|_b\) and let \(G_n : A_n \times B \to \mathbb{R}\) and \(F_n : A_n \times B \to \mathbb{R}\) be random functions. Assume: (1) \(G_n\) and \(F_n\) are continuous on \(A_n \times B\) almost sure; (2) \(\sup_{a \in A_n} \sup_{b \in B} |G_n(a, b) - F_n(a, b)|^2 = o_p(1)\); and (3) \(\inf_{a \in A_n} \sup_{b \in B} F_n^2(a, b) = O_p(1)\). It then follows:

\[
\inf_{a \in A_n} \sup_{b \in B} \sup_{G_n^2(a, b)} = \inf_{a \in A_n} \sup_{b \in B} F_n^2(a, b) + o_p(1).
\]

28
Lemma 6. For any $c > 0$, let $H_{nR,c} \equiv \{(r_\theta, r_h) \in H_{nR} : \|r_h p^{k_n}\|_s^2 \leq c, \|r_h\|_E^2 \leq c\}$. Define the following family of functions:

$$V_{k_n, c}^{\alpha_0} = \left\{ v : T \to \mathbb{R}^{d_n} \text{ s.t. } v(t) = \frac{\partial \rho (\theta_0, h_0, t)}{\partial \theta'} \cdot r_\theta + \frac{d \rho (\theta_0, h_0, t)}{dh} \left[p^{k_n}\right]' r_h, (r_\theta, r_h) \in H_{nR,c} \right\}.$$

For any real number sequence $c_n \to \infty$ and $\alpha_0 \in A_I$, we have

$$\cup V_{k_n, c_n}^{\alpha_0} = \cup V_{k_n}^{\alpha_0}. \quad (28)$$

(Proofs of Lemmas 1-6 above are given in Appendix B.)

Proof of Theorem 3.1

Without loss of generality, we assume $m(\cdot)$ and $h(\cdot)$ to be real-valued in the proof for simplicity of notation.

Define $u_i (\alpha, t) = m(y_i, \theta, h(x_i)) \cdot g(t, z_i)$.

We first consider the case where $W_n$ is chosen to be the identity matrix. Lemma 3, Assumption 3.3$(v)$, and Assumption 3.5$(ii)$ imply that for $\hat{\alpha}_n \in \arg \min_{\alpha \in A_n \cap R} \sup_{t \in T} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i (\alpha, t) \right]^2 = o_p \left( n^{-1/4} \right). \quad (29)$

Let $\delta_n = o_p \left( n^{-1/4} \right)$ such that $n^{-1/2} = o(\delta_n)$. Define the neighborhoods

$$B^{\delta_n} (\alpha_0) = \{ \alpha \in A_n : \|\alpha - \alpha_0\|_{L^2} \leq \delta_n \}$$

for all $\alpha_0 \in A_I$. Then

$$\min_{\alpha \in A_n \cap R} \sup_{t \in T} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i (\alpha, t) \right]^2 = \inf_{\alpha_0 \in A_I \cap R} \min_{\alpha \in B^{\delta_n} (\alpha_0)} \sup_{t \in T} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i (\alpha, t) \right]^2 + o_p (1). \quad (29)$$

By Lemma 1, the empirical process

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ m(y_i, \theta, h(x_i)) \cdot g(t, z_i) - E [m(Y, \theta, h(X)) \cdot g(t, Z)] \right], (\alpha, t) \in A \times T$$

is asymptotically $\| \cdot \|_{L^2}$-equicontinuous in probability w.r.t. $\alpha$, uniformly on $T$. So,
since \( \delta_n \downarrow 0 \),

\[
\sup_{\|\alpha_1 - \alpha_2\|_{L^2} \leq \delta_n} \sup_T \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ [m(y_i, \theta_1, h_1(x_i)) - m(y_i, \theta_2, h_2(x_i))] g(t, z_i) - E[m(Y, \alpha_1) - m(Y, \alpha_2)] g(t, Z) \right\} = o_p(1).
\]

(30)

Therefore,

\[
\min_{\alpha \in A_n \cap R} \sup_{t \in T_n} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(\alpha, t) \right]^2 = \inf_{\alpha_0 \in A \cap R} \min_{\alpha \in B_n(\alpha_0)} \sup_{t \in T} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(\alpha, t) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [m(y_i, \theta, h(x_i)) - m(y_i, \theta_0, h_0(x_i))] \cdot g(t, z_i) \right\}^2 + o_p(1)
\]

where the last equality in (31) follows from (30) and Lemma 5.

(31)

Notice that for any \( \alpha_0 \in A \), \( \alpha \in B_n(\alpha_0) \), we have \( \|\alpha - \alpha_0\|_{L^2} = o_p(n^{-1/4}) \). Consequently,

\[
\sqrt{n} E \left[ [m(Y, \alpha) - m(Y, \alpha_0)] \cdot g(t, Z) \right] = \sqrt{n} \frac{d E [m(Y, \alpha_0) \cdot g(t, Z)]}{d \alpha} [\alpha - \alpha_0] + \sqrt{n} \frac{d^2 E [m(Y, \tilde{\alpha}) \cdot g(t, Z)]}{d \alpha^2} [\alpha - \alpha_0, \alpha - \alpha_0] \leq \sqrt{n} \frac{d E [m(Y, \alpha_0) \cdot g(t, Z)]}{d \alpha} [\alpha - \alpha_0] + \sqrt{n} \cdot C \cdot \|\alpha - \alpha_0\|_{L^2}^2 \quad \text{for some } C > 0
\]

where \( \tilde{\alpha} \equiv \alpha_0 + s(\alpha - \alpha_0) \) for some \( s \in (0, 1) \), and the inequality in (32) follows from Assumption 3.6 (ii) and (iii).

Continuing with (31) and reapplying Lemma 5 yield the first equality in (33) below.

(32)
\[
\min_{\alpha \in \mathcal{A}_n \cap R_t \in T_n} \sup_{t} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(\alpha, t) \right]^2 \\
= \inf_{\alpha_0 \in \mathcal{A}_I \cap R_{\alpha} \in B_{\delta_n}(\alpha_0) \in T} \min_{\alpha_0} \sup_{t} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(\alpha_0, t) + \sqrt{n} dE \left[ m(Y, \alpha_0) \cdot g(t, Z) \right] \frac{dE}{d\alpha} \left[ \alpha - \alpha_0 \right] \right]^2 + o_p(1) \\
= \inf_{\alpha_0 \in \mathcal{A}_I \cap R_{\alpha} \in B_{\delta_n}(\alpha_0) \in T} \min_{\alpha_0} \sup_{t} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(\alpha_0, t) + \sqrt{n} dE \left[ m(Y, \alpha_0) \cdot g(t, Z) \right] \frac{dE}{d\alpha} \left[ \alpha - \Pi_n \alpha_0 \right] \right]^2 + o_p(1).
\]

where \(\Pi_n \alpha_0 \equiv (\theta_0, \Pi_n h_0)\) and the second equality is because

\[
\sqrt{n} dE \left[ m(Y, \alpha_0) \cdot g(t, Z) \right] \frac{dE}{d\alpha} \left[ \alpha - \alpha_0 \right] = \sqrt{n} dE \left[ m(Y, \alpha_0) \cdot g(t, Z) \right] \frac{dE}{d\alpha} \left[ \alpha - \Pi_n \alpha_0 \right] + o_p(1) \\
= \sqrt{n} dE \left[ m(Y, \alpha_0) \cdot g(t, Z) \right] \frac{dE}{d\alpha} \left[ \alpha - \Pi_n \alpha_0 \right] + o(1)
\]

where the second equality in (34) is because \(\sqrt{n} dE \left[ m(Y, h_0) \cdot w(t, Z) \right] \frac{dE}{d\alpha} \left[ \Pi_n h_0 - h_0 \right] \leq \sqrt{n} \| \Pi_n h_0 - h_0 \|_w = o(1)\) uniformly on \(\mathcal{H}_I\) according to Assumption 3.5 (ii).

