Constrained conditional moment restriction models

Victor Chernozhukov
Whitney Newey
Andres Santos

Constrained Conditional Moment Restriction Models

Victor Chernozhukov
Department of Economics
M.I.T.
vchern@mit.edu

Whitney K. Newey
Department of Economics
M.I.T.
wnewey@mit.edu

Andres Santos[*]
Department of Economics
U.C. San Diego
a2santos@ucsd.edu

September, 2015

Abstract

This paper examines a general class of inferential problems in semiparametric and nonparametric models defined by conditional moment restrictions. We construct tests for the hypothesis that at least one element of the identified set satisfies a conjectured (Banach space) "equality" and/or (a Banach lattice) "inequality" constraint. Our procedure is applicable to identified and partially identified models, and is shown to control the level, and under some conditions the size, asymptotically uniformly in an appropriate class of distributions. The critical values are obtained by building a strong approximation to the statistic and then bootstrapping a (conservatively) relaxed form of the statistic. Sufficient conditions are provided, including strong approximations using Koltchinskii's coupling.

Leading important special cases encompassed by the framework we study include: (i) Tests of shape restrictions for infinite dimensional parameters; (ii) Confidence regions for functionals that impose shape restrictions on the underlying parameter; (iii) Inference for functionals in semiparametric and nonparametric models defined by conditional moment (in)equalities; and (iv) Uniform inference in possibly nonlinear and severely ill-posed problems.

KEYWORDS: Shape restrictions, inference on functionals, conditional moment (in)equality restrictions, instrumental variables, nonparametric and semiparametric models, Banach space, Banach lattice, Koltchinskii coupling.

---

1 Introduction

Nonparametric constraints, often called shape restrictions, have played a central role in economics as both testable implications of classical theory and sufficient conditions for obtaining informative counterfactual predictions (Topkis, 1998). A long tradition in applied and theoretical econometrics has as a result studied shape restrictions, their ability to aid in identification, estimation, and inference, and the possibility of testing for their validity (Matzkin, 1994). The canonical example of this interplay between theory and practice is undoubtedly consumer demand analysis, where theoretical predictions such as Slutsky symmetry have been extensively tested for and exploited in estimation (Hausman and Newey, 1995; Blundell et al., 2012). The empirical analysis of shape restrictions, however, goes well beyond this important application with recent examples including studies into the monotonicity of the state price density (Jackwerth, 2000; Aıt-Sahalia and Duarte, 2003), the presence of ramp-up and start-up costs (Wolak, 2007; Reguant, 2014), and the existence of complementarities in demand (Gentzkow, 2007) and organizational design (Athey and Stern, 1998; Kretschmer et al., 2012).

Despite the importance of nonparametric constraints, their theoretical study has focused on a limited set of models and restrictions – a limitation that has resulted in practitioners often facing parametric modeling as their sole option. In this paper, we address this gap in the literature by developing a framework for testing general shape restrictions and exploiting them for inference in a widespread class of conditional moment restriction models. Specifically, we study nonparametric constraints in settings where the parameter of interest $\theta_0 \in \Theta$ satisfies $\mathcal{J}$ conditional moment restrictions

$$E_P[\rho_\jmath(X_i, \theta_0)|Z_{i,\jmath}] = 0 \text{ for } 1 \leq \jmath \leq \mathcal{J} \tag{1}$$

with $\rho_\jmath : \mathbf{R}^{d_x} \times \Theta \to \mathbf{R}$ possibly non-smooth functions, $X_i \in \mathbf{R}^{d_x}$, $Z_{i,\jmath} \in \mathbf{R}^{d_{z_\jmath}}$, and P denoting the distribution of $(X_i, \{Z_{i,\jmath}\}_{j=1}^{\mathcal{J}})$. As shown by Ai and Chen (2007, 2012), under appropriate choices of the parameter space and moment restrictions, this model encompasses parametric (Hansen, 1982), semiparametric (Ai and Chen, 2003), and nonparametric (Newey and Powell, 2003) specifications, as well as panel data applications (Chamberlain, 1992) and the study of plug-in functionals. By incorporating nuisance parameters into the definition of the parameter space, it is in fact also possible to view conditional moment (in)equality models as a special case of the specification we study. For example, the restriction $E_P[\tilde{\rho}(X_i, \tilde{\theta})|Z_i] \leq 0$ may be rewritten as $E_P[\tilde{\rho}(X_i, \tilde{\theta}) + \lambda(Z_i)|Z_i] = 0$ for some unknown positive function λ, which fits (1) with $\theta = (\tilde{\theta}, \lambda)$ and λ subject to the constraint $\lambda(Z_i) \geq 0$; see Example 2.4 below.

While in multiple applications identification of $\theta_0 \in \Theta$ is straightforward to establish, there also exist specifications of the model we examine for which identification can be uncertain (Canay et al., 2013; Chen et al., 2014). In order for our framework to be

2

robust to a possible lack of identification, we therefore define the identified set

$$\Theta_0(P) \equiv \{\theta \in \Theta : E_P[\rho_\jmath(X_i, \theta)|Z_{i,\jmath}] = 0 \text{ for } 1 \leq \jmath \leq \mathcal{J}\} \tag{2}$$

and employ it as the basis of our statistical analysis. Formally, for a set R of parameters satisfying a conjectured restriction, we develop a test for the hypothesis

$$H_0 : \Theta_0(P) \cap R \neq \emptyset \qquad H_1 : \Theta_0(P) \cap R = \emptyset \; ; \tag{3}$$

i.e. we device a test of whether at least one element of the identified set satisfies the posited constraints. In an identified model, a test of (3) is thus equivalent to a test of whether θ_0 satisfies the hypothesized constraint. The set R, for example, may constitute the set of functions satisfying a conjectured shape restriction, in which case a test of (3) corresponds to a test of the validity of such shape restriction. Alternatively, the set R may consist of the functions that satisfy an assumed shape restriction and for which a functional of interest takes a prescribed value – in which case test of inversion of (3) yields a confidence region for the value of the desired functional that imposes the assumed shape restriction on the underlying parameter.

The wide class of hypotheses with which we are concerned necessitates the sets R to be sufficiently general, yet be endowed with enough structure to ensure a fruitful asymptotic analysis. An important insight of this paper is that this simultaneous flexibility and structure is possessed by sets defined by "equality" restrictions on Banach space valued maps, and "inequality" restrictions on Abstract M (AM) space valued maps (an AM space is a Banach lattice whose norm obeys a particular condition).[1] We illustrate the generality granted by these sets by showing they enable us to employ tests of (3) to: (i) Conduct inference on the level of a demand function while imposing a Slutsky constraint; (ii) Construct a confidence interval in a regression discontinuity design where the conditional mean is known to be monotone in a neighborhood of, but not necessarily at, the discontinuity point; (iii) Test for the presence of complementarities in demand; and (iv) Conduct inference in semiparametric conditional moment (in)equality models. Additionally, while we do not pursue further examples in detail for conciseness, we note such sets R also allow for tests of homogeneity, supermodularity, and economies of scale or scope, as well as for inference on functionals of the identified set.

As our test statistic, we employ the minimum of a suitable criterion function over parameters satisfying the hypothesized restriction – an approach sometimes referred to as a sieve generalized method of moments J-test. Under appropriate conditions, we show that the distribution of the proposed statistic can be approximated by the law of the projection of a Gaussian process onto the image of the local parameter space under a linear map. In settings where the local parameter space is asymptotically linear and

---

[1]Due to their uncommon use in econometrics, we overview AM spaces in Appendix A.

the model is identified, the derived approximation can reduce to a standard chi-squared distribution as in Hansen (1982). However, in the presence of "binding" shape restrictions the local parameter space is often not asymptotically linear resulting in non-pivotal and potentially unreliable pointwise (in P) asymptotic approximations (Andrews, 2000, 2001). We address these challenges by projecting a bootstrapped version of the relevant Gaussian process into the image of an appropriate sample analogue of the local parameter space under an estimated linear map. Specifically, we establish that the resulting critical values provide asymptotic size control uniformly over a class of underlying distributions P. In addition, we characterize a set of alternatives for which the proposed test possesses nontrivial local power. While aspects of our analysis are specific to the conditional moment restriction model, the role of the local parameter space is solely dictated by the set R. As such, we expect the insights of our arguments to be applicable to the study of shape restrictions in alternative models as well.

The literature on nonparametric shape restrictions in econometrics has classically focused on testing whether conditional mean regressions satisfy the restrictions implied by consumer demand theory; see Lewbel (1995), Haag et al. (2009), and references therein. The related problem of studying monotone conditional mean regressions has also garnered widespread attention – recent advances on this problem includes Chetverikov (2012) and Chatterjee et al. (2013). Chernozhukov et al. (2009) propose generic methods, based on rearrangement and/or projection operators, that convert function estimators and confidence bands into monotone estimators and confidence bands, provably delivering finite-sample improvements; see Evdokimov (2010) for an application in the context of structural heterogeneity models. Additional work concerning monotonicity constraints includes Beare and Schmidt (2014) who test the monotonicity of the pricing kernel, Chetverikov and Wilhelm (2014) who study estimation of a nonparametric instrumental variable regression under monotonicity constraints, and Armstrong (2015) who develops minimax rate optimal one sided tests in a Gaussian regression discontinuity design. In related work, Freyberger and Horowitz (2012) examine the role of monotonicity and concavity or convexity constraints in a nonparametric instrumental variable regression with discrete instruments and endogenous variables. Our paper also contributes to a literature studying semiparametric and nonparametric models under partial identification (Manski, 2003). Examples of such work include Chen et al. (2011a), Chernozhukov et al. (2013), Hong (2011), Santos (2012), and Tao (2014) for conditional moment restriction models, and Chen et al. (2011b) for the maximum likelihood setting.

The remainder of the paper is organized as follows. In Section 2 we formally define the sets of restrictions we study and discuss examples that fall within their scope. In turn, in Section 3 we introduce our test statistic and basic notation that we employ throughout the paper. Section 4 obtains a rate of convergence for set estimators in conditional moment restriction models that we require for our subsequent analysis. Our

main results are contained in Sections 5 and 6, which respectively characterize and estimate the asymptotic distribution of our test statistic. Finally, Section 7 presents a brief simulation study, while Section 8 concludes. All mathematical derivations are included in a series of appendices; see in particular Appendix A for an overview of AM spaces and an outline of how Appendices B through H are organized.

2 The Hypothesis

In this section, we formally introduce the set of null hypotheses we examine as well as motivating examples that fall within their scope.

