Econometrica, Vol. 80, No. 1 (January, 2012), 213–275

INFERENCE IN NONPARAMETRIC INSTRUMENTAL VARIABLES WITH PARTIAL IDENTIFICATION

BY ANDRES SANTOS[1]

This paper develops methods for hypothesis testing in a nonparametric instrumental variables setting within a partial identification framework. We construct and derive the asymptotic distribution of a test statistic for the hypothesis that at least one element of the identified set satisfies a conjectured restriction. The same test statistic can be employed under identification, in which case the hypothesis is whether the true model satisfies the posited property. An almost sure consistent bootstrap procedure is provided for obtaining critical values. Possible applications include testing for semiparametric specifications as well as building confidence regions for certain functionals on the identified set. As an illustration we obtain confidence intervals for the level and slope of Brazilian fuel Engel curves. A Monte Carlo study examines finite sample performance.

KEYWORDS: Instrumental variables, partial identification, bootstrap.

1. INTRODUCTION

EMPIRICAL WORK IN ECONOMICS is often concerned with the estimation and analysis of econometric models that are derived from the behavior of optimizing agents. The underlying structural relations typically imply some of the regressors are endogenous, so that these models do not fit the classical regression framework, but are instead of the form

$$Y = \theta_0(X) + \varepsilon, \tag{1}$$

where $E[\varepsilon|X] \neq 0$. Instrumental variables (IV) methods have become immensely popular in econometrics, as they allow for the estimation of the unknown function θ_0.

The potential misspecification of parametric models makes it desirable to extend the IV approach to a more flexible nonparametric framework. Unfortunately, the theoretical study and empirical implementation of nonparametric IV methods has faced two important challenges. First, as originally discussed in Newey and Powell (2003), the nonparametric identification of θ_0 requires the availability of an instrument that satisfies conditions far stronger than the usual covariance restrictions needed in the parametric case. Second, few methods are available for hypothesis testing in a nonparametric IV setting. Ai and

[1]I thank Whitney Newey and three anonymous referees for comments that helped greatly improve the paper. Parts of this paper derive from my doctoral dissertation, completed under the guidance and encouragement of Frank Wolak. I also thank Brendan Beare, Xiaohong Chen, Peter Hansen, Joel Horowitz, Aprajit Mahajan, Sriniketh Nagavarapu, Azeem Shaikh, and seminar participants where this paper was presented. The financial support of the Stanford Institute for Economic and Policy Research is gratefully acknowledged.

 DOI: 10.3982/ECTA7493

Chen (2003) proposed a $\sqrt{n}$ asymptotic normal estimator for continuous functionals of θ_0 that can be employed for inference. Horowitz (2007) obtained the asymptotic normality of the estimator proposed in Hall and Horowitz (2005) and in this way was able to build confidence intervals for the level of θ_0.

In the present paper, we extend the existing results in the nonparametric IV literature by constructing a family of test statistics for a wider set of hypotheses than was previously available. In addition, since nonparametric identification may fail to be attained in applications, our analysis does not assume identification. Instead, we employ many of the prevalent ideas in the partial identification literature; see Manski (1990, 2003). For any dependent variable, endogenous regressor, and instrument triplet (Y, X, Z), we define the *identified set* to be

$$\Theta_0 \equiv \{\theta \in \Theta : E[Y - \theta(X)|Z] = 0\}, \tag{2}$$

where Θ is a nonparametric set of functions that by assumption contains θ_0. The elements of Θ_0 are those functions in Θ that are consistent with the exogeneity assumption on the instrument.

For a family of restrictions on functions θ, we show how to construct a test statistic for the null hypothesis that these restrictions are satisfied by at least one element of Θ_0. Special cases of the restrictions for which we can test include homogeneity and parametric or semiparametric specification tests against a nonparametric alternative. The testing framework developed in this paper can also be used to build confidence regions for certain functionals of identifiable parameters, as proposed in Romano and Shaikh (2008), elaborating on Imbens and Manski (2004). Such confidence regions $C_n(1-\alpha)$ for functionals $f : \Theta_0 \to \mathbf{R}^{d_f}$ and confidence level $1 - \alpha$ satisfy the coverage requirement

$$\inf_{\theta \in \Theta_0} \liminf_{n\to\infty} P(f(\theta) \in C_n(1-\alpha)) \geq 1 - \alpha. \tag{3}$$

This procedure is applicable to a wide range of functionals including the level of a function and its derivatives at a point. If the model is actually identified, then the hypothesis test reduces to whether the true model satisfies the conjectured restriction. Confidence intervals such as (3) are then for the functional evaluated at the true model θ_0.

For a fixed $\theta \in \Theta$, it is possible to test whether the residuals it implies agree with the exogeneity assumption on Z by employing standard specification tests for conditional expectations. Our test builds on this observation by examining whether any function on a sieve that satisfies the conjectured restriction generates residuals for which we are unable to reject that they are mean independent of Z. The limiting distribution of the statistic we propose is nonstandard, and for this reason, we develop a bootstrap procedure and establish its almost sure consistency. As an illustration, we study the Engel curves for gasoline and ethanol in Brazil. For both goods, we fail to reject the null hypothesis that there is at least one log-linear Engel curve in the identified set, a commonly used parametric specification. We also build and compare log-linear and nonparametric confidence regions for the level and derivative of the Engel curves at the sample mean.

The literature on nonparametric IV originated with Newey and Powell (2003), who proposed a consistent estimator for θ_0. They solved the ill-posed inverse problem by obtaining compactness through smoothness assumptions on θ_0, an insight we rely on in this paper. Darolles, Florens, and Renault (2003) and Hall and Horowitz (2005) proposed alternative estimators and established rates of convergence. Ai and Chen (2003, 2007) analyzed semiparametric specifications and obtained efficient estimators for the parametric component. Blundell, Chen, and Kristensen (2007) provided the first empirical application of these methods, conditions for identification, and rates of convergence for the nonparametric component. Severini and Tripathi (2006, 2010) studied linear functionals of θ_0 when θ_0 is not identified and derived the semiparametric efficiency bound for their estimation. Santos (2011) constructed $\sqrt{n}$ asymptotically normal estimators for such functionals. Ai and Chen (2012) derived the semiparametric efficiency bound for identified smooth functionals of θ_0 without requiring θ_0 to be identified. Their estimation procedure, however, assumes identification of θ_0. Horowitz (2006, 2008) derived a specification test for parametric and nonparametric models, respectively. In related work, Newey, Powell, and Vella (1999), Chesher (2003, 2005, 2007), and Imbens and Newey (2009) explored identification and estimation in triangular systems. Inference when instruments fail to provide identification has also been studied in the treatment effects literature. See Manski and Pepper (2000), Heckman and Vytlacil (2001), Manski (2003), Shaikh and Vytlacil (2011), and references therein.

The remainder of the paper is organized as follows. Section 2 expands on the partial identification framework while Section 3 develops the test statistic, proposes a bootstrap procedure for obtaining critical values and presents a Monte Carlo study. In Section 4, we analyze the Brazilian fuel Engel curves and we briefly conclude in Section 5. All proofs are contained in the Appendix.

2. PARTIAL IDENTIFICATION

For a random variable V with support $\mathcal{V}$, we denote the space $L^p(V)$ and the associated norm $\|\cdot\|_{L^2_p}$ by

$$L^p(V) \equiv \{\theta : \mathcal{V} \to \mathbf{R} : E[\theta^p(V)] < \infty\}, \quad \|\theta\|^p_{L^p} \equiv E[\theta^p(V)], \tag{4}$$

and as is standard, $L^\infty(V)$ is understood to be the set of bounded functions on $\mathcal{V}$ and $\|\cdot\|_\infty$ is understood to be the usual supremum norm. If we define the continuous linear operator $\Upsilon : L^2(X) \to L^2(Z)$ by $\Upsilon(\theta) \equiv E[\theta(X)|Z]$, we see that a function θ agrees with the exogeneity assumption on Z if and only if (iff)

$$E[Y|Z] = \Upsilon(\theta). \tag{5}$$

Therefore, without further assumptions on θ_0, the identified set is given by the set of solutions to the integral equation in (5). Due to the linearity of Υ, such a set can be characterized as

$$\mathbf{V}_0 \equiv \theta_0 + \mathcal{N}(\Upsilon), \tag{6}$$

where $\mathcal{N}(\Upsilon)$ denotes the null space of Υ in $L^2(X)$. Identification of θ_0 is hence equivalent to $\mathcal{N}(\Upsilon) = \emptyset$, which, as Newey and Powell (2003) originally noted, is tantamount to a completeness assumption on the conditional distribution of X given Z.

Restrictions on the parameter space Θ aid in identification as the relevant identified set is no longer $\mathbf{V}_0$, but $\mathbf{V}_0 \cap \Theta$. A common restriction in the nonparametric literature is to assume Θ is compact in $\|\cdot\|_{L^2}$. We can obtain some intuition for the effects of such an assumption by examining the parametric linear model, in which case the identified set has a simple geometric interpretation.[2]

EXAMPLE 2.1: Let $X \in \mathbf{R}^{d_x}$, $Z \in \mathbf{R}^{d_z}$, and $d_x = d_z$, and suppose the parametric model

$$Y = X'\beta_0 + \varepsilon \tag{7}$$

holds with $E[X\varepsilon] \neq 0$ but $E[Z\varepsilon] = 0$. Defining the matrix $\Upsilon_L \equiv E[ZX']$, the set of solutions to

$$E[ZY] = \Upsilon_L \beta \tag{8}$$

constitutes the identified set for β_0. The model therefore fails to be identified if Υ_L is not full rank, in which case the identified set is given by the affine subspace $\beta_0 + \mathcal{N}(\Upsilon_L)$. If β_0 is assumed to satisfy a known norm bound $\|\beta_0\| \leq B$, then the identified set can be further restricted. Geometrically, such a set is given by the intersection of the affine subspace $\beta_0 + \mathcal{N}(\Upsilon_L)$ and the ball $\{\beta : \|\beta\| \leq B\}$. This intersection may be (i) empty for B sufficiently small (model is misspecified), (ii) a singleton when $\beta_0 + \mathcal{N}(\Upsilon_L)$ is tangent to the ball, or (iii) a nonsingleton set for B sufficiently large.

In a nonparametric context, the identification problem is more subtle, as compactness of a set is no longer equivalent to it being closed and bounded.

[2]We thank Whitney Newey for suggesting this example.

However, we show that by examining the identified set in the appropriate space, the geometric interpretation of Example 2.1 translates verbatim into our setting. Toward this end, we first formally introduce the parameter space Θ.

2.1. *Smoothness Restrictions*

We follow Newey and Powell (2003) and restrict the parameter space to be a smooth set of functions. This restriction guarantees the statistics we examine are tight processes indexed by $\theta \in \Theta$.

Specifically, let X have support $\mathcal{X} \subseteq \mathbf{R}^{d_x}$ and for λ a d_x-dimensional vector of nonnegative integers, define $|\lambda| = \sum_{i=1}^{d_x} \lambda_i$, $D^\lambda \theta(x) = \partial^{|\lambda|}\theta(x)/\partial x_1^{\lambda_1} \cdots \partial x_{d_x}^{\lambda_{d_x}}$ and introduce the norms

$$\|\theta\|_s^2 \equiv \sum_{|\lambda| \le m+m_0} \int_{\mathcal{X}} [D^\lambda \theta(x)]^2 (1 + x'x)^{\delta_0}\, dx, \tag{9}$$

$$\|\theta\|_c \equiv \max_{|\lambda| \le m} \sup_{x \in \mathcal{X}} |D^\lambda \theta(x)| (1 + x'x)^{\delta/2},$$

where m and m_0 are strictly positive integers, and $\delta, \delta_0 \ge 0$. In addition, denote the corresponding weighted Sobolev spaces associated with these norms by

$$W^s(\mathcal{X}) \equiv \{\theta : \mathcal{X} \to \mathbf{R} \text{ s.t. } \|\theta\|_s < \infty\}, \tag{10}$$

$$W^c(\mathcal{X}) \equiv \{\theta : \mathcal{X} \to \mathbf{R} \text{ s.t. } \|\theta\|_c < \infty\}.$$

We impose the following requirements on the degree of smoothness and the regularity of $\mathcal{X}$.

ASSUMPTION 2.1: (i) *For $\mathcal{X}$ bounded, $\delta_0 = \delta = 0$ and $\min\{m_0, m\} > \frac{d_x}{2}$, while for $\mathcal{X}$ unbounded, $m_0 > \frac{d_x}{2}$, $\frac{d_x(m+\delta)}{m\delta} < 2$ and $\delta_0 > \delta > \frac{d_x}{2}$.* (ii) *$\mathcal{X}$ satisfies a uniform cone condition.*

The requirements on the norm parameters m, m_0, δ, and δ_0 are stated in Assumption 2.1(i). Weighting the integrand in the definition of $\|\cdot\|_s$ by $(1 + x'x)^{\delta_0}$ controls the tails of θ and its derivatives, which enables us to let $\mathcal{X}$ be an unbounded subset of $\mathbf{R}^{d_x}$. For bounded supports, however, the tail condition becomes unnecessary, and δ_0 and δ both can be set equal to zero. Assumption 2.1(ii) imposes a weak regularity condition on the shape of $\mathcal{X}$; see Paragraph 4.8 in Adams and Fournier (2003). Heuristically, Assumption 2.1(ii) is satisfied if there exists some small finite cone whose vertex can be placed in each point in the boundary of $\mathcal{X}$ in such a way that the cone is contained in $\mathcal{X}$.

We assume that the true model θ_0 satisfies $\|\theta_0\|_s \le B$ for some known finite constant B. The parameter space Θ is therefore defined to be a ball of radius B in $W^s(\mathcal{X})$ centered at the origin:

$$\Theta \equiv \{\theta \in W^s(\mathcal{X}) : \|\theta\|_s \le B\}. \tag{11}$$

As a subset of $W^s(\mathcal{X})$, the parameter space Θ therefore has a simple geometric interpretation. This offers three distinct advantages: (i) the identification discussion of Example 2.1 translates verbatim to this setting; (ii) Θ has an interior in the topology induced by $\|\cdot\|_s$, which makes the discussion of the local parameter space straightforward; (iii) the constraint $\theta \in \Theta$ is easily enforced as shown in Newey and Powell (2003). Equally important, as shown in Lemma A.2 in the Appendix, Θ is also compact under $\|\cdot\|_c$, a property that plays a crucial role in our asymptotic analysis. While it is possible to directly assume Θ is compact under $\|\cdot\|_c$ (without introducing $\|\cdot\|_s$), we would lose the clean geometric discussion of identification as well as the ease of analysis of the local parameter space (since Θ has no interior under $\|\cdot\|_c$).

REMARK 2.1: The integral equation in (5) is said to be well posed if (i) Υ is bijective and (ii) the inverse of Υ is continuous; see Kress (1999). Under identification, if the domain of Υ is $L^2(X)$, then its inverse is not continuous. As a result, estimation of θ_0 requires the use of regularization methods, such as in Darolles, Florens, and Renault (2003), Hall and Horowitz (2005), and Chen and Pouzo (2012). Alternatively, condition (ii) may be ensured by restricting the domain of Υ to be a compact set, which is the approach pursued in Newey and Powell (2003) and Ai and Chen (2003). In our setting, the problem is ill-posed due to Υ not being bijective and, hence, its inverse not existing. However, the compactness of Θ under $\|\cdot\|_c$ still plays a similar role to that in Newey and Powell (2003), as it implies that the inverse correspondence is upper hemicontinuous.

2.2. *Examples*

The assumption $\theta_0 \in \Theta$ aids the instrument Z in identifying θ_0. Blundell, Chen, and Kristensen (2007), for example, exploited the boundedness of Engel curves to attain nonparametric identification under assumptions weaker than in Newey and Powell (2003). While the restriction that $\theta_0 \in \Theta$ ideally would be enough to attain identification, this is not necessarily the case. This is clear, for example, when X is continuous and Z is discrete or when X and Z are independent.

In fact, elaborating on the geometric intuition of Example 2.1, it is possible to construct a large class of models for which identification fails. If $\|\theta_0\|_s < \infty$, then for $\mathbf{V}_0$ as defined in (6), we obtain

$$\mathbf{V}_0 \cap W^s(\mathcal{X}) = \theta_0 + \mathcal{N}(\Upsilon) \cap W^s(\mathcal{X}), \tag{12}$$

which is a nonsingleton affine subspace of $W^s(\mathcal{X})$ provided $\mathcal{N}(\Upsilon) \cap W^s(\mathcal{X}) \neq \{0\}$. Since Θ is a ball of radius B in $W^s(\mathcal{X})$, this setting is completely analogous to Example 2.1. In particular, the identified set Θ_0 is again given by the intersection of an affine subspace with a ball. Hence, identification fails whenever $\mathcal{N}(\Upsilon) \cap W^s(\mathcal{X}) \neq \{0\}$ and the radius B of Θ is sufficiently large.

Examples 2.2–2.4 illustrate the possible failure of identification by showing $\mathcal{N}(Y) \cap W^s(\mathcal{X}) \neq \{0\}$ for a rich class of distributions of (X, Z). The requirement that the radius of Θ be sufficiently large is in turn satisfied, for example, if the true model θ_0 is an interior point of the parameter space.

EXAMPLE 2.2: Let $X \in \mathcal{X} \subset \mathbf{R}^{d_x}$ and $Z \in \mathcal{Z} \subset \mathbf{R}^{d_z}$ with $\mathcal{X}, \mathcal{Z}$ compact, and assume that the conditional density of X given Z, denoted $f_{X|Z}$, is square integrable. For an appropriate set of basis functions $\{\psi_k(x)\}_{k=1}^{\infty}$ and $\{\phi_j(z)\}_{j=1}^{\infty}$ that are orthogonal in the sense that

$$(13) \qquad \int_{\mathcal{X}} \psi_k(x)\psi_i(x)\, dx = 0 \quad \forall k \neq i, \qquad \int_{\mathcal{Z}} \phi_j(z)\phi_i(z)\, dz = 0 \quad \forall j \neq i,$$

we can expand the conditional density $f_{X|Z}$ under the square error norm as a series of the form

$$(14) \qquad f_{X|Z}(x|z) = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} a_{kj} \psi_k(x) \phi_j(z).$$

For some k_0, if $a_{k_0 j} = 0$ for all j and in addition $\|\psi_{k_0}\|_s < \infty$, then $\psi_{k_0} \in \mathcal{N}(Y) \cap W^s(\mathcal{X})$. From our preceding discussion, it then follows that identification fails if θ_0 is an interior point of Θ. Specifically, the function $\theta_0 + \lambda\psi_{k_0}$ also lies in the identified set Θ_0 for λ sufficiently small.

EXAMPLE 2.3: As a special case of Example 2.2, suppose $(X, Z) \in [-1, 1]^2$ and $f_{X|Z}$ is polynomial of order $J < \infty$. For $\{P_k\}_{k=1}^{\infty}$ the Legendre polynomials,[3] we can then write

$$(15) \qquad f_{X|Z}(x|z) = \sum_{k=0}^{J} \sum_{j=0}^{J} a_{kj} P_k(x) P_j(z).$$

It follows from Example 2.2 that if $\|\theta_0\|_s < B$, then $\theta_0 + \lambda P_k \in \Theta_0$ for λ small enough and $k > J$. In more generality, we conclude that if X and Z have compact support, $f_{X|Z}$ is a polynomial of finite order, and B is sufficiently large, then θ_0 is not identified.

EXAMPLE 2.4: Suppose $Z \in \mathbf{R}^{d_z}$ and that $X \in \mathbf{R}$ is generated according to the model

$$(16) \qquad X = g(Z) + U,$$

[3]The Legendre polynomials are obtained by applying Gram–Schmidt orthogonalization to $\{u^i\}_{i=1}^{\infty}$ on $[-1, 1]$.

where U is independent of Z and has compact support that without loss of generality we assume is contained in $[0, \bar{u}]$. If U has a square integrable density f_U, then it admits for an expansion

$$f_U(u) = \sum_{k=0}^{\infty}\{a_k \sin(k\omega u) + b_k \cos(k\omega u)\}, \tag{17}$$

where $\omega = 2\pi/\bar{u}$. It may be verified by direct calculation that for any integer k,

$$E[\sin(k\omega X)|Z] = \frac{\bar{u}a_k}{2}\cos(k\omega g(Z)) + \frac{\bar{u}b_k}{2}\sin(k\omega g(Z)). \tag{18}$$

Therefore, for some k_0, if $a_{k_0} = b_{k_0} = 0$ and θ_0 is an interior point of Θ, then for λ sufficiently small, the function $\theta_0 + \lambda \sin(k_0\omega\cdot)$ also lies in the identified set Θ_0.

An interesting special case of Example 2.4 occurs when $Z \in \mathbf{R}$ and $g(Z) = \alpha Z$. By letting α be arbitrarily large, the correlation between X and Z can be made arbitrarily close to 1 and yet θ_0 will not be identified. Example 2.4 hence illustrates that the correlation between X and Z is a poor measure of the identifying power of an instrument in a nonparametric setting. It is also important to note that relationship (16) must hold unbeknownst to the econometrician. Otherwise, the triangular structure can be exploited to attain identification (Newey, Powell, and Vella (1999)).

Examples 2.2–2.4 suggest that nonparametric identification of θ_0 may be frail. In particular, Examples 2.2 and 2.3 indicate that the set of distributions for which identification fails is dense, so any distribution of (X, Z) under which θ_0 is identified is arbitrarily close to a distribution under which identification fails. We conclude this discussion by establishing this point. For this purpose, for a compact set $K \subset \mathbf{R}^{d_x+d_z}$, denote the set of continuous densities on it by

$$\mathbf{D}(K) \equiv \Big\{f : K \to \mathbf{R}, f(x, z) \geq 0, \int_K f(x, z)\,dx\,dz = 1, f \text{ is continuous}\Big\}. \tag{19}$$

We further define the subset of $\mathbf{D}(K)$ under which identification fails for some choice of B to be

$$\mathbf{D}_{\emptyset}(K) \equiv \Big\{f \in \mathbf{D}(K) : \exists\theta \neq 0 \text{ with } \int \theta(x)f(x, z)\,dx = 0\ \forall z, \text{ and } \|\theta\|_s < \infty\Big\}. \tag{20}$$

In Lemma 2.1, we show that any density $f \in \mathbf{D}(K)$ can be uniformly well approximated by some sequence of densities $\{f_n\}_{n=1}^{\infty}$ such that $f_n \in \mathbf{D}_{\emptyset}(K)$ for all n.

LEMMA 2.1: *If K is compact, then $\mathbf{D}_{\emptyset}(K)$ is dense in $\mathbf{D}(K)$ with respect to* $\|\cdot\|_{\infty}$.

2.3. *Analysis Without Identification*

As argued, even under the restrictions imposed on Θ, it is still a strong requirement that θ_0 be identified. It is therefore prudent to utilize a testing framework that is robust to a possible lack of identification. Instead of performing hypothesis tests on a function, the test statistic developed in this paper allows us to test restrictions on Θ_0. The kind of hypotheses we consider are of the form

$$H_0 : \Theta_0 \cap R \neq \emptyset, \quad H_1 : \Theta_0 \cap R = \emptyset, \tag{21}$$

where R is a set of functions that satisfy a property we wish to test for. The null hypothesis in (21) is that at least one element of Θ_0 satisfies the restrictions imposed in R. When θ_0 is identified, the null hypothesis and the alternative in (21) simplify to

$$H_0 : \theta_0 \in R, \quad H_1 : \theta_0 \notin R. \tag{22}$$

If nonparametric identification is attained, then the present testing framework reduces to tests on the true model θ_0. There are, however, three important advantages to examining the hypothesis in (21) instead of (22). First, confidence intervals for partially identified functionals that satisfy the coverage requirement in (3) may still prove to be informative under partial identification. Second, rejection of the null hypothesis in (21) is informative about the true model θ_0, as it implies $\theta_0 \notin R$. Third, the inability to reach compelling conclusions using this approach points out the limitations of the instrument to nonparametrically identify parameters of interest.

REMARK 2.2: For $\mathbf{V}_0$ as defined in (6), let $\mathbf{W}_0 \equiv \mathbf{V}_0 \cap W^s(\mathcal{X})$ denote the identified set from assuming $\theta_0 \in W^s(\mathcal{X})$ but not imposing the norm bound $\|\theta_0\|_s \leq B$. Our assumptions imply there exists a θ^* that minimizes $\|\cdot\|_s$ on $\mathbf{W}_0 \cap R$. Therefore, the hypothesis in (21) is equivalent to

$$H_0 : \mathbf{W}_0 \cap R \neq \emptyset, \quad H_1 : \mathbf{W}_0 \cap R = \emptyset \tag{23}$$

if and only if $\|\theta^*\|_s \leq B$. Hence, the norm bound restriction may cause the hypotheses in (21) and (23) to not be equivalent. Whether disagreement of these hypotheses is a concern of course depends on the validity of the assumption that the true model θ_0 satisfies $\|\theta_0\|_s \leq B$.

2.3.1. *Structure of the Set R*

We consider sets R that are characterized by its elements satisfying an equality restrictions on the image of a linear operator with domain $W^c(\mathcal{X})$. Concretely, we require the set R be of the form

$$R \equiv \{\theta \in W^c(\mathcal{X}) : L(\theta) = l\}, \tag{24}$$

where l is an element of the range of L that is required to satisfy the following assumption.

ASSUMPTION 2.2: *For $(\mathcal{L}, \|\cdot\|_{\mathcal{L}})$ a Banach space, $L : W^c(\mathcal{X}) \to \mathcal{L}$ is continuous and linear.*

Since Assumption 2.2 imposes that L be linear, it would not be without loss of generality to replace l with 0 in (24). The restriction that L be linear and continuous is balanced by the flexibility in choosing the range space $(\mathcal{L}, \|\cdot\|_{\mathcal{L}})$ and the strength of the norm $\|\cdot\|_c$, which makes continuity easy to verify. In what follows, examples are provided that illustrate the wide array of hypotheses that can be examined through sets R as in (24).