Next, we focus on the local parameters of \(\mathcal{H}_I\) of the form

\[
h = \Pi_n h_0 + \frac{p_{\alpha}^k}{\sqrt{n}}, \quad h_0 \in \mathcal{H}_I, r \in \mathbb{R}^{k_n}.
\]

Note that for any \(h \in B_{\delta_n}(h_0)\), there exists \(r \in \mathbb{R}^{k_n}\), s.t. \(h = \Pi_n h_0 + \frac{p_{\alpha}^k}{\sqrt{n}}, \quad h_0 \in \mathcal{H}_I\).

Therefore

\[
B_{\delta_n}(h_0) \subseteq \left\{ h \in \mathcal{H} : h = \Pi_n h_0 + \frac{p_{\alpha}^k}{\sqrt{n}}, r \in \mathbb{R}^{k_n} \right\}, \quad \forall h_0 \in \mathcal{H}_I.
\]

Now we proceed to show that there exists a real number sequence \(c_n \to \infty\) s.t.

\[
h = \Pi_n h_0 + \frac{p_{\alpha}^k}{\sqrt{n}} \in B_{\delta_n}(h_0),
\]

for any \(r \in \mathbb{R}^{k_n}\) that satisfies \(\|p_{\alpha}^k\|_E^2 \leq c_n^2, \quad |r|_E \leq c_n\).
To show (36), first note that

\[
\sup_{h_0 \in \mathcal{H}_I} \|h_0 - \left( \Pi_n h_0 + \frac{p^{k_n} r}{\sqrt{n}} \right) \|_{L^2} \leq \sup_{h_0 \in \mathcal{H}_I} \|h_0 - \Pi_n h_0\|_{L^2} + \left\| \frac{p^{k_n} r}{\sqrt{n}} \right\|_{L^2} \\
\leq \frac{\delta_n}{2} + \sqrt{J} \frac{c_n}{\sqrt{n}}.
\]

(37)

The first inequality in (37) follows from the triangle inequality. And the second inequality follows from Assumption 3.4(ii), Assumption 3.5(ii), and the fact that \(\|r\| \leq c_n\).

Also note that

\[
\sup_{h_0 \in \mathcal{H}_I} \|\Pi_n h_0 + \frac{p^{k_n} r}{\sqrt{n}}\|_s \leq \sup_{h_0 \in \mathcal{H}_I} \|\Pi_n h_0\|_s + \left\| \frac{p^{k_n} r}{\sqrt{n}} \right\|_s \\
\leq B - \gamma_n + \frac{c_n}{\sqrt{n}}.
\]

(38)

The second inequality in (38) follows from Assumption 3.5(iii) and the fact that \(\|p^{k_n} r\|_E^2 \leq c_n^2\).

According to (37) and (38), as long as \(\frac{c_n}{\sqrt{n}} = o(\delta_n)\) and \(\frac{c_n}{\sqrt{n}} = o(\gamma_n)\), for \(n\) large enough, we have:

\[
\sup_{h_0 \in \mathcal{H}_I} \|h_0 - \left( \Pi_n h_0 + \frac{p^{k_n} r}{\sqrt{n}} \right) \|_{L^2} \leq \delta_n
\]

(39)

and

\[
\sup_{h_0 \in \mathcal{H}_I} \|\Pi_n h_0 + \frac{p^{k_n} r}{\sqrt{n}}\|_s \leq B.
\]

(40)

Such \(c_n \searrow \infty\) exists because \(n^{-1/2} = o(\delta_n)\) and \(n^{-1/2} = o(\gamma_n)\) by Assumption 3.5(iii).

Therefore, we can pick \(c_n \searrow \infty\) s.t. \(\frac{c_n}{\sqrt{n}} = o(\delta_n)\) and \(\frac{c_n}{\sqrt{n}} = o(\gamma_n)\) (e.g. \(c_n = \ln n\) might work). As long as \(\|p^{k_n} r\|_E^2 \leq c_n^2\), \(\|r\|_E \leq c_n\), for \(h = \Pi_n h_0 + \frac{p^{k_n} r}{\sqrt{n}}\), (39) guarantees that

\[
\|h_0 - h\|_{L^2} \leq \delta_n.
\]

(41)

And (40) guarantees that

\[
\|h\|_s \leq B.
\]

(42)

(41) and (42) imply that \(h = \Pi_n h_0 + \frac{p^{k_n} r}{\sqrt{n}} \in B^{\delta_n}(h_0)\). So we prove (36). It follows that

\[
\left\{ h \in \mathcal{H} : h = \Pi_n h_0 + \frac{p^{k_n} r}{\sqrt{n}}, r \in \mathbb{R}^{k_n}, r' \Lambda_n r \leq c_n^2, \|r\| \leq c_n \right\} \subseteq B^{\delta_n}(h_0)
\]

(43)
for all \( h_0 \in \mathcal{H}_I \).

The first inequality in (44) is implied by (35). The equality repeats (33). And the final inequality is implied by (43)

\[
\inf_{h_0 \in \mathcal{H}_I, r \in \mathbb{R}^{kn}} \max_{t \in T} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(t, h_0) + \frac{d E [m(Y, h_0) \cdot w(t, Z)]}{dh} [p_{kn}]^r \right\}^2
\leq \min_{h_0 \in \mathcal{H}_I} \min_{h \in B_{\delta n}(h_0)} \max_{t \in T} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(t, h_0) + \frac{d E [m(Y, h_0) \cdot w(t, Z)]}{dh} |h - h_0| \right\}^2
\]

\[
= \min_{h_0 \in \mathcal{H}_I} \max_{t \in T} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(t, h) \right]^2 + o_p(1)
\]

\[
\leq \inf_{h_0 \in \mathcal{H}_I, r \in \mathbb{R}^{kn} \text{ s.t.} \ r' \Lambda_n r \leq c_n^2, ||r|| \leq c_n} \max_{t \in T} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(t, h_0) + \frac{d E [m(Y, h_0) \cdot w(t, Z)]}{dh} [p_{kn}]^r \right\}^2
\]

Define the families of functions:

\[
V_{\alpha,0}^{\alpha_0} = \left\{ v : T \to \mathbb{R} \ s.t. \ v(t) = \frac{\partial \gamma}{\partial \alpha_0} \cdot r_0 + \frac{d \gamma}{dh} [p_{kn}]^r \right\},
\]

for all \( \alpha \in \mathcal{H}_0 \) and \( c > 0 \). Consequently,

\[
\inf_{\alpha_0 \in \mathcal{A}_I \cap \mathcal{R}_I} \inf_{v \in V_{\alpha_0}^{\alpha_0}} \sup_{t \in T} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(\alpha_0, t) + v \right\}^2 \leq \min_{\alpha \in \mathcal{A}_I \cap \mathcal{R}_I} \sup_{t \in T} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(\alpha, t) \right]^2 + o_p(1)
\]

\[
\leq \inf_{\alpha_0 \in \mathcal{A}_I \cap \mathcal{R}_I} \inf_{v \in V_{\alpha_0}^{\alpha_0}} \sup_{t \in T} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(\alpha_0, t) + v \right\}^2
\]

Lemma 1 implies that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(\alpha, t) \overset{L}{\rightarrow} G(\alpha, t)
\]

on \( \mathcal{A}_I \times T \).