2.1 The Restriction Set

The defining elements determining the generality of the hypotheses allowed for in (3) are the choice of parameter space Θ and the set of restrictions embodied by R. In imposing restrictions on both Θ and R we aim to allow for as general a framework as possible while simultaneously ensuring enough structure for a fruitful asymptotic analysis. To this end, we require the parameter space Θ to be a subset of a Banach space $\mathbf{B}$, and consider sets R that are defined through "equality" and "inequality" restrictions. Specifically, for known maps Υ_F and Υ_G, we impose that the set R be of the form

$$R \equiv \{\theta \in \mathbf{B} : \Upsilon_F(\theta) = 0 \text{ and } \Upsilon_G(\theta) \leq 0\} . \tag{4}$$

In order to allow for hypotheses that potentially concern global properties of θ, such as shape restrictions, the maps $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ and $\Upsilon_G : \mathbf{B} \to \mathbf{G}$ are also assumed to take values on general Banach spaces $\mathbf{F}$ and $\mathbf{G}$ respectively. While no further structure on $\mathbf{F}$ is needed for testing "equality" restrictions, the analysis of "inequality" restrictions necessitates that $\mathbf{G}$ be equipped with a partial ordering – i.e. that "$\leq$" be well defined in (4). We thus impose the following requirements on Θ, and the maps Υ_F and Υ_G:

Assumption 2.1. *(i) $\Theta \subseteq \mathbf{B}$, where $\mathbf{B}$ is a Banach space with metric $\|\cdot\|_{\mathbf{B}}$.*

Assumption 2.2. *(i) $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ and $\Upsilon_G : \mathbf{B} \to \mathbf{G}$, where $\mathbf{F}$ is a Banach space with metric $\|\cdot\|_{\mathbf{F}}$, and $\mathbf{G}$ is an AM space with order unit $\mathbf{1_G}$ and metric $\|\cdot\|_{\mathbf{G}}$; (ii) The maps $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ and $\Upsilon_G : \mathbf{B} \to \mathbf{G}$ are continuous under $\|\cdot\|_{\mathbf{B}}$.*

Assumption 2.1 formalizes the requirement that the parameter space Θ be a subset of a Banach space $\mathbf{B}$. In turn, Assumption 2.2(i) similarly imposes that Υ_F take values in a Banach space $\mathbf{F}$, while the map Υ_G is required to take values in an AM space $\mathbf{G}$ – since AM spaces are not often used in econometrics, we provide an overview in Appendix

5

A. Heuristically, the essential implications of Assumption 2.2(i) for $\mathbf{G}$ are that: (i) $\mathbf{G}$ is a vector space equipped with a partial order relationship "$\leq$"; (ii) The partial order "$\leq$" and the vector space operations interact in the same manner they do on $\mathbf{R}$;[2] and (iii) The order unit $\mathbf{1_G} \in \mathbf{G}$ is an element such that for any $\theta \in \mathbf{G}$ there exists a scalar $\lambda > 0$ satisfying $|\theta| \leq \lambda \mathbf{1_G}$; see Remark 2.1 for an example. Finally, we note that in Assumption 2.2(ii) the maps Υ_F and Υ_G are required to be continuous, which ensures that the set R is closed in $\mathbf{B}$. Since the choice of maps Υ_F and Υ_G is dictated by the hypothesis of interest, verifying Assumption 2.2(ii) is often accomplished by restricting $\mathbf{B}$ to have a sufficiently "strong" norm that ensures continuity.

Remark 2.1. In applications we will often work with the space of continuous functions with bounded derivatives. Formally, for a set $A \subseteq \mathbf{R}^d$, a function $f : A \to \mathbf{R}$, a vector of positive integers $\alpha = (\alpha_1, \ldots, \alpha_d)$, and $|\alpha| = \sum_{i=1}^d \alpha_i$ we denote

$$D^\alpha f(a_0) = \frac{\partial^{|\alpha|}}{\partial a_1^{\alpha_1}, \ldots, \partial a_d^{\alpha_d}} f(a) \Big|_{a=a_0} . \tag{5}$$

For a nonnegative integer m, we may then define the space $C^m(A)$ to be given by

$$C^m(A) \equiv \{f : D^\alpha f \text{ is continuous and bounded on } A \text{ for all } |\alpha| \leq m\} , \tag{6}$$

which we endow with the metric $\|f\|_{m,\infty} \equiv \max_{|\alpha| \leq m} \sup_{a \in A} |D^\alpha f(a)|$. The space $C^0(A)$ with norm $\|f\|_{0,\infty}$ – which we denote $C(A)$ and $\|\cdot\|_\infty$ for simplicity – is then an AM space. In particular, equipping $C(A)$ with the ordering $f_1 \leq f_2$ if and only if $f_1(a) \leq f_2(a)$ for all $a \in A$ implies the constant function $\mathbf{1}(a) = 1$ for all $a \in A$ is an order unit. ∎

2.2 Motivating Examples

In order to illustrate the relevance of the introduced framework, we next discuss a number of applications based on well known models. For conciseness, we keep the discussion brief and revisit these examples in more detail in Appendix F.

We draw our first example from a long-standing literature aiming to replace parametric assumptions with shape restrictions implied by economic theory (Matzkin, 1994).

Example 2.1. (Shape Restricted Demand). Blundell et al. (2012) examine a semiparametric model for gasoline demand, in which quantity demanded Q_i given price P_i, income Y_i, and demographic characteristics $W_i \in \mathbf{R}^{d_w}$ is assumed to satisfy

$$Q_i = g_0(P_i, Y_i) + W_i' \gamma_0 + U_i . \tag{7}$$

---

[2]For example, if $\theta_1 \leq \theta_2$, then $\theta_1 + \theta_3 \leq \theta_2 + \theta_3$ for any $\theta_3 \in \mathbf{G}$.

The authors propose a kernel estimator for the function $g_0 : \mathbf{R}_+^2 \to \mathbf{R}$ under the assumption $E[U_i|P_i, Y_i, W_i] = 0$ and the hypothesis that g_0 obeys the Slutsky restriction

$$\frac{\partial}{\partial p}g_0(p,y) + g_0(p,y)\frac{\partial}{\partial y}g_0(p,y) \le 0 \ . \tag{8}$$

While in their application Blundell et al. (2012) find imposing (8) to be empirically important, their asymptotic framework assumes (8) holds strictly and thus implies the constrained an unconstrained estimators are asymptotically equivalent. In contrast, our results will enable us to test, for example, for $(p_0, y_0) \in \mathbf{R}_+^2$ and $c_0 \in \mathbf{R}$ the hypothesis

$$H_0 : g_0(p_0, y_0) = c_0 \qquad H_1 : g_0(p_0, y_0) \ne c_0 \tag{9}$$

employing an asymptotic analysis that is able to capture the finite sample importance of imposing the Slutsky restriction. To map this problem into our framework, we set $\mathbf{B} = C^1(\mathbf{R}_+^2) \times \mathbf{R}^{d_w}$, $\mathcal{J} = 1$, $Z_i = (P_i, Y_i, W_i)$, $X_i = (Q_i, Z_i)$ and $\rho(X_i, \theta) = Q_i - g(P_i, W_i) - W_i'\gamma$ for any $\theta = (g, \gamma) \in \mathbf{B}$. Letting $\mathbf{F} = \mathbf{R}$ and defining $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ by $\Upsilon_F(\theta) = g(p_0, y_0) - c_0$ for any $\theta \in \mathbf{B}$ enables us to test (9), while the Slutsky restriction can be imposed by setting $\mathbf{G} = C(\mathbf{R}_+^2)$ and defining $\Upsilon_G : \mathbf{B} \to \mathbf{G}$ to be given by

$$\Upsilon_G(\theta)(p,y) = \frac{\partial}{\partial p}g(p,y) + g(p,y)\frac{\partial}{\partial y}g(p,y) \tag{10}$$

for any $\theta \in \mathbf{B}$. Alternatively, we may also conduct inference on deadweight loss as considered in Blundell et al. (2012) building on Hausman and Newey (1995), or allow for endogeneity and quantile restrictions on U_i as pursued by Blundell et al. (2013). ∎

Our next example builds on Example 2.1 by illustrating how to exploit shape restrictions in a regression discontinuity (RD) setting; see also Armstrong (2015).[3]

Example 2.2. (Monotonic RD). We consider a sharp design in which treatment is assigned whenever a forcing variable $R_i \in \mathbf{R}$ is above a threshold which we normalize to zero. For an outcome variable Y_i and a treatment indicator $D_i = 1\{R_i \ge 0\}$, Hahn et al. (2001) showed the average treatment effect τ_0 at zero is identified by

$$\tau_0 = \lim_{r\downarrow 0} E[Y_i|R_i = r] - \lim_{r\uparrow 0} E[Y_i|R_i = r] \ . \tag{11}$$

In a number of applications it is additionally reasonable to assume $E[Y_i|R_i = r]$ is monotonic in a neighborhood of, but not necessarily at, zero. Such restriction is natural, for instance, in Lee et al. (2004) where R_i is the democratic vote share and Y_i is a measure of how liberal the elected official's voting record is, or in Black et al. (2007) where R_i and Y_i are respectively measures of predicted and actual collected unemployment benefits.

---

[3]We thank Pat Kline for suggesting this example.

In order to illustrate the applicability of our framework to this setting, we suppose we wish to impose monotonicity of $E[Y_i|R = r]$ on $r \in [-1, 0)$ and $r \in [0, 1]$ while testing

$$H_0 : \tau_0 = 0 \qquad H_1 : \tau_0 \neq 0 \ . \tag{12}$$

To this end, we let $\mathbf{B} = C^1([-1, 0]) \times C^1([0, 1])$, $X_i = (Y_i, R_i, D_i)$, $Z_i = (R_i, D_i)$, and for $\mathcal{J} = 1$ set $\rho(x, \theta) = y - g_-(r)(1 - d) - g_+(r)d$ which yields the restriction

$$E[Y_i - g_-(R_i)(1 - D_i) - g_+(R_i)D_i|R_i, D_i] = 0 \tag{13}$$

for any $\theta = (g_-, g_+) \in \mathbf{B}$. The functions g_- and g_+ are then respectively identified by $E[Y_i|R_i = r]$ for $r \in [-1, 0)$ and $r \in [0, 1]$, and hence we may test (12) by setting $\mathbf{F} = \mathbf{R}$ and $\Upsilon_F(\theta) = g_+(0) - g_-(0)$ for any $\theta = (g_-, g_+) \in \mathbf{B}$.[4] In turn, monotonicity can be imposed by setting $\mathbf{G} = C([-1, 0]) \times C([0, 1])$ and letting $\Upsilon_G(\theta) = (-g'_-, -g'_+)$ for any $\theta = (g_-, g_+) \in \mathbf{B}$. A similar construction can also be applied in fuzzy RD designs or the regression kink design studied in Card et al. (2012) and Calonico et al. (2014). ∎

While Examples 2.1 and 2.2 concern imposing shape restriction to conduct inference on functionals, in certain applications interest instead lies on the shape restriction itself. The following example is based on a model originally employed by Gentzkow (2007) in examining whether print and online newspapers are substitutes or complements.

Example 2.3. (Complementarities). Suppose an agent can buy at most one each of two goods $j \in \{1, 2\}$, and let $a = (a_1, a_2) \in \{(0, 0), (1, 0), (0, 1), (1, 1)\}$ denote the possible bundles to be purchased. We consider a random utility model

$$U(a, Z_i, \epsilon_i) = \sum_{j=1}^{2} (W'_i \gamma_{0,j} + \epsilon_{ij}) 1\{a_j = 1\} + \delta_0(Y_i) 1\{a_1 = 1, a_2 = 1\} \tag{14}$$

where $Z_i = (W_i, Y_i)$ are observed covariates, $Y_i \in \mathbf{R}^{d_y}$ can be a subvector of $W_i \in \mathbf{R}^{d_w}$, $\delta_0 \in C(\mathbf{R}^{d_y})$ is an unknown function, and $\epsilon_i = (\epsilon_{i,1}, \epsilon_{i,2})$ follows a parametric distribution $G(\cdot|\alpha_0)$ with $\alpha_0 \in \mathbf{R}^{d_\alpha}$; see Fox and Lazzati (2014) for identification results. In (14), $\delta_0 \in C(\mathbf{R}^{d_y})$ determines whether the goods are complements or substitutes and we may consider, for example, a test of the hypothesis that they are always substitutes

$$H_0 : \delta_0(y) \leq 0 \text{ for all } y \qquad H_1 : \delta_0(y) > 0 \text{ for some } y \ . \tag{15}$$

In this instance, $\mathbf{B} = \mathbf{R}^{2d_w + d_\alpha} \times C(\mathbf{R}^{d_y})$, and for any $\theta = (\gamma_1, \gamma_2, \alpha, \delta) \in \mathbf{B}$ we map (15) into our framework by letting $\mathbf{G} = C(\mathbf{R}^{d_y})$, $\Upsilon_G(\theta) = \delta$, and imposing no equality

---

[4]Here, with some abuse of notation, we identify $g_- \in C([-1, 0])$ with the function $E[Y_i|R_i = r]$ on $r \in [-1, 0)$ by letting $g_-(0) = \lim_{r \uparrow 0} E[Y_i|R_i = r]$ which exists by assumption.

restrictions. For observed choices A_i the conditional moment restrictions are then

$$P(A_i = (1,0)|Z_i)$$
$$= \int 1\{W_i'\gamma_1 + \epsilon_1 \geq 0, \ W_i'\gamma_2 + \delta(Y_i) + \epsilon_2 \leq 0, \ W_i'\gamma_1 + \epsilon_1 \geq W_i'\gamma_2 + \epsilon_2\}dG(\epsilon|\alpha) \quad (16)$$

and, exploiting $\delta(Y_i) \leq 0$ under the null hypothesis, the two additional conditions

$$P(A_i = (0,0)|Z_i) = \int 1\{\epsilon_1 \leq -W_i'\gamma_1, \ \epsilon_2 \leq -W_i'\gamma_2\}dG(\epsilon|\alpha) \quad (17)$$

$$P(A_i = (1,1)|Z_i) = \int 1\{\epsilon_1 + \delta(Y_i) \geq -W_i'\gamma_1, \ \epsilon_2 + \delta(Y_i) \geq -W_i'\gamma_2\}dG(\epsilon|\alpha) \quad (18)$$

so that in this model $\mathcal{J} = 3$. An analogous approach may also be employed to conduct inference on interaction effects in discrete games as in De Paula and Tang (2012). ■

The introduced framework can also be employed to study semiparametric specifications in conditional moment (in)equality models – thus complementing a literature that, with the notable exception of Chernozhukov et al. (2013), has been largely parametric (Andrews and Shi, 2013). Our final example illustrates such an application in the context of a study of hospital referrals by Ho and Pakes (2014).