2.3.2. *Confidence Regions for Identifiable Functionals*

In many instances, we are not interested in the function θ_0 itself, but rather in a functional of it. The identified set for a functional $f : \Theta \to \mathbf{R}^{d_f}$ is given by

$$\mathcal{F}_0 \equiv \{f(\theta) : \theta \in \Theta_0\}. \tag{25}$$

Using a family of hypotheses as in (21), it is possible to construct confidence regions that satisfy the coverage requirement in (3). This is a weaker coverage requirement than building a confidence region for the whole set $\mathcal{F}_0$, as discussed in Chernozhukov, Hong, and Tamer (2007) and Romano and Shaikh (2010). To construct the desired confidence region, we consider the null hypotheses

$$H_0(\gamma) : \Theta_0 \cap R(\gamma) \neq \emptyset, \quad R(\gamma) = \{\theta \in W^c(\mathcal{X}) : f(\theta) = \gamma\}. \tag{26}$$

In Section 3 we show how to construct tests for (26) that control the probability of a Type I error. Therefore, if $C_n(1-\alpha)$ is the set of γ such that $H_0(\gamma)$ is not rejected, then $C_n(1-\alpha)$ has the desired coverage requirement in (3) by the duality of confidence intervals and hypothesis testing.

Examples 2.5 and 2.6 illustrate functionals f such that $R(\gamma)$ in (26) can be expressed as in (24).

EXAMPLE 2.5—Function at a Point: Suppose the object of interest is the value of the function θ_0 at a point x_0. Letting $(\mathcal{L}, \|\cdot\|_{\mathcal{L}}) = (\mathbf{R}, \|\cdot\|)$ and $L(\theta) = \theta(x_0)$, it follows that the set

$$R(\gamma) = \{\theta \in W^c(\mathcal{X}) : \theta(x_0) = \gamma\} \tag{27}$$

can be expressed as in (24). The same argument also applies to hypothesis testing on other linear functionals, such as consumer surplus and derivatives up to order m.

EXAMPLE 2.6—Price Elasticity: In Example 2.5, the functional of interest is itself linear. To illustrate that this is not a necessary condition for expressing $R(\gamma)$ as in (24), consider constructing a confidence interval for price elasticity of demand. For θ a demand function, let $\theta(p, x)$ denote quantity demanded under price p and consumer covariates x. We can then define

$$R(\gamma) = \left\{\theta \in W^c(\mathcal{X}) : -p_0 \frac{\partial\theta(p_0, x_0)}{\partial p} \frac{1}{\theta(p_0, x_0)} = \gamma\right\} \tag{28}$$

for some point (p_0, x_0). For $(\mathcal{L}, \|\cdot\|_{\mathcal{L}}) = (\mathbf{R}, \|\cdot\|)$ and the functional $L(\theta) = -p_0 \frac{\partial\theta(p_0,x_0)}{\partial p} - \gamma\theta(p_0, x_0)$, which is linear and continuous under $\|\cdot\|_c$, the set $R(\gamma)$ in (28) is then equivalent to

$$R(\gamma) = \{\theta \in W^c(\mathcal{X}) : L(\theta) = 0\}. \tag{29}$$

Therefore, the restriction set $R(\gamma)$ can be expressed in the form of (24) as desired.

REMARK 2.3: The norm bound $\|\theta_0\|_s \leq B$ can play an important role in determining $\mathcal{F}_0$. For instance, setting B to be infinite in Example 2.5 implies that $\mathcal{F}_0$ equals either a singleton or the entire real line. Hence, to the extent that the assumption $\|\theta_0\|_s \leq B$ is correct, this additional information is instrumental in obtaining informative bounds under partial identification.

2.3.3. *Tests of Global Restrictions*

While assuming linearity of the operator L is restrictive, this is compensated by the flexibility in selecting the range space. Examples 2.7 and 2.8 illustrate this point by showing how restrictions of the form $L(\theta) = l$ can be used to test for global properties of θ_0.

EXAMPLE 2.7—Homogeneity: Consider the problem of testing whether Θ_0 contains a production function θ that is homogenous of degree α; that is, whether $\theta(\lambda k, \lambda l) = \lambda^\alpha \theta(k, l)$ for some $\theta \in \Theta_0$. By Euler's theorem, we can characterize homogeneity as the set of functions that satisfy

$$k\frac{\partial\theta(k, l)}{\partial k} + l\frac{\partial\theta(k, l)}{\partial l} = \alpha\theta(k, l). \tag{30}$$

Suppose $\mathcal{X}$ is compact, and let $(\mathcal{L}, \|\cdot\|_{\mathcal{L}}) = (L^\infty(X), \|\cdot\|_\infty)$ and $L(\theta) =$

$k\frac{\partial\theta(k,l)}{\partial k}+l\frac{\partial\theta(k,l)}{\partial l}-\alpha\theta(k,l)$. We can then test the desired hypothesis by letting $R=\{\theta\in W^c(\mathcal{X}):L(\theta)=0\}$.

EXAMPLE 2.8—Parametric Specification: Let $\mathcal{X}$ be compact and let $\{\psi_j\}_{j=1}^K\in W^c(\mathcal{X})$ with K finite. Suppose we are interested in testing whether Θ_0 intersects with the parametric family

$$\mathcal{P}\equiv\left\{\theta\in W^c(\mathcal{X}):\theta(x)=\sum_{j=1}^K\beta_j\psi_j(x)\right\}. \tag{31}$$

Let $(\mathcal{L},\|\cdot\|_{\mathcal{L}})=(L^2(X),\|\cdot\|_{L^2})$ and let $P_{\mathcal{P}}(\theta)$ denote the projection of $\theta\in L^2(X)$ onto $\mathcal{P}$ (under $\|\cdot\|_{L^2}$). Since $\mathcal{P}$ is a vector subspace of $L^2(X)$, it follows that $P_{\mathcal{P}}$ is a continuous linear operator and, hence, so is $L:W^c(\mathcal{X})\to L^2(X)$ pointwise defined by $L(\theta)=\theta-P_{\mathcal{P}}(\theta)$. We can then conduct the specification test by letting $R=\{\theta\in W^c(\mathcal{X}):L(\theta)=0\}$. Notice that the same argument implies we can also test for a partially linear specification of the form $\theta(x_1,x_2)=x_1\beta+\psi(x_2)$.

3. TEST STATISTIC

The null hypothesis in (21) states the existence of a $\theta\in\Theta\cap R$ whose implied residual is mean independent of the instrument. As a general principle, we can therefore construct a test of this null hypothesis by examining whether there is any $\theta\in\Theta\cap R$ such that the conditional expectation of its implied residual, given the instrument, has norm zero. Different choices of norms, however, lead to alternative test statistics with varying asymptotic properties. Options include $\|\cdot\|_{L^2}$ and its variants (Hall (1984), Horowitz (2006)) as well as norms based on unconditional moments (Newey (1985), Bierens (1990)). We opt in this paper to follow the latter literature, which is closely linked to the use of series methods in estimation. We leave the development of alternative test statistics based on other norms such as Hall (1984) or Horowitz (2006) to future work.

As in Bierens (1990), we characterize the conditional moment restriction $E[\varepsilon|Z]=0$ as an infinite number of appropriate unconditional moment restrictions. In particular, for $\mathcal{Z}$ the support of Z and $T\subset\mathbf{R}^{d_t}$ a known compact set, we require a weight function $w:T\times\mathcal{Z}\to\mathbf{R}$ such that for all random variables $V\in\mathbf{R}$ with $E[|V|]<\infty$, we have

$$E[V|Z]=0\quad\text{if and only if}\quad E[Vw(t,Z)]=0\quad\forall t\in T. \tag{32}$$

For example, Bierens (1990) showed that if Z is bounded almost surely, then $w(t,z)=e^{t'z}$ satisfies (32) for any set T with positive Lebesgue measure. Stinchcombe and White (1998) provided numerous other examples of valid choices for functions w and sets T.

We impose the following assumption on the weight function:

ASSUMPTION 3.1: (i) $w : T \times \mathcal{Z} \to \mathbf{R}$ *is uniformly continuous, is bounded, and satisfies* (32) *for* $T \subset \mathbf{R}^{d_t}$ *compact*. (ii) *For some* $M > 0$, $|w(t_1, z) - w(t_2, z)| \leq M\|t_1 - t_2\|$ *for all* $z \in \mathcal{Z}$.

Assumption 3.1 is easily satisfied by a wide number of choices for w. For example, let $\mathcal{Z} \subset \mathbf{R}^{d_z}$ and let $\Psi : \mathcal{Z} \to \mathbf{R}^{d_z}$ be a bounded Borel measurable injective function. It then follows by Corollary 3.9 in Stinchcombe and White (1998) that for $t \equiv (t_0, t_1) \in \mathbf{R} \times \mathbf{R}^{d_z}$, and $g : \mathbf{R} \to \mathbf{R}$ analytic and nonpolynomial, the choice $w(t, z) = g(t_0 + t_1'\Psi(z))$ satisfies (32) for any set $T \subset \mathbf{R}^{d_z+1}$ with positive Lebesgue measure. Additional possible choices for g include the normal cumulative distribution function (c.d.f.) or probability density function; see also Theorem 1 in Bierens and Ploberger (1997) and Theorem 3.10 in Stinchcombe and White (1998).

Lemma 3.1 shows that Assumption 3.1 indeed yields a simple characterization of the identified set Θ_0 and, hence, also of the null hypothesis in (21).

LEMMA 3.1: *If* $E[Y^2] < \infty$ *and Assumptions* 2.1(i) *and* (ii) *and* 3.1(i) *and* (ii) *hold, then it follows that*

$$\theta \in \Theta_0 \quad \text{iff} \quad \max_{t \in T} \big(E\big[(Y - \theta(X))w(t, Z)\big]\big)^2 = 0.$$

Furthermore, if Assumption 2.2 *also holds, then we can conclude*

$$\Theta_0 \cap R \neq \emptyset \quad \text{iff} \quad \min_{\theta \in \Theta \cap R} \max_{t \in T} \big(E\big[(Y - \theta(X))w(t, Z)\big]\big)^2 = 0.$$

Since $\Theta_0 \subseteq \Theta$, the null hypothesis in (21) is equivalent to stating there is a function $\theta \in \Theta \cap R$ that is also in Θ_0. Given the characterization of such $\theta \in \Theta_0$ obtained in the first claim of Lemma 3.1, the second claim then follows by establishing the compactness of $\Theta \cap R$ under $\|\cdot\|_c$. We conclude that under the assumptions of Lemma 3.1, we can restate the null and alternative hypotheses as

$$H_0 : \min_{\theta \in \Theta \cap R} \max_{t \in T} \big(E\big[(Y - \theta(X))w(t, Z)\big]\big)^2 = 0,$$

$$H_1 : \min_{\theta \in \Theta \cap R} \max_{t \in T} \big(E\big[(Y - \theta(X))w(t, Z)\big]\big)^2 > 0.$$

This equivalent representation of the null and alternative hypotheses is analogous in spirit to the Anderson and Rubin (1949) test. A similar type of equivalence was also employed in Romano and Shaikh (2008) to derive confidence intervals for functions of the identifiable parameters.

The obtained alternative representation of the hypotheses suggests employing the test statistic

$$I_n(R) \equiv \min_{\theta \in \Theta_n \cap R} \max_{t \in T_n} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (y_i - \theta(x_i)) w(t, z_i) \right)^2, \tag{33}$$

where $\{\Theta_n \cap R\}_{n=1}^{\infty}$ and $\{T_n\}_{n=1}^{\infty}$ are sieves that grow to be dense in $\Theta \cap R$ and T, respectively.

REMARK 3.1: As in Chen and Pouzo (2012), adding a strictly convex lower-semicompact penalty to the objective function in (33) would enable us to dispense with the requirement that $\|\theta_0\|_s \le B$. As argued in Chen and Pouzo (2012), such modification can yield a consistent estimator for the unique minimizer of the penalty on the set $\mathbf{V}_0 \cap W^s(\mathcal{X}) \cap R$. While we consider such an extension important, for ease of exposition, we do not pursue it at present.

REMARK 3.2: In models where the parameter space Θ is parametric and the identified set Θ_0 is finite dimensional, Chernozhukov, Hong, and Tamer (2007) and Romano and Shaikh (2010) provided methods for constructing confidence sets $\hat{\mathcal{C}}_n$ that satisfy the coverage requirement

$$\liminf_{n \to \infty} P(\Theta_0 \subseteq \hat{\mathcal{C}}_n) \ge 1 - \alpha. \tag{34}$$

In this setting, a test that rejects whenever $\hat{\mathcal{C}}_n \cap R = \emptyset$ controls size, since if $\Theta_0 \cap R \neq \emptyset$, then

$$\liminf_{n \to \infty} P(\hat{\mathcal{C}}_n \cap R \neq \emptyset) \ge \liminf_{n \to \infty} P(\Theta_0 \subseteq \hat{\mathcal{C}}_n) \ge 1 - \alpha. \tag{35}$$

The generalization of this approach to our present nonparametric context, however, presents important challenges. In particular, a set $\hat{\mathcal{C}}_n$ that satisfies (34) must be itself nonparametric and can, therefore, not be a subset of a finite-dimensional sieve. In contrast, we note that the use of sieves in (33) does not present similar problems for our approach.

3.1. *Definitions and Notation*

In the definition of $I_n(R)$, the minimum is computed not over the whole parameter space $\Theta \cap R$, but over an approximating sieve $\{\Theta_n \cap R\}_{n=1}^{\infty}$ that we now introduce. Let $\{p_j\}_{j=1}^{\infty}$ be a set of known basis functions with $p_j \in W^s(\mathcal{X})$ for all j and denote $p^{k_n}(x) = (p_1(x), \ldots, p_{k_n}(x))'$. The functions $\{p_j\}_{j=1}^{k_n}$ then span the vector subspace of $W^s(\mathcal{X})$,

$$W_n^s(\mathcal{X}) \equiv \{\theta \in W^s(\mathcal{X}) : \theta(x) = p^{k_n'}(x) h \text{ for some } h \in \mathbf{R}^{k_n}\}. \tag{36}$$

Recalling the definition of Θ in (11) and of R in (24), a natural sieve $\{\Theta_n \cap R\}_{n=1}^{\infty}$ is then given by

$$\Theta_n \cap R \equiv \{\theta \in W_n^s(\mathcal{X}) : \|\theta\|_s \leq B, L(\theta) = l\}. \tag{37}$$

As pointed out in Newey and Powell (2003), imposing $\|\theta\|_s \leq B$ is straightforward in $W_n^s(\mathcal{X})$, since

$$\|p^{k_n'} h\|_s^2 = h' \Lambda_n h, \quad \text{where} \tag{38}$$
$$\Lambda_n \equiv \sum_{|\lambda| \leq m+m_0} \int \left[D^\lambda p^{k_n}(x) D^\lambda p^{k_n'}(x)\right](1 + x'x)^{\delta_0}\, dx.$$

To derive the asymptotic distribution of $I_n(R)$, it is necessary to study the local parameter space to each $\theta \in \Theta_0 \cap R$. The local parameters are those functions of the form

$$\theta_n = \Pi_n \theta + \frac{p^{k_n'} h}{\sqrt{n}}, \tag{39}$$

where $\Pi_n \theta$ is a projection of $\theta \in \Theta_0 \cap R$ onto $\Theta_n \cap R$ and $h \in \mathbf{R}^{k_n}$ is such that $\theta_n \in \Theta_n \cap R$. The set of $h \in \mathbf{R}^{k_n}$ that indexes local functions as in (39) is contained in the set

$$\mathcal{H}_{k_n} \equiv \left\{h \in \mathbf{R}^{k_n} : L\left(p^{k_n'} h\right) = 0\right\}. \tag{40}$$

We note that since the functional L is linear, $\mathcal{H}_{k_n}$ is itself a vector subspace of $\mathbf{R}^{k_n}$.

Each $h \in \mathcal{H}_{k_n}$ induces a function $v \in L^\infty(T)$ via the mapping $h \mapsto E[w(\cdot, Z) p^{k_n'}(X) h]$. The resulting subspaces, and their closure, play an important role in the asymptotic distribution of $I_n(R)$. For this reason, we first define the class of functions

$$V_{k_n}(T) \equiv \left\{v : T \to \mathbf{R} \text{ s.t. } v(t) = E\left[w(t, Z) p^{k_n'}(X) h\right] \text{ for some } h \in \mathcal{H}_{k_n}\right\}, \tag{41}$$

and note that since $w(t, z)$ is uniformly bounded and $p^{k_n} h \in W^c(\mathcal{X})$ for all h, it indeed follows that $V_{k_n}(T) \subset L^\infty(T)$. We can then define $V_\infty(T)$ to be the closure of $\bigcup V_{k_n}(T)$ under $\|\cdot\|_\infty$:

$$V_\infty(T) \equiv \left\{v \in L^\infty(T) : \exists v_{k_n} \in V_{k_n}(T) \text{ with } \left\|v_{k_n} - v\right\|_\infty = o(1)\right\}. \tag{42}$$

The properties of $V_\infty(T)$ are determined by the distribution of (X, Z) as well as by the functional L. For example, while in most applications we expect $V_\infty(T)$ to be infinite dimensional, it may have finite dimension for particular distributions of (X, Z), as when $X \perp Z$, or certain choices of L, as in parametric specification tests (Example 2.8).

We conclude by introducing a pseudometric $\|\cdot\|_w$ which is induced by our choice of criterion function. As shown in previous work (see, for example, Ai and Chen (2003) and Chen and Pouzo (2012)), a crucial role is played in nonparametric IV problems by the properties of such a pseudonorm. Specifically, in the present context, the pseudometric $\|\cdot\|_w$ is given by

$$\|\theta\|_w \equiv \max_{t\in T}\bigl|E[w(t, Z)\theta(X)]\bigr|. \tag{43}$$

Since $w(t, z)$ is uniformly bounded, Jensen's inequality implies $\|\cdot\|_w \lesssim \|\cdot\|_{L^2}$ and similarly that

$$\|\theta\|_w \lesssim \|\theta\|_{wo}, \quad \|\theta\|_{wo} \equiv \sqrt{E\bigl[\bigl(E[\theta(X)|Z]\bigr)^2\bigr]}, \tag{44}$$

which is of particular interest as $\|\cdot\|_{wo}$ is the standard weak norm considered in the literature (Ai and Chen (2003)). As a result, the $\|\cdot\|_w$ approximation error introduced from employing a sieve is no larger than the corresponding error under either $\|\cdot\|_{L^2}$ or $\|\cdot\|_{wo}$.

3.2. *Asymptotic Distribution*

We introduce the following assumptions so as to obtain the asymptotic distribution of $I_n(R)$.

ASSUMPTION 3.2: (i) $\{y_i, x_i, z_i\}_{i=1}^n$ *is independent and identically distributed* (*i.i.d.*) *from* (1) *with* $Y \in \mathbf{R}$, $X \in \mathcal{X} \subseteq \mathbf{R}^{d_x}$, $Z \in \mathcal{Z} \subseteq \mathbf{R}^{d_z}$, *and* $\theta_0 \in \Theta$. (ii) X *has bounded density and* Y *satisfies* $E[Y^2] < \infty$.

ASSUMPTION 3.3: (i) *The eigenvalues of* $E[p^{k_n}(X)p^{k'_n}(X)]$ *are uniformly bounded*. (ii) $\{T_n\}_{n=1}^{\infty} \subseteq T$ *are closed and there is a* $\Pi_n t \in T_n$ *such that* $\sup_{t\in T}\|t - \Pi_n t\| = o(n^{-1/2})$.

ASSUMPTION 3.4: *There is* $\Pi_n\theta \in \Theta_n \cap R$ *that satisfies*: (i) $\sup_{\theta\in\Theta\cap R}\|\theta - \Pi_n\theta\|_{L^2} = o(1)$, (ii) $\sup_{\theta\in\Theta_0\cap R}\|\theta - \Pi_n\theta\|_w = o(n^{-1/2})$, *and* (iii) $\sup_{\theta\in\Theta_0\cap R}\|\Pi_n \times \theta\|_s \le B - \gamma_n$ *for some* $\gamma_n \downarrow 0$ *with* $n^{1/2} \times \gamma_n \uparrow \infty$.

In Assumption 3.2(i), the requirement $\Theta_0 \neq \emptyset$ can be tested by employing the test statistic $I_n(R)$ with $R = \Theta$. Horowitz (2008) also provided a procedure for testing this hypothesis. Assumption 3.3(i) ensures that $\|p^{k'_n}h\|_{L^2} \lesssim \|h\|$ uniformly in n; see Newey (1997) and Huang (1998, 2003). The approximating requirements on the sieve $\{T_n\}_{n=1}^{\infty}$, which may be set equal to T, are stated in Assumption 3.3(ii). Under Assumption 3.4(i), $\{\Theta_n \cap R\}_{n=1}^{\infty}$ must approximate $\Theta \cap R$ uniformly well under $\|\cdot\|_{L^2}$; see Chen (2007). Assumption 3.4(ii) and (iii) needs only to hold under the null hypothesis. Assumption 3.4(ii) controls the bias under $\|\cdot\|_w$ from using a sieve, which is required to decay at a sufficiently fast rate so that $I_n(R)$ does not diverge to infinity when the null hypothesis is true. Since $\|\cdot\|_w \lesssim \|\cdot\|_{wo} \lesssim \|\cdot\|_{L^2}$, Assumption 3.4(ii) can be verified using results for $\|\cdot\|_{L^2}$ or $\|\cdot\|_{wo}$ approximations; see Newey (1997), Chen (2007), and Chen and Pouzo (2012). Finally, Assumption 3.4(iii) ensures that the local parameter space is the same for elements that are on the boundary and those that are in the interior.[4] It is important to emphasize that Assumption 3.4(iii) does not require that all $\theta \in \Theta_0 \cap R$ satisfy $\|\theta\|_s < B$, but rather that they can be well approximated from the interior of Θ. In Section 3.2.1, we discuss how Assumption 3.4 can be verified and note that, in some instances, we can set $\Theta_n \cap R = \Theta \cap R$ for all n.

Assumptions 2.1, 2.2, and 3.1–3.4 are sufficient to establish the asymptotic behavior of $I_n(R)$.

THEOREM 3.1: *Let Assumptions 2.1, 2.2, 3.1, 3.2, 3.3, and 3.4 hold. If $\Theta_0 \cap R \neq \emptyset$, then*

$$I_n(R) \xrightarrow{\mathcal{L}} \inf_{\theta \in \Theta_0 \cap R, v \in V_\infty(T)} \|G(\cdot, \theta) - v\|_\infty^2,$$

where G is a tight Gaussian process on $L^\infty(T \times \Theta_0)$. Moreover, if $\Theta_0 \cap R = \emptyset$, then:

$$n^{-1} I_n(R) \xrightarrow{a.s.} \min_{\theta \in \Theta \cap R} \big\| E\big[(Y - \theta(X))w(\cdot, Z)\big]\big\|_\infty^2.$$

The test statistic $I_n(R)$ can be interpreted as an infinite-dimensional version of a generalized method of moments (GMM) overidentification test. Hansen (1982) established that the minimum of the GMM objective function converges to a $\mathcal{X}^2_{r-k}$, where r is the number of restrictions and k is the number of parameters. Theorem 3.1 offers an interesting parallel to this result. While in Hansen (1982), the criterion evaluated at the true model converges to the squared norm of a multivariate normal, in our setting, the convergence is to the squared supremum norm of a Gaussian process. The $r - k$ degrees of freedom adjustment from Hansen (1982) is then reflected in the asymptotic distribution of $I_n(R)$ being the squared norm of the approximation error of G by the function space $V_\infty(T)$.

[4]In the local specification $\Pi_n \theta + p^{k'_n} h/\sqrt{n}$, Assumptions 3.4(iii) ensures that h is asymptotically unconstrained by the restrictions $\|\Pi_n \theta + p^{k'_n} h/\sqrt{n}\|_s \leq B$ even if $\theta \in \Theta \cap R$ is such that $\|\theta\|_s = B$.

REMARK 3.3: For $c_{1-\alpha}$ the $1-\alpha$ quantile of the random variable $\inf_{\theta\in\Theta_0\cap R}\|G(\cdot,\theta)\|_\infty^2$, it is possible to show that for any set R that is closed under $\|\cdot\|_c$,

$$\limsup_{n\to\infty} P(I_n(R)\geq c_{1-\alpha})\leq\alpha. \tag{45}$$

Such a result can be employed to conduct hypothesis testing for a wider class of sets R than allowed for in (24). However, as Theorem 3.1 implies, such a procedure may prove conservative.

3.2.1. *Sufficient Conditions*

In this section we discuss sufficient conditions for $\{\Theta_n\cap R\}_{n=1}^\infty$ to satisfy Assumption 3.4. Given the strict convexity of $\Theta_0\cap R$, we only need to consider two cases under the null hypothesis: (i) there exists some $\theta^*\in\Theta_0\cap R$ such that $\|\theta^*\|_s<B$ or (ii) $\Theta_0\cap R=\{\theta^*\}$ for some θ^* with $\|\theta^*\|_s=B$.[5]

The case where $\|\theta^*\|_s<B$ for some $\theta^*\in\Theta_0\cap R$ is of primary interest because it encompasses all settings in which $\Theta_0\cap R$ is not a singleton as well as situations where θ_0 is identified and an interior point of Θ. Our first example shows that in this important context, all that is needed is that the approximation error of the sieve under $\|\cdot\|_s$ tends to zero and that the sieve grows sufficiently fast.

EXAMPLE 3.1—Interior Point: Suppose there exists some $\theta^*\in\Theta_0\cap R$ such that $\|\theta^*\|_s<B$ and that $\sup_{\theta\in\Theta\cap R}\|\theta-\tilde{\Pi}_n\theta\|_s=o(1)$ for some $\tilde{\Pi}_n\theta\in\Theta_n\cap R$. Since $\|\cdot\|_w\lesssim\|\cdot\|_s$, by letting $k_n\uparrow\infty$ sufficiently fast, we can ensure that $\sup_{\theta\in\Theta_0\cap R}\|\theta-\tilde{\Pi}_n\theta\|_w=o(n^{-1/2})$. For any $\theta\in\Theta_0\cap R$, define

$$\Pi_n\theta\equiv\lambda_n\tilde{\Pi}_n\theta^*+(1-\lambda_n)\tilde{\Pi}_n\theta \tag{46}$$

for some $0\leq\lambda_n\leq 1$. Notice that $\Pi_n\theta$ belongs to $\Theta_n\cap R$ due to convexity. Assumptions 3.4(i) and (ii) then hold provided $\lambda_n\downarrow 0$, since $\sup_{\theta\in\Theta\cap R}\|\theta-\Pi_n\theta\|_{L^2}=o(1)$ and $\sup_{\theta\in\Theta_0\cap R}\|\theta-\Pi_n\theta\|_w=o(n^{-1/2})$ due to $\Theta_0\cap R$ being an equivalence class under $\|\cdot\|_w$. In addition, it also follows that

$$\sup_{\theta\in\Theta_0\cap R}\|\Pi_n\theta\|_s\leq\lambda_n\|\tilde{\Pi}_n\theta^*\|_s+(1-\lambda_n)B, \tag{47}$$

and, hence, setting $\lambda_n\geq\gamma_n/(B-\|\tilde{\Pi}_n\theta^*\|_s)$ verifies that Assumption 3.4(iii) is satisfied. Such a choice of λ_n is compatible with $\lambda_n\downarrow 0$ for any $\gamma_n\downarrow 0$ due to $\|\theta^*-\tilde{\Pi}_n\theta^*\|_s=o(1)$ and $\|\theta^*\|_s<B$.