(46) in turn implies that

\[
|| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(\alpha, t) - G(\alpha, t) ||_\infty = o_p(1).
\]

(48) is implied by Lemma 4 and Theorem 1.11.1 (Extended continuous mapping) in
van der Vaart and Wellner (1996) whose conditions are satisfied by (47).

\[
\inf_{\alpha_0 \in A \cap R \cap V_{k_n \cap R}^{\alpha_0}} \sup_{t \in T} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i (\alpha_0, t) + v \right\}^2 \xrightarrow{L} \inf_{\alpha \in A \cap R \cap V_{\infty}^{\alpha}} \sup_{t \in T} \{ G(\alpha, t) + v(t) \}^2. \tag{48}
\]

(49) is implied by Lemma 4 and Theorem 1.11.1 in van der Vaart and Wellner (1996) whose conditions are satisfied by (47) and Lemma 6.

\[
\inf_{\alpha_0 \in A \cap R \cap V_{k_n \cap R}^{\alpha_0}} \sup_{t \in T} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i (\alpha_0, t) + v \right\}^2 \xrightarrow{L} \inf_{\alpha \in A \cap R \cap V_{\infty}^{\alpha}} \sup_{t \in T} \{ G(\alpha, t) + v(t) \}^2. \tag{49}
\]

(45), (48) and (49) imply that

\[
\min_{\alpha \in A \cap R \cap T_{n}} \sup_{t \in T} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i (\alpha, t) \right\}^2 \xrightarrow{L} \inf_{\alpha \in A \cap R \cap V_{\infty}^{\alpha}} \sup_{t \in T} \{ G(\alpha, t) + v(t) \}^2. \tag{50}
\]

For other choices of weighting matrix \( W_n (\alpha, t) \) that satisfy Assumption 3.2, it can be shown via similar steps that

\[
\min_{\alpha \in A \cap R \cap T_{n}} \sup_{t \in T} \left\{ W^{1/2}_n (\alpha, t) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i (\alpha, t) \right\}^2 \xrightarrow{L} \inf_{\alpha \in A \cap R \cap V_{\infty}^{\alpha}} \max_{t \in T} \left\{ W^{1/2}_n (\alpha, t) [G(\alpha, t) + v(t)] \right\}^2. \tag{51}
\]

\[\blacksquare\]

**Proof of Theorem 3.2**

Let \( \delta_n = o_p (n^{-1/4}) \) as in the proof of Theorem 1. According to Theorem 2.4.1 in van der Vaart and Wellner (1996),

\[
\frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) g(t, z_i) \xrightarrow{a.s.} E [m(Y, \theta, h(X)) g(t, Z)] \tag{52}
\]

as a process on \( A \times T \). Applying the theorem of maximum and the continuous mapping theorem yields

\[
\min_{\alpha \in A} \sup_{t \in T} \left\{ \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) g(t, z_i) \right\}^2 \xrightarrow{a.s.} \min_{\alpha \in A} \sup_{t \in T} [E [m(Y, \theta, h(X)) g(t, Z)]]^2. \tag{53}
\]
The first inequality in (55) follows from
\[
\sup_{\alpha \in A, t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) - E [u (\alpha, t)] \right| \overset{a.s.}{\to} 0. \tag{54}
\]

Let \( \hat{\alpha}_n \in \arg \minsup_{\alpha \in A_n} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i (\alpha, t) \right] \) and \( \Pi_n \hat{\alpha}_n = (\hat{\theta}_n, \Pi_n \hat{h}_n) \) where \( \Pi_n \hat{h}_n \) is as in Assumption 3.5. In addition, let \( \hat{t}_n \in \arg \max_{t \in T} \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\Pi_n \hat{\alpha}_n, t) \right] \). For \( n \) large enough, we have
\[
\begin{align*}
&\minsup_{\alpha \in A_n} \sup_{t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right|^2 - \minsup_{\alpha \in A_n} \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right]^2 \\
&\leq \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\Pi_n \hat{\alpha}_n, \hat{t}_n) \right]^2 - \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\hat{\alpha}_n, \hat{t}_n) \right]^2 \\
&\leq \sup_{t \in T, ||\alpha - \alpha||_{L^2} \leq \delta_n} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_1, t) \right|^2 - \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_2, t) \right]^2. \tag{55}
\end{align*}
\]
The first inequality in (55) follows from
\[
\minsup_{\alpha \in A_n} \sup_{t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right|^2 \leq \sup_{t \in T} \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\Pi_n \hat{\alpha}_n, \hat{t}_n) \right]^2 = \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\Pi_n \hat{\alpha}_n, \hat{t}_n) \right]^2 \tag{56}
\]
and
\[
\minsup_{\alpha \in A} \sup_{t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right|^2 = \sup_{t \in T} \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\Pi_n \hat{\alpha}_n, t) \right]^2 \geq \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\hat{\alpha}_n, \hat{t}_n) \right]^2. \tag{57}
\]
The second inequality in (55) holds for \( n \) large enough because \( ||\hat{h}_n - \Pi_n \hat{h}_n||_{L^2} = o (\delta_n) \) due to Assumption 3.5(ii).

Similarly, it can be shown that
\[
\begin{align*}
&\minsup_{\alpha \in A} \sup_{t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right|^2 - \minsup_{\alpha \in A} \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right]^2 \\
&\leq \sup_{t \in T, ||\alpha - \alpha||_{L^2} \leq \delta_n} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_1, t) \right|^2 - \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_2, t) \right]^2. \tag{58}
\end{align*}
\]
(55) and (58) imply that

$$| \min_{\alpha \in \mathcal{A}_n} \sup_{t \in T} \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right] - \min_{\alpha \in \mathcal{A}} \sup_{t \in T} \left[ \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right] | \leq \sup_{t \in T, ||\alpha_1 - \alpha_2||_{L^2} \leq \delta_n} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_1, t) \right|^2 - \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_2, t) \right|^2 \right| \right|_{L^2}.$$  

(59)

Note that

$$\left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_1, t) \right|^2 - \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_2, t) \right|^2 \right| \leq \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_1, t) - \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_2, t) \right| \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_1, t) \right| + \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_2, t) \right|$$

$$\leq \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_1, t) - \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_2, t) \right| \cdot 2 \sup_{\alpha \in \mathcal{A}, t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right|$$

$$\leq \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_1, t) - E[u(\alpha_1, t)] \right| \cdot 2 \sup_{\alpha \in \mathcal{A}, t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right|$$

$$+ \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_2, t) - E[u(\alpha_2, t)] \right| \cdot 2 \sup_{\alpha \in \mathcal{A}, t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right|$$

$$\leq \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_1, t) - E[u(\alpha_1, t)] \right| \cdot 2 \sup_{\alpha \in \mathcal{A}, t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right|$$

$$+ \left| E[u(\alpha_1, t)] - E[u(\alpha_2, t)] \right| \cdot 2 \sup_{\alpha \in \mathcal{A}, t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right|$$

$$\leq 4 \sup_{\alpha \in \mathcal{A}, t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) - E[u(\alpha, t)] \right| \cdot \sup_{\alpha \in \mathcal{A}, t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right|$$

$$+ 2 \cdot |E[u(\alpha_1, t)] - E[u(\alpha_2, t)]| \cdot \sup_{\alpha \in \mathcal{A}, t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right|$$

$$\leq 4 \sup_{\alpha \in \mathcal{A}, t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) - E[u(\alpha, t)] \right| \cdot \sup_{\alpha \in \mathcal{A}, t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha, t) \right|$$

$$+ 2 \cdot \max_{y \in \mathcal{Y}, ||l|| \leq B} |m_1^l (y, l)| \cdot P \cdot ||h_1 - h_2||_{L^2} \cdot \sup_{h \in \mathcal{H}, t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (t, h) \right|.$$

(60)