Example 2.4. (Testing Parameter Components in Moment (In)Equalities).
We consider the problem of estimating how an insurer assigns patients to hospital within its network $\mathcal{H}$. Suppose each observation i consists of two individuals $j \in \{1,2\}$ of similar characteristics for whom we know the hospital $H_{ij} \in \mathcal{H}$ to which they were referred, as well as the cost of treatment $P_{ij}(h)$ and the distance $D_{ij}(h)$ to any hospital $h \in \mathcal{H}$. Under certain assumptions, Ho and Pakes (2014) then derive the moment restriction

$$E[\sum_{j=1}^{2}\{\gamma_0(P_{ij}(H_{ij}) - P_{ij}(H_{ij'})) + g_0(D_{ij}(H_{ij})) - g_0(D_{ij}(H_{ij'}))\}|Z_i] \leq 0 \quad (19)$$

where $\gamma_0 \in \mathbf{R}$ denotes the insurer's sensitivity to price, $g_0 : \mathbf{R}_+ \to \mathbf{R}_+$ is an unknown monotonically increasing function reflecting a preference for referring patients to nearby hospitals, $Z_i \in \mathbf{R}^{d_z}$ is an appropriate instrument, and $j' \equiv \{1,2\} \setminus \{j\}$.[5] Employing our proposed framework we may, for example, test for some $c_0 \in \mathbf{R}$ the null hypothesis

$$H_0 : \gamma_0 = c_0 \qquad H_1 : \gamma_0 \neq c_0 \quad (20)$$

without imposing parametric restrictions on g_0 but instead requiring it to be monotone.

---

[5] In other words, $j' = 2$ when $j = 1$, and $j' = 1$ when $j = 2$.

To this end, let $X_i = (\{\{P_{ij}(h), D_{ij}(h)\}_{h \in \mathcal{H}}, H_{ij}\}_{j=1}^2, Z_i)$, and define the function

$$\psi(X_i, \gamma, g) \equiv \sum_{j=1}^2 \{\gamma(P_{ij}(H_{ij}) - P_{ij}(H_{ij'})) + g(D_{ij}(H_{ij})) - g(D_{ij}(H_{ij'}))\} \qquad (21)$$

for any $(\gamma, g) \in \mathbf{R} \times C^1(\mathbf{R}_+)$. The moment restriction in (19) can then be rewritten as

$$E[\psi(X_i, \gamma_0, g_0) + \lambda_0(Z_i)|Z_i] = 0 \qquad (22)$$

for some unknown function λ_0 satisfying $\lambda_0(Z_i) \geq 0$. Thus, (19) may be viewed as a conditional moment restriction model with parameter space $\mathbf{B} = \mathbf{R} \times C^1(\mathbf{R}_+) \times \ell^\infty(\mathbf{R}^{d_z})$ in which $\rho(x, \theta) = \psi(x, \gamma, g) + \lambda(z)$ for any $\theta = (\gamma, g, \lambda) \in \mathbf{B}$. The monotonicity restriction on g and positivity requirement on λ can in turn be imposed by setting $\mathbf{G} = \ell^\infty(\mathbf{R}_+) \times \ell^\infty(\mathbf{R}^{d_z})$ and $\Upsilon_G(\theta) = -(g', \lambda)$, while the null hypothesis in (20) may be tested by letting $\mathbf{F} = \mathbf{R}$ and defining $\Upsilon_F(\theta) = \gamma - c_0$ for any $\theta = (\gamma, g, \lambda) \in \mathbf{B}$. An analogous construction can similarly be applied to extend conditional moment inequality models with parametric specifications to semiparametric or nonparametric ones;[6] see Ciliberto and Tamer (2009), Pakes (2010) and references therein. ∎

3   Basic Setup

Having formally stated the hypotheses we consider, we next develop a test statistic and introduce basic notation and assumptions that will be employed throughout the paper.

3.1   Test Statistic

We test the null hypothesis in (3) by employing a sieve-GMM statistic that may be viewed as a generalization of the overidentification test of Sargan (1958) and Hansen (1982). Specifically, for the instrument $Z_{i,\jmath}$ of the $\jmath^{th}$ moment restriction, we consider a set of transformations $\{q_{k,n,\jmath}\}_{k=1}^{k_{n,\jmath}}$ and let $q_{n,\jmath}^{k_{n,\jmath}}(z_\jmath) \equiv (q_{1,n,\jmath}(z_\jmath), \ldots, q_{k_{n,\jmath},n,\jmath}(z_\jmath))'$. Setting $Z_i \equiv (Z_{i,1}', \ldots, Z_{i,\mathcal{J}}')'$ to equal the vector of all instruments, $k_n \equiv \sum_{j=1}^{\mathcal{J}} k_{n,\jmath}$ the total number of transformations, $q_n^{k_n}(z) \equiv (q_{n,1}^{k_{n,1}}(z_1)', \ldots, q_{n,\mathcal{J}}^{k_{n,\mathcal{J}}}(z_\mathcal{J})')'$ the vector of all transformations, and $\rho(x, \theta) \equiv (\rho_1(x, \theta), \ldots, \rho_\mathcal{J}(x, \theta))'$ the vector of all generalized residuals, we then construct for each $\theta \in \Theta$ the $k_n \times 1$ vector of scaled sample moments

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \rho(X_i, \theta) * q_n^{k_n}(Z_i) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n (\rho_1(X_i, \theta) q_{n,1}^{k_{n,1}}(Z_{i,1})', \ldots, \rho_\mathcal{J}(X_i, \theta) q_{n,\mathcal{J}}^{k_{n,\mathcal{J}}}(Z_{i,\mathcal{J}})')'$$

(23)

---

[6]Alternatively, through test inversion we may employ the framework of this example to construct confidence regions for functionals of a semi or non-parametric identified set (Romano and Shaikh, 2008).

where for partitioned vectors a and b, $a * b$ denotes their Khatri-Rao product[7] – i.e. the vector in (23) consists of the scaled sample averages of the product of each generalized residual $\rho_{\jmath}(X_i, \theta)$ with the transformations of its respective instrument $Z_{i,\jmath}$. Clearly, if (23) is evaluated at a parameter θ in the identified set $\Theta_0(P)$, then its mean will be zero. As noted by Newey (1985), however, for any fixed dimension k_n the expectation of (23) may still be zero even if $\theta \notin \Theta_0(P)$. For this reason, we conduct an asymptotic analysis in which k_n diverges to infinity, and note the choice of transformations $\{q_{k,n,\jmath}\}_{k=1}^{k_{n,\jmath}}$ is allowed to depend on n to accommodate the use of splines or wavelets.

Intuitively, we test (3) by examining whether there is a parameter $\theta \in \Theta$ satisfying the hypothesized restrictions and such that (23) has mean zero. To this end, for any $r \geq 2$, vector $a \equiv (a^{(1)}, \ldots, a^{(d)})'$, and $d \times d$ positive definite matrix A we define

$$\|a\|_{A,r} \equiv \|Aa\|_r \qquad\qquad \|a\|_r^r \equiv \sum_{i=1}^{d} |a^{(i)}|^r \ , \tag{24}$$

with the usual modification $\|a\|_\infty \equiv \max_{1 \leq i \leq d} |a^{(i)}|$. For any possibly random $k_n \times k_n$ positive definite matrix $\hat{\Sigma}_n$, we then construct a function $Q_n : \Theta \to \mathbf{R}_+$ by

$$Q_n(\theta) \equiv \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho(X_i, \theta) * q_n^{k_n}(Z_i) \|_{\hat{\Sigma}_n, r} \ . \tag{25}$$

Heuristically, the criterion Q_n should diverge to infinity when evaluated at any $\theta \notin \Theta_0(P)$ and remain "stable" when evaluated at a $\theta \in \Theta_0(P)$. We therefore employ the minimum of Q_n over R to examine whether there exists a θ that simultaneously makes Q_n "stable" ($\theta \in \Theta_0(P)$) and satisfies the conjectured restriction ($\theta \in R$). Formally, we employ

$$I_n(R) \equiv \inf_{\theta \in \Theta_n \cap R} Q_n(\theta) \ , \tag{26}$$

where $\Theta_n \cap R$ is a sieve for $\Theta \cap R$ – i.e. $\Theta_n \cap R$ is a finite dimensional subset of $\Theta \cap R$ that grows dense in $\Theta \cap R$. Since the choice of $\Theta_n \cap R$ depends on $\Theta \cap R$, we leave it unspecified though note common choices include flexible finite dimensional specifications, such as splines, polynomials, wavelets, and neural networks; see (Chen, 2007).

3.2 Notation and Assumptions

3.2.1 Notation

Before stating our next set of assumptions, we introduce basic notation that we employ throughout the paper. For conciseness, we let $V_i \equiv (X_i', Z_i')'$ and succinctly refer to the

---

[7]For partitioned vectors $a = (a_1', \ldots, a_{\mathcal{J}}')'$ and $b = (b_1', \ldots, a_{\mathcal{J}}')'$, $a * b \equiv ((a_1 \otimes b_1)', \ldots, (a_{\mathcal{J}} \otimes b_{\mathcal{J}})')'$.

set of $P \in \mathbf{P}$ satisfying the null hypothesis in (3) by employing the notation

$$\mathbf{P}_0 \equiv \{P \in \mathbf{P} : \Theta_0(P) \cap R \neq \emptyset\} . \tag{27}$$

We also view any $d \times d$ matrix A as a map from $\mathbf{R}^d$ to $\mathbf{R}^d$, and note that when $\mathbf{R}^d$ is equipped with the norm $\|\cdot\|_r$ it induces on A the operator norm $\|\cdot\|_{o,r}$ given by

$$\|A\|_{o,r} \equiv \sup_{a \in \mathbf{R}^d : \|a\|_r = 1} \|Aa\|_r . \tag{28}$$

For instance, $\|A\|_{o,1}$ corresponds to the largest maximum absolute value column sum of A, and $\|A\|_{o,2}$ corresponds to the square root of the largest eigenvalue of $A'A$.