[5]Suppose $\theta_1,\theta_2\in\Theta_0\cap R$ with $\|\theta_1\|_s=\|\theta_2\|_s=B$ and $\theta_1\neq\theta_2$. Defining $\theta_\lambda=\lambda\theta_1+(1-\lambda)\theta_2$ for $0<\lambda<1$, it then follows that $\theta_\lambda\in\Theta_0\cap R$ and, by Lemma 3.1 in Luenberger (1969), that $\|\theta_\lambda\|_s<B$ unless $\theta_1=\pm\theta_2$. However, $\theta_1=\theta_2$ is ruled out and $\theta_1=-\theta_2$ implies $0\in\Theta_0\cap R$, which trivially satisfies $\|0\|_s<B$.

Heuristically, Example 3.1 relies on Θ_0 being an equivalence class under $\|\cdot\|_w$ to show that $\tilde{\Pi}_n\theta$ may be "shrunken" toward the interior of $\Theta_n \cap R$, which affects the approximation rate under $\|\cdot\|_s$ but not under $\|\cdot\|_w$. In this manner, we can ensure that Assumption 3.4(iii) is satisfied without affecting the sieve approximation error under $\|\cdot\|_w$. Remarkably, this implies that if there indeed exists a $\theta^* \in \Theta_0 \cap R$ satisfying $\|\theta^*\|_s < B$, then Assumption 3.4 imposes no upper bound on the rate of growth of the sieve. The only constraint on k_n is that it grows sufficiently fast for Assumption 3.4(ii) to hold. In this case, the role of $\{\Theta_n \cap R\}_{n=1}^\infty$ is as a computational aid and k_n may diverge to infinity arbitrarily fast. In particular, Assumption 3.4 is satisfied when $\Theta_n \cap R = \Theta \cap R$ for all n.

In situations where $\Theta_0 \cap R \neq \emptyset$ and there is no $\theta \in \Theta_0 \cap R$ such that $\|\theta\|_s < B$, the strict convexity of Θ implies that $\Theta_0 \cap R = \{\theta^*\}$ for some θ^* satisfying $\|\theta^*\|_s = B$. When $\Pi_n\theta^*$ is the projection (under $\|\cdot\|_s$) of θ^* into a linear subspace of $W^s(\mathcal{X})$, we obtain

$$\|\Pi_n\theta^*\|_s^2 = \|\theta^*\|_s^2 - \|\theta^* - \Pi_n\theta^*\|_s^2 = B^2 - \|\theta^* - \Pi_n\theta^*\|_s^2. \tag{48}$$

If θ^* does not belong to the approximating space, then $\|\Pi_n\theta^*\|_s < B$ and, hence, $\Pi_n\theta^*$ approaches θ^* from the interior of Θ. Moreover, from (48), we also obtain that γ_n in Assumption 3.4(iii) can be set with $\gamma_n \asymp \|\theta^* - \Pi_n\theta^*\|_s^2$. Hence, provided $\|\theta^* - \Pi_n\theta^*\|_w$ tends to zero sufficiently fast relative to $\|\theta^* - \Pi_n\theta^*\|_s$, we can select k_n so that, as required by Assumption 3.4,

$$\sqrt{n}\|\theta^* - \Pi_n\theta^*\|_w \to 0, \quad \sqrt{n}\gamma_n \asymp \sqrt{n}\|\theta^* - \Pi_n\theta^*\|_s^2 \to \infty. \tag{49}$$

In Example 3.2, we employ a particular choice of sieve to formalize this intuition and to show that Assumption 3.4 can still be satisfied when $\Theta_0 \cap R = \{\theta^*\}$ for some θ^* such that $\|\theta^*\|_s = B$.

EXAMPLE 3.2—Boundary Point: Suppose Example 3.1 does not apply, so that $\Theta_0 \cap R = \{\theta^*\}$ for some $\theta^* \in \Theta$ with $\|\theta^*\|_s = B$. Let $\langle\cdot,\cdot\rangle_s$ denote the inner product in $W^s(\mathcal{X})$, let $\mathcal{N}(L)$ be the null space of $L : W^s(\mathcal{X}) \to \mathcal{L}$, and let $\overline{\mathcal{N}(L)}$ denote its closure under $\|\cdot\|_{L^2}$. The identity operator

$$I : W^s(\mathcal{X}) \cap \mathcal{N}(L) \to L^2(X) \cap \overline{\mathcal{N}(L)} \tag{50}$$

is then compact due to $W^s(\mathcal{X})$ being compactly embedded in $L^2(X)$ (see Lemma A.2 in the Appendix). For I^* the adjoint of I, the eigenfunctions of I^*I then yield an orthonormal basis $\{\varphi_j\}_{j=1}^\infty$ for $W^s(\mathcal{X}) \cap \mathcal{N}(L)$ which is also orthogonal in $L^2(X)$. In addition, $\{\varphi_j\}_{j=1}^\infty$ also satisfies

$$\|\varphi_j\|_{L^2}^2 = \lambda_j^2, \tag{51}$$

where $\lambda_1^2 \geq \lambda_2^2 \geq \cdots > 0$ are the ordered eigenvalues of I^*I. Moreover, for $\mathcal{N}^\perp(L)$ the orthogonal complement of $\mathcal{N}(L)$ (in $W^s(\mathcal{X})$), there exists a unique

$\theta_l \in \mathcal{N}^{\perp}(L)$ such that $L(\theta_l) = l$. Hence, letting $\varphi_0 = \theta_l / \|\theta_l\|_s$, $\{p_j\}_{j=0}^{\infty} = \{\varphi_0\} \cup \{\varphi_j\}_{j=1}^{\infty}$, and $\Pi_n \theta = \sum_{j=0}^{k_n} \langle \theta, p_j \rangle_s p_j$, it is possible to verify by direct calculation that $\Pi_n \theta \in \Theta_n \cap R$ provided $\theta \in \Theta \cap R$.[6] In addition, since

$$\|\Pi_n \theta^*\|_s^2 = B^2 - \sum_{j=k_n+1}^{\infty} \langle \theta^*, \varphi_j \rangle_s^2, \quad \|\theta^* - \Pi_n \theta^*\|_{L^2}^2 = \sum_{j=k_n+1}^{\infty} \lambda_j^2 \langle \theta^*, \varphi_j \rangle_s^2, \tag{52}$$

it is possible to satisfy Assumption 3.4 for an appropriate choice of k_n provided $\lambda_j^2 \downarrow 0$ sufficiently fast. For example, Lemma A.10 establishes that $\lambda_j^2 \lesssim j^{-2(m_0+m)/d_x}$ when $\mathcal{X}$ is bounded. Therefore, if for some $1 < \bar{\nu} \leq \underline{\nu}$, we have $j^{-\underline{\nu}} \lesssim \langle \theta^*, \varphi_j \rangle_s^2 \lesssim j^{-\bar{\nu}}$ and there is a β satisfying $\frac{d_x}{2(m+m_0)+d_x(\bar{\nu}-1)} < \beta < \frac{1}{2(\underline{\nu}-1)}$, then setting $k_n \asymp n^{\beta}$ yields $\|\theta^* - \Pi_n \theta^*\|_{L^2} = o(n^{-1/2})$, verifying Assumption 3.4(i) and (ii), while Assumption 3.4(iii) holds with $\gamma_n \asymp \|\theta^* - \Pi_n \theta^*\|_s^2 \gtrsim k_n^{-(\underline{\nu}-1)}$.

3.3. *Bootstrap Critical Values*

The asymptotic results of Theorem 3.1 cannot be implemented for inference without a consistent estimator for the appropriate critical values. Toward this end, in this section we develop a bootstrap procedure and establish its almost sure consistency. For notational convenience, we denote

$$U(t, \theta) \equiv (Y - \theta(X)) w(t, Z), \tag{53}$$

and let $\mathcal{L}^*$, p^*, and E^* denote law, probability, and expectation statements evaluated under the empirical measures. Similarly, we also define $\{u_i^*(t, \theta)\}_{i=1}^n$ to be an i.i.d. sample of $U(t, \theta)$ distributed according to the empirical measure.

By Theorem 3.1, the limiting law of $I_n(R)$ under the null hypothesis is given by

$$\inf_{\theta \in \Theta_0 \cap R, v \in V_\infty(T)} \|G(\cdot, \theta) - v\|_\infty^2. \tag{54}$$

Obtaining a consistent bootstrap therefore requires us to (i) find an appropriate estimator for $V_\infty(T)$ and (ii) estimate the law of the Gaussian process G on $L^\infty(T \times \Theta_0)$. For the first goal, we define

$$\mathcal{H}_{b_n}^{B_n} \equiv \left\{ h \in \mathcal{H}_{b_n} : \|h\| \leq B_n \right\} \tag{55}$$

for some sequences $b_n \uparrow \infty$, $B_n \uparrow \infty$, and $\mathcal{H}_{b_n}$ as in (40). The norm constraint $\|h\| \leq B_n$ allows us to establish a uniform strong law of large numbers that

[6] If $l = 0$, then $\theta_l = 0$ and we can set $\{p_j\}_{j=1}^{\infty} = \{\varphi_j\}_{j=1}^{\infty}$ rather than $\{p_j\}_{j=0}^{\infty} = \{\varphi_0\} \cup \{\varphi_j\}_{j=1}^{\infty}$.

implies that the class of functions

$$\hat{V}_{b_n}(T) \equiv \left\{ v \in L^\infty(T) : v(t) = \frac{1}{n}\sum_{i=1}^{n} w(t, z_i) p^{k_n'}(x_i) h \right. \\ \left. \text{for some } h \in \mathcal{H}_{b_n}^{B_n} \right\} \tag{56}$$

provides a suitable estimator for $V_\infty(T)$.

Since the set of residuals induced by $(t, \theta) \in T \times \Theta$ forms a Donsker class, it follows from Gine and Zinn (1990) that the law of G can be consistently estimated by the law of the process

$$G_n^*(t, \theta) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ u_i^*(t, \theta) - E^*[U(t, \theta)] \}. \tag{57}$$

However, because $G_n^*(t, \theta)$ is properly centered for all $(t, \theta) \in T \times \Theta$, it converges to a tight process on all $L^\infty(T \times \Theta)$, not just on the subspace $L^\infty(T \times \Theta_0)$. We ensure that the bootstrap process G_n^* is asymptotically evaluated on $T \times \Theta_0$ by introducing the penalty function

$$P_n^*(t, \theta) \equiv \left(\frac{1}{n} \sum_{i=1}^{n} (y_i - \theta(x_i)) w(t, z_i) \right)^2, \tag{58}$$

which converges almost surely to zero for all $(t, \theta) \in T \times \Theta_0$ and has a nonzero limit for some $t \in T$ whenever $\theta \notin \Theta_0$. For an appropriate sequence $0 < \lambda_n \uparrow \infty$, our bootstrap statistic is then

$$I_n^*(R) \equiv \inf_{\theta \in \Theta_n \cap R, v \in \hat{V}_{b_n}(T)} \max_{t \in T_n} \{ (G_n^*(t, \theta) - v(t))^2 + \lambda_n P_n^*(t, \theta) \}. \tag{59}$$

The penalty $\lambda_n P_n^*$ ensures that the maximum diverges to infinity for all $\theta \notin \Theta_0$ but remains stable for all $\theta \in \Theta_0$. As a result, the infimum in (59) is asymptotically attained in a shrinking neighborhood of $\Theta_0 \cap R$, on which the law of G_n^* provides a consistent estimator for the law of G.

We require the following assumption to establish the consistency of the bootstrap procedure.

ASSUMPTION 3.5: (i) *The penalty* λ_n *satisfies* $\lambda_n \uparrow \infty$ *and* $\lambda_n = o(n/\log(\log(n)))$. (ii) b_n *and* B_n *satisfy* $b_n \uparrow \infty$, $B_n \uparrow \infty$, *and* $b_n \times B_n^2 \times \xi_n^2 \times \log(n) = o(n)$, *where* $\xi_n = \sup_{x \in \mathcal{X}} \|p^{b_n}(x)\|$. (iii) *There is* $\Pi_n \theta \in \Theta_n \cap R$ *with* $\sup_{\theta \in \Theta_0 \cap R} \|\theta - \Pi_n \theta\|_w = o(\lambda_n^{-1/2})$.

Assumption 3.5(i) states the feasible rates for the penalty weight λ_n. Employing a compact law of iterated logarithms, we show that such a rate restriction implies that $\lambda_n P_n^*(t,\theta)$ converges almost surely to zero uniformly on $(t,\theta) \in (T \times \Theta_0)$. The rate requirements in Assumption 3.5(ii) imply that minimizing over the random set of functions $\hat{V}_{b_n}(T)$ is asymptotically equivalent to minimizing over $V_\infty(T)$. The relationship between ξ_n and b_n is well known though sieve specific. For example, $\xi_n \lesssim b_n^{d_x/2}$ for tensor product univariate splines and $\xi_n \lesssim b_n^{d_x}$ for polynomials; see Newey (1997), Huang (1998), and Chen (2007). The approximation error on the sieve imposed in Assumption 3.5(iii) is necessary so that $I_n^*(R)$ does not diverge to infinity when the null hypothesis is true. This requirement needs only to hold under the null hypothesis and is weaker than in Assumption 3.4(ii) because the divergence of $I_n^*(R)$ is governed by the penalty $\lambda_n P_n^*$, while the divergence of $I_n(R)$ is the result of improper centering of the empirical process.

Under the stated assumptions, the proposed bootstrap procedure is consistent almost surely (a.s.).

THEOREM 3.2: *Let Assumptions 2.1, 2.2, 3.1, 3.2, 3.3(ii), 3.4(i), and 3.5 hold. If $\Theta_0 \cap R \neq \emptyset$, then*

$$I_n^*(R) \xrightarrow{\mathcal{L}^*} \inf_{\theta \in \Theta_0 \cap R, v \in V_\infty(T)} \|G(\cdot,\theta) - v\|_\infty^2 \quad a.s.,$$

where G has the same law as in Theorem 3.1. Under the same assumptions, if $\Theta_0 \cap R = \emptyset$, then

$$\lambda_n^{-1} I_n^*(R) \xrightarrow{p^*} \min_{\theta \in \Theta \cap R} \left\| E\left[(Y - \theta(X))w(\cdot, Z)\right]\right\|_\infty^2 \quad a.s.$$

It follows from the first claim of Theorem 3.2 that the $1-\alpha$ quantile of $I_n^*(R)$ under the empirical measure can be employed as a critical value to control the size of the test. We therefore define

$$\hat{c}_{1-\alpha} \equiv \inf\left\{u : P^*(I_n^*(R) \leq u) \geq 1 - \alpha\right\}. \tag{60}$$

Alternatively, the second claim of Theorem 3.2 implies that $\hat{c}_{1-\alpha}$ diverges to infinity if $\Theta_0 \cap R = \emptyset$. By Theorem 3.1, however, the divergence of $\hat{c}_{1-\alpha}$ is slower than that of $I_n(R)$. As a result, under the alternative hypothesis, $\hat{c}_{1-\alpha} < I_n(R)$ with probability tending to 1 and the null hypothesis will be asymptotically rejected. Corollary 3.1 formalizes these results.

COROLLARY 3.1: *Let Assumptions 2.1, 2.2, and 3.1–3.5 hold. If $\Theta_0 \cap R \neq \emptyset$ and, in addition, the asymptotic distribution of $I_n(R)$ is continuous and strictly* *increasing at its* $1-\alpha$ *quantile, then*

$$\lim_{n\to\infty} P(I_n(R) \le \hat{c}_{1-\alpha}) = 1-\alpha.$$

Furthermore, under the same assumptions, if $\Theta_0 \cap R = \emptyset$ *and* $\alpha > 0$, *then it follows that*

$$\lim_{n\to\infty} P(I_n(R) > \hat{c}_{1-\alpha}) = 1.$$

REMARK 3.4: Extending results in Chernozhukov, Hong, and Tamer (2007) to more general metric spaces, Santos (2011) developed consistency results for estimators of sets of functions. In the present context, for $\varepsilon_n \downarrow 0$ at an appropriate rate, a consistent estimator for Θ_0 is given by

$$\hat{\Theta}_0 \equiv \left\{\theta \in \Theta_n : \max_{t\in T}\left|\frac{1}{n}\sum_{i=1}^{n}(y_i - \theta(x_i))w(t, z_i)\right| \le \varepsilon_n\right\}. \tag{61}$$

Interestingly, it can be shown that if $\Theta_0 \cap R \neq \emptyset$ and $\varepsilon_n \asymp \lambda_n^{-1/2}$, then almost surely

$$I_n^*(R) = \inf_{\theta\in\hat{\Theta}_0\cap R}\ \inf_{v\in\hat{V}_{b_n}(T)} \|G_n^*(\cdot, \theta) - v\|_\infty^2 + o_{p^*}(1). \tag{62}$$

Therefore, the proposed bootstrap procedure can be interpreted as a "plug-in" bootstrap where $\Theta_0 \cap R$ is estimated by $\hat{\Theta}_0 \cap R$, $V_\infty(T)$ is estimated by $\hat{V}_{b_n}(T)$, and the law of G is estimated by that of G_n^*.

3.4. *Monte Carlo*

In this section, we conduct a limited Monte Carlo study to explore the small sample performance of the proposed test statistic as well as to illustrate its implementation.[7] We generate an i.i.d. sample by

$$\begin{pmatrix} X^* \\ Z^* \\ \varepsilon^* \end{pmatrix} \sim N\left(0, \begin{bmatrix} 1 & 0.5 & 0.3 \\ 0.5 & 1 & 0 \\ 0.3 & 0 & 1 \end{bmatrix}\right) \tag{63}$$

and then employ the latent variables $(X^*, Z^*, \varepsilon^*)$ to construct (X, Z, ε) defined by

$$X = 2(\Phi(X^*/3) - 0.5), \quad Z = 2(\Phi(Z^*/3) - 0.5), \quad \varepsilon = \varepsilon^*, \tag{64}$$

[7]Code is available online (Santos (2012)).

where $\Phi(\cdot)$ is the c.d.f. of a standard normal random variable. As constructed, (X, Z) has support on $[-1, 1]^2$, while ε has full support. We use the transformation $2(\Phi(u/3) - 0.5)$ instead of $2(\Phi(u) - 0.5)$ so as to obtain random variables with unimodal rather than uniform distributions. Finally, the dependent variable Y is then generated according to the design

$$Y = 2\sin(X\pi) + \varepsilon. \tag{65}$$

As in most nonparametric problems, the procedure is sensitive to the choice of smoothing parameters. We primarily focus on studying the effect of the penalty weight λ_n and the norm constraints B_n and B. For this purpose, we consider the family of null hypotheses

$$H_0 : \Theta_0 \cap R(\gamma) \neq \emptyset, \quad R(\gamma) \equiv \{\theta \in W^c([-1, 1]) : \theta(0) = \gamma\}. \tag{66}$$

Under the specification in (63)–(66), the model is nonparametrically identified. Hence, the hypotheses in (66) can be used through test inversion to construct a confidence interval for the level of θ_0 at zero. We employ an identified specification to highlight that under identification, the proposed procedure automatically performs inference on the true parameter. However, it is important to note that this specification is favorable to the procedure in that confidence regions for identifiable functionals are, of course, larger under set identification than under point identification.

We set $m = 1$ and $m_0 = 1$, which satisfies the requirement $\min\{m_0, m\} > d_x/2$ of Assumption 2.1(i). Since X has compact support, tail control on the function θ_0 is unnecessary and following Assumption 2.1(i), we set $\delta = \delta_0 = 0$. The sieve was chosen to be B-splines of order 3 with knot sequence $\{-1, -1, -1, 0, 1, 1, 1\}$, which implies $k_n = 4$. For the weight function $w(t, z)$, we selected

$$w(t, z) = \phi\left(\frac{t_1 - z}{t_2}\right), \tag{67}$$

where $\phi(\cdot)$ is the density of a standard normal random variable. The sieve T_n was set to be the grid $(t_1, t_2) \in \{-0.8, -0.4, 0, 0.4, 0.8\} \times \{0.05, 0.2\}$. Finally, for simplicity, we set the bootstrap sieve equal to the sample sieve with $b_n = k_n$ and employed 500 bootstrap evaluations to calculate critical values. The Monte Carlo study consisted of 500 replications of samples of 500 observations. Given the design in (65), the null hypothesis in (66) is true for $\gamma = 0$ and is false otherwise.

In Table I, we report the simulated size of the test as a function of nominal size (α), penalty weight (λ_n), and norm constraints (B_n, B). The choice $\lambda_n = 0$ is not warranted by Theorem 3.2 and is considered to illustrate the extreme case of selecting too small a penalty weight. As predicted by the theory, setting $\lambda_n = 0$ leads to severe overrejection. The values $(n^{1/3}, n^{1/2}, n^{2/3}) \approx (7.9, 22.4, 63.0)$

TABLE I

SIZE AS A FUNCTION OF (α, λ_n)

	Norm Constraint $B = 50$			Norm Constraint $B = 10^2$			Norm Constraint $B = 10^3$		
	$B_n = 50$	$B_n = 10^2$	$B_n = 10^3$	$B_n = 50$	$B_n = 10^2$	$B_n = 10^3$	$B_n = 50$	$B_n = 10^2$	$B_n = 10^3$
$\alpha = 0.1, \lambda_n = 0$	0.430	0.464	0.508	0.450	0.482	0.516	0.452	0.492	0.510
$\alpha = 0.1, \lambda_n = n^{1/3}$	0.224	0.242	0.268	0.198	0.228	0.244	0.202	0.224	0.240
$\alpha = 0.1, \lambda_n = n^{1/2}$	0.162	0.186	0.206	0.152	0.156	0.182	0.146	0.160	0.180
$\alpha = 0.1, \lambda_n = n^{2/3}$	0.108	0.128	0.154	0.102	0.106	0.126	0.104	0.116	0.126
$\alpha = 0.05, \lambda_n = 0$	0.310	0.340	0.406	0.316	0.368	0.406	0.326	0.382	0.408
$\alpha = 0.05, \lambda_n = n^{1/3}$	0.128	0.144	0.174	0.120	0.132	0.146	0.124	0.138	0.144
$\alpha = 0.05, \lambda_n = n^{1/2}$	0.088	0.102	0.132	0.082	0.096	0.104	0.084	0.100	0.102
$\alpha = 0.05, \lambda_n = n^{2/3}$	0.060	0.076	0.096	0.054	0.068	0.076	0.052	0.062	0.068
$\alpha = 0.01, \lambda_n = 0$	0.108	0.140	0.186	0.122	0.156	0.188	0.126	0.164	0.188
$\alpha = 0.01, \lambda_n = n^{1/3}$	0.042	0.054	0.064	0.040	0.052	0.054	0.042	0.048	0.050
$\alpha = 0.01, \lambda_n = n^{1/2}$	0.016	0.028	0.042	0.014	0.026	0.034	0.018	0.034	0.036
$\alpha = 0.01, \lambda_n = n^{2/3}$	0.006	0.008	0.020	0.010	0.012	0.012	0.010	0.016	0.016

represent a wide spectrum of choices for the penalty λ_n. Overall, $\lambda_n = n^{2/3}$ provides good size control across the different choices of (B_n, B), while $\lambda_n = n^{1/3}$ appears to be too small a choice of penalty weight, leading to severe size distortions in most specifications. For the penalty weight choices that yield adequate size control ($\lambda_n = n^{1/2}$, $\lambda_n = n^{2/3}$), the results appear to be robust to the selection of norm bound B. The choice of bootstrap bound B_n, however, has a more pronounced effect, with larger values increasing rejection probabilities as would be expected.

Figure 1 illustrates rejection probabilities of the null hypothesis in (66) as a function of $\gamma \in [-1, 1]$ for different choices of penalty weight (λ_n) and norm constraints (B_n, B). For conciseness we only consider a subset of the parameter specifications in Table I for which size control was adequate. As expected, higher penalty choices (λ_n) lead to smaller rejection rates as do smaller bootstrap norm constraints (B_n). These differences are on the order of the size distortions in Table I for small values of γ, but are significantly larger for alternatives far away from the null hypothesis. A smaller choice of norm constraint B also seems to increases the power of the test for alternatives far away from the null hypothesis.

4. BRAZILIAN FUEL ENGEL CURVES

In response to the oil shocks of the 1970s, Brazil embarked in 1975 on a national program to substitute gasoline consumption with ethanol processed from sugar cane. Today, ethanol accounts for an important fraction of the transport fuel market. In this section, we study the Engel curves for ethanol and

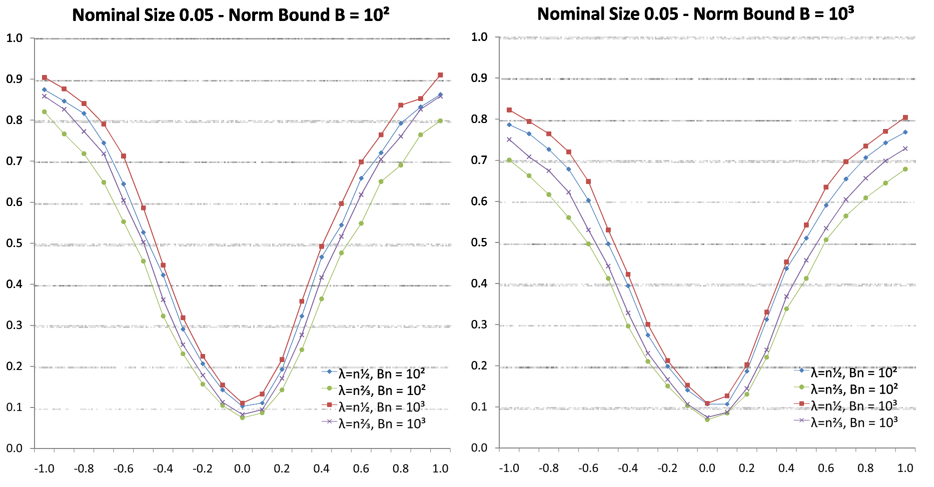

FIGURE 1.—Power curves as a function of λ_n, B_n, and B.

gasoline in Brazil using data from Pesquisa de Orçamentos Familiares 2002–2003 (POF). The POF is similar to the United States Bureau of Labor Statistics Consumer Expenditure Survey, but is conducted more sporadically (the previous study was 1995–1996) and more extensively (total of 48,470 households).