(54) and (60) together imply

$$\sup_{t \in T, ||\alpha_1 - \alpha_2||_{L^2} \leq \delta_n} \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_1, t) \right|^2 - \left| \frac{1}{n} \sum_{i=1}^{n} u_i (\alpha_2, t) \right|^2 \overset{a.s.}{\rightarrow} 0.$$  

(61)
(59) and (61) together imply
\[
\min_{\alpha \in A \cap R} \left\{ \sum_{i=1}^{n} u_i(\alpha, t) \right\}^2 - \min_{\alpha \in A \cap R} \left\{ \sum_{i=1}^{n} u_i(\alpha, t) \right\}^2 \xrightarrow{a.s.} 0. \tag{62}
\]

(53) and (62) imply that
\[
\min_{\alpha \in A \cap R} \left\{ \sum_{i=1}^{n} u_i(\alpha, t) \right\}^2 \xrightarrow{a.s.} \min_{\alpha \in A \cap R} \left[ E \left[ m(Y, \theta, h(X)) g(t, Z) \right] \right]^2. \tag{63}
\]

Assumption 3.2 and (63) imply that
\[
\min_{\alpha \in A \cap R} \left\{ \sum_{i=1}^{n} u_i(\alpha, t) \right\}^2 \xrightarrow{a.s.} \min_{\alpha \in A \cap R} \left[ E \left[ m(Y, \theta, h(X)) g(t, Z) \right] \right]^2, \tag{64}
\]

which completes the proof of Theorem 3.2. ■

**Proof of Theorem 3.3**

Recall that the bootstrap test statistic takes the form
\[
S_n^* = \min_{\alpha \in A \cap R} \left\{ \sup_{t \in T_n} \left[ J_n^* (\alpha, t) \right] W_n^* (\alpha, t) \left[ J_n^* (\alpha, t) \right] + \lambda_n P_n (\alpha, t) \right\}, \tag{65}
\]

where
\[
J_n^* (\alpha, t) = \frac{1}{n} \sum_{i=1}^{n} \left[ m(y_i^*, \theta, h(x_i^*)) \cdot g(t, z_i^*) - \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) g(t, z_i) \right]. \tag{66}
\]

According to Lemma 1,
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( m(y_i, h(x_i)) \cdot g(t, z_i) - E \left[ m(Y, h(X)) \cdot g(t, Z) \right] \right) \xrightarrow{L} G(\alpha, t). \]

Consequently,
\[
J_n^* (\alpha, t) \xrightarrow{L} G(\alpha, t) \quad \text{a.s.} \tag{67}
\]

according to Theorem 3.6.3 in van der Vaart and Wellner (1996).

It can be shown through essentially the same argument as in the proof of Lemma .17 in Santos (2011) that
\[
S_n^* = \min_{\alpha \in A \cap A_n \cap R} \left\{ \sup_{t \in T_n} \left[ J_n^* (\alpha, t) \right] W_n^* (\alpha, t) \left[ J_n^* (\alpha, t) \right] \right\} + o_p(1), \tag{68}
\]
which, together with (67), imply that
\[ S^*_n \Rightarrow D' \equiv \inf_{\alpha \in A_I \cap R} \sup_{t \in T} \{ G(\alpha, t)^t W(\alpha, t) G(\alpha, t) \}. \] (69)

Recall that the null asymptotic distribution of \( S_n \) is
\[ D = \inf_{\alpha \in A_I \cap R} \inf_{v \in V^\alpha} \sup_{t \in T} \{ [G(\alpha, t) + v(t)]^t W(\alpha, t) [G(\alpha, t) + v(t)] \}. \] (70)

According to Definition 3.4, the function \( v(t) \equiv 0 \) belongs to \( V^\alpha_\infty \) for all \( \alpha \in A_I \cap R \). Therefore, \( D' \) stochastically dominates \( D \) (in a nonstrict sense), which proves the first part of Theorem 3.3. The second part follows from the same argument as in the proof of Theorem 3.2. ■

**Proof of Theorem 3.4**

Define by \( C(1 - \rho) \) the \((1 - \rho)\)th quantile of \( D' \). Then \( C^*_n (1 - \rho) \overset{a.s.}{\rightarrow} C(1 - \rho) \) according to Theorem 3.3.

The first equality in (71) below follows from Theorem 3.1. The inequality in (71) follows from Theorem 3.3.

\[
\limsup_{n \to \infty} \Pr (S_n > C^*_n (1 - \rho) | H_0) = \Pr (D > C(1 - \rho)) \leq \Pr (D' > C(1 - \rho)) = \rho \] (71)

The first part of Theorem 3.4 follows from (71).

When \( A_I \cap R = \emptyset \), it follows from Theorem 3.2 that \( S_n = O_p(n) \), and it follows from Theorem 3.3 that \( S^*_n = O_p(\lambda_n) \). The second part of Theorem 3.4 follows from the fact that \( \lambda_n = o(n) \) which is guaranteed by Assumption 3.7. ■

**Proof of Theorem 4.1**

According to (17) which defines our confidence sets, for any \( \theta \in \Theta_I \),
\[ \Pr (\theta \in CS_n (1 - \rho)) = \Pr (S_n (\theta) \leq C^*_n (1 - \rho, \theta)). \] (72)
Taking \( \lim \inf \) on both sides of (72) yields

\[
\lim_{n \to \infty} \Pr(\theta \in \text{CS}_n (1 - \rho)) = \lim_{n \to \infty} \Pr(S_n (\theta) \leq C^*_n (1 - \rho, \theta)) \geq 1 - \rho
\]  

(73)

where the inequality follows directly from Theorem 3.4.

Suppose \( \lim \inf_{n \to \infty} \inf_{\theta \in \Theta} \Pr(\theta \in \text{CS}_n (1 - \rho)) < 1 - \rho \). Then \( \exists \epsilon > 0 \) s.t.

\[
\lim_{n \to \infty} \inf_{\theta \in \Theta} \Pr(\theta \in \text{CS}_n (1 - \rho)) = 1 - \rho - \epsilon.
\]  

(74)

Consequently, \( \exists \) sequence \( \{\theta_n\}_{n=1}^{\infty} \subseteq \Theta \) s.t.

\[
\lim_{n \to \infty} \Pr(\theta_n \in \text{CS}_n (1 - \rho)) = 1 - \rho.
\]  

(75)

According to the definition of \( \Theta_I \), \( \exists \{h_n\}_{n=1}^{\infty} \subseteq H \) s.t. \( \{\theta_n, h_n\}_{n=1}^{\infty} \subseteq A_I \).

Because \( \Theta \) is compact under \( \| \cdot \|_E \), and \( H \) is compact under \( \| \cdot \|_{c} \), there is subsequence \( \{\theta_{n_l}, h_{n_l}\}_{n_l=1}^{\infty} \) and \( \tilde{\alpha} = (\tilde{\theta}, \tilde{h}) \) s.t. \( (\theta_{n_l}, h_{n_l}) \to \tilde{\alpha} \) under \( \| \cdot \|_E + \| \cdot \|_c \), which implies that \( (\theta_{n_l}, h_{n_l}) \to \tilde{\alpha} \) under \( \| \cdot \|_w \) because \( \| \cdot \|_w \leq \| \cdot \|_E + \| \cdot \|_c \). Therefore, \( d_w (\tilde{\alpha}, A_I) = 0 \).

It follows from Assumption 3.3 (iv) that

\[
\max_{t \in T} \| E[m(Y, \theta, h(X)) \cdot g(t, Z)] \|_E \leq c_2 d_w (\tilde{\alpha}, A_I).
\]  

(76)

Therefore, \( \tilde{\alpha} \in A_I \) because \( \max_{t \in T} \| E[m(Y, \theta, h(X)) \cdot g(t, Z)] \|_E = 0 \). Consequently, \( \tilde{\theta} \in \Theta_I \) and \( \theta_{n_l} \to \tilde{\theta} \) under \( \| \cdot \|_E \).