Our analysis relies heavily on empirical process theory, and we therefore borrow extensively from the literature's notation (van der Vaart and Wellner, 1996). In particular, for any function f of V_i we for brevity sometimes write its expectation as

$$Pf \equiv E_P[f(V_i)] . \tag{29}$$

In turn, we denote the empirical process evaluated at a function f by $\mathbb{G}_{n,P}f$ – i.e. set

$$\mathbb{G}_{n,P}f \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{f(V_i) - Pf\} . \tag{30}$$

We will often need to evaluate the empirical process at functions generated by the maps $\theta \mapsto \rho_\jmath(\cdot, \theta)$ and the sieve $\Theta_n \cap R$, and for convenience we therefore define the set

$$\mathcal{F}_n \equiv \{\rho_\jmath(\cdot, \theta) : \theta \in \Theta_n \cap R \text{ and } 1 \leq \jmath \leq \mathcal{J}\} . \tag{31}$$

The "size" of $\mathcal{F}_n$ plays a crucial role, and we control it through the bracketing integral

$$J_{[]}(\delta, \mathcal{F}_n, \|\cdot\|_{L_P^2}) \equiv \int_0^\delta \sqrt{1 + \log N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{L_P^2})} d\epsilon , \tag{32}$$

where $N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{L_P^2})$ is the smallest number of brackets of size ϵ (under $\|\cdot\|_{L_P^2}$) required to cover $\mathcal{F}_n$.[8] Finally, we let $\mathbb{W}_{n,P}$ denote the isonormal process on L_P^2 – i.e. $\mathbb{W}_{n,P}$ is a Gaussian process satisfying for all $f, g \in L_P^2$, $E[\mathbb{W}_{n,P}f] = E[\mathbb{W}_{n,P}g] = 0$ and

$$E[\mathbb{W}_{n,P}f\mathbb{W}_{n,P}g] = E_P[(f(V_i) - E_P[f(V_i)])(g(V_i) - E_P[g(V_i)])] . \tag{33}$$

It will prove useful to denote the vector subspace generated by the sieve $\Theta_n \cap R$ by

$$\mathbf{B}_n \equiv \overline{\text{span}}\{\Theta_n \cap R\} , \tag{34}$$

---

[8]An ϵ bracket under $\|\cdot\|_{L_P^2}$ is a set of the form $\{f \in \mathcal{F}_n : L(v) \leq f(v) \leq U(v)\}$ with $\|U - L\|_{L_P^2} < \epsilon$. Here, as usual, the L_P^q spaces are defined by $L_P^q \equiv \{f : \|f\|_{L_P^q} < \infty\}$ where $\|f\|_{L_P^q}^q \equiv E_P[|f|^q]$.

where $\overline{\text{span}}\{C\}$ denotes the closure under $\|\cdot\|_{\mathbf{B}}$ of the linear span of any set $C \subseteq \mathbf{B}$. Since $\mathbf{B}_n$ will further be assumed to be finite dimensional, all well defined norms on it will be equivalent in the sense that they generate the same topology. Hence, if $\mathbf{B}_n$ is a subspace of two different Banach spaces $(\mathbf{A}_1, \|\cdot\|_{\mathbf{A}_1})$ and $(\mathbf{A}_2, \|\cdot\|_{\mathbf{A}_2})$, then the modulus of continuity of $\|\cdot\|_{\mathbf{A}_1}$ with respect to $\|\cdot\|_{\mathbf{A}_2}$, which we denote by

$$\mathcal{S}_n(\mathbf{A}_1, \mathbf{A}_2) \equiv \sup_{b \in \mathbf{B}_n} \frac{\|b\|_{\mathbf{A}_1}}{\|b\|_{\mathbf{A}_2}} \ , \tag{35}$$

will be finite for any n though possibly diverging to infinity with the dimension of $\mathbf{B}_n$. For example, if $\mathbf{B}_n \subseteq L_P^\infty$, then $\mathcal{S}_n(L_P^2, L_P^\infty) \leq 1$, while $\mathcal{S}_n(L_P^\infty, L_P^2)$ is the smallest constant such that $\|b\|_{L_P^\infty} \leq \|b\|_{L_P^2} \times \mathcal{S}_n(L_P^\infty, L_P^2)$ for all $b \in \mathbf{B}_n$.

3.2.2 Assumptions

The following assumptions introduce a basic structure we employ throughout the paper.

Assumption 3.1. *(i)* $\{V_i\}_{i=1}^\infty$ *is an i.i.d. sequence with* $V_i \sim P \in \mathbf{P}$.

Assumption 3.2. *(i) For all* $1 \leq \jmath \leq \mathcal{J}$, $\sup_{1 \leq k \leq k_{n,\jmath}} \sup_{P \in \mathbf{P}} \|q_{k,n,\jmath}\|_{L_P^\infty} \leq B_n$ *with* $B_n \geq 1$; *(ii) The largest eigenvalue of* $E_P[q_{n,\jmath}^{k_{n,\jmath}}(Z_{i,\jmath}) q_{n,\jmath}^{k_{n,\jmath}}(Z_{i,\jmath})']$ *is bounded uniformly in* $1 \leq \jmath \leq \mathcal{J}$, n, *and* $P \in \mathbf{P}$; *(iii) The dimension of* $\mathbf{B}_n$ *is finite for any* n.

Assumption 3.3. *The classes* $\mathcal{F}_n$: *(i) Are closed under* $\|\cdot\|_{L_P^2}$; *(ii) Have envelope* F_n *with* $\sup_{P \in \mathbf{P}} E_P[F_n^2(V_i)] < \infty$; *(iii) Satisfy* $\sup_{P \in \mathbf{P}} J_{[]}(\|F_n\|_{L_P^2}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) \leq J_n$.

Assumption 3.4. *(i) For each* $P \in \mathbf{P}$ *there is a* $\Sigma_n(P) > 0$ *with* $\|\hat{\Sigma}_n - \Sigma_n(P)\|_{o,r} = o_p(1)$ *uniformly in* $P \in \mathbf{P}$; *(ii) The matrices* $\Sigma_n(P)$ *are invertible for all* n *and* $P \in \mathbf{P}$; *(iii)* $\|\Sigma_n(P)\|_{o,r}$ *and* $\|\Sigma_n(P)^{-1}\|_{o,r}$ *are uniformly bounded in* n *and* $P \in \mathbf{P}$.

Assumption 3.1 imposes that the sample $\{V_i\}_{i=1}^n$ be i.i.d. with P belonging to a set of distributions $\mathbf{P}$ over which our results will hold uniformly. In Assumption 3.2(i) we require the functions $\{q_{k,n,\jmath}\}_{k=1}^{k_{n,\jmath}}$ to be bounded by a constant B_n possibly diverging to infinity with the sample size. Hence, Assumption 3.2(i) accommodates both transformations that are uniformly bounded in n, such as trigonometric series, and those with diverging bound, such as b-splines, wavelets, and orthogonal polynomials (after orthonormalization). The bound on eigenvalues imposed in Assumption 3.2(ii) guarantees that $\{q_{k,n,\jmath}\}_{n=1}^{k_{n,\jmath}}$ are Bessel sequences uniformly in n, while Assumption 3.2(iii) formalizes that the sieve $\Theta_n \cap R$ be finite dimensional. In turn, Assumption 3.3 controls the "size" of the class $\mathcal{F}_n$, which is crucial in studying the induced empirical process. We note that the entropy integral is allowed to diverge with the sample size and thus accommodates non-compact parameter spaces Θ as in Chen and Pouzo (2012). Alternatively, if the class $\mathcal{F} \equiv \bigcup_{n=1}^\infty \mathcal{F}_n$ is restricted to be Donsker, then Assumptions 3.3(ii)-(iii) can hold

with uniformly bounded J_n and $\|F_n\|_{L_P^2}$. Finally, Assumption 3.4 imposes requirements on the weighting matrix $\hat{\Sigma}_n$ – namely, that it converge to an invertible matrix $\Sigma_n(P)$ possibly depending on P. Assumption 3.4 can of course be automatically satisfied under nonstochastic weights.

4   Rate of Convergence

As a preliminary step towards approximating the finite sample distribution of $I_n(R)$, we first aim to characterize the asymptotic behavior of the minimizers of Q_n on $\Theta_n \cap R$. Specifically, for any sequence $\tau_n \downarrow 0$ we study the probability limit of the set

$$\hat{\Theta}_n \cap R \equiv \{\theta \in \Theta_n \cap R : \frac{1}{\sqrt{n}} Q_n(\theta) \leq \inf_{\theta \in \Theta_n \cap R} \frac{1}{\sqrt{n}} Q_n(\theta) + \tau_n\} , \qquad (36)$$

which constitutes the set of exact ($\tau_n = 0$) or near ($\tau_n > 0$) minimizers of Q_n. We study the general case with $\tau_n \downarrow 0$ because results for both exact and near minimizers are needed in our analysis. In particular, the set of exact and near minimizers will be employed to respectively characterize and estimate the distribution of $I_n(R)$.

While it is natural to view $\Theta_0(P) \cap R$ as the candidate probability limit for $\hat{\Theta}_n \cap R$, it is in fact more fruitful to instead consider $\hat{\Theta}_n \cap R$ as consistent for the sets[9]

$$\Theta_{0n}(P) \cap R \equiv \arg\min_{\theta \in \Theta_n \cap R} \|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_r . \qquad (37)$$

Heuristically, $\Theta_{0n}(P) \cap R$ is the set of minimizers of a population version of Q_n where the number of moments k_n has been fixed and the parameter space has been set to $\Theta_n \cap R$ (instead of $\Theta \cap R$). As we will show, a suitable rate of convergence towards $\Theta_{0n}(P) \cap R$ suffices for establishing size control, and can in fact be obtained under weaker requirements than those needed for convergence towards $\Theta_0(P) \cap R$.

Following the literature on set estimation in finite dimensional settings (Chernozhukov et al., 2007; Beresteanu and Molinari, 2008; Kaido and Santos, 2014), we study set consistency under the Hausdorff metric. In particular, for any sets A and B we define

$$\overrightarrow{d}_H(A, B, \|\cdot\|) \equiv \sup_{a \in A} \inf_{b \in B} \|a - b\| \qquad (38)$$

$$d_H(A, B, \|\cdot\|) \equiv \max\{\overrightarrow{d}_H(A, B, \|\cdot\|), \overrightarrow{d}_H(B, A, \|\cdot\|)\} , \qquad (39)$$

which respectively constitute the directed Hausdorff distance and the Hausdorff distance under the metric $\|\cdot\|$. In contrast to finite dimensional problems, however, in the present

---

[9]Assumptions 3.3(i) and 3.3(iii) respectively imply $\mathcal{F}_n$ is closed and totally bounded under $\|\cdot\|_{L_P^2}$ and hence compact. It follows that the minimum in (37) is attained and $\Theta_{0n}(P) \cap R$ is well defined.

setting we emphasize the metric under which the Hausdorff distance is computed due to its importance in determining a rate of convergence; see also Santos (2011).

4.1 Consistency

We establish the consistency of $\hat{\Theta}_n \cap R$ under the following additional assumption:

Assumption 4.1. *(i)* $\sup_{P \in \mathbf{P}_0} \inf_{\theta \in \Theta_n \cap R} \|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_r \leq \zeta_n$ *for some* $\zeta_n \downarrow 0$; *(ii) Let* $(\Theta_{0n}(P) \cap R)^\epsilon \equiv \{\theta \in \Theta_n \cap R : \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}(P) \cap R, \|\cdot\|_\mathbf{B}) < \epsilon\}$ *and set*

$$S_n(\epsilon) \equiv \inf_{P \in \mathbf{P}_0} \inf_{\theta \in (\Theta_n \cap R) \setminus (\Theta_{0n}(P) \cap R)^\epsilon} \|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_r \tag{40}$$

for any $\epsilon > 0$. *Then* $\{\zeta_n + k_n^{1/r}\sqrt{\log(k_n)}J_n B_n/\sqrt{n}\} = o(S_n(\epsilon))$ *for any* $\epsilon > 0$.

Assumption 4.1(i) requires that the sieve $\Theta_n \cap R$ be such that the infimum

$$\inf_{\theta \in \Theta_n \cap R} \|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_r \tag{41}$$

converges to zero uniformly over $P \in \mathbf{P}_0$. Heuristically, since for any $P \in \mathbf{P}_0$ the infimum in (41) over the entire parameter space equals zero ($\Theta_0(P) \cap R \neq \emptyset$), Assumption 4.1(i) can be interpreted as demanding that the sieve $\Theta_n \cap R$ provide a suitable approximation to $\Theta \cap R$. In turn, the parameter $S_n(\epsilon)$ introduced in Assumption 4.1(ii) measures how "well separated" the infimum in (41) is (see (40)), while the quantity

$$\frac{k_n^{1/r}\sqrt{\log(k_n)}J_n B_n}{\sqrt{n}} \tag{42}$$

represents the rate at which the scaled criterion $Q_n/\sqrt{n}$ converges to its population analogue; see Lemma B.2. Thus, Assumption 4.1(ii) imposes that the rate at which "well separatedness" is lost ($S_n(\epsilon) \downarrow 0$) be slower than the rate at which $Q_n/\sqrt{n}$ converges to its population counterpart – a condition originally imposed in estimation problems with non compact parameter spaces by Chen and Pouzo (2012) who also discuss sufficient conditions for it; see Remark 4.1.