We let Y_e and Y_g be the share of total nondurable expenditures spent on ethanol and gasoline, respectively, let X denote the log of total nondurable expenditures, and let Z be total household income. The Engel curves for ethanol and gasoline are assumed to satisfy the additively separable specification

$$Y_m = \theta_m(X) + \varepsilon_m, \tag{68}$$

where $m \in \{e, g\}$ and ε_m is unobservable heterogeneity. We condition on households composed of cohabitating couples in urban areas who have children and exhibited positive consumption. These restrictions yield a data set of 4994 observations for gasoline and 467 observations for ethanol.

We specify $\|\cdot\|_s$ and Θ as in (9) and (11), respectively, setting $\mathcal{X} = [7, 13]$, which includes all observations for both gasoline and ethanol. Since $\mathcal{X}$ is compact, no control on the tail behavior of θ_0 is necessary and we let $\delta_0 = \delta = 0$. Further, we assume $m = m_0 = 1$, which allows for inference on the first derivative of the Engel curves, and specify $B = 10^2$. For the sieve, we employ polynomials of order 5 for gasoline and order 4 for ethanol. Adding additional terms to the sieve failed to significantly change the value of the unconstrained minimum $I_n(\Theta)$. The weight function $w(t, z)$ was set as in (67), with $T_n = \{5, 5.5, \ldots, 9.5, 10\} \times \{0.05, 0.1, 0.2\}$ for gasoline and $T_n = \{6, 6.5, \ldots, 9.5, 9\} \times \{0.05, 0.2\}$ for ethanol. Finally, the bootstrap was computed with 2000 simu-

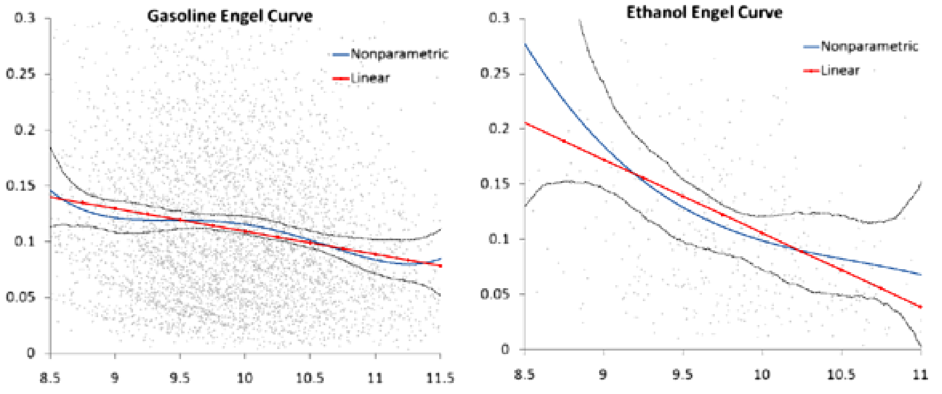

FIGURE 2.—Nonparametric and linear estimates.

lations, setting $B_n = 10^4$ and $B_n = 10^3$ for gasoline and ethanol, respectively. Following the results from the Monte Carlo study, we conduct our hypothesis tests with $\lambda_n = n^{1/2}$ and $\lambda_n = n^{2/3}$.

Engel curves are commonly parametrized as either linear or quadratic in the log of total nondurable expenditures. For example, Banks, Blundell, and Lewbel (1997) found that while a quadratic term seems to be present in the Engel curves for clothing and alcohol, a linear relationship suffices to adequately describe the Engel curves for food and fuel. Defining the set

$$R_L \equiv \{\theta \in W^s(\mathcal{X}) : \theta(x) = \alpha + x\beta \text{ for some } (\alpha, \beta) \in \mathbf{R}^2\}, \tag{69}$$

we plot in Figure 2 the minimizer of $I_n(R_L)$ (best linear fit) and the minimizer of $I_n(\Theta)$ (implied nonparametric estimator). The dashed lines represent pointwise 95% bands obtained from bootstrapping the minimizer of $I_n(\Theta)$. These are not valid confidence intervals and are just meant to illustrate the variability present in the data. Figure 2 suggests that a log-linear specification for the Engel curves is compatible with the data. Using the methodology developed in Section 3 we test

$$H_0 : \Theta_0 \cap R_L \neq \emptyset, \quad H_1 : \Theta_0 \cap R_L = \emptyset, \tag{70}$$

and fail to reject the null for both gasoline and ethanol. Employing $\lambda_n = n^{1/2}$ and $\lambda_n = n^{2/3}$ yielded p-values equal to 0.403 and 0.441 for gasoline and 0.371 and 0.433 for ethanol.[8] Interestingly, a GMM-based J-test constructed using the moments $w(t, z)$ for $t \in T_n$ also fails to reject the null that the model is properly specified, with p-values of 0.392 and 0.602 for gasoline and ethanol, respectively.

[8]Here, by p-value we mean $P^*(I_n^*(R_L) > I_n(R_L))$.

Despite failing to reject the null hypothesis that there are log-linear Engel curves in the identified set, we should not necessarily feel comfortable using such a parametric specification. If the model is not identified, then even when there indeed are log-linear specifications in Θ_0, there is no guarantee that the true model is one of them. As a result, confidence intervals constructed assuming log-linearity may asymptotically exclude the true parameter of interest. We therefore further examine the robustness of a log-linear specification by comparing confidence regions for the level and derivative of the Engel curve at the sample average $\bar{X}$ constructed with and without a log-linearity assumption.

In Table II, we report confidence intervals for different choices of penalty weight λ_n and coverage requirement α. For the level of the Engel curve, we obtained nonparametric confidence intervals through test inversion of $I_n(R(\gamma))$ (for $R(\gamma)$ defined in (66)), while for the log-linear specification, we inverted $I_n(R(\gamma) \cap R_L)$. Confidence intervals for the level of the derivative were calculated in a similar manner. For comparison purposes, standard instrumental variable confidence intervals are also reported. All three approaches yield similar conclusions regarding $\theta_g(\bar{X})$. In contrast, while all procedures agree on the upper end of the confidence interval for $\theta_e(\bar{X})$, they differ substantially on the lower end. Similarly, while all three procedures yield comparable answers for the confidence interval for $\theta'_g(\bar{X})$, they agree more on the lower bound than the upper bound. The difference in conclusions attainable through the three approaches is strongest for $\theta'_g(\bar{X})$, with the nonparametric procedure yielding a substantially larger confidence interval. Differences between the gasoline and the ethanol analysis are likely due to the important disparity in their respective sample sizes.

5. CONCLUSION

We have developed a flexible framework that allows us to test a wide array of hypotheses in nonparametric instrumental variables problems. Through examples, we have shown that identification may fail even under smoothness assumptions on the underlying model. An appealing feature of the proposed method is that it is robust to a possible lack of identification.

APPENDIX

A. *Notation and Definitions*

The following table of notation and definitions will be used throughout the Appendix:

TABLE II

CONFIDENCE INTERVALS FOR $\theta_m(\bar{X})$ AND $\theta'_m(\bar{X})$

	Gasoline CI for $\theta_g(\bar{X})$			Ethanol CI for $\theta_e(\bar{X})$		
	Nonparametric	Linear	Standard IV	Nonparametric	Linear	Standard IV
$\lambda_n = n^{1/2}$, $\alpha = 0.90$	[0.107, 0.120]	[0.110, 0.115]	[0.111, 0.115]	[0.097, 0.125]	[0.107, 0.127]	[0.112, 0.127]
$\lambda_n = n^{2/3}$, $\alpha = 0.90$	[0.106, 0.121]	[0.110, 0.115]	[0.111, 0.115]	[0.095, 0.128]	[0.107, 0.128]	[0.112, 0.127]
$\lambda_n = n^{1/2}$, $\alpha = 0.95$	[0.106, 0.121]	[0.110, 0.116]	[0.111, 0.115]	[0.095, 0.129]	[0.106, 0.128]	[0.111, 0.129]
$\lambda_n = n^{2/3}$, $\alpha = 0.95$	[0.104, 0.122]	[0.110, 0.116]	[0.111, 0.115]	[0.092, 0.132]	[0.105, 0.130]	[0.111, 0.129]
$\lambda_n = n^{1/2}$, $\alpha = 0.99$	[0.102, 0.123]	[0.109, 0.117]	[0.110, 0.116]	[0.089, 0.135]	[0.102, 0.132]	[0.108, 0.131]
$\lambda_n = n^{2/3}$, $\alpha = 0.99$	[0.100, 0.125]	[0.109, 0.117]	[0.110, 0.116]	[0.087, 0.137]	[0.100, 0.134]	[0.108, 0.131]
	Gasoline CI for $\theta'_g(\bar{X})$			Ethanol CI for $\theta'_e(\bar{X})$		
	Nonparametric	Linear	Standard IV	Nonparametric	Linear	Standard IV
$\lambda_n = n^{1/2}$, $\alpha = 0.90$	[−0.028, −0.001]	[−0.024, −0.011]	[−0.025, −0.018]	[−0.109, 0.034]	[−0.091, −0.042]	[−0.080, −0.037]
$\lambda_n = n^{2/3}$, $\alpha = 0.90$	[−0.031, 0.000]	[−0.024, −0.011]	[−0.025, −0.018]	[−0.118, 0.053]	[−0.092, −0.040]	[−0.080, −0.037]
$\lambda_n = n^{1/2}$, $\alpha = 0.95$	[−0.032, 0.001]	[−0.025, −0.009]	[−0.026, −0.018]	[−0.120, 0.049]	[−0.098, −0.036]	[−0.084, −0.042]
$\lambda_n = n^{2/3}$, $\alpha = 0.95$	[−0.035, 0.003]	[−0.025, −0.009]	[−0.026, −0.018]	[−0.129, 0.074]	[−0.100, −0.035]	[−0.084, −0.042]
$\lambda_n = n^{1/2}$, $\alpha = 0.99$	[−0.041, 0.006]	[−0.027, −0.006]	[−0.027, −0.016]	[−0.146, 0.078]	[−0.117, −0.025]	[−0.090, −0.035]
$\lambda_n = n^{2/3}$, $\alpha = 0.99$	[−0.043, 0.009]	[−0.027, −0.006]	[−0.027, −0.016]	[−0.160, 0.100]	[−0.120, −0.023]	[−0.090, −0.036]

$a \lesssim b$	$a \leq Mb$ for some constant M which is universal in the context of the proof,
$\\|\cdot\\|$	when the context is clear, it denotes the Euclidean norm; otherwise, it denotes a generic norm,
$\\|\theta\\|_s$	the norm $[\sum_{\|\lambda\|\leq m+m_0}\int_{\mathcal{X}}[D^\lambda\theta(x)]^2(1+x'x)^{\delta_0}\,dx]^{1/2}$,
$\\|\theta\\|_c$	the norm $\max_{\|\lambda\|\leq m}\sup_{x\in\mathcal{X}}\|D^\lambda\theta(x)\|(1+x'x)^{\delta/2}$,
$\mathcal{H}_{b_n}^{B_n}$	the set $\{h\in\mathbf{R}^{b_n}: L(p^{b_n'}h)=0, \\|h\\|\leq B_n\}$; without B_n bound, we denote $\mathcal{H}_{b_n}$,
$N(\mathcal{F},\\|\cdot\\|,\varepsilon)$	covering numbers of size ε for $\mathcal{F}$ under the norm $\\|\cdot\\|$,
$N_{[]}(\mathcal{F},\\|\cdot\\|,\varepsilon)$	bracketing numbers of size ε for $\mathcal{F}$ under the norm $\\|\cdot\\|$,
$u(t,\theta)$	the implied residual $(y-\theta(x))w(t,z)$.

In these definitions, θ denotes an arbitrary function, not necessarily an element of Θ.

REMARK A.1: Probability statements are meant to hold in outer measure. The proof of Theorem 3.1 establishes convergence in distribution in the sense of Chapter 1.3 in van der Vaart and Wellner (1996). However, $I_n(R)$ is measurable, as the theorem of the maximum implies $I_n(R)$ is a continuous function of the data. Hence, Theorem 3.1 need not be interpreted in terms of outer measures. Theorem 3.2 and Corollary 3.1, however, hold in outer measure almost surely and in outer probability, respectively, because the law of $I_n^*(R)$ under $\mathcal{L}^*$ need not be measurable.

B. *Proofs*

PROOF OF LEMMA 2.1: Fix $\varepsilon > 0$, $f \in \mathbf{D}(K)$, and for $\delta > 0$, pointwise define the adjusted truncated function by

$$f_\delta(x,z) \equiv \frac{\max\{f(x,z),\delta\}}{\int_K \max\{f(x,z),\delta\}\,dx\,dz}, \tag{71}$$

which is by construction an element of $\mathbf{D}(K)$ for all δ. We can then obtain by compactness of K that

$$\begin{aligned}\lim_{\delta\downarrow 0}\left|\int_K \max\{f(x,z),\delta\}\,dx\,dz - 1\right| &= \lim_{\delta\downarrow 0}\left|\int_K \big\{\max\{f(x,z),\delta\} - f(x,z)\big\}\,dx\,dz\right| \\ &\leq \lim_{\delta\downarrow 0}\int_K \delta\,dx\,dz = 0.\end{aligned} \tag{72}$$

Therefore, by the triangle inequality, f bounded due to it being continuous and K being compact, and (72), we obtain

$$\lim_{\delta\downarrow 0}\|f-f_\delta\|_\infty \le \lim_{\delta\downarrow 0}\|f\|_\infty \times \left|1-\frac{1}{\int_K \max\{f(x,z),\delta\}\,dx\,dz}\right| + \lim_{\delta\downarrow 0}\frac{\|f-\max\{\delta,f\}\|_\infty}{\int_K \max\{f(x,z),\delta\}\,dx\,dz} = 0. \tag{73}$$

From (73), there exists $\delta^* > 0$ such that $\|f - f_{\delta^*}\|_\infty < \frac{\varepsilon}{2}$. Since f_{δ^*} is continuous on K and since K is compact, the Stone–Weierstrass theorem implies the existence of a sequence of polynomials $\{P_n\}_{n=1}^\infty$ such that

$$\lim_{n\to\infty}\|f_{\delta^*}-P_n\|_\infty = 0. \tag{74}$$

For each P_n, define $f_n(x,z) \equiv P_n(x,z)/\int_K |P_n(x,z)|\,dx\,dz$, and note that arguing as in (72) and (73), we can show

$$\lim_{n\to\infty}\|f_{\delta^*}-f_n\|_\infty = 0. \tag{75}$$

Since f_{δ^*} is bounded away from zero, (75) implies $f_n \in \mathbf{D}(K)$ for n sufficiently large. Selecting n^* so that $f_{n^*} \in \mathbf{D}(K)$ and $\|f_{\delta^*} - f_{n^*}\| < \frac{\varepsilon}{2}$, we then obtain by the triangle inequality that

$$\|f - f_{n^*}\|_\infty < \varepsilon. \tag{76}$$

Let k be the order of the polynomial f_{n^*} and observe that by Gram–Schmidt orthogonalization, there is a polynomial $p^{k+1} : \mathcal{X} \to \mathbf{R}$ of order $k+1$ such that $\int_K p^{k+1}(x) f_{n^*}(x,z)\,dx = 0$ for all z. Therefore, as $\|p^{k+1}\|_s < \infty$, we conclude that $f_{n^*} \in \mathbf{D}_\emptyset(K)$ and the lemma follows. *Q.E.D.*

LEMMA A.1: *Suppose* $(H_1, \|\cdot\|_1)$ *and* $(H_2, \|\cdot\|_2)$ *are separable Hilbert spaces with* $(H_1, \|\cdot\|_1)$ *compactly embedded in* $(H_2, \|\cdot\|_2)$. *If* $\Omega \equiv \{h \in H_1 : \|h\|_1 \le B\}$ *for* $B < \infty$, *then it follows that* Ω *is closed in* $(H_2, \|\cdot\|_2)$.

PROOF: Let $I : H_1 \to H_2$ denote the identity operator and note that it is compact since $(H_1, \|\cdot\|_1)$ is compactly embedded in $(H_2, \|\cdot\|_2)$. By Theorem 4.10 in Kress (1999), the adjoint I^* is compact and therefore so is I^*I. Hence, by Theorem 15.16 in Kress (1999), the singular value decomposition of I^*I generates a sequence $\{\varphi_i\}_{i=1}^\infty$ with

$$I^*I(\varphi_i) = \lambda_i^2\varphi_i, \quad \langle\varphi_i,\varphi_j\rangle_1 = \delta_{ij}, \quad \left\langle\frac{\varphi_i}{\lambda_i},\frac{\varphi_j}{\lambda_j}\right\rangle_2 = \delta_{ij}, \tag{77}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots > 0$, $\delta_{ij} = 1$ if $i = j$ and zero otherwise, and $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$ denote the inner products on H_1 and H_2, respectively. Moreover, since $\mathcal{N}(I^*I) = \mathcal{N}(I) = \{0\}$, it follows that I^*I is injective and hence Theorem 15.16 in Kress (1999) additionally implies $\{\varphi_i\}_{i=1}^\infty$ is an orthonormal basis for H_1. Similarly, since $\{\varphi_i/\lambda_i\}_{i=1}^\infty$ is dense in H_1 under $\|\cdot\|_2$, it is also dense in the closure of H_1 under $\|\cdot\|_2$ and therefore, by Theorem 3.4.2 in Christensen (2003), $\{\varphi_i/\lambda_i\}_{i=1}^\infty$ is an orthonormal basis for $\bar{H}_1$ (the closure of H_1 in H_2 under $\|\cdot\|_2$).

To establish the lemma, we aim to show that if $\|h_n - h_0\|_2 = o(1)$ and $h_n \in \Omega$ for all n, then $h_0 \in \Omega$. Toward this end, observe that since $h_n \in H_1 \subset H_2$ for all n, we can derive from the definition of the adjoint and (77) that

$$\langle h_n, \varphi_i \rangle_2 = \langle I(h_n), I(\varphi_i) \rangle_2 = \langle h_n, I^*I(\varphi_i) \rangle_1 = \lambda_i^2 \langle h_n, \varphi_i \rangle_1 \tag{78}$$

for all φ_i. Since $h_0 \in \bar{H}_1$ and $\{\varphi_i/\lambda_i\}_{i=1}^\infty$ is an orthonormal basis for $\bar{H}_1$, we also obtain from $\|h_n - h_0\|_2 = o(1)$,

$$\begin{aligned} 0 &= \lim_{n\to\infty} \|h_n - h_0\|_2^2 = \lim_{n\to\infty} \left\| \sum_{i=1}^\infty \left\langle h_n - h_0, \frac{\varphi_i}{\lambda_i} \right\rangle_2 \frac{\varphi_i}{\lambda_i} \right\|_2^2 \\ &= \lim_{n\to\infty} \sum_{i=1}^\infty \frac{\langle h_n - h_0, \varphi_i \rangle_2^2}{\lambda_i^2}, \end{aligned} \tag{79}$$

where the final equality follows by Parseval's equality. Moreover, since $h_n \in \Omega$ for all n, we also have

$$B^2 \geq \|h_n\|_1^2 = \sum_{i=1}^\infty \langle h_n, \varphi_i \rangle_1^2 = \sum_{i=1}^\infty \frac{\langle h_n, \varphi_i \rangle_2^2}{\lambda_i^4}, \tag{80}$$

where the first equality follows from $\{\varphi_i\}_{i=1}^\infty$ being an orthonormal basis for H_1 and Parseval's equality, while the second equality is implied by (78). Since (80) holds for all n, it follows that

$$\sum_{i=1}^\infty \frac{\langle h_0, \varphi_i \rangle_2^2}{\lambda_i^4} \leq B^2. \tag{81}$$

To see this, note that if (81) fails to hold, then $\sum_{i=1}^{K_0} \langle h_0, \varphi_i \rangle_2^2 / \lambda_i^4 > B^2$ for K_0 sufficiently large. Together with (79), this implies $\sum_{i=1}^{K_0} \langle h_{n_0}, \varphi_i \rangle_2^2 / \lambda_i^4 > B^2$ for n_0 large enough, contradicting (80). Letting $\ell^2 \equiv \{\{a_j\}_{j=1}^\infty : \sum_{j=1}^\infty a_j^2 < \infty\}$, we, in particular, conclude from (81) that $\{\langle h_0, \varphi_i \rangle_2 / \lambda_i^2\}_{i=1}^\infty \in \ell^2$. Hence, by Theorem 3.2.3 in Christensen (2003),

$$\lim_{K\to\infty} \left\| \tilde{h}_0 - \sum_{i=1}^K \frac{\langle h_0, \varphi_i \rangle_2}{\lambda_i^2} \varphi_i \right\|_1 = 0 \tag{82}$$

for some $\tilde{h}_0 \in H_1$. Moreover, by (81), $\{\varphi_i\}_{i=1}^{\infty}$ being an orthonormal basis for H_1, and Parseval's equality, it follows that $\|\tilde{h}_0\|_1 \leq B$, which additionally implies $\tilde{h}_0 \in \Omega$. Finally, since $\{\varphi_i/\lambda_i\}_{i=1}^{\infty}$ is orthonormal in H_2, we obtain

$$\left\langle \tilde{h}_0, \frac{\varphi_i}{\lambda_i} \right\rangle_2 = \left\langle \sum_{i=1}^{\infty} \frac{\langle h_0, \varphi_i \rangle_2}{\lambda_i^2} \varphi_i, \frac{\varphi_i}{\lambda_i} \right\rangle_2 = \left\langle h_0, \frac{\varphi_i}{\lambda_i} \right\rangle_2. \tag{83}$$

Since $\tilde{h}_0, h_0 \in \bar{H}_1$ and $\{\varphi_i/\lambda_i\}_{i=1}^{\infty}$ is an orthonormal basis for $\bar{H}_1$, we conclude $\|\tilde{h}_0 - h_0\|_2 = 0$. It follows that $\tilde{h}_0 \in \Omega$ is in the same $\|\cdot\|_2$ equivalence class as h_0 and therefore Ω is closed in $(H_2, \|\cdot\|_2)$, establishing the lemma. *Q.E.D.*

LEMMA A.2: *Under Assumption* 2.1(i) *and* (ii), *the parameter space Θ is compact under $\|\cdot\|_c$.*

PROOF: If $\mathcal{X}$ is bounded, then Assumption 2.1(i) and (ii) and Theorem 6.3 Part II of Adams and Fournier (2003) imply $W^s(\mathcal{X})$ is compactly embedded in $W^c(\mathcal{X})$. Alternatively, if $\mathcal{X}$ is unbounded, then the arguments in Lemma A.4 of Gallant and Nychka (1987) applied to $\mathcal{X}$ instead of $\mathbf{R}^{d_x}$ verbatim imply that $W^c(\mathcal{X})$ is compactly embedded in $W^s(\mathcal{X})$ thanks to Assumption 2.1(i) and (ii). It follows that Θ is relatively compact in $W^c(\mathcal{X})$. However, since $\|\cdot\|_{L^2} \leq \|\cdot\|_c$, it also follows that $W^s(\mathcal{X})$ is compactly embedded in $L^2(X)$ as well, and by Lemma A.1 that Θ is closed in $\|\cdot\|_{L^2}$ and therefore also under the norm $\|\cdot\|_c$. We conclude that Θ is compact under $\|\cdot\|_c$ as claimed. *Q.E.D.*

PROOF OF LEMMA 3.1: Since $\theta \in \Theta$ are uniformly bounded and $E[Y^2] < \infty$, it follows by Assumption 3.1(ii) that

$$\begin{aligned} &\big|E\big[(Y - \theta(X))w(t_1, Z)\big] - E\big[(Y - \theta(X))w(t_2, Z)\big]\big| \\ &\quad \leq E\big[|Y - \theta(X)|\big] \times M\|t_1 - t_2\| \lesssim \|t_1 - t_2\|. \end{aligned} \tag{84}$$

Therefore, $(E[(Y - \theta(X))w(t, Z)])^2$ is continuous in t. It follows from T being compact that the maximum is indeed attained. The first claim of the lemma is then a direct consequence of Assumption 3.1(i).

For the second claim of the lemma, notice that $E[(Y - \theta(X))w(t, Z)]$ is jointly continuous on $T \times \Theta$ with respect to the product topology of the Euclidean norm and $\|\cdot\|_c$. Therefore, by the theorem of the maximum,

$$\max_{t \in T}\big(E\big[(Y - \theta(X))w(t, Z)\big]\big)^2 \tag{85}$$

is continuous in Θ with respect to $\|\cdot\|_c$. By Lemma A.2, Θ is compact under $\|\cdot\|_c$ and since R is closed under $\|\cdot\|_c$ by continuity of L, $\Theta \cap R$ is compact as well. Thus, continuity implies the minimum is attained and the result then follows by the first claim of the lemma. *Q.E.D.*

LEMMA A.3: *Under Assumption 2.1, there exists a constant* $K > 0$ *such that for all* ε *sufficiently small*, (i) *if* $\mathcal{X}$ *is unbounded, then* $\log N(\Theta, \|\cdot\|_\infty, \varepsilon) \leq K(\frac{1}{\varepsilon})^{(m+\delta)d_x/(\delta m)}$; (ii) *if* $\mathcal{X}$ *is bounded, then* $\log N(\Theta, \|\cdot\|_\infty, \varepsilon) \leq K(\frac{1}{\varepsilon})^{d_x/m}$.