Lemma 1, (75), and stochastic equicontinuity imply that there is \( N > 0 \), s.t.

\[
\Pr(\tilde{\theta} \in \text{CS}_n (1 - \rho)) \leq 1 - \rho - \frac{\epsilon}{2}.
\]  

(77)

(77) implies that \( \lim \inf_{n \to \infty} \Pr(\tilde{\theta} \in \text{CS}_n (1 - \rho)) < 1 - \rho \), which contradicts with (73).

Therefore,

\[
\lim_{n \to \infty} \inf_{\theta \in \Theta_I} \Pr(\theta \in \text{CS}_n (1 - \rho)) \geq 1 - \rho.
\]  

(78)

For any \( \theta \notin \Theta_I \),

\[
\lim_{n \to \infty} \Pr(\theta \in \text{CS}_n (1 - \rho)) = \lim_{n \to \infty} \Pr(S_n (\theta) \leq C^*_n (1 - \rho, \theta)) = 0
\]  

(79)

where the second equality follows from Theorem 3.4.  ■
Proof of Theorem 4.2 follows the same reasoning and very similar steps as in the proof of Theorem 4.1. Therefore, it is skipped here.

Appendix B. Proofs of Lemmas 1-6

In this appendix we prove Lemmas 1-6.

Proof of Lemma 1:

If Assumption 3.1(ii) (a) is satisfied, \( g(t, z) = 1 \{ z \leq t \} \) for \( t \in \mathbb{Z} \). \{\{z : z \leq t\} : t \in \mathbb{Z}\} is a Vapnik–Chervonenkis (VC) class of sets. Consequently, being a family of indicator functions of the VC class of sets, \( \{g(t, \cdot) : t \in \mathbb{Z}\} \) forms a Donsker class according to Thm 2.6.4 of van der Vaart and Wellner (1996). (Or, for a simpler exhibition, see the extension of the type I classes of functions discussed in Andrews (1994).) Also note that in this case \(|g(t, \cdot)| \leq 1\). Therefore, \( \{g(t, \cdot) : t \in \mathbb{Z}\} \) is a uniformly bounded Donsker class.

If Assumption 3.1(ii) (b) is satisfied, \( \{g(t, \cdot) : t \in T\} \) becomes a type II class defined in Andrews (1994). According to Thm 2 in Andrews (1994) and the compactness of \( T \), \( \{g(t, z) : t \in T\} \) is a uniformly bounded Donsker class.

Therefore, Assumption 3.1(ii) guarantees that \( \{g(t, z) : t \in T\} \) forms a uniformly bounded Donsker class.

According to Assumption 3.3(i), \( \{m(\cdot, \alpha) : \alpha \in A\} \) is a uniformly bounded Donsker class.

Therefore, it follows directly from Example 2.10.8 of van der Vaart and Wellner (1996) that \( \{m(\cdot, \alpha)g(t, \cdot) : (\alpha, t) \in A \times T\} \), being the pairwise products \( \{m(\cdot, \alpha) : \alpha \in A\} \cdot \{g(t, z) : t \in T\} \), forms a Donsker class. ■

Proof of Lemma 2:

The proof of Lemma 2 uses the following lemma (stated in general terms):

**Lemma 2.1.** Suppose the following conditions hold: (1) \( Q(\theta) \geq 0 \) and \( \Theta_I = \{\theta \in \Theta : Q(\theta) = 0\} \); (2) \( \Theta_n \subseteq \Theta \) are closed and \( \exists \Pi_n \theta \in \Theta_n \) for each \( \theta \in \Theta \) s.t. \( \|\Pi_n \theta - \theta\| = o(1) \) and \( \sigma_n \equiv \sup_{\theta_0 \in \Theta_I} \|\Pi_n \theta_0 - \theta_0\| = o(1) \); (3) \( \sup_{\theta \in \Theta_n} |Q_n(\theta) - Q(\theta)| = O_p(n^{-1/2}) \); (4) \( \exists \) positive constants \( c_1, c_2 \) and \( \delta \) s.t. \( c_1 \left( \inf_{\theta_0 \in \Theta_I} \|\theta - \theta_0\| \wedge \delta \right) \leq Q(\theta) \leq c_2 \inf_{\theta_0 \in \Theta_I} \|\theta - \theta_0\| \) for all \( \theta \in \Theta \).
Then, for $\hat{\theta}_n \in \text{arg min}_{\theta \in \Theta_n} Q_n(\theta)$ it follows that $\inf_{\theta_0 \in \Theta_0} ||\hat{\theta}_n - \theta_0|| = O_p\left(\max\{\sigma_n, n^{-1/2}\}\right)$.

(Proof: Assumption (3) implies that

$$\sup_{\theta \in \Theta_n} |Q_n(\theta) - Q(\theta)| \leq L \cdot n^{-1/2} \tag{80}$$

for some positive constant $L < \infty$. For any fixed $\theta_0 \in \Theta_0$, $\exists$ sequence $\{\theta_{0n}\}$ with $\theta_{0n} \in \Theta_n$ s.t. $||\theta_{0n} - \theta_0|| \leq K \cdot \sigma_n$ for some positive $K < \infty$ (which does not depend on $\theta_0$ because of Assumption (2)). Let $\delta_n = \max\left\{\frac{4L}{c_1}, \frac{2c_2K}{c_1}\right\} \cdot \max\{\sigma_n, n^{-1/2}\}$. Let $\Theta_0^{\delta_n}$ denote an open $\delta_n$-enlargement of $\Theta_0$ under $|| \cdot ||$. By Assumption (4), we have:

$$\Delta_n \equiv \inf_{\theta \in (\Theta_0^{\delta_n})^c \cap \Theta} Q(\theta) \geq c_1 \cdot \delta_n > 0 \text{ for } n \text{ large enough.}$$

$$Q\left(\hat{\theta}_n\right) - Q(\theta_{0n}) = \left[Q\left(\hat{\theta}_n\right) - Q_n\left(\hat{\theta}_n\right)\right] + \left[Q_n\left(\hat{\theta}_n\right) - Q_n(\theta_{0n})\right] + \left[Q_n(\theta_{0n}) - Q(\theta_{0n})\right]$$

$$\leq \left[Q\left(\hat{\theta}_n\right) - Q_n\left(\hat{\theta}_n\right)\right] + \left[Q_n(\theta_{0n}) - Q(\theta_{0n})\right]$$

$$\leq |Q\left(\hat{\theta}_n\right) - Q_n\left(\hat{\theta}_n\right)| + |Q(\theta_{0n}) - Q_n(\theta_{0n})| \tag{81}$$

It follows that

$$\Pr\left(Q\left(\hat{\theta}_n\right) < Q(\theta_{0n}) + \frac{\Delta_n}{2}\right) = \Pr\left(Q\left(\hat{\theta}_n\right) - Q(\theta_{0n}) < \frac{\Delta_n}{2}\right)$$

$$\geq \Pr\left(Q\left(\hat{\theta}_n\right) - Q(\theta_{0n}) < \frac{1}{2}c_1\delta_n\right)$$

$$\geq \Pr\left(|Q\left(\hat{\theta}_n\right) - Q_n\left(\hat{\theta}_n\right)| + |Q(\theta_{0n}) - Q_n(\theta_{0n})| < 2L \cdot n^{-1/2}\right)$$

$$\rightarrow 1 \tag{82}$$

where the last step is implied by (80).