Given the introduced assumption, Lemma 4.1 establishes the consistency of $\hat{\Theta}_n \cap R$.

Lemma 4.1. *Let Assumptions 3.1, 3.2(i), 3.3, 3.4, and 4.1 hold. (i) If the sequence* $\{\tau_n\}$ *satisfies* $\tau_n = o(S_n(\epsilon))$ *for all* $\epsilon > 0$, *then it follows that uniformly in* $P \in \mathbf{P}_0$

$$\overrightarrow{d}_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \|\cdot\|_\mathbf{B}) = o_p(1) . \tag{43}$$

(ii) Moreover, if in addition $\{\tau_n\}$ *is such that* $(\frac{k_n^{1/r}\sqrt{\log(k_n)}J_n B_n}{\sqrt{n}} + \zeta_n) = o(\tau_n)$, *then*

$$\liminf_{n \to \infty} \inf_{P \in \mathbf{P}_0} P(\Theta_{0n}(P) \cap R \subseteq \hat{\Theta}_n \cap R) = 1 . \tag{44}$$

The first claim of Lemma 4.1 shows that, provided $\tau_n \downarrow 0$ sufficiently fast, $\hat{\Theta}_n \cap R$ is contained in arbitrary neighborhoods of $\Theta_{0n}(P) \cap R$ with probability approaching one. This conclusion will be of use when characterizing the distribution of $I_n(R)$. In turn, the second claim of Lemma 4.1 establishes that, provided $\tau_n \downarrow 0$ slowly enough, $\Theta_{0n}(P) \cap R$ is contained in $\hat{\Theta}_n \cap R$ with probability approaching one. This second conclusion will be of use when employing $\hat{\Theta}_n \cap R$ to construct an estimator of the distribution of $I_n(R)$.

Remark 4.1. Under Assumption 3.2(ii), it is possible to show there is a $C < \infty$ with

$$\inf_{\theta \in (\Theta_n \cap R) \backslash (\Theta_{0n}(P) \cap R)^{\epsilon}} \|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_r$$

$$\leq C \times \inf_{\theta \in (\Theta_n \cap R) \backslash (\Theta_0(P) \cap R)^{\epsilon}} \{\sum_{j=1}^{\mathcal{J}} \{E_P[(E_P[\rho_j(X_i, \theta)|Z_{i,j}])^2]\}^{\frac{1}{2}}\} ; \quad (45)$$

see Lemma C.5. Therefore, if the problem is ill-posed and the sieve $\Theta_n \cap R$ grows dense in $\Theta \cap R$, result (45) implies that $S_n(\epsilon) = o(1)$ for all $\epsilon > 0$ as in Chen and Pouzo (2012). In contrast, Newey and Powell (2003) address the ill-posed inverse problem by imposing compactness of the parameter space. Analogously, in our setting it is possible to show that if $\Theta \cap R$ is compact and $\{q_{k,n,j}\}_{k=1}^{k_{n,j}}$ are suitable dense, then

$$\liminf_{n \to \infty} \inf_{\theta \in (\Theta \cap R) \backslash (\Theta_{0n}(P) \cap R)^{\epsilon}} \|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)\|_r > 0 \quad (46)$$

for any $P \in \mathbf{P}_0$. Hence, under compactness of $\Theta \cap R$, it is possible for $\liminf S_n(\epsilon) > 0$, in which case Assumption 4.1(ii) follows from 4.1(i) and $k_n^{1/r}\sqrt{\log(k_n)}B_n J_n = o(\sqrt{n})$. ∎

4.2 Convergence Rate

The consistency result in Lemma 4.1 enables us to derive a rate of convergence by exploiting the local behavior of the population criterion function in a neighborhood of $\Theta_{0n}(P) \cap R$. We do not study a rate convergence under the original norm $\|\cdot\|_{\mathbf{B}}$, however, but instead introduce a potentially weaker norm we denote by $\|\cdot\|_{\mathbf{E}}$.

Heuristically, the need to introduce $\|\cdot\|_{\mathbf{E}}$ arises from the "strength" of $\|\cdot\|_{\mathbf{B}}$ being determined by the requirement that the maps $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ and $\Upsilon_G : \mathbf{B} \to \mathbf{G}$ be continuous under $\|\cdot\|_{\mathbf{B}}$; recall Assumption 2.2(ii). On the other hand, in approximating the distribution of $I_n(R)$ we must also rely on a metric under which the empirical process is stochastically equicontinuous – a purpose for which $\|\cdot\|_{\mathbf{B}}$ is often "too strong" with its use leading to overly stringent assumptions. Thus, while $\|\cdot\|_{\mathbf{B}}$ ensures continuity of the maps Υ_F and Υ_G, we employ a weaker norm $\|\cdot\|_{\mathbf{E}}$ to guarantee the stochastic equicontinuity of the empirical process – here "$\mathbf{E}$" stands for equicontinuity. The following assumption formally introduces $\|\cdot\|_{\mathbf{E}}$ and enables us to obtain a rate of convergence under the induced Hausdorff distance.

Assumption 4.2. *(i) For a Banach Space $\mathbf{E}$ with norm $\|\cdot\|_{\mathbf{E}}$ satisfying $\mathbf{B}_n \subseteq \mathbf{E}$ for all n, there is an $\epsilon > 0$ and sequence $\{\nu_n\}_{n=1}^{\infty}$ with $\nu_n^{-1} = O(1)$ such that*

$$\nu_n^{-1} \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{E}}) \leq \{\|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_r + O(\zeta_n)\}$$

for all $P \in \mathbf{P}_0$ and $\theta \in (\Theta_{0n}(P) \cap R)^{\epsilon} \equiv \{\theta \in \Theta_n \cap R : \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{B}}) < \epsilon\}$.

Intuitively, Assumption 4.2 may be interpreted as a generalization of a classical local identification condition. In particular, the parameter ν_n^{-1} measures the strength of identification, with large/small values of ν_n^{-1} indicating how quickly/slowly the criterion grows as θ moves away from the set of minimizers $\Theta_{0n}(P) \cap R$. The strength of identification, however, may decrease with n for at least two reasons. First, in ill-posed problems ν_n^{-1} decreases with the dimension of the sieve, reflecting that local identification is attained in finite dimensional subspaces but not on the entire parameter space. Second, the strength of identification is affected by the choice of norm $\|\cdot\|_r$ employed in the construction of Q_n. While the norms $\|\cdot\|_r$ are equivalent on any fixed finite dimensional space, their modulus of continuity can decrease with the number of moments which in turn affects ν_n^{-1}; see Remark 4.2.

The following Theorem exploits Assumption 4.2 to obtain a rate of convergence.

Theorem 4.1. *Let Assumptions 3.1, 3.2(i), 3.3, 3.4, 4.1, and 4.2 hold, and let*

$$\mathcal{R}_n \equiv \nu_n \{ \frac{k_n^{1/r} \sqrt{\log(k_n)} J_n B_n}{\sqrt{n}} + \zeta_n \} \ . \tag{47}$$

(i) If $\{\tau_n\}$ satisfies $\tau_n = o(S_n(\epsilon))$ for any $\epsilon > 0$, then it follows that uniformly in $P \in \mathbf{P}_0$

$$\overrightarrow{d}_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{E}}) = O_p(\mathcal{R}_n + \nu_n \tau_n) \ . \tag{48}$$

(ii) Moreover, if in addition $(\frac{k_n^{1/r} \sqrt{\log(k_n)} J_n B_n}{\sqrt{n}} + \zeta_n) = o(\tau_n)$, then uniformly in $P \in \mathbf{P}_0$

$$d_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{E}}) = O_p(\mathcal{R}_n + \nu_n \tau_n) \ . \tag{49}$$

Together, Lemma 4.1 and Theorem 4.1 establish the consistency (in $\|\cdot\|_{\mathbf{B}}$) and rate of convergence (in $\|\cdot\|_{\mathbf{E}}$) of the set estimator $\hat{\Theta}_n \cap R$. While we exploit these results in our forthcoming analysis, it is important to emphasize that in specific applications alternative assumptions that are better suited for the particular structure of the model may be preferable. In this regard, we note that Assumptions 4.1 and 4.2 are not needed in our analysis beyond their role in delivering consistency and a rate of convergence result through Lemma 4.1 and Theorem 4.1 respectively. In particular, if an alternative rate of convergence $\mathcal{R}_n$ is derived under different assumptions, then such a result can still be combined with our forthcoming analysis to establish the validity of the proposed

17

inferential methods – i.e. our inference results remain valid if Assumptions 4.1 and 4.2 are instead replaced with a high level condition that $\hat{\Theta}_n \cap R$ be consistent (in $\|\cdot\|_{\mathbf{B}}$) with an appropriate rate of convergence $\mathcal{R}_n$ (in $\|\cdot\|_{\mathbf{E}}$).

Remark 4.2. In models in which $\Theta_{0n}(P) \cap R$ is a singleton, Assumption 4.2 is analogous to a standard local identification condition (Chen et al., 2014). In particular, suppose $\{b_j\}_{j=1}^{j_n}$ is a basis for $\mathbf{B}_n$ and for each $1 \leq j \leq j_n$ and $\theta \in (\Theta_{0n}(P) \cap R)^\epsilon$ define

$$A_{P,n}^{(j)}(\theta) \equiv \frac{\partial}{\partial \tau} E_P[\rho(X_i, \theta + \tau b_j) * q_n^{k_n}(Z_i)]\Big|_{\tau=0} \tag{50}$$

and set $A_{P,n}(\theta) \equiv [A_{P,n}^{(1)}(\theta), \ldots, A_{P,n}^{(j_n)}(\theta)]$. Further let $\alpha : \mathbf{B}_n \to \mathbf{R}^{j_n}$ be such that

$$b = \sum_{j=1}^{j_n} \alpha_j(b) \times b_j \ , \tag{51}$$

for any $b \in \mathbf{B}_n$ and $\alpha(b) = (\alpha_1(b), \ldots, \alpha_{j_n}(b))'$. If the smallest singular value of $A_{P,n}(\theta)$ is bounded from below by some $\vartheta_n > 0$ uniformly in $\theta \in (\Theta_{0n}(P) \cap R)^\epsilon$ and $P \in \mathbf{P}_0$, then it is straightforward to show Assumption 4.2 holds with $\nu_n = \vartheta_n^{-1} \times k_n^{1/2-1/r}$ and the norm $\|b\|_{\mathbf{E}} = \|\alpha(b)\|_2$ – a norm that is closely related to $\|\cdot\|_{L_P^2}$ when $\mathbf{B} \subseteq L_P^2$. ∎

5 Strong Approximation

In this section, we exploit the rate of convergence derived in Theorem 4.1 to obtain a strong approximation to the proposed test statistic $I_n(R)$. We proceed in two steps. First, we construct a preliminary local approximation involving the norm of a Gaussian process with an unknown "drift". Second, we refine the initial approximation by linearizing the "drift" while accommodating possibly severely ill-posed problems.

5.1 Local Approximation

The first strong approximation to our test statistic relies on the following assumptions:

Assumption 5.1. *(i)* $\sup_{f \in \mathcal{F}_n} \|\mathbb{G}_{n,P} f q_n^{k_n} - \mathbb{W}_{n,P} f q_n^{k_n}\|_r = o_p(a_n)$ *uniformly in* $P \in \mathbf{P}$, *where* $\{a_n\}_{n=1}^\infty$ *is some known bounded sequence.*

Assumption 5.2. *(i) There exist* $\kappa_\rho > 0$ *and* $K_\rho < \infty$ *such that for all* n, $P \in \mathbf{P}$, *and all* $\theta_1, \theta_2 \in \Theta_n \cap R$ *we have that* $E_P[\|\rho(X_i, \theta_1) - \rho(X_i, \theta_2)\|_2^2] \leq K_\rho^2 \|\theta_1 - \theta_2\|_{\mathbf{E}}^{2\kappa_\rho}$.