PROOF: We first establish claim (i) of the lemma. Fix $\varepsilon > 0$. By Lemma A.2, Θ is compact in $\|\cdot\|_c$ and hence is totally bounded. It follows that for some constant C, we have that for all $\theta \in \Theta$,

$$\max_{|\lambda|\leq m} \sup_{x\in\mathcal{X}} |D^\lambda \theta(x)|(1 + x'x)^{\delta/2} \leq C. \tag{86}$$

Let $\mathcal{X}_{\varepsilon/2} \equiv \{x \in \mathcal{X} : (1 + x'x)^{\delta/2}\varepsilon/2 \leq C\}$. By (86), it follows that for any $\theta_1, \theta_2 \in \Theta$, we must have

$$\sup_{x\in\mathcal{X}^c_{\varepsilon/2}} |\theta_1(x) - \theta_2(x)| \leq \sup_{x\in\mathcal{X}^c_{\varepsilon/2}} |\theta_1(x)| + \sup_{x\in\mathcal{X}^c_{\varepsilon/2}} |\theta_2(x)| < \varepsilon. \tag{87}$$

This implies that $\sup_{x\in\mathcal{X}} |\theta_1(x) - \theta_2(x)| < \varepsilon$ if and only if $\sup_{x\in\mathcal{X}_{\varepsilon/2}} |\theta_1(x) - \theta_2(x)| < \varepsilon$. Therefore, without loss of generality, when calculating $N(\Theta, \|\cdot\|_\infty, \varepsilon)$, we can assume that $\theta(x) = 0$ for all $x \in \mathcal{X}^c_{\varepsilon/2}$ and $\theta \in \Theta$.

For a domain $\Omega \subseteq \mathbf{R}^{d_x}$, define the Sobolev spaces $W^{k,p}(\Omega) \equiv \{\theta : \Omega \to \mathbf{R} \text{ s.t. } \sum_{|\lambda|\leq k} \int_\Omega |D^\lambda \theta(x)|^p \, dx < \infty\}$ with the usual modification for $p = \infty$. By Theorem 5.28 in Adams and Fournier (2003) and Assumption 2.1(ii), there exists a simple extension operator $E : W^{m+m_0,2}(\mathcal{X}) \to W^{m+m_0,2}(\mathbf{R}^{d_x})$ such that $E(\theta)(x) = \theta(x)$ for all $x \in \mathcal{X}$ and all $\theta \in W^{m+m_0,2}(\mathcal{X})$. In addition, Theorem 4.12, Part I, Case A of Adams and Fournier (2003) implies that $W^{m+m_0,2}(\mathbf{R}^{d_x})$ is embedded in $W^{m,\infty}(\mathbf{R}^{d_x})$. Since $W^s(\mathcal{X}) \subseteq W^{m+m_0,2}(\mathcal{X})$, we conclude that for any $\theta \in \Theta$,

$$\begin{aligned} \sup_{x\in\mathbf{R}^{d_x}} \sum_{|\lambda|\leq m} |D^\lambda E(\theta)(x)|^2 &\leq K_0 \sum_{|\lambda|\leq m+m_0} \int_{\mathbf{R}^{d_x}} [D^\lambda E(\theta)(x)]^2 \, dx \\ &\leq K_1 \sum_{|\lambda|\leq m+m_0} \int_{\mathcal{X}} [D^\lambda \theta(x)]^2 \, dx \leq K_1 B^2, \end{aligned} \tag{88}$$

where the final inequality follows from $(1 + x'x)^{\delta_0} \geq 1$ and the constants are independent of θ. Let $\operatorname{co}(\mathcal{X}_{\varepsilon/2})$ denote the closed convex hull of $\mathcal{X}_{\varepsilon/2}$ and let $\bar{\Theta}$ denote a ball of radius $\sqrt{K_1}B$ in $W^{m,\infty}(\operatorname{co}(\mathcal{X}_{\varepsilon/2}))$. By (88), for any $\theta \in \Theta$, the restriction of $E(\theta)$ to the domain $\operatorname{co}(\mathcal{X}_{\varepsilon/2})$ then maps into an element of $\bar{\Theta}$. Hence, in lieu of (87), we obtain by Theorem 2.7.1 in van der Vaart and Wellner (1996), that for some K depending only on m, d_x, and $\sqrt{K_1}B$,

$$\log N(\Theta, \|\cdot\|_\infty, \varepsilon) \leq \log N(\bar{\Theta}, \|\cdot\|_\infty, \varepsilon) \leq K\lambda\big(\operatorname{co}(\mathcal{X}_{\varepsilon/2})^1\big)\left(\frac{1}{\varepsilon}\right)^{d_x/m}, \tag{89}$$

where $\lambda(\cdot)$ is the Lebesgue measure and $\mathrm{co}(\mathcal{X}_{\varepsilon/2})^1 \equiv \{x \in \mathbf{R}^{d_x} : \inf_{\tilde{x} \in \mathrm{co}(\mathcal{X}_{\varepsilon/2})} \|x - \tilde{x}\| < 1\}$. If $x \in \mathrm{co}(\mathcal{X}_{\varepsilon/2})^1$, then for some $\tilde{x} \in \mathrm{co}(\mathcal{X}_{\varepsilon/2})$, $\|x\| \leq \|x - \tilde{x}\| + \|\tilde{x}\| < 1 + [(2C/\varepsilon)^{2/\delta} - 1]^{1/2}$, which implies $\mathrm{co}(\mathcal{X}_{\varepsilon/2})^1 \subseteq \{x \in \mathbf{R}^{d_x} : \|x\| \leq (4C/\varepsilon)^{1/\delta}\}$ for ε sufficiently small. Hence, we can conclude that

$$\lambda\big(\mathrm{co}(\mathcal{X}_{\varepsilon/2})^1\big) \lesssim \left(\frac{1}{\varepsilon}\right)^{d_x/\delta}. \tag{90}$$

Combining this result with (89), it then follows that for ε sufficiently small, we must have

$$\log N(\Theta, \|\cdot\|_\infty, \varepsilon) \lesssim \left(\frac{1}{\varepsilon}\right)^{d_x/\delta} \times \left(\frac{1}{\varepsilon}\right)^{d_x/m} = \left(\frac{1}{\varepsilon}\right)^{(m+\delta)d_x/(\delta m)}, \tag{91}$$

which establishes the first claim of the lemma. The second claim is immediate by noting that if $\mathcal{X}$ is bounded, then $\mathcal{X} \subseteq \mathrm{co}(\mathcal{X}_{\varepsilon/2})$ for ε sufficiently small, and the result follows from (89). *Q.E.D.*

LEMMA A.4: *Let $\mathcal{F} \equiv \{f : \mathbf{R} \times \mathcal{X} \times \mathcal{Z} \to \mathbf{R} : f(y, x, z) = (y - \theta(x))w(t, z)$ for some $(\theta, t) \in \Theta \times T\}$ and let Assumptions* 2.1(i) *and* (ii), 3.1(i) *and* (ii) *hold. Then there is a $C > 0$ such that $F(y) \equiv (|y| + 1)C$ is an envelope for $\mathcal{F}$ and, in addition,*

$$\begin{aligned} &\big|(y - \theta_1(x))w(t_1, z) - (y - \theta_2(x))w(t_2, z)\big| \\ &\quad \leq F(y) \times \{\|\theta_1 - \theta_2\|_\infty + \|t_1 - t_2\|\}. \end{aligned}$$

Moreover, there exists a constant K such that for all norms $\|\cdot\|$ with $\|F\| < \infty$ and ε sufficiently small,

$$N_{[]}(\mathcal{F}, \|\cdot\|, \varepsilon\|F\|) \leq K \times \exp\left\{\left(\frac{4}{\varepsilon}\right)^\nu\right\} \times \left(\frac{\operatorname{diam} T}{\varepsilon}\right)^{d_t},$$

where $\nu = (m + \delta)d_x/(\delta m)$ if $\mathcal{X}$ is unbounded and $\nu = d_x/m$ if $\mathcal{X}$ is bounded.

PROOF: For the first claim, use the fact that $\theta \in \Theta$ and $w(t, z)$ are uniformly bounded to conclude

$$\begin{aligned} \big|(y - \theta(x))w(t, z)\big| &\leq \Big\{|y| + \sup_{\theta \in \Theta} \|\theta\|_\infty\Big\} \times \sup_{(t,z) \in T \times \mathcal{Z}} |w(t, z)| \\ &\leq B_1\{|y| + 1\} \end{aligned} \tag{92}$$

for some $B_1 > 0$. Next notice that the first and second inequalities in (93) follow by direct calculation and Assumption 3.1(ii), while the third holds for some

$B_2 > 0$:

$$(93)\qquad \bigl|(y-\theta_1(x))w(t_1,z)-(y-\theta_2(x))w(t_2,z)\bigr|$$
$$\le \bigl|(y-\theta_1(x))(w(t_1,z)-w(t_2,z))\bigr| + \bigl|w(t_2,z)(\theta_1(x)-\theta_2(x))\bigr|$$
$$\le \Bigl\{|y| + \sup_{\theta\in\Theta}\|\theta\|_\infty\Bigr\} \times M\|t_1-t_2\| + \sup_{(t,z)\in T\times\mathcal{Z}} |w(t,z)| \times \|\theta_1-\theta_2\|_\infty$$
$$\le B_2\{|y|+1\} \times \{\|t_1-t_2\| + \|\theta_1-\theta_2\|_\infty\}.$$

It follows from (93) that the class $\mathcal{F}$ is Lipschitz in $\Theta \times T$ with respect to the norm $\|\cdot\|_\infty + \|\cdot\|$. Theorem 2.7.11 in van der Vaart and Wellner (1996) then implies the first inequality in (94). The second inequality then holds for ε sufficiently small by Lemma A.3 and $N(\Theta\times T, \|\cdot\|_\infty + \|\cdot\|, \varepsilon) \le N(\Theta, \|\cdot\|_\infty, \varepsilon/2) \times N(T, \|\cdot\|, \varepsilon/2)$:

$$(94)\qquad N_{[]}(\mathcal{F}, \|\cdot\|, \varepsilon\|F\|) \le N(\Theta\times T, \|\cdot\|_\infty + \|\cdot\|, \varepsilon/2)$$
$$\le K \exp\biggl\{\Bigl(\frac{4}{\varepsilon}\Bigr)^\nu\biggr\} \times \Bigl(\frac{\operatorname{diam} T}{\varepsilon/4}\Bigr)^{d_t}.$$

Absorbing constants into K and letting $C > \max\{B_1, B_2\}$ then concludes the proof the lemma. *Q.E.D.*

LEMMA A.5: *Assume* (i) $Q(\theta) \ge 0$ *and* $\Theta_0 = \{\theta \in \Theta : Q(\theta) = 0\}$ *with* Θ *compact with respect to* (*w.r.t.*) $\|\cdot\|$. (ii) $\Theta_n \subseteq \Theta$ *are closed and* $\sup_{\theta\in\Theta}\inf_{\theta_n\in\Theta_n}\|\theta - \theta_n\| = o(1)$. (iii) Q *and* Q_n *are continuous w.r.t.* $\|\cdot\|$ *in* Θ *and* Θ_n, *respectively*. (iv) $\sup_{\theta\in\Theta_n}|Q_n(\theta) - Q(\theta)| = o_p(1)$. *Then for* $\hat\theta_n \in \arg\min_{\theta\in\Theta_n} Q_n(\theta)$, *it follows that* $\min_{\theta\in\Theta_0}\|\hat\theta_n - \theta\| = o_p(1)$.

PROOF: Let Θ_0^δ denote an open δ enlargement of Θ_0 under $\|\cdot\|$. By compactness of Θ and continuity of Q, we have

$$(95)\qquad \Delta \equiv \min_{\theta\in(\Theta_0^\delta)^c\cap\Theta} Q(\theta) > 0.$$

Fix $\theta_0 \in \Theta_0$ and let $\theta_{0n} \in \Theta_n$ be such that $\|\theta_{0n} - \theta_0\| = o(1)$. Since $Q_n(\hat\theta_n) \le Q_n(\theta_{0n})$, it follows from (iv) that

$$(96)\qquad \lim_{n\to\infty} P\biggl(Q(\hat\theta_n) < Q(\theta_{0n}) + \frac{\Delta}{2}\biggr) = 1.$$

By continuity, $Q(\theta_0) = 0$, and $\|\theta_{0n} - \theta_0\| = o(1)$, we have $Q(\theta_{0n}) < \frac{\Delta}{2}$ for n sufficiently large. Therefore, (96) implies

$$(97)\qquad \lim_{n\to\infty} P(Q(\hat\theta_n) < \Delta) = 1.$$

By (95) and (97), with probability tending to 1, $\min_{\theta\in\Theta_0}\|\hat{\theta}_n-\theta\|\leq\delta$ and the lemma follows. *Q.E.D.*

LEMMA A.6: *If Assumptions* 2.1(i) *and* (ii), 3.1(i) *and* (ii), *and* 3.2(i) *and* (ii) *hold*, *then uniformly in* $(t,\theta)\in T\times\Theta$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{(y_i-\theta(x_i))w(t,z_i)-E[(Y-\theta(X))w(t,Z)]\}\xrightarrow{\mathcal{L}}G(t,\theta),$$

where G is a tight Gaussian process on $L^\infty(T\times\Theta)$.

PROOF: Since $\theta\in\Theta$ and $w(t,z)$ are uniformly bounded by Assumption 3.1(i), we get by Assumption 3.2(ii),

$$E[(Y-\theta(X))^2w^2(t,Z)]<\infty. \tag{98}$$

Hence, the central limit theorem implies the convergence of marginals. To verify uniform asymptotic equicontinuity, let $\mathcal{F}\equiv\{f:\mathbf{R}\times\mathcal{X}\times\mathcal{Z}\to\mathbf{R}:f(y,x,z)=(y-\theta(x))w(t,z)\text{ for some }(t,\theta)\in T\times\Theta\}$ and $F(y)$ as in Lemma A.4. The first inequality in (94), $E[F^2(Y)]<\infty$, and $T\times\Theta$ totally bounded imply $N_{[]}(\mathcal{F},\|\cdot\|_{L^2},\varepsilon)=1$ for ε large. Thus, the first and second inequalities in (99) follow for some D by Lemma A.4 and the change of variables $u=\varepsilon/\|F\|_{L^2}$,

$$\begin{aligned}\int_0^\infty\sqrt{\log N_{[]}(\mathcal{F},\|\cdot\|_{L^2},\varepsilon)}\,d\varepsilon&\leq\int_0^D\sqrt{\log N_{[]}(\mathcal{F},\|\cdot\|_{L^2},\varepsilon)}\,d\varepsilon\\&\lesssim\int_0^{D/\|F\|_{L^2}}\sqrt{u^{-\nu}-d_t\log u}\,du,\end{aligned} \tag{99}$$

where $\nu=(m+\delta)d_x/(\delta m)$ if $\mathcal{X}$ is unbounded and $\nu=d_x/m$ if $\mathcal{X}$ is bounded. Since the integrand is dominated by the term $u^{-\nu}$ in a neighborhood of 0, $\nu<2$ by Assumption 2.1(i) implies that the integral in (99) is finite. Theorem 2.5.6 in van der Vaart and Wellner (1996) then implies that $\mathcal{F}$ is Donsker. *Q.E.D.*

LEMMA A.7: *Let A_n and B_n be sets with norms $\|\cdot\|_a$ and $\|\cdot\|_b$, and let $G_n:A_n\times B_n\to\mathbf{R}$, $F_n:A_n\times B_n\to\mathbf{R}$ be random functions. Assume* (i) *G_n and F_n are continuous in $(a,b)\in A_n\times B_n$ with probability* 1 *and B_n is compact*. (ii) *Further assume* $\sup_{a\in A_n}\max_{b\in B_n}(G_n(a,b)-F_n(a,b))^2=o_p(1)$ *and* (iii) $\inf_{a\in A_n}\max_{b\in B_n}F_n^2(a,b)=O_p(1)$. *Then*

$$\inf_{a\in A_n}\max_{b\in B_n}G_n^2(a,b)=\inf_{a\in A_n}\max_{b\in B_n}F_n^2(a,b)+o_p(1).$$

PROOF: Note that due to the compactness assumptions on B_n and continuity assumption on G_n and F_n, all maximums are indeed attained. Expanding the square and using Assumption (ii) we first obtain that

$$\begin{aligned}(100)\qquad &\inf_{a\in A_n}\max_{b\in B_n} G_n^2(a,b)\\ &\le \inf_{a\in A_n}\max_{b\in B_n}\big\{(G_n(a,b)-F_n(a,b))^2+F_n^2(a,b)\\ &\quad +2|F_n(a,b)||G_n(a,b)-F_n(a,b)|\big\}\\ &=\inf_{a\in A_n}\max_{b\in B_n}\big\{F_n^2(a,b)+2|F_n(a,b)||G_n(a,b)-F_n(a,b)|\big\}+o_p(1).\end{aligned}$$

Next we fix $\varepsilon_n \downarrow 0$ and choose $\hat{a}_n \in A_n$ such that the inequality

$$(101)\qquad \inf_{a\in A_n}\max_{b\in B_n} F_n^2(a,b) \ge \max_{b\in B_n} F_n^2(\hat{a}_n,b)-\varepsilon_n$$

holds. Combining (100) and (101) and using assumptions (ii) and (iii), we can then conclude that

$$\begin{aligned}(102)\qquad &\inf_{a\in A_n}\max_{b\in B_n} G_n^2(a,b)\\ &\le \max_{b\in B_n}\big\{F_n^2(\hat{a}_n,b)+2|F_n(\hat{a}_n,b)||G_n(\hat{a}_n,b)-F_n(\hat{a}_n,b)|\big\}+o_p(1)\\ &\le \Big\{\inf_{a\in A_n}\max_{b\in B_n} F_n^2(a,b)+\varepsilon_n\Big\}\\ &\quad +2\Big\{\inf_{a\in A_n}\max_{b\in B_n} F_n^2(a,b)+\varepsilon_n\Big\}^{1/2}\times o_p(1)+o_p(1)\\ &\le \inf_{a\in A_n}\max_{b\in B_n} F_n^2(a,b)+o_p(1).\end{aligned}$$

To conclude, notice that assumptions (ii) and (iii) imply $\inf_{a\in A_n}\max_{b\in B_n} G_n^2(a,b)=O_p(1)$. Following the same arguments as in (100) and (102), it is then possible to establish that

$$(103)\qquad \inf_{a\in A_n}\max_{b\in B_n} F_n^2(a,b) \le \inf_{a\in A_n}\max_{b\in B_n} G_n^2(a,b)+o_p(1).$$

The conclusion of the lemma then follows from (102) and (103). *Q.E.D.*

LEMMA A.8: *Let A and B be sets with norms $\|\cdot\|_a$ and $\|\cdot\|_b$, and let $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq L^\infty(A)$ with $\mathcal{F}_\infty$ the closure of $\bigcup \mathcal{F}_n$ under $\|\cdot\|_\infty$. If $g\in L^\infty(A\times B)$ and for some sequence $\{g_n\}_{n=1}^\infty \in L^\infty(A\times B)$ we have $\|g_n-g\|_\infty = o(1)$, then*

$$\lim_{n\to\infty}\inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g_n(a,b)-f(a)| = \inf_{b\in B, f\in\mathcal{F}_\infty}\sup_{a\in A}|g(a,b)-f(a)|.$$

PROOF: Applying the triangle inequality, we first decompose the problem by noting

$$\begin{aligned}(104)\qquad &\Big|\inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g_n(a,b)-f(a)| - \inf_{b\in B, f\in\mathcal{F}_\infty}\sup_{a\in A}|g(a,b)-f(a)|\Big| \\ &\quad\le \Big|\inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g(a,b)-f(a)| - \inf_{b\in B, f\in\mathcal{F}_\infty}\sup_{a\in A}|g(a,b)-f(a)|\Big| \\ &\qquad + \Big|\inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g_n(a,b)-f(a)| - \inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g(a,b)-f(a)|\Big|.\end{aligned}$$

Fix $\varepsilon > 0$ and let $\hat{f}_\infty \in \mathcal{F}_\infty$ satisfy $\inf_{b\in B, f\in\mathcal{F}_\infty}\sup_{a\in A}|g(a,b)-f(a)| \ge \inf_{b\in B}\sup_{a\in A}|g(a,b)-\hat{f}_\infty(a)| - \varepsilon$. Also let $\hat{f}_n^p \in \mathcal{F}_n$ be such that $\inf_{f\in\mathcal{F}_n}\|f-\hat{f}_\infty\|_\infty \ge \|\hat{f}_n^p - \hat{f}_\infty\|_\infty - \varepsilon$. It then follows by $\inf_{f\in\mathcal{F}_n}\|f-\hat{f}_\infty\|_\infty = o(1)$ that

$$\begin{aligned}(105)\qquad &\inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g(a,b)-f(a)| \\ &\quad\le \inf_{b\in B}\sup_{a\in A}|g(a,b)-\hat{f}_n^p(a)| \\ &\quad\le \inf_{b\in B}\Big\{\sup_{a\in A}|g(a,b)-\hat{f}_\infty(a)| + \sup_{a\in A}|\hat{f}_\infty(a)-\hat{f}_n^p(a)|\Big\} \\ &\quad\le \inf_{b\in B, f\in\mathcal{F}_\infty}\sup_{a\in A}|g(a,b)-f(a)| + 2\varepsilon + o(1).\end{aligned}$$

Since $\mathcal{F}_n \subseteq \mathcal{F}_\infty$ by hypothesis for all n and since ε was arbitrary, it then follows from (105) that

$$(106)\qquad \inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g(a,b)-f(a)| = \inf_{b\in B, f\in\mathcal{F}_\infty}\sup_{a\in A}|g(a,b)-f(a)| + o(1).$$

Next observe that since $\|g_n - g\|_\infty = o(1)$, we are able to conclude that

$$\begin{aligned}(107)\qquad &\inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g_n(a,b)-f(a)| \\ &\quad\le \inf_{b\in B, f\in\mathcal{F}_n}\Big\{\sup_{a\in A}|g(a,b)-f(a)| + \sup_{a\in A}|g_n(a,b)-g(a,b)|\Big\} \\ &\quad\le \inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g(a,b)-f(a)| + o(1).\end{aligned}$$

By similar manipulations,

$$\inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g(a,b)-f(a)| \le \inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g_n(a,b)-f(a)| + o(1)$$

and, hence,

$$(108)\qquad \inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g_n(a,b)-f(a)| = \inf_{b\in B, f\in\mathcal{F}_n}\sup_{a\in A}|g(a,b)-f(a)| + o(1).$$

The claim of the lemma then follows by (104), (106), and (108). *Q.E.D.*

LEMMA A.9: *Let Assumptions* 2.1(i) *and* (ii), 2.2, 3.1(i) *and* (ii), 3.2(i) *and* (ii), 3.3(ii), *and* 3.4(i) *and* (ii) *hold, and let* $\Theta_0 \cap R \neq \emptyset$. *Define*

$$\hat{\theta}_n \in \arg \min_{\theta \in \Theta_n \cap R} \max_{t \in T} \left(\frac{1}{n} \sum_{i=1}^{n} (y_i - \theta(x_i)) w(t, z_i) \right)^2 .$$

Then (i) $\min_{\theta \in \Theta_0 \cap R} \|\hat{\theta}_n - \theta\|_{L^2} = o_p(1)$ *and* (ii) $\min_{\theta \in \Theta_n \cap R} \max_{t \in T_n} (\frac{1}{\sqrt{n}} \sum_i u_i(t, \theta))^2 = \max_{t \in T} (\frac{1}{\sqrt{n}} \sum_i u_i(t, \hat{\theta}_n))^2 + o_p(1)$.

PROOF: Throughout the proof, note that continuity and compactness imply that all minimums and maximums are indeed attained. To establish the first claim of the lemma, we define the criterion functions

$$(109) \qquad Q(\theta) \equiv \max_{t \in T} \big(E\big[(Y - \theta(X)) w(t, Z)\big]\big)^2,$$

$$Q_n(\theta) \equiv \max_{t \in T} \left(\frac{1}{n} \sum_{i=1}^{n} (y_i - \theta(x_i)) w(t, z_i) \right)^2 .$$

By Lemma 3.1, $\Theta_0 \cap R = \{\theta \in \Theta \cap R : Q(\theta) = 0\}$. By Lemma A.2, Θ is compact under $\|\cdot\|_c$ and since R is closed by continuity of L, it follows that $\Theta \cap R$ is compact in $\|\cdot\|_c$ and hence in $\|\cdot\|_{L^2}$. Let $\mathcal{F} \equiv \{f : \mathbf{R} \times \mathcal{X} \times \mathcal{Z} \to \mathbf{R} : f(y, x, z) = (y - \theta(x)) w(t, z)$ for some $(t, \theta) \in T \times \Theta\}$, and note that by Lemma A.4, we have $N_{[]}(\mathcal{F}, \|\cdot\|_{L^2}, \varepsilon) < \infty$ for any $\varepsilon > 0$. Hence, Theorem 2.4.1 in van der Vaart and Wellner (1996) implies that as a process in $L^\infty(T \times \Theta)$,

$$(110) \qquad \frac{1}{n} \sum_{i=1}^{n} (y_i - \theta(x_i)) w(t, z_i) = E\big[(Y - \theta(X)) w(t, Z)\big] + o_p(1).$$

Define the mapping $M : L^\infty(T \times \Theta) \to L^\infty(\Theta)$ by $M(g)(\theta) = \max_{t \in T} |g(\theta, t)|$. Since $g \mapsto M(g)$ is continuous, we have by the continuous mapping theorem and (110), that as a process in $L^\infty(\Theta)$,

$$(111) \qquad \max_{t \in T} \left| \frac{1}{n} \sum_{i=1}^{n} (y_i - \theta(x_i)) w(t, z_i) \right| = \max_{t \in T} \big| E\big[(Y - \theta(X)) w(t, Z)\big]\big| + o_p(1).$$

The first claim of the lemma then follows by Lemma A.5 and Assumption 3.4(i).