By Assumption (4) and $||\theta_{0n} - \theta_0|| = O(\delta_n)$, we have $Q(\theta_{0n}) \leq c_2 ||\theta_{0n} - \theta_0|| < c_2 \cdot K \cdot \sigma_n < \frac{\Delta_n}{2}$ for $n$ large enough. Therefore:

$$\Pr\left(Q\left(\hat{\theta}_n\right) < \Delta_n\right) \rightarrow 1. \tag{83}$$

This implies that $\Pr\left(d\left(\hat{\theta}_n, \Theta_0\right) \leq \delta_n\right) \rightarrow 1$. Therefore, $\inf_{\theta_0 \in \Theta_0} ||\hat{\theta}_n - \theta_0|| \leq O_p(\delta_n) = O_p\left(\max\{\sigma_n, n^{-1/2}\}\right)$, which completes our proof.)
Define \( Q(\alpha) = \sup_{t \in T} \|W^{1/2}(\alpha, t) \cdot E[m(Y, \theta, h(X)) \cdot g(t, Z)]\|_E \), and \( Q_n(\alpha) = \sup_{t \in T} \|W^{1/2}_n(\alpha, t) \cdot \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) \cdot g(t, z_i)\|_E \).

Obviously, \( \arg \min_{\alpha \in A_n \cap R} \left\{ \sup_{t \in T} [J_n(\alpha, t)]' W_n(\alpha, t) [J_n(\alpha, t)] \right\} = \arg \min_{\alpha \in A_n \cap R} Q_n(\alpha) \).

Assumption 3.1(ii) implies condition (1) in Lemma 3.1 holds for \( \| \cdot \|_w \).

Assumption 3.4 (iii) and (iv) implies condition (2) holds for \( A_n \).

Lemma 1 implies that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{m(y_i, \theta, h(x_i)) \cdot g(t, z_i) - E[m(Y, \theta, h(X)) \cdot g(t, Z)]\} \overset{L}{\rightarrow} G(\alpha, t) \tag{84}
\]
on \( T \times \mathcal{H} \).

Consequently,

\[
\sqrt{n} |Q_n(\alpha) - Q(\alpha)| = \sqrt{n} \sup_{t \in T} \|W^{1/2}_n(\alpha, t) \cdot \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) \cdot g(t, z_i)\|_E - \sup_{t \in T} \|W^{1/2}(\alpha, t) E[m(Y, \theta, h(X)) \cdot g(t, Z)]\|_E |
\]

\[
\leq \sqrt{n} \sup_{t \in T} \|W^{1/2}_n(\alpha, t) \cdot \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) \cdot g(t, z_i) - W^{1/2}(\alpha, t) E[m(Y, \theta, h(X)) \cdot g(t, Z)]\|_E
\]

\[
\leq \sqrt{n} \sup_{t \in T} \|W^{1/2}(\alpha, t) \left[ \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) \cdot g(t, z_i) - E[m(Y, \theta, h(X)) \cdot g(t, Z)] \right]\|_E
\]

\[
+ \sqrt{n} \sup_{t \in T} \|W^{1/2}(\alpha, t) - W^{1/2}_n(\alpha, t)\|_E \cdot \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) \cdot g(t, z_i)\|_E
\]

\[
\leq \sqrt{n} \sup_{t \in T} \|W^{1/2}(\alpha, t) \cdot \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) \cdot g(t, z_i) - E[m(Y, \theta, h(X)) \cdot g(t, Z)]\|_E
\]

\[
+ I \cdot \sup_{t \in T} \|W^{1/2}_n(\alpha, t) \|_E \cdot \frac{1}{n} \sum_{i=1}^{n} m(y_i, \theta, h(x_i)) \cdot g(t, z_i)\|_E \quad \text{for some } I > 0
\]

\[
\overset{L}{\rightarrow} \sup_{t \in T} \|G(t, \alpha)\|_E + I \cdot \sup_{t \in T} \|E[m(Y, \theta, h(X)) \cdot g(t, Z)]\|_E \tag{85}
\]

Therefore, w.p. \( \rightarrow 1 \),

\[
\sqrt{n} \sup_{\alpha \in A_n \cap R} |Q_n(\alpha) - Q(\alpha)| \leq \sup_{\alpha \in A_n \cap R} \sup_{t \in T} \|G(t, \alpha)\|_E + I \cdot \sup_{\alpha \in A_n \cap R} \sup_{t \in T} \|E[m(Y, \theta, h(X)) \cdot g(t, Z)]\|_E \tag{86}
\]

(86) implies that \( \sqrt{n} \sup_{\alpha \in A_n \cap R} |Q_n(\alpha) - Q(\alpha)| = O_p(1) \), which implies condition (3) in Lemma 3.1.

Assumption 3.3 (iv) states that condition (4) holds.
With condition (1) - (4) of Lemma 3.1 being verified, Lemma 2 is valid according to Lemma 2.1. ■

Proof of Lemma 3:
For any $\alpha_0 \in A_I$,
\[
\|\hat{\alpha}_n - \alpha_0\|_{L^2} \\
\leq \|\hat{\alpha}_n - P(\alpha_0)\|_{L^2} + \|P(\alpha_0) - \alpha_0\|_{L^2} \\
\leq \psi_n \cdot \|\hat{\alpha}_n - \alpha_0\|_w + \delta_{s,n} \\
\leq \psi_n \cdot (\|\hat{\alpha}_n - \alpha_0\|_w + \delta_{w,n}).
\] (87)

According to (87),
\[
\|\hat{\alpha}_n - \alpha_0\|_{L^2} \leq \psi_n \cdot (\|\hat{\alpha}_n - \alpha_0\|_w + \delta_{w,n}).
\] (88)

Taking $\inf_{\alpha \in A_n \cap R}$ on both sides of (88) yields
\[
d_{L^2}(\hat{\alpha}_n, A_I \cap R) \leq \psi_n \cdot (d_w(\hat{\alpha}_n, A_I \cap R) + \delta_{w,n}),
\] (89)

which, together with Lemma 2 implies lemma 3. ■

Proof of Lemma 4:
The triangle inequality implies that :
\[
\left| \inf_{b \in B} \inf_{f \in \mathcal{F}_\infty^b} \sup_{a \in A} |g(a, b) - f(a)| - \inf_{b \in B} \inf_{f \in \mathcal{F}_\infty^b} \sup_{a \in A} |g_n(a, b) - f(a)| \right| \\
\leq \left| \inf_{b \in B} \inf_{f \in \mathcal{F}_\infty^b} \sup_{a \in A} |g(a, b) - f(a)| - \inf_{b \in B} \inf_{f \in \mathcal{F}_\infty^b} \sup_{a \in A} |g(a, b) - f(a)| \right| \\
+ \left| \inf_{b \in B} \inf_{f \in \mathcal{F}_\infty^b} \sup_{a \in A} |g(a, b) - f(a)| - \inf_{b \in B} \inf_{f \in \mathcal{F}_\infty^b} \sup_{a \in A} |g_n(a, b) - f(a)| \right|. 
\] (90)

Fix $\epsilon > 0$ and there exists $\left\{\hat{f}_{b,\infty}\right\}_{b \in B}$ with $\hat{f}_{b,\infty} \in \mathcal{F}_\infty^b$ for all $b \in B$ s.t.:
\[
\inf_{b \in B} \inf_{f \in \mathcal{F}_\infty^b} \sup_{a \in A} |g(a, b) - f(a)| \geq \inf_{b \in B} \sup_{a \in A} |g(a, b) - \hat{f}_{b,\infty}| - \epsilon.
\]
Also there exists \( \hat{b}_{n} \in F_{n}^{b} \) s.t. \( \inf_{f \in F_{n}^{b}} \| f - \hat{b}_{n} \|_{\infty} \geq \| \hat{b}_{n} - \hat{b}_{\infty} \|_{\infty} - \epsilon \), for each \( b \in B \) and \( n = 1, 2, 3, \ldots \).