Assumption 5.3. *(i)* $k_n^{1/r}\sqrt{\log(k_n)} B_n \sup_{P \in \mathbf{P}} J_{[]}(\mathcal{R}_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$; *(ii)* $\sqrt{n}\zeta_n = o(a_n)$; *(iii)* $\|\hat{\Sigma}_n - \Sigma_n(P)\|_{o,r} = o_p(a_n\{k_n^{1/r}\sqrt{\log(k_n)} B_n J_n\}^{-1})$ *uniformly in* $P \in \mathbf{P}$.

18

Assumption 5.1(i) requires that the empirical process $\mathbb{G}_{n,P}$ be approximated by an isonormal Gaussian process $\mathbb{W}_{n,P}$ uniformly in $P \in \mathbf{P}$. Intuitively, Assumption 5.1(i) replaces the traditional requirement of convergence in distribution by a strong approximation, which is required to handle the asymptotically non-Donsker setting that arises naturally in our case and other related problems; see Chernozhukov et al. (2013) for further discussion. The sequence $\{a_n\}_{n=1}^{\infty}$ in Assumption 5.1(i) denotes a bound on the rate of convergence of the coupling to the empirical process, which will in turn characterize the rate of convergence of our strong approximation to $I_n(R)$. We provide sufficient conditions for verifying Assumption 5.1(i) based on Koltchinskii (1994)'s coupling in Corollary G.1 in Appendix E. These results could be of independent interest. Alternatively, Assumption 5.1(i) can be verified by employing methods based on Rio (1994)'s coupling or Yurinskii (1977)'s couplings; see e.g., Chernozhukov et al. (2013). Assumption 5.2(i) is a Hölder continuity condition on the map $\rho(\cdot, X_i) : \Theta_n \cap R \to \{L_P^2\}^{\mathcal{J}}$ with respect to the norm $\| \cdot \|_{\mathbf{E}}$, and thus ensures that $\mathbb{W}_{n,P} f q_n^{k_n}$ is equicontinuous with respect to the index θ under $\| \cdot \|_{\mathbf{E}}$ for fixed n. However, this process gradually looses its equicontinuity property as n diverges infinity due to the addition of moments and increasing complexity of the class $\mathcal{F}_n$. Hence, Assumption 5.3(i) demands that the $\| \cdot \|_{\mathbf{E}}$-rate of convergence ($\mathcal{R}_n$) be sufficiently fast to overcome the loss of equicontinuity at a rate no slower than a_n (as in Assumption 5.1(i)). Finally, Assumption 5.3(ii) ensures the test statistic is asymptotically properly centered under the null hypothesis, while Assumption 5.3(iii) controls the rate of convergence of the weighting matrix.

Together, the results of Section 4 and Assumptions 5.1, 5.2, and 5.3 enable us to obtain a strong approximation to the test statistic $I_n(R)$. To this end, we define

$$V_n(\theta, \ell) \equiv \{ \frac{h}{\sqrt{n}} \in \mathbf{B}_n : \theta + \frac{h}{\sqrt{n}} \in \Theta_n \cap R \text{ and } \| \frac{h}{\sqrt{n}} \|_{\mathbf{E}} \le \ell \} , \tag{52}$$

which for any $\theta \in \Theta_n \cap R$ constitutes the collection of local deviations from θ that remain in the constrained sieve $\Theta_n \cap R$. Thus the local parameter space $V_n(\theta, \ell)$ is indexed by θ which runs over $\Theta_n \cap R$ and parameterized by deviations $h/\sqrt{n}$ from θ; this follows previous uses in Chernozhukov et al. (2007) and Santos (2007). The normalization by $\sqrt{n}$ plays no particular role here, since ℓ can grow and merely visually emphasizes localization. By Theorem 4.1, it then follows that in studying $I_n(R)$ we need not consider the infimum over the entire sieve (see (26)) but may instead examine the infimum over local deviations to $\Theta_{0n}(P) \cap R$ – i.e. the infimum over parameters

$$(\theta_0, \frac{h}{\sqrt{n}}) \in (\Theta_{0n}(P) \cap R, V_n(\theta_0, \ell_n)) \tag{53}$$

with the neighborhood $V_n(\theta_0, \ell_n)$ shrinking at an appropriate rate ($\mathcal{R}_n = o(\ell_n)$). In turn, Assumptions 5.1, 5.2, and 5.3 control the relevant stochastic processes over the localized space (53) and allow us to characterize the distribution of $I_n(R)$.

The following Lemma formalizes the preceding discussion.

Lemma 5.1. *Let Assumptions 3.1, 3.2(i), 3.3, 3.4, 4.1, 4.2, 5.1, 5.2, and 5.3 hold. It then follows that for any sequence $\{\ell_n\}$ satisfying $\mathcal{R}_n = o(\ell_n)$ and $k_n^{1/r} \sqrt{\log(k_n)} B_n \times \sup_{P \in \mathbf{P}} J_{[]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \| \cdot \|_{L_P^2}) = o(a_n)$, we have uniformly in $P \in \mathbf{P}_0$ that*

$$I_n(R) = \inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \| \mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \sqrt{n} P \rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n} \|_{\Sigma_n(P), r} + o_p(a_n)$$

Lemma 5.1 establishes our first strong approximation and further characterizes the rate of convergence to be no slower than a_n (as in Assumption 5.1). Thus, for a consistent coupling we only require that Assumptions 5.1 and 5.3 hold with $\{a_n\}_{n=1}^\infty$ a bounded sequence. In certain applications, however, successful estimation of critical values will additionally require us to impose that a_n be logarithmic or double logarithmic; see Section 6.3. We further note that for the conclusion of Lemma 5.1 to hold, the neighborhoods $V_n(\theta, \ell_n)$ must shrink at a rate ℓ_n satisfying two conditions. First, ℓ_n must decrease to zero slowly enough to ensure the infimum over the entire sieve is indeed equivalent to the infimum over the localized space ($\mathcal{R}_n = o(\ell_n)$). Second, ℓ_n must decrease to zero sufficiently fast to overcome the gradual loss of equicontinuity of the isonormal process $\mathbb{W}_{n,P}$ – notice $\mathbb{W}_{n,P}$ is evaluated at $\rho(\cdot, \theta_0) * q_n^{k_n}$ in place of $\rho(\cdot, \theta_0 + h/\sqrt{n}) * q_n^{k_n}$. The existence of a sequence ℓ_n satisfying these requirements is guaranteed by Assumption 5.3(i). However, as we next discuss, the approximation in Lemma 5.1 must be further refined before it can be exploited for inference.

5.2 Drift Linearization

A challenge arising from Lemma 5.1, is the need for a tractable expression for the term

$$\sqrt{n} E_P[\rho(X_i, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n}(Z_i)] , \tag{54}$$

which we refer to as the local "drift" of the isonormal process. Typically, the drift is approximated by a linear function of the local parameter h by requiring an appropriate form of differentiability of the moment functions. In this section, we build on this approach by requiring differentiability of the maps $m_{P,j} : \Theta \cap R \to L_P^2$ defined by

$$m_{P,j}(\theta)(Z_{i,j}) \equiv E_P[\rho_j(X_i, \theta)|Z_{i,j}] . \tag{55}$$

In the same manner that a norm $\| \cdot \|_{\mathbf{E}}$ was needed to ensure stochastic equicontinuity of the empirical process, we now introduce a final norm $\| \cdot \|_{\mathbf{L}}$ to deliver differentiability of the maps $m_{P,j}$ – here "$\mathbf{L}$" stands for linearization. Thus, $\| \cdot \|_{\mathbf{B}}$ is employed to ensure

smoothness of the maps Υ_F and Υ_G, $\|\cdot\|_{\mathbf{E}}$ guarantees the stochastic equicontinuity of $\mathbb{G}_{n,P}$, and $\|\cdot\|_{\mathbf{L}}$ delivers the smoothness of the maps $m_{P,\jmath}$. Formally, we impose:

Assumption 5.4. *For a Banach space $\mathbf{L}$ with norm $\|\cdot\|_{\mathbf{L}}$ and $\mathbf{B}_n \subseteq \mathbf{L}$ for all n, there are $K_m < \infty$, $M_m < \infty$, $\epsilon > 0$ such that for all $1 \le \jmath \le \mathcal{J}$, $n \in \mathbb{N}$, $P \in \mathbf{P}_0$, and $\theta_1 \in (\Theta_{0n}(P) \cap R)^\epsilon$, there is a linear $\nabla m_{P,\jmath}(\theta_1) : \mathbf{B} \to L_P^2$ satisfying for all $h \in \mathbf{B}_n$: (i) $\|m_{P,\jmath}(\theta_1 + h) - m_{P,\jmath}(\theta_1) - \nabla m_{P,\jmath}(\theta_1)[h]\|_{L_P^2} \le K_m \|h\|_{\mathbf{L}} \|h\|_{\mathbf{E}}$; (ii) $\|\nabla m_{P,\jmath}(\theta_1)[h] - \nabla m_{P,\jmath}(\theta_0)[h]\|_{L_P^2} \le K_m \|\theta_1 - \theta_0\|_{\mathbf{L}} \|h\|_{\mathbf{E}}$; and (iii) $\|\nabla m_{P,\jmath}(\theta_0)[h]\|_{L_P^2} \le M_m \|h\|_{\mathbf{E}}$.*

Heuristically, Assumption 5.4(i) simply demands that the functions $m_{P,\jmath} : \Theta \cap R \to L_P^2$ be locally well approximated under $\|\cdot\|_{L_P^2}$ by linear maps $\nabla m_{P,\jmath} : \mathbf{B} \to L_P^2$. Moreover, the approximation error is required to be controlled by the product of the $\|\cdot\|_{\mathbf{E}}$ and $\|\cdot\|_{\mathbf{L}}$ norms; see Remark 5.1 for a leading example. We emphasize, however, that Assumption 5.4(i) does not require the generalized residuals $\rho_\jmath(X_i, \cdot) : \Theta \cap R \to \mathbf{R}$ themselves to be differentiable, and thus accommodates models such as the nonparametric quantile IV regression of Chernozhukov and Hansen (2005). In addition, we note that whenever $\rho_\jmath(X_i, \theta)$ is linear in θ, such as in the nonparametric IV regression of Newey and Powell (2003), Assumption 5.4(i) is automatically satisfied with $K_m = 0$. Finally, Assumptions 5.4(ii) and 5.4(iii) respectively require the derivatives $\nabla m_{P,\jmath}(\theta) : \mathbf{B} \to L_P^2$ to be Lipschitz continuous in θ with respect to $\|\cdot\|_{\mathbf{L}}$ and norm bounded uniformly on $\theta \in \Theta_{0n}(P) \cap R$ and $P \in \mathbf{P}_0$. The latter two assumptions are not required for the purposes of refining the strong approximation of Lemma 5.1, but will be needed for the study of our inferential procedure in Section 6.

Given Assumption 5.4, we next aim to approximate the local drift in Lemma 5.1 (see (54)) by a linear map $\mathbb{D}_{n,P}(\theta_0) : \mathbf{B}_n \to \mathbf{R}^{k_n}$ pointwise defined by

$$\mathbb{D}_{n,P}(\theta_0)[h] \equiv E_P[\nabla m_P(\theta_0)[h](Z_i) * q_n^{k_n}(Z_i)] . \tag{56}$$

where $\nabla m_P(\theta_0)[h](Z_i) \equiv (\nabla m_{P,1}(\theta_0)[h](Z_{i,1}), \dots, \nabla m_{P,\mathcal{J}}(\theta_0)[h](Z_{i,\mathcal{J}}))'$. Regrettably, it is well understood that, particularly in severely ill-posed problems, the rate of convergence may be too slow for $\mathbb{D}_{n,P}(\theta_0)$ to approximate the drift uniformly over the local parameter space in nonlinear models (Chen and Pouzo, 2009; Chen and Reiss, 2011). However, while such a complication can present important challenges when employing the asymptotic distribution of estimators for inference, severely ill-posed problems can still be accommodated in our setting. Specifically, instead of considering the entire local parameter space, as in Lemma 5.1, we may restrict attention to an infimum over the subset of the local parameter space for which an approximation of the drift by $\mathbb{D}_{n,P}(\theta_0)$ is indeed warranted. The resulting bound for $I_n(R)$ is potentially conservative in nonlinear models when the rate of convergence $\mathcal{R}_n$ is not sufficiently fast, but remains asymptotically equivalent to $I_n(R)$ in the remaining settings.