To verify the second claim of the lemma, set $\delta_n \downarrow 0$ so that $\sup_{t\in T}\|t-\Pi_n t\| = o(\delta_n/\sqrt{n})$, which is possible due to Assumption 3.3(ii). It then follows that

$$
\begin{aligned}
(112)\qquad & \sup_{\theta\in\Theta_n\cap R}\ \sup_{\|t_1-t_2\|\le\delta_n/\sqrt{n}} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-\theta(x_i))(w(t_1,z_i)-w(t_2,z_i))\right)^2 \\
&\le 2\sup_{\theta\in\Theta_n\cap R}\ \sup_{\|t_1-t_2\|\le\delta_n/\sqrt{n}}\big(\sqrt{n}E\big[(Y-\theta(X))(w(t_1,Z)-w(t_2,Z))\big]\big)^2 \\
&\quad + o_p(1) \\
&\le \sup_{\|t_1-t_2\|\le\delta_n/\sqrt{n}} 2\big(\sqrt{n}E[F(Y)]\times\|t_1-t_2\|\big)^2 + o_p(1),
\end{aligned}
$$

where the first and second inequalities follow from Lemma A.6 and Lemma A.4, respectively, and $F(y)$ is as defined in Lemma A.4 and hence satisfies $E[F(Y)]<\infty$. Next, let us define

$$
(113)\qquad \hat{\theta}_n^s \in \arg\min_{\theta\in\Theta_n\cap R}\ \max_{t\in T_n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-\theta(x_i))w(t,z_i)\right)^2
$$

and noting that expanding the square, exploiting (112), and nothing that $\delta_n \downarrow 0$, we can then obtain the result

$$
\begin{aligned}
(114)\qquad & \min_{\theta\in\Theta_n\cap R}\ \max_{t\in T}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-\theta(x_i))w(t,z_i)\right)^2 \\
&\le \max_{t\in T}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-\hat{\theta}_n^s(x_i))w(\Pi_n t,z_i)\right)^2 \\
&\quad + 2\max_{t\in T}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-\hat{\theta}_n^s(x_i))w(\Pi_n t,z_i)\right| \\
&\quad \times \max_{t\in T}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-\hat{\theta}_n^s(x_i))(w(t,z_i)-w(\Pi_n t,z_i))\right| + o_p(1).
\end{aligned}
$$

Next fix $\tilde{\theta}\in\Theta_0\cap R$ and let $\Pi_n\tilde{\theta}$ be as in Assumption 3.4(i). It then follows from $\Pi_n\tilde{\theta}\in\Theta_n\cap R$ that

$$
\begin{aligned}
(115)\qquad & \min_{\theta\in\Theta_n\cap R}\ \max_{t\in T_n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-\theta(x_i))w(t,z_i)\right)^2 \\
&\le 2\max_{t\in T}\big(\sqrt{n}E\big[(Y-\Pi_n\tilde{\theta}(X))w(t,Z)\big]\big)^2
\end{aligned}
$$

$$+2\max_{t\in T}\Bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big\{(y_i-\Pi_n\tilde{\theta}(x_i))w(t,z_i)$$
$$-E\big[(Y-\Pi_n\tilde{\theta}(X))w(t,Z)\big]\big\}\Bigg)^2.$$

Since by Assumption 3.4(ii), we have $\|\tilde{\theta}-\Pi_n\tilde{\theta}\|_w=o(n^{-1/2})$, we conclude from (115) and Lemma A.6 that

$$\min_{\theta\in\Theta_n\cap R}\max_{t\in T_n}\Bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-\theta(x_i))w(t,z_i)\Bigg)^2=O_p(1). \tag{116}$$

Thus, combining, (112), (114), (116), and $\Pi_n t\in T_n$ implies the inequality in (117). The equality in turn follows by definition of $\hat{\theta}_n^s$ in (113) and noting that $T_n\subseteq T$ implies that $T_n=\{\Pi_n t: t\in T\}$:

$$\begin{aligned}
&\min_{\theta\in\Theta_n\cap R}\max_{t\in T}\Bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-\theta(x_i))w(t,z_i)\Bigg)^2 \\
&\quad\le\max_{t\in T}\Bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-\hat{\theta}_n^s(x_i))w(\Pi_n t,z_i)\Bigg)^2+o_p(1) \\
&\quad=\min_{\theta\in\Theta_n\cap R}\max_{t\in T_n}\Bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i-\theta(x_i))w(t,z_i)\Bigg)^2+o_p(1).
\end{aligned} \tag{117}$$

Since $T_n\subseteq T$, (117) establishes the second claim of the lemma. *Q.E.D.*

PROOF OF THEOREM 3.1: Without further mention, we note that compactness of $\Theta_n\cap R$, $\Theta\cap R$, T_n, and T under $\|\cdot\|_\infty$ and $\|\cdot\|$, respectively, implies all relevant mins and maxes are indeed attained. To establish the asymptotic distribution, we need to analyze the test statistic in terms of local parameters to $\Theta_0\cap R$. As a first step, define

$$\hat{\theta}_n\in\arg\min_{\theta\in\Theta_n\cap R}\max_{t\in T}\Bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_i(t,\theta)\Bigg)^2. \tag{118}$$

For any random variable Z_n recall that $|Z_n|=o_p(1)$ implies the existence of a $\delta_n\downarrow 0$ such that $|Z_n|=o_p(\delta_n)$. Therefore, $\min_{\theta\in\Theta_0\cap R}\|\hat{\theta}_n-\theta\|_{L^2}=o_p(1)$ and

$\sup_{\theta\in\Theta\cap R}\|\theta-\Pi_n\theta\|_{L^2}=o(1)$ by Lemma A.9(i) and Assumption 3.4(i), imply

$$\begin{aligned}(119)\qquad &\min_{\theta\in\Theta_0\cap R}\|\hat{\theta}_n-\theta\|_{L^2}=o_p(\delta_n),\\ &\sup_{\theta\in\Theta\cap R}\|\theta-\Pi_n\theta\|_{L^2}=o(\delta_n),\\ &n^{-1/2}=o(\delta_n)\end{aligned}$$

for some $\delta_n\downarrow 0$ sufficiently slowly. Furthermore, also note that Assumption 3.4(iii) implies there is a $\gamma_n\downarrow 0$ such that

$$(120)\qquad \sup_{\theta\in\Theta_0\cap R}\|\Pi_n\theta\|_s\le B-\gamma_n.$$

Defining $B_n^{\delta_n}(\tilde{\theta})\equiv\{\theta\in\Theta_n\cap R:\|\tilde{\theta}-\theta\|_{L^2}\le\delta_n\}$ for all $\tilde{\theta}\in\Theta_0\cap R$, Lemma A.9(ii) and (119) then imply that

$$\begin{aligned}(121)\qquad &\min_{\theta\in\Theta_n\cap R}\max_{t\in T_n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i(t,\theta)\right)^2\\ &\quad=\min_{\theta\in\Theta_n\cap R}\max_{t\in T}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i(t,\theta)\right)^2+o_p(1)\\ &\quad=\min_{\tilde{\theta}\in\Theta_0\cap R}\min_{\theta\in B_n^{\delta_n}(\tilde{\theta})}\max_{t\in T}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i(t,\theta)\right)^2+o_p(1).\end{aligned}$$

By Lemma A.6, the empirical process induced by $\mathcal{F}\equiv\{f:\mathbf{R}\times\mathcal{X}\times\mathcal{Z}\to\mathbf{R}:f(y,x,z)=(y-\theta(x))w(t,z)$ with $(\theta,t)\in\Theta\times T\}$ is asymptotically uniformly $\|\cdot\|_{L^2}$-equicontinuous in probability. In particular, since $\delta_n\downarrow 0$, it follows that

$$\begin{aligned}(122)\qquad &\sup_{\|\theta_1-\theta_2\|_{L^2}\le\delta_n}\sup_{t\in T}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\big\{(\theta_1(x_i)-\theta_2(x_i))w(t,z_i)\right.\\ &\qquad\left.-E\big[(\theta_1(X)-\theta_2(X))w(t,Z)\big]\big\}\right|=o_p(1).\end{aligned}$$

Next, using simple manipulations and (121), we can derive the first equality in (123). The second equality in (123) is then implied by Lemma A.7, whose conditions are satisfied by (122), (116), and (117). Similarly, since by Assumption 3.4(i), we have that $\sup_{\tilde{\theta}\in\Theta_0\cap R}\|\tilde{\theta}-\Pi_n\tilde{\theta}\|_w=o(n^{-1/2})$, Lemma A.7 establishes the final equality in (123):

$$\begin{aligned}
(123)\qquad & \min_{\theta\in\Theta_n\cap R}\max_{t\in T_n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_i(t,\theta)\right)^2 \\
&= \min_{\tilde\theta\in\Theta_0\cap R}\min_{\theta\in B_n^{\delta_n}(\tilde\theta)}\max_{t\in T}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_i(t,\tilde\theta)\right. \\
&\qquad \left.+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde\theta(x_i)-\theta(x_i))w(t,z_i)\right)^2+o_p(1) \\
&= \min_{\tilde\theta\in\Theta_0\cap R}\min_{\theta\in B_n^{\delta_n}(\tilde\theta)}\max_{t\in T}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_i(t,\tilde\theta)\right. \\
&\qquad \left.+\sqrt{n}E\big[(\tilde\theta(X)-\theta(X))w(t,Z)\big]\right)^2+o_p(1) \\
&= \min_{\tilde\theta\in\Theta_0\cap R}\min_{\theta\in B_n^{\delta_n}(\tilde\theta)}\max_{t\in T}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_i(t,\tilde\theta)\right. \\
&\qquad \left.+\sqrt{n}E\big[(\Pi_n\tilde\theta(X)-\theta(X))w(t,Z)\big]\right)^2+o_p(1).
\end{aligned}$$

We now study the right-hand side of (123) in terms of local parameters. Note that any $\theta\in B_n^{\delta_n}(\tilde\theta)$ is of the form $\theta=\Pi_n\tilde\theta+p^{k_n'}h/\sqrt{n}$ for some $h\in\mathbf{R}^{k_n}$. Let us define a subset of $\mathcal{H}_{k_n}$ given by

$$(124)\qquad \tilde{\mathcal{H}}_{k_n}^{C_n}\equiv\big\{h\in\mathcal{H}_{k_n}:h'\Lambda_n h\le C_n^2,\ \|h\|\le C_n\big\}.$$

We first proceed by arguing that for an appropriately chosen sequence $C_n\uparrow\infty$, it will follow that for all $\tilde\theta\in\Theta_0\cap R$,

$$(125)\qquad \theta\in B_n^{\delta_n}(\tilde\theta)\quad \forall\theta=\Pi_n\tilde\theta+\frac{p^{k_n'}h}{\sqrt{n}}\quad\text{with}\quad h\in\tilde{\mathcal{H}}_{k_n}^{C_n}.$$

Any $h\in\tilde{\mathcal{H}}_{k_n}^{C_n}$ satisfies $\|h\|\le C_n$ and hence $\sup_{\tilde\theta\in\Theta_0\cap R}\|\tilde\theta-\Pi_n\tilde\theta\|_{L^2}=o(\delta_n)$ by (119), and the eigenvalues of $E[p^{k_n}(X)p^{k_n'}(X)]$ being uniformly bounded by Assumption 3.3(i) yield that for some $M>0$ and n sufficiently large,

$$(126)\qquad \sup_{\tilde\theta\in\Theta_0\cap R}\left\|\tilde\theta-\Pi_n\tilde\theta-\frac{p^{k_n'}h}{\sqrt{n}}\right\|_{L^2}\le\frac{\delta_n}{2}+\left\|\frac{p^{k_n'}h}{\sqrt{n}}\right\|_{L^2}\le\frac{\delta_n}{2}+M\frac{C_n}{\sqrt{n}}.$$

Similarly, noting that $\|p^{k_n'}h\|_s^2 = h'\Lambda_n h \le C_n^2$ for any $h \in \tilde{\mathcal{H}}_{k_n}^{C_n}$ and using $\sup_{\tilde\theta\in\Theta_0\cap R}\|\Pi_n\tilde\theta\|_s \le B-\gamma_n$, we obtain

$$(127)\qquad \sup_{\tilde\theta\in\Theta_0\cap R}\left\|\Pi_n\tilde\theta + \frac{p^{k_n'}h}{\sqrt{n}}\right\|_s \le B-\gamma_n + \left\|\frac{p^{k_n'}h}{\sqrt{n}}\right\|_s \le B-\gamma_n+\frac{C_n}{\sqrt{n}}.$$

Since $n^{-1/2}\gamma_n^{-1} = o(1)$ and $n^{-1/2}\delta_n^{-1} = o(1)$ by Assumption 3.4(iii) and (119), we can set $C_n \uparrow \infty$ slow enough so that

$$(128)\qquad \frac{C_n}{\sqrt{n}} = o(\delta_n), \quad \frac{C_n}{\sqrt{n}} = o(\gamma_n).$$

For such C_n and n large, (i) inequalities (126) imply $\sup_{\tilde\theta\in\Theta_0\cap R}\|\tilde\theta - \Pi_n\tilde\theta - p^{k_n'}h/\sqrt{n}\|_{L^2} \le \delta_n$, and (ii) (127) yields $\sup_{\tilde\theta\in\Theta_0\cap R}\|\Pi_n\tilde\theta + p^{k_n'}h/\sqrt{n}\|_s \le B$, and (iii) by linearity of L, $L(\Pi_n\tilde\theta + p^{k_n'}h/\sqrt{n}) = l$ iff $L(p^{k_n'}h) = 0$. We conclude that for n large, $\Pi_n\tilde\theta + p^{k_n'}h/\sqrt{n} \in \Theta_n\cap R$ for all $(\tilde\theta, h)\in\Theta_0\cap R\times\tilde{\mathcal{H}}_{k_n}^{C_n}$, which establishes (125). Therefore, exploiting (123),

$$(129)\qquad \inf_{\tilde\theta\in\Theta_0\cap R, h\in\mathcal{H}_{k_n}}\max_{t\in T}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i(t,\tilde\theta) + E\big[w(t,Z)p^{k_n'}(X)h\big]\right)^2$$
$$\le \min_{\theta\in\Theta_n\cap R}\max_{t\in T_n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i(t,\theta)\right)^2 + o_p(1)$$
$$\le \min_{\tilde\theta\in\Theta_0\cap R, h\in\tilde{\mathcal{H}}_{k_n}^{C_n}}\max_{t\in T}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i(t,\tilde\theta) + E\big[w(t,Z)p^{k_n'}(X)h\big]\right)^2,$$

where the lower bound follows by $L(\Pi_n\tilde\theta + p^{k_n'}h/\sqrt{n}) = l$ if and only if $L(p^{k_n'}h) = 0$. Define the set of functions $V_{k_n}^{C_n}(T) \equiv \{v: T\to\mathbf{R}\text{ s.t. } v(t) = E[w(t,Z)p^{k_n'}(X)h], h\in\tilde{\mathcal{H}}_{k_n}^{C_n}\}$, and note that (129) may alternatively be rewritten

$$(130)\qquad \inf_{\tilde\theta\in\Theta_0\cap R, v\in V_{k_n}(T)}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i(\cdot,\tilde\theta) - v\right\|_\infty^2$$
$$\le \min_{\theta\in\Theta_n\cap R}\max_{t\in T_n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i(t,\theta)\right)^2 + o_p(1)$$
$$\le \inf_{\tilde\theta\in\Theta_0\cap R, v\in V_{k_n}^{C_n}(T)}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i(\cdot,\tilde\theta) - v\right\|_\infty^2.$$

Moreover, by Lemma A.6, it also follows that for G a Gaussian process on $L^\infty(T \times \Theta_0 \cap R)$, we must have that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(t,\theta) \xrightarrow{\mathcal{L}} G(t,\theta). \tag{131}$$

Also notice that since $V_{k_n}(T) \subseteq V_{k_n+1}(T)$, it follows from $C_n \uparrow \infty$ that $\bigcup V_{k_n}^{C_n}(T) = \bigcup V_{k_n}(T)$. Therefore, while $V_\infty(T)$ is the closure of $\bigcup V_{k_n}(T)$ under $\|\cdot\|_\infty$ by definition, it also follows that it is the closure of $\bigcup V_{k_n}^{C_n}(T)$. Hence, applying Theorem 1.11.1 in van der Vaart and Wellner (1996), Lemma A.8, and result (131), we can conclude that

$$\begin{aligned} &\inf_{\tilde\theta \in \Theta_0 \cap R, v \in V_{k_n}(T)} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(\cdot,\tilde\theta) - v \right\|_\infty^2 \\ &\quad \xrightarrow{\mathcal{L}} \inf_{\tilde\theta \in \Theta_0 \cap R, v \in V_\infty(T)} \|G(\cdot,\tilde\theta) - v\|_\infty^2, \\ &\inf_{\tilde\theta \in \Theta_0 \cap R, v \in V_{k_n}^{C_n}(T)} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i(\cdot,\tilde\theta) - v \right\|_\infty^2 \\ &\quad \xrightarrow{\mathcal{L}} \inf_{\tilde\theta \in \Theta_0 \cap R, v \in V_\infty(T)} \|G(\cdot,\tilde\theta) - v\|_\infty^2. \end{aligned} \tag{132}$$

Thus, the first claim of the theorem follows from the inequalities in (130) and the results in (132).

To establish the second claim of the theorem, let $\mathcal{F}$ and $F(y)$ be as in Lemma A.4 and notice that Lemma A.4 implies $N_{[]}(\mathcal{F}, \|\cdot\|_{L^1}, \varepsilon) < \infty$ for all $\varepsilon > 0$. It then follows by Theorem 2.4.1 in van der Vaart and Wellner (1996) that

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \theta(x_i)) w(t, z_i) \xrightarrow{\text{a.s.}} E\big[(Y - \theta(X)) w(t, Z)\big] \tag{133}$$

as a process in $L^\infty(T \times \Theta)$. The continuous mapping theorem therefore implies

$$\begin{aligned} &\min_{\theta \in \Theta \cap R} \max_{t \in T} \left(\frac{1}{n} \sum_{i=1}^{n} (y_i - \theta(x_i)) w(t, z_i) \right)^2 \\ &\quad \xrightarrow{\text{a.s.}} \min_{\theta \in \Theta \cap R} \max_{t \in T} \big(E\big[(Y - \theta(X)) w(t, Z)\big]\big)^2. \end{aligned} \tag{134}$$

Next let $\hat\theta_n \in \arg\min_{\theta \in \Theta \cap R} \max_{t \in T} (\frac{1}{n} \sum_i u_i(t,\theta))^2$ and let $\Pi_n \hat\theta_n$ be as in Assumption 3.4(i). Further defining the maximizer $\hat t_n \in \arg\max_{t \in T_n} (\frac{1}{n} \sum_i u_i(t, \Pi_n \hat\theta_n))^2$

and using that $T_n \subseteq T$, we then obtain

$$\begin{aligned}(135)\qquad &\min_{\theta\in\Theta_n\cap R}\max_{t\in T_n}\left(\frac{1}{n}\sum_{i=1}^n u_i(t,\theta)\right)^2 - \min_{\theta\in\Theta\cap R}\max_{t\in T}\left(\frac{1}{n}\sum_{i=1}^n u_i(t,\theta)\right)^2\\ &\quad\le \left(\frac{1}{n}\sum_{i=1}^n u_i(\hat{t}_n, \Pi_n\hat{\theta}_n)\right)^2 - \left(\frac{1}{n}\sum_{i=1}^n u_i(\hat{t}_n,\hat{\theta}_n)\right)^2\\ &\quad\le \sup_{\|t_1-t_2\|\le\delta_n, \|\theta_1-\theta_2\|_{L^2}\le\delta_n}\left|\left(\frac{1}{n}\sum_{i=1}^n u_i(t_1,\theta_1)\right)^2 - \left(\frac{1}{n}\sum_{i=1}^n u_i(t_2,\theta_2)\right)^2\right|,\end{aligned}$$

where we set $\delta_n \downarrow 0$ sufficiently slowly so that $\sup_{t\in T}\|t - \Pi_n t\| = o(\delta_n)$ and $\sup_{\theta\in\Theta\cap R}\|\theta - \Pi_n\theta\|_{L^2} = o(\delta_n)$, which is possible by Assumptions 3.3(ii) and 3.4(i). Furthermore, by similar manipulations as in (135), we can obtain

$$\begin{aligned}(136)\qquad &\min_{\theta\in\Theta_n\cap R}\max_{t\in T_n}\left(\frac{1}{n}\sum_{i=1}^n u_i(t,\theta)\right)^2 - \min_{\theta\in\Theta\cap R}\max_{t\in T}\left(\frac{1}{n}\sum_{i=1}^n u_i(t,\theta)\right)^2\\ &\quad\ge -\sup_{\|t_1-t_2\|\le\delta_n, \|\theta_1-\theta_2\|_{L^2}\le\delta_n}\left|\left(\frac{1}{n}\sum_{i=1}^n u_i(t_1,\theta_1)\right)^2 - \left(\frac{1}{n}\sum_{i=1}^n u_i(t_2,\theta_2)\right)^2\right|.\end{aligned}$$

In turn, applying the triangle inequality, we obtain the bound

$$\begin{aligned}(137)\qquad &\left|\left(\frac{1}{n}\sum_{i=1}^n u_i(t_1,\theta_1)\right)^2 - \left(\frac{1}{n}\sum_{i=1}^n u_i(t_2,\theta_2)\right)^2\right|\\ &\quad\le 2\sup_{\theta\in\Theta, t\in T}\left|\frac{1}{n}\sum_{i=1}^n u_i(t,\theta) - E[U(t,\theta)]\right| \times 2\sup_{\theta\in\Theta, t\in T}\left|\frac{1}{n}\sum_{i=1}^n u_i(t,\theta)\right|\\ &\quad\quad + \left|E[U(t_1,\theta_1)] - E[U(t_2,\theta_2)]\right| \times 2\sup_{\theta\in\Theta, t\in T}\left|\frac{1}{n}\sum_{i=1}^n u_i(t,\theta)\right|.\end{aligned}$$

Notice that $E[U(t,\theta)]$ is continuous in $(t,\theta)\in T\times\Theta$ with respect to the norm $\|\cdot\| + \|\cdot\|_{L^2}$ and that $T\times\Theta$ is compact. Hence $(t,\theta)\mapsto E[U(t,\theta)]$ is uniformly continuous, and, furthermore, as the class $\mathcal{F}$ is Glivenko–Cantelli and $\max_{\theta\in\Theta, t\in T}|E[U(t,\theta)]| < \infty$ by compactness and continuity, we obtain from (137) that

$$(138)\qquad \sup_{\|t_1-t_2\|\le\delta_n, \|\theta_1-\theta_2\|_{L^2}\le\delta_n}\left|\left(\frac{1}{n}\sum_{i=1}^n u_i(t_1,\theta_1)\right)^2 - \left(\frac{1}{n}\sum_{i=1}^n u_i(t_2,\theta_2)\right)^2\right| \xrightarrow{\text{a.s.}} 0.$$

Therefore, combining (135), (136), and (138), it follows that

$$\min_{\theta\in\Theta_n\cap R}\max_{t\in T_n}\left(\frac{1}{n}\sum_{i=1}^{n}u_i(t,\theta)\right)^2 \xrightarrow{\text{a.s.}} \min_{\theta\in\Theta\cap R}\max_{t\in T}\left(\frac{1}{n}\sum_{i=1}^{n}u_i(t,\theta)\right)^2. \tag{139}$$

The second claim of the theorem then follows from (134) and (139). *Q.E.D.*

LEMMA A.10: *Let Assumptions* 2.1 *and* 3.2(ii) *hold, let* $\mathcal{X}$ *be bounded, and let* $\mathcal{N}(L)$ *be the null space of* $L:W^s(\mathcal{X})\to L^2(X)$. *Then* (i) *the identity* $I:\mathcal{N}(L)\to L^2(X)$ *is compact and* (ii) *the eigenvalues* $\{\lambda_j^2\}_{j=1}^{\infty}$ *of* I^*I *satisfy* $\lambda_j^2\lesssim j^{-2(m+m_0)/d_x}$.