Finally, there exists \( \tilde{b} \in B \) s.t.

\[
\inf \sup_{b \in B} | g(a, b) - \hat{b}_{n}(a) | \geq \sup_{a \in A} | g(a, \tilde{b}) - \hat{b}_{n}(a) | - \epsilon.
\]

By the definition of \( F_{\infty}^{b} \), \( \inf_{f \in F_{n}^{b}} \| f - \hat{b}_{\infty} \|_{\infty} = o(1) \), for each \( b \in B \).

Therefore,

\[
\inf \inf_{b \in B} \sup_{f \in F_{n}^{b}} \sup_{a \in A} | g(a, b) - f(a) | \leq \inf \sup_{b \in B} \sup_{a \in A} | g(a, b) - \hat{b}_{n}(a) | \leq \inf \left\{ \sup_{a \in A} \sup_{b \in B} | g(a, b) - \hat{b}_{\infty}(a) | + \sup_{a \in A} | \hat{b}_{\infty}(a) - \hat{b}_{n}(a) | \right\} = \inf \left\{ \sup_{a \in A} \sup_{b \in B} | g(a, b) - \hat{b}_{\infty}(a) | + \| \hat{b}_{\infty} - \hat{b}_{n} \|_{\infty} \right\} \leq \inf \left\{ \sup_{a \in A} \sup_{b \in B} | g(a, \tilde{b}) - \hat{b}_{\infty}(a) | + \inf_{f \in F_{n}^{b}} \| f - \hat{b}_{\infty} \|_{\infty} + \epsilon \right\} \leq \sup_{a \in A} \left( g(a, \tilde{b}) - \hat{b}_{\infty}(a) \right) + \inf_{f \in F_{n}^{b}} \| f - \hat{b}_{\infty} \|_{\infty} + \epsilon \leq \inf \sup_{b \in B} | g(a, b) - \hat{b}_{\infty}(a) | + o(1) + 2\epsilon \leq \inf \inf_{b \in B} \sup_{f \in F_{n}^{b}} \sup_{a \in A} | g(a, b) - f(a) | + o(1) + 3\epsilon
\]

(91)

Since \( \epsilon \) is arbitrary, we have:

\[
\inf \inf_{b \in B} \sup_{f \in F_{n}^{b}} \sup_{a \in A} | g(a, b) - f(a) | \leq \inf \inf_{b \in B} \sup_{f \in F_{n}^{b}} \sup_{a \in A} | g(a, b) - \hat{b}_{\infty}(a) | + o(1) + 3\epsilon
\]

(92)

where the first inequality follows from \( F_{n}^{b} \subseteq F_{n}^{b} \).

(92) implies that,

\[
\inf \inf_{b \in B} \sup_{f \in F_{n}^{b}} \sup_{a \in A} | g(a, b) - f(a) | = \inf \inf_{b \in B} \sup_{f \in F_{n}^{b}} \sup_{a \in A} | g(a, b) - \hat{b}_{\infty}(a) | + o(1). \tag{93}
\]

Next, \( \| g_{n} - g \|_{\infty} = o(1) \) implies that:

\[
44
\]
\[ \inf_{b \in B} \inf_{f \in F_n} \sup_{a \in A} |g(a, b) - f(a)| \leq \inf_{b \in B} \inf_{f \in F_n} \left[ \sup_{a \in A} |g(a, b) - g_n(a, b)| + \sup_{a \in A} |g_n(a, b) - f(a)| \right] \]
\[ \leq o(1) + \inf_{b \in B} \inf_{f \in F_n} \sup_{a \in A} |g_n(a, b) - f(a)| \]  
\[ \text{(94)} \]

Similar, we can show that:

\[ \inf_{b \in B} \inf_{f \in F_n} \sup_{a \in A} |g_n(a, b) - f(a)| \leq \inf_{b \in B} \inf_{f \in F_n} \sup_{a \in A} |g_n(a, b) - f(a)| + o(1) \]  
\[ \text{(95)} \]

Therefore, by (94) and (95), we have

\[ \inf_{b \in B} \inf_{f \in F_n} \sup_{a \in A} |g(a, b) - f(a)| = \inf_{b \in B} \inf_{f \in F_n} \sup_{a \in A} |g_n(a, b) - f(a)| + o(1). \]  
\[ \text{(96)} \]

(90), (93), and (96) complete the proof. \[ \blacksquare \]

**Proof of Lemma 5:**

Fix \( \varepsilon_n \searrow 0 \). There exists a sequence \( \{\hat{a}_n\} \) s.t. \( \hat{a}_n \in A_n \) and the following inequality holds:

\[ \inf_{a \in A_n} \max_{b \in B} F_n^2(a, b) \geq \max_{b \in B} F_n^2(\hat{a}_n, b) + \varepsilon_n. \]  
\[ \text{(97)} \]

Then step 5 in (98) follows from (97) and \( \sup_{a \in A_n} \sup_{b \in B} [G_n(a, b) - F_n(a, b)]^2 = o_p(1) \):
\[
\inf_{a \in A_n} \max_{b \in B} G_n^2(a, b) \\
= \inf_{a \in A_n} \max_{b \in B} \{[G_n(a, b) - F_n(a, b)] + F_n(a, b)\}^2 \\
\leq \inf_{a \in A_n} \max_{b \in B} \{[G_n(a, b) - F_n(a, b)]^2 + F_n^2(a, b) + 2|F_n(a, b)| \cdot |G_n(a, b) - F_n(a, b)|\} \\
\leq \max_{b \in B} \{[G_n(\hat{a}_n, b) - F_n(\hat{a}_n, b)]^2 + F_n^2(\hat{a}_n, b) + 2|F_n(\hat{a}_n, b)| \cdot |G_n(\hat{a}_n, b) - F_n(\hat{a}_n, b)|\} \\
\leq \max_{b \in B} [G_n(\hat{a}_n, b) - F_n(\hat{a}_n, b)]^2 + \max_{b \in B} F_n^2(\hat{a}_n, b) \\
+ 2\max_{b \in B} |F_n(\hat{a}_n, b)| \cdot \max_{b \in B} |G_n(\hat{a}_n, b) - F_n(\hat{a}_n, b)| \\
\leq o_p(1) + \inf_{a \in A_n} \max_{b \in B} F_n^2(a, b) - \varepsilon_n + 2\sqrt{\inf_{a \in A_n} \max_{b \in B} F_n^2(a, b) - \varepsilon_n} \cdot o_p(1) \\
= \inf_{a \in A_n} \max_{b \in B} F_n^2(a, b) + o_p(1) \\
\] (98)

Similarly, it can be shown that

\[
\inf_{a \in A_n} \max_{b \in B} F_n^2(a, b) = \inf_{a \in A_n} \max_{b \in B} G_n^2(a, b) + o_p(1). \\
\] (99)

(98) and (99) together complete the proof. ■

**Proof of Lemma 6:**

According to Definition 3.4, \(V_{k_n,c_n}^{\alpha_0} \subseteq V_{k_n}^{\alpha_0}\). Therefore

\[
\cup_{V_{k_n,c_n}^{\alpha_0}} \subseteq \cup \cup_{V_{k_n}^{\alpha_0}}. \\
\] (100)

Next, for any positive integer \(l\) and \(v \in V_{k_n}^{\alpha_0}\), there is \(r \in H_{IR}\) s.t.

\[
v(t) = \frac{\partial \rho(\alpha_0, t)}{\partial \theta'} \cdot r_\theta + \frac{d \rho(\alpha_0, t)}{dh} [p_{k_i}'] r_h. \\
\]

Since \(c_n \to \infty\), there is \(N > n\) s.t.