21

Figure 1: Local Drift Linearization
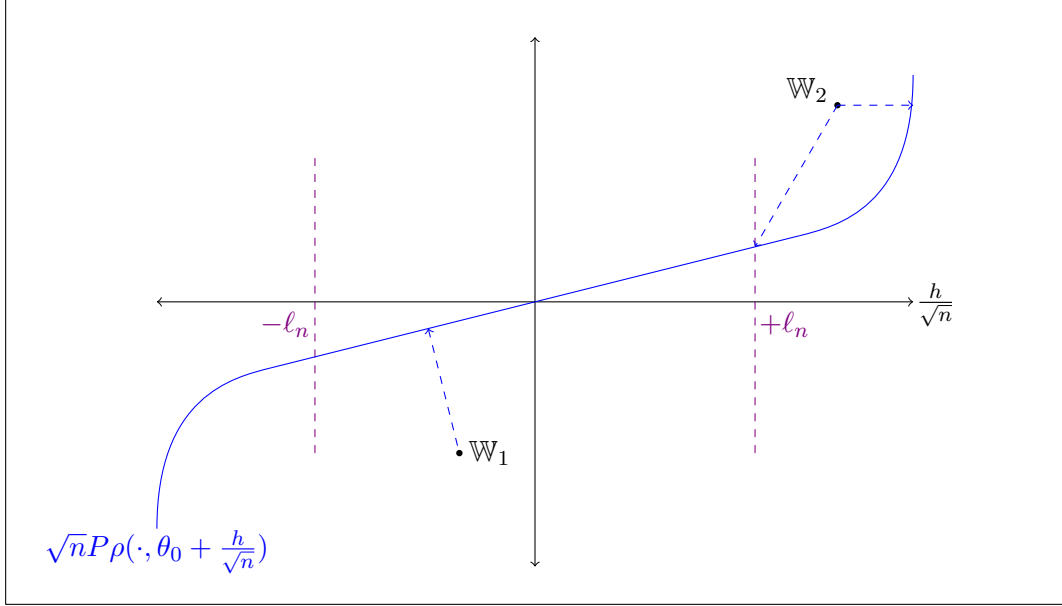
Our next theorem characterizes the properties of the described strong approximation. It is helpful to note here that the notation $\mathcal{S}_n(\mathbf{A}_1, \mathbf{A}_2)$ is defined in Section 3.2.1 as the modulus of continuity (on $\mathbf{B}_n$) between the norms on two spaces $\mathbf{A}_1$ and $\mathbf{A}_2$ (see (35)).

Theorem 5.1. *Let Assumptions 3.1, 3.2, 3.3, 3.4, 4.1, 4.2, 5.1, 5.2, 5.3, and 5.4(i) hold. (i) Then, for any sequence $\{\ell_n\}$ satisfying $K_m \ell_n^2 \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$ and $k_n^{1/r} \sqrt{\log(k_n)} B_n \times \sup_{P \in \mathbf{P}} J_{[]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$ it follows that*

$$I_{n,P}(R) \leq \inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \|\mathbb{W}_{n,P}\rho(\cdot, \theta_0) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_0)[h]\|_{\Sigma_n(P), r} + o_p(a_n) ,$$

uniformly in $P \in \mathbf{P}_0$. (ii) Moreover, if in addition $K_m \mathcal{R}_n^2 \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$, then the sequence $\{\ell_n\}$ may be chosen so that uniformly in $P \in \mathbf{P}_0$

$$I_{n,P}(R) = \inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \|\mathbb{W}_{n,P}\rho(\cdot, \theta_0) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_0)[h]\|_{\Sigma_n(P), r} + o_p(a_n) .$$

The conclusion of Theorem 5.1 can be readily understood through Figure 1, which illustrates the special case in which $\mathcal{J} = 1$, $k_n = 2$, $\mathbf{B}_n = \mathbf{R}$, and $V_n(\theta_0, +\infty) = \mathbf{R}$. In this context, the Gaussian process $\mathbb{W}_{n,P}\rho(\cdot, \theta_0) * q_n^{k_n}$ is simply a bivariate normal random variable in $\mathbf{R}^2$ that we denote by $\mathbb{W}$ for conciseness. In turn, the drift is a surface on $\mathbf{R}^2$ that is approximately linear (equal to $\mathbb{D}_{n,P}(\theta_0)[h]$) in a neighborhood of zero. According to Lemma 5.1, $I_n(R)$ is then asymptotically equivalent to the distance between $\mathbb{W}$ and the surface representing the drift. Intuitively, Theorem 5.1(i) then bounds $I_n(R)$ by the distance between $\mathbb{W}$ and the restriction of the drift surface to the region where it is linear – a bound that may be equal to or strictly larger than $I_n(R)$ as illustrated

by the realizations $\mathbb{W}_1$ and $\mathbb{W}_2$ respectively. However, if the rate of convergence $\mathcal{R}_n$ is sufficiently fast or $\rho(X_i, \theta)$ is linear in θ ($K_m = 0$), then Theorem 5.1(ii) establishes $I_n(R)$ is in fact asymptotically equivalent to the derived bound – i.e. the realizations of $\mathbb{W}$ behave in the manner of $\mathbb{W}_1$ and not that of $\mathbb{W}_2$.

Remark 5.1. In an important class of models studied by Newey and Powell (2003), $\mathbf{B} \subseteq L_P^2$ and the generalized residual function $\rho(X_i, \theta)$ has the structure

$$\rho(X_i, \theta) = \tilde{\rho}(X_i, \theta(V_i)) \tag{57}$$

for a known map $\tilde{\rho} : \mathbf{R}^{d_x} \times \mathbf{R} \to \mathbf{R}$. Suppose $\tilde{\rho}(X_i, \cdot) : \mathbf{R} \to \mathbf{R}$ is differentiable for all X_i with derivative denoted $\nabla_\theta \tilde{\rho}(X_i, \cdot)$ and satisfying for some $L_m < \infty$

$$|\nabla_\theta \tilde{\rho}(X_i, u_1) - \nabla_\theta \tilde{\rho}(X_i, u_2)| \le L_m |u_1 - u_2| . \tag{58}$$

It is then straightforward to verify that Assumptions 5.4(i)-(ii) hold with $K_m = L_m$, $\| \cdot \|_{\mathbf{E}} = \| \cdot \|_{L_P^2}$, and $\| \cdot \|_{\mathbf{L}} = \| \cdot \|_{L_P^\infty}$, while Assumption 5.4(iii) is satisfied provided $\nabla_\theta \tilde{\rho}(X_i, \theta_0(V_i))$ is bounded uniformly in (X_i, V_i), $\theta_0 \in \Theta_{0n}(P) \cap R$, and $P \in \mathbf{P}_0$. ∎

6   Bootstrap Inference

The results of Section 5 establish a strong approximation to our test statistic and thus provide us with a candidate distribution whose quantiles may be employed to conduct valid inference. In this section, we develop an estimator for the approximating distribution derived in Section 5 and study its corresponding critical values.

6.1   Bootstrap Statistic

Theorem 5.1 indicates a valid inferential procedure can be constructed by comparing the test statistic $I_n(R)$ to the quantiles of the distribution of the random variable

$$U_{n,P}(R) \equiv \inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \| \mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_0)[h] \|_{\Sigma_n(P), r} . \tag{59}$$

In particular, as a result of Theorem 5.1(i), we may expect that employing the quantiles of $U_{n,P}(R)$ as critical values for $I_n(R)$ can control asymptotic size even in severely ill-posed nonlinear problems. Moreover, as a result of Theorem 5.1(ii), we may further expect the asymptotic size of the resulting test to equal its significance level at least in linear problems ($K_m = 0$) or when the rate of convergence ($\mathcal{R}_n$) is sufficiently fast.

In what follows, we construct an estimator of the distribution of $U_{n,P}(R)$ by replacing the population parameters in (59) with suitable sample analogues. To this end, we note

that by Theorem 4.1 and Assumption 3.4(i), the set $\Theta_{0n}(P) \cap R$ and weighting matrix $\Sigma_n(P)$ may be estimated by $\hat{\Theta}_n \cap R$ and $\hat{\Sigma}_n$ respectively. Thus, in mimicking (59), we only additionally require sample analogues $\hat{\mathbb{W}}_n$ for the isonormal process $\mathbb{W}_{n,P}$, $\hat{\mathbb{D}}_n(\theta)$ for the derivative $\mathbb{D}_{n,P}(\theta)$, and $\hat{V}_n(\theta, \ell_n)$ for the local parameter space $V_n(\theta, \ell_n)$. Given such analogues we may then approximate the distribution of $U_{n,P}(R)$ by that of

$$\hat{U}_n(R) \equiv \inf_{\theta \in \hat{\Theta}_n \cap R} \inf_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, \ell_n)} \|\hat{\mathbb{W}}_n \rho(\cdot, \theta) * q_n^{k_n} + \hat{\mathbb{D}}_n(\theta)[h]\|_{\hat{\Sigma}_n, r} . \tag{60}$$

In the next two sections, we first propose standard estimators for the isonormal process $\mathbb{W}_{n,P}$ and derivative $\mathbb{D}_{n,P}(\theta)$, and subsequently address the more challenging task of constructing an appropriate sample analogue for the local parameter space $V_n(\theta, \ell_n)$.

6.1.1 The Basics

We approximate the law of the isonormal process $\mathbb{W}_{n,P}$ by relying on the multiplier bootstrap (Ledoux and Talagrand, 1988). Specifically, for an i.i.d. sample $\{\omega_i\}_{i=1}^n$ with ω_i following a standard normal distribution and independent of $\{V_i\}_{i=1}^n$ we set

$$\hat{\mathbb{W}}_n f \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \omega_i \{f(V_i) - \frac{1}{n} \sum_{j=1}^n f(V_j)\} \tag{61}$$

for any function $f \in L_P^2$. Since $\{\omega_i\}_{i=1}^n$ are standard normal random variables drawn independently of $\{V_i\}_{i=1}^n$, it follows that conditionally on $\{V_i\}_{i=1}^n$ the law of $\hat{\mathbb{W}}_n f$ is also Gaussian, has mean zero, and in addition satisfies for any f and g (compare to (33))

$$E[\hat{\mathbb{W}}_n f \hat{\mathbb{W}}_n g | \{V_i\}_{i=1}^n] = \frac{1}{n} \sum_{i=1}^n (f(V_i) - \frac{1}{n} \sum_{j=1}^n f(V_j))(g(V_i) - \frac{1}{n} \sum_{j=1}^n g(V_j)) . \tag{62}$$

Hence, $\hat{\mathbb{W}}_n$ can be simply viewed as a Gaussian process whose covariance kernel equals the sample analogue of the unknown covariance kernel of $\mathbb{W}_{n,P}$.

In order to estimate the derivative $\mathbb{D}_{n,P}(\theta)$ we for concreteness adopt a construction that is applicable to nondifferentiable generalized residuals $\rho(X_i, \cdot) : \Theta \cap R \to \mathbf{R}^{\mathcal{J}}$. Specifically, we employ a local difference of the empirical process by setting

$$\hat{\mathbb{D}}_n(\theta)[h] \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n (\rho(X_i, \theta + \frac{h}{\sqrt{n}}) - \rho(X_i, \theta)) * q_n^{k_n}(Z_i) \tag{63}$$

for any $\theta \in \Theta_n \cap R$ and $h \in \mathbf{B}_n$; see also Hong et al. (2010) for a related study on numerical derivatives. We note, however, that while we adopt the estimator in (63) due to its general applicability, alternative approaches may be preferable in models where the generalized residual $\rho(X_i, \theta)$ is actually differentiable in θ; see Remark 6.1.