PROOF: Since $\mathcal{N}(L)\subseteq W^s(\mathcal{X})$ and $W^s(\mathcal{X})$ is compactly embedded in $L^2(X)$ by Lemma A.2, we immediately obtain that $I:\mathcal{N}(L)\to L^2(X)$ is indeed a compact operator. By Theorem 4.10 in Kress (1999), its adjoint I^* is also compact and hence so is $I^*I:\mathcal{N}(L)\to\mathcal{N}(L)$. For a linear operator P, the jth approximation number $a_j(P)$ is defined by

$$a_{j+1}(P)\equiv\inf\{\|P-S\|_o:\mathrm{rank}(S)\le j\}, \tag{140}$$

where $\|\cdot\|_o$ denotes the usual operator norm. By Theorem 2.1 in Gohberg and Krein (1998), we have $\lambda_j^2=a_j(I^*I)$. Since approximation numbers are submultiplicative and $a_j(I^*)=a_j(I)$ by Theorem 2.1 in Hutton (1974),

$$\lambda_{2j-1}^2=a_{2j-1}(I^*I)\le a_j(I^*)a_j(I)=a_j^2(I). \tag{141}$$

Let $I_E:W^s(\mathcal{X})\to L^2(X)$ denote the identity operator and note that I is the restriction of I_E to $\mathcal{N}(L)$. In particular, $\mathcal{N}(L)\subseteq W^s(\mathcal{X})$ implies that $a_j(I)\le a_j(I_E)$ for all j. Hence, since the eigenvalues are ordered, (141) implies

$$\lambda_{2j}^2\le\lambda_{2j-1}^2\le a_j^2(I)\le a_j^2(I_E). \tag{142}$$

To conclude, exploit Theorem 3.2.5 in Edmunds and Triebel (1992) and X having bounded density to conclude that $a_j^2(I_E)\lesssim j^{-2(m+m_0)/d_x}$. The claim of the lemma then follows from (142). *Q.E.D.*

LEMMA A.11: *If Assumptions* 2.1(i) *and* (ii), 3.1(i) *and* (ii), 3.2(i) *and* (ii), *and* 3.5(i) *hold, then it follows that*

$$\sup_{\theta\in\Theta,t\in T}\left|\frac{\sqrt{\lambda_n}}{n}\sum_{i=1}^{n}u_i(t,\theta)-E[U(t,\theta)]\right|\xrightarrow{a.s.}0.$$

PROOF: Endow $T\times\Theta$ with the norm $\|\cdot\|+\|\cdot\|_\infty$ and let $C(T\times\Theta)$ denote the set of continuous functions on $T\times\Theta$, which is a separable Banach space under $\|\cdot\|_\infty$. Also define the map $\Psi:\mathbf{R}\times\mathcal{X}\times\mathcal{Z}\to C(T\times\Theta)$ pointwise by

$$\Psi(y,x,z)(t,\theta)=(y-\theta(x))w(t,z). \tag{143}$$

Since Θ is compact under $\|\cdot\|_\infty$ by Lemma A.2, it follows from the Arzela–Ascoli theorem that the $\theta \in \Theta$ are equicontinuous and uniformly bounded. It then follows from $w(\cdot,\cdot)$ being uniformly continuous and bounded that the mapping Ψ is continuous and thus measurable. Therefore, as $U(\cdot,\cdot) = \Psi(Y, X, Z)$, we conclude that $U(\cdot,\cdot)$ is a random variable in $C(T\times\Theta)$. Furthermore, for $F(y)$ as defined in Lemma A.4, we have $E[\|U(\cdot,\cdot)\|_\infty^2] \leq E[F^2(Y)] < \infty$ and hence since $U(\cdot,\cdot)$ obeys a central limit theorem in $C(T\times\Theta)$ by Lemma A.6 and Theorem 1.3.10 in van der Vaart and Wellner (1996), Theorem 4.1 in Goodman, Kuelbs, and Zinn (1981) establishes that $U(\cdot,\cdot)$ satisfies a compact law of iterated logarithms. The claim of the lemma then follows from $\sqrt{\lambda_n/n} = o((\log(\log(n)))^{-1/2})$. *Q.E.D.*

LEMMA A.12: *Let Assumptions* 3.1(i) *and* (ii), 3.2(i) *and* (ii), *and* 3.5(ii) *hold. Then it follows that*

$$\max_{h\in\mathcal{H}_{b_n}^{B_n}, t\in T}\left|\frac{1}{n}\sum_{i=1}^{n} w(t, z_i)p^{b_n'}(x_i)h - E\big[w(t, Z)p^{b_n'}(X)h\big]\right| \xrightarrow{a.s.} 0.$$

PROOF: By Assumption 3.1(i) and (ii), there exists an $M > 0$ such that uniformly in z,

$$|w(t_1, z) - w(t_2, z)| \leq M\|t_1 - t_2\| \tag{144}$$

and, in addition, $w(t, z)$ is bounded by M uniformly in (t, z). Next define the class of functions

$$\mathcal{F}_n \equiv \left\{f : \mathcal{X}\times\mathcal{Z} \to \mathbf{R} : f(x, z) = \frac{w(t, z)}{2M\xi_{b_n}} p^{b_n'}(x)\frac{h}{B_n} + \frac{1}{2}\right. \\ \left. \text{for some } t \in T, h \in \mathbf{R}^{b_n}, \|h\| \leq B_n\right\}. \tag{145}$$

By the Cauchy–Schwarz inequality, it follows that $|p^{b_n'}(x)h| \leq \xi_{b_n}\|h\|$ uniformly in x. Therefore, (144) and the triangle inequality imply that for any two functions in $\mathcal{F}_n$, we must have

$$\begin{aligned} &\left|\frac{w(t_1, z)}{2M\xi_{b_n}} p^{b_n'}(x)\frac{h_1}{B_n} - \frac{w(t_2, z)}{2M\xi_{b_n}} p^{b_n'}(x)\frac{h_2}{B_n}\right| \\ &\quad \leq \frac{1}{2M}|w(t_1, z) - w(t_2, z)|\left|p^{b_n'}(x)\frac{h_1}{\xi_{b_n}B_n}\right| \\ &\qquad + \frac{1}{2M}|w(t_2, z)|\left|p^{b_n'}(x)\frac{h_1 - h_2}{\xi_{b_n}B_n}\right| \\ &\quad \leq \frac{1}{2}\left\{\|t_1 - t_2\| + \left\|\frac{h_1}{B_n} - \frac{h_2}{B_n}\right\|\right\}. \end{aligned} \tag{146}$$

Hence, from (146), the class $\mathcal{F}_n$ is Lipschitz in the parameter $(t, \tilde{h}) \in T \times U(b_n)$, where $U(b_n)$ is the unit ball in $\mathbf{R}^{b_n}$. From Theorem 2.7.11 in van der Vaart and Wellner (1996), it then follows that for $\varepsilon \leq 1$ and $b_n \geq d_t$,

$$
\begin{aligned}
N_{[]}(\mathcal{F}_n, \|\cdot\|_{L^2}, \varepsilon) &\leq N(\varepsilon, T \times U(b_n), \|\cdot\|) \\
&\lesssim \left(\frac{2\,\mathrm{diam}\, T}{\varepsilon}\right)^{d_t} \times \left(\frac{2}{\varepsilon}\right)^{b_n} \\
&\lesssim \left(\frac{2}{\varepsilon}\right)^{2b_n}.
\end{aligned} \tag{147}
$$

Furthermore, the same calculations as in (146) reveal that if $f \in \mathcal{F}_n$, then the image of f is contained in $[0, 1]$. Hence, the definition of $\mathcal{F}_n$ and Theorem 2.14.9 in van der Vaart and Wellner (1996) imply that for some $C > 0$,

$$
\begin{aligned}
&P\left(\max_{h \in \mathcal{H}_{b_n}^{B_n}, t \in T}\left|\frac{1}{n}\sum_{i=1}^{n} w(t, z_i) p^{b_n'}(x_i) h - E[w(t, Z) p^{b_n}(X) h]\right| > \varepsilon\right) \\
&\quad = P\left(\max_{f \in \mathcal{F}_n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f(x_i, z_i) - E[f(X, Z)]\right| > \frac{\varepsilon\sqrt{n}}{2M\xi_{b_n} B_n}\right) \\
&\quad \leq \left(\frac{C\varepsilon\sqrt{n}}{2M\xi_{b_n} B_n \sqrt{2b_n}}\right)^{2b_n} \times \exp\left\{-\frac{2\varepsilon^2 n}{4M^2 \xi_{b_n}^2 B_n^2}\right\}.
\end{aligned} \tag{148}
$$

Since by Assumption 3.5(ii), we have $n^{b_n+\alpha} \times \exp\{-\frac{2\varepsilon^2 n}{4M^2 \xi_n^2 B_n^2}\} \downarrow 0$ for any $\alpha > 0$, from (148) we can obtain

$$
\sum_{n=1}^{\infty} P\left(\max_{h \in \mathcal{H}_{b_n}^{B_n}, t \in T}\left|\frac{1}{n}\sum_{i=1}^{n} w(t, z_i) p^{b_n'}(x_i) h - E[w(t, Z) p^{b_n}(X) h]\right| > \varepsilon\right) < \infty. \tag{149}
$$

Hence, the Borel–Cantelli lemma and (149) establish the claim of the lemma. *Q.E.D.*

LEMMA A.13: *If Assumptions* 2.1(i) *and* (ii), 3.1(i) *and* (ii), *and* 3.2(i) *and* (ii) *hold, then as elements of* $L^\infty(T \times \Theta)$,

$$
G_n^* \xrightarrow{\mathcal{L}^*} G \quad a.s.,
$$

where G_n^* *is pointwise defined in* (57) *and* G *is the same limiting Gaussian process as in Lemma* A.6.

PROOF: By Lemma A.6, the class $\mathcal{F} \equiv \{f : \mathbf{R} \times \mathcal{X} \times \mathcal{Z} \to \mathbf{R} : f(y, x, z) = (y - \theta(x))w(t, z)$ for some $(\theta, t) \in \Theta \times T\}$ is Donsker. Further, by Lemma A.4 and Assumption 3.2(ii), $\mathcal{F}$ has envelope $F(y)$ satisfying $E[F^2(Y)] < \infty$. The conclusion then follows from Theorem 3.6.2 in van der Vaart and Wellner (1996). *Q.E.D.*

LEMMA A.14: *Define* $(\Theta_0^{\varepsilon_n})^c \equiv \{\theta \in \Theta : \|\theta - \theta_0\|_w > \varepsilon_n\}$. *If Assumptions* 2.1(i) *and* (ii), 3.1(i) *and* (ii), 3.2(i) *and* (ii), 3.3(ii), *and* 3.5(i) *hold, and, in addition,* $\lambda_n \times \varepsilon_n^2 \uparrow \infty$, *it then follows that*

$$\inf_{\theta \in \Theta_n \cap R \cap (\Theta_0^{\varepsilon_n})^c, h \in \mathcal{H}_{b_n}^{B_n}} \max_{t \in T_n} \big\{\big(G_n^*(t, \theta) - E^*\big[w(t, Z) p^{b_n'}(X) h\big]\big)^2 + \lambda_n P_n^*(t, \theta)\big\} \xrightarrow{p^*} \infty \quad a.s.$$

PROOF: Let $(\overline{\Theta_0^{\varepsilon_n}})^c$ denote the closure of $(\Theta_0^{\varepsilon_n})^c$ under $\|\cdot\|_c$. Since Θ is compact under $\|\cdot\|_c$ by Lemma A.2, it follows that $(\overline{\Theta_0^{\varepsilon_n}})^c$ is itself compact. Thus, we note without further mention that minimums are indeed attained. Next notice that since $\Theta_n \cap R \cap (\Theta_0^{\varepsilon_n})^c \subseteq (\overline{\Theta_0^{\varepsilon_n}})^c$, it follows that

$$\inf_{\theta \in \Theta_n \cap R \cap (\Theta_0^{\varepsilon_n})^c, h \in \mathcal{H}_{b_n}^{B_n}} \max_{t \in T_n} \big\{\big(G_n^*(t, \theta) - E^*[w(t, Z) p^{b_n'}(X) h]\big)^2 + \lambda_n P_n^*(t, \theta)\big\} \geq \min_{\theta \in (\overline{\Theta_0^{\varepsilon_n}})^c} \max_{t \in T_n} \lambda_n P_n^*(t, \theta). \tag{150}$$

From the definition of $P_n^*(t, \theta)$ in (58), the inequality $a^2/2 - (a-b)^2 \leq b^2$, and Lemma A.11, we further obtain

$$\begin{aligned} &\min_{\theta \in (\overline{\Theta_0^{\varepsilon_n}})^c} \max_{t \in T_n} \lambda_n P_n^*(t, \theta) \\ &\quad \geq \min_{\theta \in (\overline{\Theta_0^{\varepsilon_n}})^c} \max_{t \in T_n} \left\{ \frac{\lambda_n}{2} \big(E[U(t, \theta)]\big)^2 - \lambda_n \left(\frac{1}{n} \sum_{i=1}^{n} u_i(t, \theta) - E[U(t, \theta)] \right)^2 \right\} \\ &\quad \geq \min_{\theta \in (\overline{\Theta_0^{\varepsilon_n}})^c} \max_{t \in T_n} \frac{\lambda_n}{2} \big(E[U(t, \theta)]\big)^2 + o_{as}(1). \end{aligned} \tag{151}$$

For $\Pi_n t$ as in Assumption 3.3(ii), note that $T_n \subseteq T$ implies $T_n = \{\Pi_n t : t \in T\}$. This yields the equality in (152). The two inequalities then follow from $a^2/2 -$

$(a-b)^2 \le b^2$, the definition of $(\Theta_0^{\varepsilon_n})^c$, and Lemma A.4:

$$\begin{aligned}
(152)\qquad & \min_{\theta\in(\overline{\Theta_0^{\varepsilon_n}})^c} \max_{t\in T_n} \lambda_n \big(E[U(t,\theta)]\big)^2 \\
&= \min_{\theta\in(\overline{\Theta_0^{\varepsilon_n}})^c} \max_{t\in T} \lambda_n \big(E\big[(Y-\theta(X))w(\Pi_n t, Z)\big]\big)^2 \\
&\ge \min_{\theta\in(\overline{\Theta_0^{\varepsilon_n}})^c} \max_{t\in T} \frac{\lambda_n}{2} \big(E\big[(Y-\theta(X))w(t,Z)\big]\big)^2 \\
&\quad - \max_{\theta\in(\overline{\Theta_0^{\varepsilon_n}})^c} \max_{t\in T} \lambda_n \big(E\big[(Y-\theta(X))(w(t,Z)-w(\Pi_n t, Z))\big]\big)^2 \\
&\ge \frac{\lambda_n}{2} \times \varepsilon_n^2 - \lambda_n \times \sup_{t\in T} \|t-\Pi_n t\|^2 \times \big(E[F(Y)]\big)^2.
\end{aligned}$$

By Assumption 3.3(ii), $\sup_{t\in T}\|t-\Pi_n t\|^2 = o(n^{-1})$, while Assumption 3.5(i) implies that $\lambda_n \times n^{-1} \downarrow 0$. Hence, $\lambda_n \times \sup_{t\in T}\|t-\Pi_n t\|^2 = o(1)$, and combining (150), (151), and (152), it follows that

$$\begin{aligned}
(153)\qquad & \inf_{\theta\in\Theta_n\cap R\cap(\Theta_0^{\varepsilon_n})^c, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T_n} \big\{\big(G_n^*(t,\theta) - E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2 \\
&\quad + \lambda_n P_n^*(t,\theta)\big\} \ge \frac{\lambda_n}{4} \times \varepsilon_n^2 + o_{as}(1).
\end{aligned}$$

Therefore, since $\lambda_n \times \varepsilon_n^2 \uparrow \infty$ by hypothesis, the claim of the lemma follows. *Q.E.D.*

LEMMA A.15: *Define* $\Theta_0^{\varepsilon_n} \equiv \{\theta\in\Theta : \|\theta-\theta_0\|_w \le \varepsilon_n\}$. *Let Assumptions* 2.1(i) *and* (ii), 2.2, 3.1(i) *and* (ii), 3.2(i) *and* (ii), 3.3(ii), *and* 3.5(ii) *hold, and suppose* $\Theta_0 \cap R \ne \emptyset$. *If, in addition,* $\varepsilon_n \downarrow 0$, *then it follows that*

$$\begin{aligned}
& \min_{\theta\in\Theta_0^{\varepsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T_n} \big(G_n^*(t,\theta) - E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2 \\
&= \min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T} \big(G_n^*(t,\theta) - E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2 \\
&\quad + o_{p^*}(1) \quad a.s.
\end{aligned}$$

PROOF: Fix $\delta > 0$ and define $\Theta_{0\|\cdot\|_{L^2}}^{\delta} \equiv \{\theta\in\Theta : \inf_{\theta\in\Theta_0}\|\theta-\theta_0\|_{L^2} \le \delta\}$. By compactness of $\Theta\cap R$ under $\|\cdot\|_{L^2}$ and continuity of $\max_{t\in T}|E[(Y-\theta(X))w(t,Z)]|$ in θ under $\|\cdot\|_{L^2}$ (due to the theorem of the maximum), it follows that

$$(154)\qquad \eta \equiv \inf_{\theta\in(\Theta_{0\|\cdot\|_{L^2}}^{\delta})^c\cap R} \max_{t\in T} \big|E\big[(Y-\theta(X))w(t,Z)\big]\big| > 0.$$

By (154), if $\max_{t\in T}|E[(Y-\theta(X))w(t,Z)]|<\eta$ for $\theta\in\Theta\cap R$, then $\inf_{\tilde\theta\in\Theta_0}\|\theta-\tilde\theta\|_{L^2}\leq\delta$. Therefore, for every $\theta\in\Theta_0^{\varepsilon_n}\cap R$, letting $\Pi_0\theta$ denote its projection onto $\Theta_0\cap R$ under $\|\cdot\|_{L^2}$, we obtain

$$\sup_{\theta\in\Theta_0^{\varepsilon_n}\cap R}\|\theta-\Pi_0\theta\|_{L^2}=o(1). \tag{155}$$

Next let $(\hat\theta_n,\hat h_n)\in\arg\min_{\theta\in\Theta_0^{\varepsilon_n}\cap R,h\in\mathcal{H}_{b_n}^{B_n}}\max_{t\in T_n}(G_n^*(t,\theta)-E^*[w(t,Z)\times p^{b_n'}(X)h])^2$ so as to obtain

$$\begin{aligned}
&\min_{\theta\in\Theta_0^{\varepsilon_n}\cap R,h\in\mathcal{H}_{b_n}^{B_n}}\max_{t\in T_n}\big(G_n^*(t,\theta)-E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2\\
&\quad\geq\max_{t\in T_n}\big(G_n^*(t,\Pi_0\hat\theta_n)-E^*\big[w(t,Z)p^{b_n'}(X)\hat h_n\big]\big)^2\\
&\qquad-\max_{t\in T_n}2|G_n^*(t,\Pi_0\hat\theta_n)-G_n^*(t,\hat\theta_n)|\\
&\qquad\times\big|G_n^*(t,\Pi_0\hat\theta_n)-E^*\big[w(t,Z)p^{b_n'}(X)\hat h_n\big]\big|\\
&\qquad-\max_{t\in T_n}(G_n^*(t,\Pi_0\hat\theta_n)-G_n^*(t,\hat\theta_n))^2.
\end{aligned} \tag{156}$$

Moreover, it follows from Lemma A.13 that G_n^* is asymptotically equicontinuous almost surely. Therefore, employing (155), $0\in\mathcal{H}_{b_n}^{B_n}$, and the fact that $(\hat\theta_n,\hat h_n)$ minimize $\max_{t\in T_n}(G_n^*(t,\theta)-E^*[w(t,Z)p^{b_n'}(X)h])^2$, we conclude

$$\begin{aligned}
&\max_{t\in T_n}|G_n^*(t,\Pi_0\hat\theta_n)-G_n^*(t,\hat\theta_n)||G_n^*(t,\Pi_0\hat\theta_n)-E^*[w(t,Z)p^{b_n'}(X)\hat h_n]|\\
&\quad\leq\max_{t\in T_n}(G_n^*(t,\Pi_0\hat\theta_n)-G_n^*(t,\hat\theta_n))^2+\max_{t\in T_n}|G_n^*(t,\Pi_0\hat\theta_n)-G_n^*(t,\hat\theta_n)|\\
&\qquad\times\max_{t\in T_n}|G_n^*(t,\hat\theta_n)-E^*[w(t,Z)p^{b_n'}(X)\hat h_n]|\\
&\quad\leq\max_{\theta\in\Theta,t\in T}|G_n^*(t,\theta)|\times o_{p^*}(1)+o_{p^*}(1).
\end{aligned} \tag{157}$$

Because Lemma A.13 further implies $\max_{\theta\in\Theta,t\in T}(G_n^*(t,\theta))^2=O_{p^*}(1)$ almost surely, using (156), (157), and that G_n^* is asymptotically equicontinuous almost surely together with $(\Pi_0\hat\theta_n,\hat h_n)\in\Theta_0\cap R\times\mathcal{H}_{b_n}^{B_n}$ establishes that

$$\begin{aligned}
&\min_{\theta\in\Theta_0^{\varepsilon_n}\cap R,h\in\mathcal{H}_{b_n}^{B_n}}\max_{t\in T_n}\big(G_n^*(t,\theta)-E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2\\
&\quad\geq\min_{\theta\in\Theta_0\cap R,h\in\mathcal{H}_{b_n}^{B_n}}\max_{t\in T_n}\big(G_n^*(t,\theta)-E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2+o_{p^*}(1).
\end{aligned} \tag{158}$$

Thus, since $\Theta_0 \subseteq \Theta_0^{\epsilon_n}$, and inequalities (156), (157), and (158) hold almost surely, we deduce that almost surely

$$\min_{\theta\in\Theta_0^{\epsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T_n} \bigl(G_n^*(t,\theta) - E^*\bigl[w(t,Z)p^{b_n'}(X)h\bigr]\bigr)^2 = \min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T_n} \bigl(G_n^*(t,\theta) - E^*\bigl[w(t,Z)p^{b_n'}(X)h\bigr]\bigr)^2 + o_{p^*}(1). \tag{159}$$

For $\Pi_n t$ as in Assumption 3.3(ii), the first inequality in (160) holds for M as in (144), while the second inequality is implied by the Cauchy–Schwarz inequality and $h \in \mathcal{H}_{b_n}^{B_n}$ being uniformly bounded by B_n:

$$\begin{aligned}\sup_{t\in T, h\in\mathcal{H}_{b_n}^{B_n}} &\bigl|E^*\bigl[w(t,Z)p^{b_n'}(X)h\bigr] - E^*\bigl[w(\Pi_n t,Z)p^{b_n'}(X)h\bigr]\bigr| \\ &\le \sup_{t\in T} M\|t-\Pi_n t\| \times \sup_{h\in\mathcal{H}_{b_n}^{B_n}, x\in\mathcal{X}} \bigl|p^{b_n'}(x)h\bigr| \\ &\le \sup_{t\in T} M\|t-\Pi_n t\| \times \xi_{b_n} \times B_n.\end{aligned} \tag{160}$$

In addition, note that since $\sup_{t\in T}\|t-\Pi_n t\| = o(n^{-1/2})$ due to Assumption 3.3(ii) and $\xi_{b_n} \times B_n \times n^{-1/2} \downarrow 0$ due to Assumption 3.5(ii), it follows that $\sup_{t\in T}\|t-\Pi_n t\| \times \xi_{b_n} \times B_n = o(1)$. Furthermore, since by Lemma A.13, the process G_n^* is asymptotically equicontinuous almost surely on $T \times \Theta$, we conclude that almost surely

$$\begin{aligned}\sup_{\theta\in\Theta, t\in T, h\in\mathcal{H}_{b_n}^{B_n}} &\bigl|\bigl(G_n^*(t,\theta) - E^*\bigl[w(t,Z)p^{b_n'}(X)h\bigr]\bigr) \\ &\quad - \bigl(G_n^*(\Pi_n t,\theta) - E^*\bigl[w(\Pi_n t,Z)p^{b_n'}(X)h\bigr]\bigr)\bigr| \\ &\le \sup_{\theta\in\Theta, t\in T} |G_n^*(t,\theta) - G_n^*(\Pi_n t,\theta)| \\ &\quad + \sup_{t\in T, h\in\mathcal{H}_{b_n}^{B_n}} \bigl|E^*\bigl[w(t,Z)p^{b_n'}(X)h\bigr] - E^*\bigl[w(\Pi_n t,Z)p^{b_n'}(X)h\bigr]\bigr| \\ &= o_{p^*}(1).\end{aligned} \tag{161}$$

Employing the same arguments as in (114), we can then exploit (161) to obtain that almost surely

$$\begin{aligned}\min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} &\max_{t\in T} \bigl(G_n^*(t,\theta) - E^*\bigl[w(t,Z)p^{b_n'}(X)h\bigr]\bigr)^2 \\ &\le \min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T} \bigl(G_n^*(\Pi_n t,\theta) - E^*\bigl[w(\Pi_n t,Z)p^{b_n'}(X)h\bigr]\bigr)^2\end{aligned} \tag{162}$$

$$+2 \min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\left|G_n^*(\Pi_n t,\theta) - E^*\left[w(\Pi_n t, Z)p^{b_n'}(X)h\right]\right|$$
$$\times o_{p^*}(1) + o_{p^*}(1).$$

Furthermore, since $\min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T} |G_n^*(\Pi_n t,\theta) - E^*[w(\Pi_n t, Z) \times p^{b_n'}(X)h]| \leq \sup_{t\in T,\theta\in\Theta} |G_n^*(t,\theta)| = O_{p^*}(1)$ almost surely due to Lemma A.13, we conclude from (162) and $\{\Pi_n t : t\in T\} = T_n$ that

$$\min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\left(G_n^*(t,\theta) - E^*\left[w(t,Z)p^{b_n'}(X)h\right]\right)^2 \leq \min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T_n}\left(G_n^*(t,\theta) - E^*\left[w(t,Z)p^{b_n'}(X)h\right]\right)^2 + o_{p^*}(1). \tag{163}$$

Since $T_n \subseteq T$, then (159) and (163) establish the claim of the lemma. *Q.E.D.*

LEMMA A.16: *Let* $\Theta_0^{\varepsilon_n} \equiv \{\theta\in\Theta : \|\theta-\theta_0\|_w \leq \varepsilon_n\}$. *If Assumptions* 2.1(i) *and* (ii), 2.2, 3.1(i) *and* (ii), 3.2(i) *and* (ii), 3.3(ii), 3.4(i), *and* 3.5(ii) *and* (iii) *hold*, $\Theta_0\cap R \neq \emptyset$, $\varepsilon_n^2\times\lambda_n \downarrow 0$, *and* $\sup_{\theta\in\Theta_0\cap R}\|\theta - \Pi_n\theta\|_w = o(\varepsilon_n)$, *then a.s.*

$$\min_{\theta\in\Theta_n\cap\Theta_0^{\varepsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\left\{\left(G_n^*(t,\theta) - E^*[w(t,Z)p^{b_n'}(X)h]\right)^2 + \lambda_n P_n^*(t,\theta)\right\} = \min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\left(G_n^*(t,\theta) - E^*\left[w(t,Z)p^{b_n'}(X)h\right]\right)^2 + o_{p^*}(1). \tag{164}$$

PROOF: First observe that Assumption 3.5(iii) implies there indeed exists a $\varepsilon_n \downarrow 0$ such that $\varepsilon_n^2\times\lambda_n\downarrow 0$ and $\sup_{\theta\in\Theta_0\cap R}\|\theta-\Pi_n\theta\|_w = o(\varepsilon_n)$. Since $\Theta_0\cap R\neq\emptyset$ by hypothesis, it therefore follows that $\Theta_n\cap R\cap\Theta_0^{\varepsilon_n}\neq\emptyset$ for n sufficiently large. Next exploit the definition of $P_n^*(t,\theta)$, $\Theta_0^{\varepsilon_n}\subseteq\Theta$, and Lemma A.11 to conclude

$$\max_{\theta\in\Theta_0^{\varepsilon_n}, t\in T} \lambda_n P_n^*(t,\theta) \leq \max_{\theta\in\Theta, t\in T} 2\left(\frac{\sqrt{\lambda_n}}{n}\sum_{i=1}^n\left\{u_i(t,\theta) - E[U(t,\theta)]\right\}\right)^2 + 2\max_{\theta\in\Theta_0^{\varepsilon_n}, t\in T}\lambda_n\left(E[U(t,\theta)]\right)^2 \leq 2\lambda_n\times\varepsilon_n^2 + o_{as}(1). \tag{165}$$

Therefore, since by assumption $\varepsilon_n^2 \times \lambda_n \downarrow 0$, we conclude from the inequalities in (165) that almost surely

$$\min_{\theta\in\Theta_n\cap\Theta_0^{\varepsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big\{\big(G_n^*(t,\theta) - E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2 + \lambda_n P_n^*(t,\theta)\big\}$$
$$= \min_{\theta\in\Theta_n\cap\Theta_0^{\varepsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big(G_n^*(t,\theta) - E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2 + o_{p^*}(1). \tag{166}$$