\[
c_N \geq \max \left\{ \|r_N' p_{k_n}^{\alpha_0} \|^2_s, \|r_N\|_E \right\}. \\
\] (101)

Let \(r_N \in H_{NR}\) be \(r_N' = (r', 0, ..., 0)\). Then (101) implies that

\[
v \in \cup V_{k_N,c_N}^{\alpha_0}. \\
\] (102)

46
Since (102) holds for any \( v \in V_{k_l}^{\alpha_0} \), we have
\[
V_{k_l}^{\alpha_0} \subseteq V_{k_N, c_N}^{\alpha_0}.
\] (103)

Since for any positive integer \( l \) there is an appropriate \( N \) s.t. (103) holds, we have
\[
\cup V_{k_n}^{\alpha_0} \subseteq \cup V_{k_n, c_n}^{\alpha_0}.
\] (104)

(100) and (104) implies that \( \cup V_{k_n, c_n}^{\alpha_0} = \cup V_{k_n}^{\alpha_0} \). ■

Proof of Assumption 3.3(ii) Being Satisfied When \( m(\cdot) \) Is Pointwise Lipschitz Continuously in \( \alpha \) and Continuous in \( y \) with Compact \( X \times Y \):

The following lemmas are used in the proof:

**Lemma .1** Let Assumption 2.1(ii). There exists a constant \( K > 0 \) s.t. for all \( \varepsilon \) sufficiently small
\[
\log N (\mathcal{H}, \|\cdot\|_{\infty}, \varepsilon) \leq K \frac{d_x}{m}.
\]

(Proof: Lemma .1 is the same as the Lemma .3 in Santos (2011). And a proof is given there.)

**Lemma .2** Let \( F \equiv \{ f : X \times Y \rightarrow \mathbb{R} : f(x, y) = m(y, \theta, h(x)) \text{ for some } (\theta, h) \in A \} \). There exist \( B' < \infty \) s.t. \( F(y) \equiv \max_{\theta \in \Theta, \|l\|_E \leq B'} \|m(y, \theta, l)\|_E \) is an envelop for \( F \). Moreover, there exists constants \( K_0, K > 0 \) s.t. for all norm with \( \|F\| < \infty \) and \( \varepsilon \) sufficiently small:
\[
N_{[\varepsilon]} (F, \|\cdot\|, \varepsilon|\|F||) \leq K_0 \cdot \left( \frac{\text{diam} \Theta}{\varepsilon} \right)^d \cdot \exp \left[ K \cdot \left( \frac{4}{\varepsilon} \right) \frac{d_x}{m} \right].
\]

(Proof: The compactness of \( \mathcal{H} \), which is implied by Assumption 2.1(ii), implies that for any \( h \in \mathcal{H} \) there exist \( B' < \infty \) s.t. \( \|h\|_c \leq B' \). In turn, \( \|h\|_\infty \leq \|h\|_c \leq B' \) for any \( h \in \mathcal{H} \). Therefore, for any \( f(x, y) = m(y, \theta, h(x)) \in F \),
\[
\|f(x, y)\|_E \leq \max_{\theta \in \Theta, \|l\|_E \leq B'} \|m(y, \theta, l)\|_E
\]
\[
= F(y).
\]

47
The pointwise Lipschitz continuity of $m$ implies
\[
\| m(y, \theta_1, h_1(x)) - m(y, \theta_2, h_2(x)) \|_E \\
\leq M(y) \| (\theta_1, h_1(x))' - (\theta_2, h_2(x))' \|_E \\
\leq M(y) (\| \theta_1 - \theta_2 \|_E + \| h_1 - h_2 \|_\infty). \tag{105}
\]

(105) shows that the class $\mathcal{F}$ is Lipschitz in $\mathcal{A}$ w.r.t. the norm $\| \cdot \|_E + \| \cdot \|_\infty$, which in turn implies the first inequality in (106) below. For $\varepsilon$ small enough, and some $K' < \infty$, the third inequality in (106) follows from Lemma .1 and the compactness of $\Theta$.

\[
N \left( \mathcal{F}, \| \cdot \|, \varepsilon \| F \| \right) \leq N \left( \mathcal{A}, \| \cdot \|_E + \| \cdot \|_\infty, \frac{\varepsilon}{2} \right) \\
\leq N \left( \Theta, \| \cdot \|_E, \frac{\varepsilon}{4} \right) : N \left( \mathcal{H}, \| \cdot \|_\infty, \frac{\varepsilon}{4} \right) \\
\leq \left( \frac{K' \cdot \text{diam}\Theta}{\varepsilon} \right)^{d_\theta} \cdot \exp \left( K \cdot \left( \frac{4}{\varepsilon} \right)^{d_x} \right) \\
= K_0 \cdot \left( \frac{\text{diam}\Theta}{\varepsilon} \right)^{d_\theta} \cdot \exp \left[ K \cdot \left( \frac{4}{\varepsilon} \right)^{d_x} \right]. \tag{106}
\]

Setting $K_0 = (4K')^{d_\theta}$ justifies the final step in (106), which completes the proof of Lemma 2.2.)

Equipped with Lemma .1 and Lemma .2, the proof proceeds as follows:

The boundedness of $\mathcal{A}$ w.r.t. $\| \cdot \|_E + \| \cdot \|_\infty$, the Lipschitz continuity of $m$ in $\alpha$, the compactness of $\mathcal{Y}$, and the continuity of $m$ in $y$ together imply the uniform boundedness of $m$, which in turn implies that, for any $(\theta, h) \in \mathcal{A},$

\[
E \left[ \| m(Y, \theta, h(X)) \|_E^2 \right] < \infty. \tag{107}
\]

Therefore, the central limit theorem implies the convergence in distribution pointwise on $\mathcal{A}$. To verify uniform asymptotic equicontinuity, let $\mathcal{F}$ and $F(y)$ be as in Lemma .2. $F(\cdot)$ is continuous according to the theorem of maximum. Then compactness of $\mathcal{Y}$ implies
boundedness of $F(\cdot)$, which implies that

$$||F||_{L^2}^2 = E \left[ (F(Y))^2 \right] < \infty.$$  \hspace{1cm} (108)

(108) guarantees that

$$N_{||.||_{L^2}, \varepsilon}^{(||.||_{L^2})} = 1,$$  \hspace{1cm} (109)

for some $D > 0$ and any $\varepsilon \geq D$.

(109) implies the first step in (110) below. The second step in (110) is implied by Lemma 2 whose condition is satisfied by (108). The change of variable $u = \frac{\varepsilon}{||F||_{L^2}}$ yields the third step. The final step is justified by $\frac{dx}{m} < 2$, which is implied by Assumption 2.1(ii), and that $q^m > \log q$ for $q$ large enough.

\begin{align*}
\int_0^\infty \sqrt{\log N_{||.||_{L^2}, \varepsilon}^{(||.||_{L^2})}} \, d\varepsilon &= \int_0^D \sqrt{\log N_{||.||_{L^2}, \varepsilon}^{(||.||_{L^2})}} \, d\varepsilon \\
&\leq \int_0^D \left[ K \cdot \left( \frac{4||F||_{L^2}}{\varepsilon} \right)^m + \log K_0 + d_\theta \log (\text{diam}\Theta) - d_\theta \log ||F||_{L^2} \right] \, d\varepsilon \\
&= \int_0^{\frac{D}{||F||_{L^2}}} \left[ K \cdot \left( \frac{4}{u} \right)^m - d_\theta \log u + \log K_0 + d_\theta \log (\text{diam}\Theta) \right] \, du \\
&\leq \infty.
\end{align*}

(110)

According to Theorem 2.5.6 in van der Vaart and Wellner (1996), whose condition is satisfied by (110), $F$ is a Donsker class. □

References