Since $\sup_{\theta\in\Theta_0\cap R}\|\theta - \Pi_n\theta\|_w = o(\varepsilon_n)$, we have $\{\Pi_n\theta : \theta\in\Theta_0\cap R\}\subseteq\Theta_0^{\varepsilon_n}$ for n large. Hence, defining

$$(\hat\theta_n, \hat h_n) \in \arg\min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big(G_n^*(t,\theta) - E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2, \tag{167}$$

we have $(\Pi_n\hat\theta_n, \hat h_n)\in\Theta_n\cap\Theta_0^{\varepsilon_n}\cap R\times\mathcal{H}_{b_n}^{B_n}$ for n large. The first and second inequalities in (168) then hold almost surely due to Lemma A.13 implying G_n^* is asymptotically equicontinuous almost surely and Assumption 3.4(i):

$$\min_{\theta\in\Theta_n\cap\Theta_0^{\varepsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big(G_n^*(t,\theta) - E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2$$
$$\le \max_{t\in T}\big(G_n^*(t,\Pi_n\hat\theta_n) - E^*\big[w(t,Z)p^{b_n'}(X)\hat h_n\big]\big)^2$$
$$\le \max_{t\in T}\big(G_n^*(t,\hat\theta_n) - E^*\big[w(t,Z)p^{b_n'}(X)\hat h_n\big]\big)^2$$
$$+ 2\max_{t\in T}\big|G_n^*(t,\hat\theta_n) - E^*\big[w(t,Z)p^{b_n'}(X)\hat h_n\big]\big| \times o_{p^*}(1) + o_{p^*}(1). \tag{168}$$

Arguing as in (157), it follows that $\max_{t\in T}(G_n^*(t,\hat\theta_n) - E^*[w(t,Z)p^{b_n'}(X)]\times \hat h_n)^2 = O_{p^*}(1)$ almost surely. Hence, (168) and the definition of $(\hat\theta_n,\hat h_n)$ in (167) imply that almost surely we also have

$$\min_{\theta\in\Theta_n\cap\Theta_0^{\varepsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big(G_n^*(t,\theta) - E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2$$
$$\le \min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big(G_n^*(t,\theta) - E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2 + o_{p^*}(1). \tag{169}$$

Alternatively, since $\Theta_n\cap\Theta_0^{\varepsilon_n}\cap R\subseteq\Theta_0^{\varepsilon_n}\cap R$ and $T_n\subseteq T$, we can conclude from Lemma A.15 that

$$\min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big(G_n^*(t,\theta) - E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2$$
$$= \min_{\theta\in\Theta_0^{\varepsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T_n}\big(G_n^*(t,\theta) - E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2 + o_{p^*}(1) \tag{170}$$

$$\leq \min_{\theta\in\Theta_n\cap\Theta_0^{\varepsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big(G_n^*(t,\theta)-E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2+o_{p^*}(1).$$

The claim of the lemma then follows since (166), (169), and (170) hold almost surely. *Q.E.D.*

LEMMA A.17: *Let Assumptions* 2.1(i) *and* (ii), 2.2, 3.1(i) *and* (ii), 3.2(i) *and* (ii), 3.3(ii), 3.4(i), *and* 3.5(i) *and* (iii) *hold*, *and*, *let* $\Theta_0\cap R\neq\emptyset$. *Then*

$$I_n^*(R)=\min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big(G_n^*(t,\theta)-E\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2 + o_{p^*}(1)\quad a.s.$$

PROOF: The proof proceeds by studying a shrinking neighborhood of Θ_0. Let $\Theta_0^{\varepsilon_n}\equiv\{\theta\in\Theta:\|\theta-\theta_0\|_w\leq\varepsilon_n\}$ and choose $\varepsilon_n\downarrow 0$ such that $\varepsilon_n^2\times\lambda_n\uparrow\infty$. Next notice that since $P_n^*(t,\theta)\geq 0$ and $\Theta_n\cap\Theta_0^{\varepsilon_n}\cap R\subseteq\Theta_0^{\varepsilon_n}\cap R$, the inequality in (171) follows and the equality then holds almost surely due to Lemma A.15:

$$\begin{aligned}(171)\quad &\min_{\theta\in\Theta_n\cap\Theta_0^{\varepsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T_n}\big\{\big(G_n^*(t,\theta)-E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2+\lambda_n P_n^*(t,\theta)\big\}\\ &\geq \min_{\theta\in\Theta_0^{\varepsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T_n}\big(G_n^*(t,\theta)-E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2\\ &= \min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big(G_n^*(t,\theta)-E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2+o_{p^*}(1).\end{aligned}$$

Next select $\tilde{\varepsilon}_n\downarrow 0$ such that $\tilde{\varepsilon}_n^2\times\lambda_n\downarrow 0$ and $\sup_{\theta\in\Theta_0\cap R}\|\theta-\Pi_n\theta\|_w=o(\tilde{\varepsilon}_n)$, which is possible due to Assumption 3.5(iii). Since $\varepsilon_n^2\times\lambda_n\uparrow\infty$, it then follows that $\tilde{\varepsilon}_n=o(\varepsilon_n)$ and hence that $\Theta_0^{\tilde{\varepsilon}_n}\subseteq\Theta_0^{\varepsilon_n}$ for n large. Since $T_n\subseteq T$, the inequality in (172) then follows, while the equality holds almost surely by Lemma A.16:

$$\begin{aligned}(172)\quad &\min_{\theta\in\Theta_n\cap\Theta_0^{\varepsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T_n}\big\{\big(G_n^*(t,\theta)-E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2+\lambda_n P_n^*(t,\theta)\big\}\\ &\leq \min_{\theta\in\Theta_n\cap\Theta_0^{\tilde{\varepsilon}_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big\{\big(G_n^*(t,\theta)-E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2\\ &\quad+\lambda_n P_n^*(t,\theta)\big\}\\ &= \min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big(G_n^*(t,\theta)-E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2+o_{p^*}(1).\end{aligned}$$

Since $0\in\mathcal{H}_{b_n}^{B_n}$ and $\sup_{\theta\in\Theta,t\in T}(G_n^*(t,\theta))^2=O_{p^*}(1)$ almost surely due to Lemma A.13, it follows that the right-hand side of (172) is $O_{p^*}(1)$. Therefore, Lemma A.14 implies that the first equality in (173) holds while the second equality in (173) follows by combining (171) and (172); Lemmas A.7 and A.12 then imply the final equality in (173):

$$
\begin{aligned}
(173)\qquad I_n^*(R) &= \min_{\theta\in\Theta_n\cap\Theta_0^{\varepsilon_n}\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T_n}\Big\{\big(G_n^*(t,\theta)-E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2 \\
&\quad + \lambda_n P_n^*(t,\theta)\Big\} + o_{p^*}(1) \\
&= \min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big(G_n^*(t,\theta)-E^*\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2 + o_{p^*}(1) \\
&= \min_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \max_{t\in T}\big(G_n^*(t,\theta)-E\big[w(t,Z)p^{b_n'}(X)h\big]\big)^2 + o_{p^*}(1).
\end{aligned}
$$

This establishes the claim of the lemma. *Q.E.D.*

PROOF OF THEOREM 3.2: To establish the first claim of the theorem, we use the bounded Lipschitz metric to metrize weak convergence. For two laws L_1 and L_2 on a metric space $(\mathcal{A}, \|\cdot\|_A)$, this metric is given by

$$
\begin{aligned}
&\|L_1-L_2\|_{\mathrm{BL}_1} \equiv \sup_{f\in \mathrm{BL}_1(\mathcal{A})}\left|\int f\,dL_1 - \int f\,dL_2\right|, \\
&\mathrm{BL}_1(\mathcal{A}) \equiv \big\{f:\mathcal{A}\to\mathbf{R}\ \text{s.t.}\ \|f\|_\infty\le 1, |f(x)-f(y)|\le\|x-y\|_A\big\}.
\end{aligned}
$$

We also need to define the mappings $\mathcal{P}_n : L^\infty(T\times\Theta_0)\to\mathbf{R}$ and $\mathcal{P}_\infty : L^\infty(T\times\Theta_0)\to\mathbf{R}$ as

$$
\begin{aligned}
(174)\qquad \mathcal{P}_n(g) &\equiv \inf_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \sup_{t\in T}\big|g(t,\theta)-E\big[w(t,Z)p^{b_n'}(X)h\big]\big|, \\
\mathcal{P}_\infty(g) &\equiv \inf_{\theta\in\Theta_0\cap R, v\in V_\infty(T)} \sup_{t\in T}|g(t,\theta)-v(t)|.
\end{aligned}
$$

The maps $\mathcal{P}_n$ are Lipschitz. To observe this, apply the triangle inequality and (174) to conclude

$$
\begin{aligned}
(175)\qquad \mathcal{P}_n(g_1) &\le \inf_{\theta\in\Theta_0\cap R, h\in\mathcal{H}_{b_n}^{B_n}} \sup_{t\in T}\Big\{\big|g_2(t,\theta)-E\big[w(t,Z)p^{b_n'}(X)h\big]\big| \\
&\quad + |g_1(t,\theta)-g_2(t,\theta)|\Big\} \le \mathcal{P}_n(g_2)+\|g_1-g_2\|_\infty.
\end{aligned}
$$

Hence, since the same manipulations in (175) also show that $\mathcal{P}_n(g_2)\le\mathcal{P}_n(g_1)+\|g_1-g_2\|_\infty$, it follows that

$$
(176)\qquad |\mathcal{P}_n(g_1)-\mathcal{P}_n(g_2)|\le\|g_1-g_2\|_\infty.
$$

First we establish the almost sure consistency of the bootstrap for $\sqrt{I_n^*(R)}$ and then exploit that result to prove the first claim of the theorem. Notice that

Lemma A.17 implies $\sqrt{I_n^*(R)} = \mathcal{P}_n(G_n^*) + o_{p^*}(1)$ almost surely and hence

$$
\begin{aligned}
(177)\qquad & \sup_{f\in \mathrm{BL}_1(\mathbf{R})} \bigl|E^*\bigl[f(\sqrt{I_n^*(R)})\bigr] - E\bigl[f(\mathcal{P}_\infty(G))\bigr]\bigr| \\
&\le \sup_{f\in \mathrm{BL}_1(\mathbf{R})} \bigl|E^*\bigl[f(\mathcal{P}_n(G_n^*))\bigr] - E\bigl[f(\mathcal{P}_\infty(G))\bigr]\bigr| \\
&\quad + \sup_{f\in \mathrm{BL}_1(\mathbf{R})} \bigl|E^*\bigl[f(\sqrt{I_n^*(R)})\bigr] - E^*\bigl[f(\mathcal{P}_n(G_n^*))\bigr]\bigr| \\
&\le \sup_{f\in \mathrm{BL}_1(\mathbf{R})} \bigl|E^*\bigl[f(\mathcal{P}_n(G_n^*))\bigr] - E\bigl[f(\mathcal{P}_n(G))\bigr]\bigr| \\
&\quad + \sup_{f\in \mathrm{BL}_1(\mathbf{R})} \bigl|E\bigl[f(\mathcal{P}_n(G))\bigr] - E\bigl[f(\mathcal{P}_\infty(G))\bigr]\bigr| + o_{as}(1).
\end{aligned}
$$

Further notice that if $f \in \mathrm{BL}_1(\mathbf{R})$, then (176) implies that $f \circ \mathcal{P}_n \in \mathrm{BL}_1(L^\infty(T \times \Theta_0))$, which establishes the first inequality in (178); the second result then follows by Lemma A.13 and $\|\cdot\|_{\mathrm{BL}_1}$ metrizing weak convergence:

$$
\begin{aligned}
(178)\qquad & \sup_{f\in \mathrm{BL}_1(\mathbf{R})} \bigl|E^*\bigl[f(\mathcal{P}_n(G_n^*))\bigr] - E\bigl[f(\mathcal{P}_n(G))\bigr]\bigr| \\
&\le \sup_{h\in \mathrm{BL}_1(L^\infty(T\times\Theta_0))} \bigl|E^*[h(G_n^*)] - E[h(G)]\bigr| \\
&= o_{as}(1).
\end{aligned}
$$

Moreover, arguing as in (132), but with $V_{b_n}^{B_n}(T)$ in place of $V_{k_n}^{C_n}(T)$, we obtain from $b_n \uparrow \infty$, $B_n \uparrow \infty$, and Lemma A.8 that $\mathcal{P}_n(G) \xrightarrow{\mathcal{L}} \mathcal{P}_\infty(G)$. Therefore, combining results (177) and (178), we conclude that

$$
(179)\qquad \sup_{f\in \mathrm{BL}_1(\mathbf{R})} \bigl|E^*\bigl[f(\sqrt{I_n^*(R)})\bigr] - E\bigl[f(\mathcal{P}_\infty(G))\bigr]\bigr| \xrightarrow{\text{a.s.}} 0.
$$

Next notice that since $\|f\|_\infty \le 1$ for all $f \in \mathrm{BL}_1(\mathbf{R}_+)$, it follows that for any random variable $U \in \mathbf{R}_+$ and constant $K > 0$, we can conclude $|E[f(U) - f(U \wedge K^2)]| \le 2P(U > K^2)$. We hence obtain

$$
\begin{aligned}
(180)\qquad & \sup_{f\in \mathrm{BL}_1(\mathbf{R}_+)} \bigl|E^*\bigl[f(I_n^*(R))\bigr] - E\bigl[f(\mathcal{P}_\infty^2(G))\bigr]\bigr| \\
&\le \sup_{f\in \mathrm{BL}_1(\mathbf{R}_+)} \bigl|E^*\bigl[f(I_n^*(R) \wedge K^2)\bigr] - E\bigl[f(\mathcal{P}_\infty^2(G) \wedge K^2)\bigr]\bigr| \\
&\quad + 2P(\mathcal{P}_\infty(G) \ge K) + 2P^*(\sqrt{I_n^*(R)} \ge K).
\end{aligned}
$$

Since, in addition, for any $u_1, u_2 \in \mathbf{R}_+$ we have $|(u_1^2 \wedge K^2) - (u_2^2 \wedge K^2)| \leq 2K|u_1 - u_2|$, we obtain from (179) that

$$\text{(181)} \qquad \sup_{f \in \mathrm{BL}_1(\mathbf{R}_+)} \left| E^*\big[f(I_n^*(R) \wedge K^2)\big] - E\big[f(\mathcal{P}_\infty^2(G) \wedge K^2)\big]\right|$$
$$\leq \sup_{f \in \mathrm{BL}_{2K}(\mathbf{R}_+)} \left| E^*\big[f(\sqrt{I_n^*(R)})\big] - E\big[f(\mathcal{P}_\infty(G))\big]\right|$$
$$= 2K \sup_{f \in \mathrm{BL}_1(\mathbf{R}_+)} \left| E^*\big[f(\sqrt{I_n^*(R)})\big] - E\big[f(\mathcal{P}_\infty(G))\big]\right| = o_{as}(1).$$

Furthermore, since $\mathcal{P}_\infty(G)$ is a tight random variable, for any $\varepsilon > 0$, there exists a K^* such that

$$\text{(182)} \qquad P(\mathcal{P}_\infty(G) \geq K^*) \leq \varepsilon.$$

Similarly, by the portmanteau lemma, (179), and $\|\cdot\|_{\mathrm{BL}_1}$ metrizing weak convergence together with (182), we get

$$\text{(183)} \qquad \limsup_{n \to \infty} P^*(\sqrt{I_n^*(R)} \geq K^*) \leq P(\mathcal{P}_\infty(G) \geq K^*) \leq \varepsilon \quad \text{a.s.}$$

The first claim of the theorem then follows from (180), (181), (182), (183), and $\|\cdot\|_{\mathrm{BL}_1}$ metrizing weak convergence.

To establish the second claim of the theorem, note that $0 \in \mathcal{H}_{b_n}^{B_n}$, $\max_{t \in T, \theta \in \Theta} |G_n^*(t, \theta)| = O_{p^*}(1)$ almost surely as a result of Lemma A.13, and $\lambda_n \uparrow \infty$ imply the inequalities

$$\text{(184)} \qquad \min_{\theta \in \Theta_n \cap R} \max_{t \in T_n} P_n^*(t, \theta)$$
$$\leq \min_{\theta \in \Theta_n \cap R, h \in \mathcal{H}_{b_n}^{B_n}} \max_{t \in T_n} \Big\{ \lambda_n^{-1}\big(G_n^*(t, \theta) - E^*\big[w(t, Z) p^{b_n'}(X) h\big]\big)^2$$
$$+ P_n^*(t, \theta)\Big\}$$
$$\leq \max_{\theta \in \Theta, t \in T} \lambda_n^{-1}(G_n^*(t, \theta))^2 + \min_{\theta \in \Theta_n \cap R} \max_{t \in T_n} P_n^*(t, \theta)$$
$$= \min_{\theta \in \Theta_n \cap R} \max_{t \in T_n} P_n^*(t, \theta) + o_{p^*}(1) \quad \text{a.s.}$$

Hence, exploiting (184) and the definition of $I_n^*(R)$, $I_n(R)$, and P_n^*, it immediately follows that

$$\text{(185)} \qquad \lambda_n^{-1} I_n^*(R) = n^{-1} I_n(R) + o_{p^*}(1) \quad \text{a.s.}$$

The conclusion of the theorem then follows from (185) and the second claim of Theorem 3.1. *Q.E.D.*

PROOF OF COROLLARY 3.1: Let F be the limiting distribution of Theorem 3.1. Further let $c_{1-\alpha}$ satisfy $F(c_{1-\alpha}) = 1-\alpha$, which is uniquely defined since by assumption F is continuous and strictly increasing at its $1-\alpha$ quantile. By Theorem 3.2 and Theorem 1.2.1 in Politis, Romano, and Wolf (1999), it then follows that

$$\hat{c}_{1-\alpha} \xrightarrow{\text{a.s.}} c_{1-\alpha}. \tag{186}$$

Therefore, by Slutsky's theorem, Theorem 3.1, and $c_{1-\alpha}$ being a continuity point of F, we conclude that

$$\lim_{n\to\infty} P(I_n(R) \le \hat{c}_{1-\alpha}) = F(c_{1-\alpha}) = 1-\alpha. \tag{187}$$

For the second claim of the corollary, use the definition of $I_n^*(R)$ and $0 \in \mathcal{H}_{b_n}^{B_n}$ to obtain the first inequality in (188). The second inequality then follows by noting that $\min_{\theta\in\Theta_n\cap R}\max_{t\in T_n} P_n^*(t,\theta) = n^{-1}I_n(R)$, while the final result is implied by Theorem 3.1, $n-\lambda_n \to \infty$, and $\max_{\theta\in\Theta,t\in T}|G_n^*(t,\theta)| = O_{p^*}(1)$ almost surely by Lemma A.13:

$$\begin{aligned} & P^*(I_n^*(R) \le I_n(R)) \\ & \quad \ge P^*\Bigl(\min_{\theta\in\Theta_n\cap R}\max_{t\in T_n}\bigl\{(G_n^*(t,\theta))^2 + \lambda_n P_n^*(t,\theta)\bigr\} \le I_n(R)\Bigr) \\ & \quad \ge P^*\Bigl(\max_{\theta\in\Theta,t\in T}(G_n^*(t,\theta))^2 \le \frac{n-\lambda_n}{n} I_n(R)\Bigr) \xrightarrow{\text{a.s.}} 1. \end{aligned} \tag{188}$$

Result (188), the definition of $\hat{c}_{1-\alpha}$, and $\alpha > 0$ then imply the second claim of the corollary. *Q.E.D.*

REFERENCES

ADAMS, R. A., AND J. J. FOURNIER (2003): *Sobolev Spaces*. The Netherlands: Elsevier. [217,245, 246]

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843. [213-215,218,228]

——— (2007): "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models With Different Conditioning Variables," *Journal of Econometrics*, 141, 5–43. [215]

——— (2012): "Semiparametric Efficiency Bound for Models of Sequential Moment Restrictions Containing Unknown Functions," Discussion Paper 1731, Cowles Foundation. *Journal of Econometrics* (forthcoming). [215]

ANDERSON, T. W., AND H. RUBIN (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, 20, 46–63. [225]

BANKS, J., R. BLUNDELL, AND A. LEWBEL (1997): "Quadratic Engel Curves and Consumer Demand," *Review of Economics and Statistics*, 79, 527–539. [239]

BIERENS, H. J. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458. [224]
BIERENS, H. J., AND W. PLOBERGER (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1151. [225]
BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613–1669. [215,218]
CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics 6B*, ed. by J. J. Heckman and E. E. Leamer. Amsterdam: North-Holland. [228,229,234]
CHEN, X., AND D. POUZO (2012): "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Moments," *Econometrica*, 80, 277–321. [218,226,228,229]
CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [222,226,235]
CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441. [215]
——— (2005): "Nonparametric Identification Under Discrete Variation," *Econometrica*, 73, 1525–1550. [215]
——— (2007): "Instrumental Values," *Journal of Econometrics*, 139, 15–34. [215]
CHRISTENSEN, O. (2003): *An Introduction to Frames and Riesz Bases*. Boston: Birkhäuser. [244]
DAROLLES, S., J. FLORENS, AND E. RENAULT (2003): "Nonparametric Instrumental Regression," Working Paper, GREMAQ. [215,218]
EDMUNDS, D. E., AND H. TRIEBEL (1992): "Entropy Numbers and Approximation Numbers in Function Spaces II," *Proceedings of the London Mathematical Society*, 64, 153–169. [260]
GALLANT, A. R., AND D. W. NYCHKA (1987): "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363–390. [245]
GINE, E., AND J. ZINN (1990): "Bootstrapping General Empirical Measures," *The Annals of Probability*, 18, 851–869. [233]
GOHBERG, I. C., AND M. G. KREIN (1998): *Introduction to the Theory of Linear Nonself-Adjoint Operators in Hilbert Spaces*. Providence: American Mathematical Society. [260]
GOODMAN, V., J. KUELBS, AND J. ZINN (1981): "Some Results on the LIL in Banach Space With Applications to Weighted Empirical Processes," *The Annals of Probability*, 9, 713–752. [261]
HALL, P. (1984): "Integrated Square Error Properties of Kernel Estimators of Regression Functions," *The Annals of Statistics*, 12, 241–260. [224]
HALL, P., AND J. HOROWITZ (2005): "Nonparametric Methods for Inference in the Presense of Instrumental Variables," *The Annals of Statistics*, 33, 2904–2929. [214,215,218]
HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [229]
HECKMAN, J., AND E. J. VYTLACIL (2001): "Local Instrumental Variables," in *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, ed. by K. M. C. Hsiao and J. Powell. Cambridge: Cambridge University Press. [215]
HOROWITZ, J. (2006): "Testing a Parametric Model Against a Nonparametric Alternative With Identification Through Instrumental Variables," *Econometrica*, 74, 521–538. [215,224]
——— (2007): "Asymptotic Normality of a Nonparametric Instrumental Variables Estimator," *International Economic Review*, 48, 1329–1349. [214]
——— (2008): "Specification Testing in Nonparametric Instrumental Variables Estimation," Working Paper, Northwestern University. [215,228]
HUANG, J. Z. (1998): "Projection Estimation in Multiple Regression With Applications to Functional ANOVA Models," *The Annals of Statistics*, 26, 242–272. [228,234]
——— (2003): "Local Asymptotics for Polynomial Spline Regression," *The Annals of Statistics*, 31, 1600–1635. [228]
HUTTON, C. (1974): "On the Approximation Numbers of an Operator and Its Adjoint," *Mathematische Annalen*, 210, 277–280. [260]

IMBENS, G. W., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [214]

IMBENS, G. W., AND W. K. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equation Models Without Additivity," *Econometrica*, 77, 1481–1512. [215]

KRESS, R. (1999): *Linear Integral Equations*. New York: Springer. [218,243,244,260]

LUENBERGER, D. G. (1969): *Optimization by Vector Space Methods*. New York: Wiley. [230]

MANSKI, C. F. (1990): "Nonparametric Bounds on Treatment Effects," *American Economic Review Papers and Proceedings*, 80, 319–323. [214]

——— (2003): *Partial Identification of Probability Distributions*. New York: Springer-Verlag. [214,215]

MANSKI, C. F., AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010. [215]

NEWEY, W. K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047–1070. [224]

——— (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168. [228,229,234]

NEWEY, W. K., AND J. POWELL (2003): "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578. [213,215-218,227]

NEWEY, W. K., J. POWELL, AND F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equation Models," *Econometrica*, 67, 565–603. [215,220]

POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. New York: Springer-Verlag. [273]

ROMANO, J. P., AND A. M. SHAIKH (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138, 2786–2807. [214,225]

——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–211. [222,226]

SANTOS, A. (2011): "Instrumental Variable Methods for Recovering Continuous Linear Functionals," *Journal of Econometrics*, 161, 129–146. [215,235]

——— (2012): "Supplement to 'Inference in Nonparametric Instrumental Variables With Partial Identification'," *Econometrica Supplemental Material*, 80, http://www.econometricsociety.org/ecta/Supmat/7493_data and programs-1.zip; http://www.econometricsociety.org/ecta/Supmat/7493_data and programs-2.zip. [235]

SEVERINI, T. A., AND G. TRIPATHI (2006): "Some Identification Issues in Nonparametric Linear Models With Endogenous Regressors," *Econometric Theory*, 22, 258–278. [215]

——— (2010): "Efficiency Bounds for Estimating Linear Functionals of Nonparametric Regression Models With Endogenous Regressors," Working paper, University of Connecticut. [215]

SHAIKH, A. M., AND E. VYTLACIL (2011): "Partial Identification in Triangular Systems of Equations With Binary Dependent Variables," *Econometrica*, 79, 949–955. [215]

STINCHCOMBE, M. B., AND H. WHITE (1998): "Consistent Specification Testing With Nuisance Parameters Present Only Under the Alternative," *Econometric Theory*, 14, 295–325. [224,225]

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer. [242,246,248,249,252,258,261-263]

Dept. of Economics, University of California–San Diego, 9500 Gilman Drive 0508, La Jolla, CA 92093-0508, U.S.A.; a2santos@ucsd.edu.

Manuscript received October, 2007; final revision received December, 2010.