Remark 6.1. In settings in which the generalized residual $\rho(X_i, \theta)$ is pathwise partially differentiable in θ P-almost surely, we may instead define $\hat{\mathbb{D}}_n(\theta)[h]$ to be

$$\hat{\mathbb{D}}_n(\theta)[h] \equiv \frac{1}{n}\sum_{i=1}^{n} \nabla_\theta \rho(X_i, \theta)[h] * q_n^{k_n}(Z_i) , \tag{64}$$

where $\nabla_\theta \rho(x, \theta)[h] \equiv \frac{\partial}{\partial \tau}\rho(x, \theta + \tau h)|_{\tau=0}$. It is worth noting that, when applicable, employing (64) in place of (63) is preferable because the former is linear in h, and thus the resulting bootstrap statistic $\hat{U}_n(R)$ (as in (60)) is simpler to compute. ∎

6.1.2 The Local Parameter Space

The remaining component we require to obtain a bootstrap approximation is a suitable sample analogue for the local parameter space. We next develop such a sample analogue, which may be of independent interest as it is more broadly applicable to hypothesis testing problems concerning general equality and inequality restrictions in settings beyond the conditional moment restriction model; see Appendix E for the relevant results.

6.1.2.1 Related Assumptions

The construction of an approximation to the local parameter space first requires us to impose additional conditions on the sieve $\Theta_n \cap R$ and the restriction maps Υ_F and Υ_G.

Assumption 6.1. *(i) For some $K_b < \infty$, $\|h\|_{\mathbf{E}} \le K_b\|h\|_{\mathbf{B}}$ for all n, $h \in \mathbf{B}_n$; (ii) For some $\epsilon > 0$, $\bigcup_{P \in \mathbf{P}_0}\{\theta \in \mathbf{B}_n : \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{B}}) < \epsilon\} \subseteq \Theta_n$ for all n.*

Assumption 6.2. *There exist $K_g < \infty$, $M_g < \infty$, and $\epsilon > 0$ such that for all n, $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}(P) \cap R$, and $\theta_1, \theta_2 \in \{\theta \in \mathbf{B}_n : \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{B}}) < \epsilon\}$: (i) There is a linear map $\nabla \Upsilon_G(\theta_1) : \mathbf{B} \to \mathbf{G}$ satisfying $\|\Upsilon_G(\theta_1) - \Upsilon_G(\theta_2) - \nabla \Upsilon_G(\theta_1)[\theta_1 - \theta_2]\|_{\mathbf{G}} \le K_g\|\theta_1 - \theta_2\|_{\mathbf{B}}^2$; (ii) $\|\nabla \Upsilon_G(\theta_1) - \nabla \Upsilon_G(\theta_0)\|_o \le K_g\|\theta_1 - \theta_0\|_{\mathbf{B}}$; (iii) $\|\nabla \Upsilon_G(\theta_1)\|_o \le M_g$.*

Assumption 6.1(i) imposes that the norm $\|\cdot\|_{\mathbf{B}}$ be weakly stronger than the norm $\|\cdot\|_{\mathbf{E}}$ uniformly on the sieve $\Theta_n \cap R$.[10] We note that even though the local parameter space $V_n(\theta, \ell)$ is determined by the $\|\cdot\|_{\mathbf{E}}$ norm (see (52)), Assumption 6.1(i) implies restricting the norm $\|\cdot\|_{\mathbf{B}}$ instead can deliver a subset of $V_n(\theta, \ell)$. In turn, Assumption 6.1(ii) demands that $\Theta_{0n}(P) \cap R$ be contained in the interior of Θ_n uniformly in $P \in \mathbf{P}$. We emphasize, however, that such a requirement does not rule out binding parameter space restrictions. Instead, Assumption 6.1(ii) simply requires that all such restrictions be explicitly stated through the set R; see Remarks 6.2 and 6.3. Finally, Assumption

---

[10]Since $\mathbf{B}_n$ is finite dimensional, there always exists a constant K_n such that $\|h\|_{\mathbf{E}} \le K_n\|h\|_{\mathbf{B}}$ for all $h \in \mathbf{B}_n$. Thus, the main content of Assumption of 6.1(i) is that K_b does not depend on n.

6.2 imposes that $\Upsilon_G : \mathbf{B} \to \mathbf{G}$ be Fréchet differentiable in a neighborhood of $\Theta_{0n}(P) \cap R$ with locally Lipschitz continuous and norm bounded derivative $\nabla \Upsilon_G(\theta) : \mathbf{B} \to \mathbf{G}$.

In order to introduce analogous requirements for the map $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ we first define

$$\mathbf{F}_n \equiv \overline{\operatorname{span}}\{ \bigcup_{\theta \in \mathbf{B}_n} \Upsilon_F(\theta) \} , \tag{65}$$

where recall for any set C, $\overline{\operatorname{span}}\{C\}$ denotes the closure of the linear span of C – i.e. $\mathbf{F}_n$ denotes the closed linear span of the range of $\Upsilon_F : \mathbf{B}_n \to \mathbf{F}$. In addition, for any linear map $\Gamma : \mathbf{B} \to \mathbf{F}$ we denote its null space by $\mathcal{N}(\Gamma) \equiv \{h \in \mathbf{B} : \Gamma(h) = 0\}$. Given these definitions, we next impose the following requirements on Υ_F and its relation to Υ_G:

Assumption 6.3. *There exist $K_f < \infty$, $M_f < \infty$, and $\epsilon > 0$ such that for all n, $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}(P) \cap R$, and $\theta_1, \theta_2 \in \{\theta \in \mathbf{B}_n : \vec{d}_H(\{\theta\}, \Theta_{0n}(P) \cap R, \| \cdot \|_{\mathbf{B}}) < \epsilon\}$: (i) There is a linear map $\nabla \Upsilon_F(\theta_1) : \mathbf{B} \to \mathbf{F}$ satisfying $\|\Upsilon_F(\theta_1) - \Upsilon_F(\theta_2) - \nabla \Upsilon_F(\theta_1)[\theta_1 - \theta_2]\|_{\mathbf{F}} \leq K_f\|\theta_1 - \theta_2\|_{\mathbf{B}}^2$; (ii) $\|\nabla \Upsilon_F(\theta_1) - \nabla \Upsilon_F(\theta_0)\|_o \leq K_f\|\theta_1 - \theta_0\|_{\mathbf{B}}$; (iii) $\|\nabla \Upsilon_F(\theta_1)\|_o \leq M_f$; (iv) $\nabla \Upsilon_F(\theta_1) : \mathbf{B}_n \to \mathbf{F}_n$ admits a right inverse $\nabla \Upsilon_F(\theta_1)^-$ with $K_f\|\nabla \Upsilon_F(\theta_1)^-\|_o \leq M_f$.*

Assumption 6.4. *Either (i) $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ is linear, or (ii) There are constants $\epsilon > 0$, $K_d < \infty$ such that for every $P \in \mathbf{P}_0$, n, and $\theta_0 \in \Theta_{0n}(P) \cap R$ there exists a $h_0 \in \mathbf{B}_n \cap \mathcal{N}(\nabla \Upsilon_F(\theta_0))$ satisfying $\Upsilon_G(\theta_0) + \nabla \Upsilon_G(\theta_0)[h_0] \leq -\epsilon \mathbf{1}_{\mathbf{G}}$ and $\|h_0\|_{\mathbf{B}} \leq K_d$.*

Assumptions 6.3 and 6.4 mark an important difference between hypotheses in which Υ_F is linear and those in which Υ_F is nonlinear – in fact, in the former case Assumptions 6.3 and 6.4 are always satisfied. This distinction reflects that when Υ_F is linear its impact on the local parameter space is known and hence need not be estimated. In contrast, when Υ_F is nonlinear its role in determining the local parameter space depends on the point of evaluation $\theta_0 \in \Theta_{0n}(P) \cap R$ and is as a result unknown.[11] In particular, we note that while Assumptions 6.3(i)-(iii) impose smoothness conditions analogous to those required of Υ_G, Assumption 6.3(iv) additionally demands that the derivative $\nabla \Upsilon_F(\theta) : \mathbf{B}_n \to \mathbf{F}_n$ posses a norm bounded right inverse for all θ in a neighborhood of $\Theta_{0n}(P) \cap R$. Existence of a right inverse is equivalent to the surjectivity of the derivative $\nabla \Upsilon_F(\theta) : \mathbf{B}_n \to \mathbf{F}_n$ and hence amounts to the classical rank condition (Newey and McFadden, 1994). In turn, the requirement that the right inverse's operator norm be uniformly bounded is imposed for simplicity.[12] Finally, Assumption 6.4(ii) specifies the relation between Υ_F and Υ_G when the former is nonlinear. Heuristically, Assumption 6.4(ii) requires the existence of a local perturbation to $\theta_0 \in \Theta_{0n}(P) \cap R$ that relaxes the "active" inequality constraints without a first order effect on the equality restriction.

---

[11]For linear Υ_F, the requirement $\Upsilon_F(\theta + h/\sqrt{n}) = 0$ is equivalent to $\Upsilon_F(h) = 0$ for any $\theta \in R$. In contrast, if Υ_F is nonlinear, then the set of $h \in \mathbf{B}_n$ for which $\Upsilon_F(\theta + h/\sqrt{n}) = 0$ can depend on $\theta \in R$.

[12]Recall for a linear map $\Gamma : \mathbf{B}_n \to \mathbf{F}_n$, its right inverse is a map $\Gamma^- : \mathbf{F}_n \to \mathbf{B}_n$ such that $\Gamma\Gamma^-(h) = h$ for any $h \in \mathbf{B}_n$. The right inverse Γ^- need not be unique if Γ is not bijective, in which case Assumption 6.3(iv) is satisfied as long as it holds for some right inverse of $\nabla \Upsilon_F(\theta) : \mathbf{B}_n \to \mathbf{F}_n$.

Remark 6.2. Certain parameter space restrictions can be incorporated through the map $\Upsilon_G : \mathbf{B} \to \mathbf{G}$. Newey and Powell (2003), for example, address estimation in ill-posed inverse problems by requiring the parameter space Θ to be compact. In our present context, and assuming $X_i \in \mathbf{R}$ for notational simplicity, their smoothness requirements correspond to setting $\mathbf{B}$ to be the Hilbert space with inner product

$$\langle \theta_1, \theta_2 \rangle_{\mathbf{B}} \equiv \sum_{j \leq J} \int \{\nabla_x^j \theta_1(x)\}\{\nabla_x^j \theta_2(x)\}(1 + x^2)^\delta dx \tag{66}$$

for some integer $J > 0$ and $\delta > 1/2$, and letting $\Theta = \{\theta \in \mathbf{B} : \|\theta\|_{\mathbf{B}} \leq B\}$. It is then straightforward to incorporate this restriction through the map $\Upsilon_G : \mathbf{B} \to \mathbf{R}$ by letting $\mathbf{G} = \mathbf{R}$ and defining $\Upsilon_G(\theta) = \|\theta\|_{\mathbf{B}}^2 - B^2$. Moreover, given these definitions, Assumption 6.2 is satisfied with $\nabla \Upsilon_G(\theta)[h] = 2\langle \theta, h \rangle_{\mathbf{B}}$, $K_g = 2$, and $M_g = B$. ∎

Remark 6.3. The consistency result of Lemma 4.1 and Assumption 6.1(ii) together imply that the minimum of $Q_n(\theta)$ over $\Theta_n \cap R$ is attained on the interior of $\Theta_n \cap R$ relative to $\mathbf{B}_n \cap R$. Therefore, if the restriction set R is convex and $Q_n(\theta)$ is convex in θ and well defined on $\mathbf{B}_n \cap R$ (rather than $\Theta_n \cap R$), then it follows that

$$I_n(R) = \inf_{\theta \in \mathbf{B}_n \cap R} \|\frac{1}{\sqrt{n}} \sum_{i=1}^n \rho(X_i, \theta) * q_n^{k_n}(Z_i)\|_{\hat{\Sigma}_n, r} + o_p(a_n) \tag{67}$$

uniformly in $P \in \mathbf{P}_0$ – i.e. the constraint $\theta \in \Theta_n$ can be omitted in computing $I_n(R)$. ∎

### 6.1.2.2 Construction and Intuition

Given the introduced assumptions, we next construct a sample analogue for the local parameter space $V_n(\theta_0, \ell_n)$ of an element $\theta_0 \in \Theta_{0n}(P) \cap R$. To this end, we note that by Assumption 6.1(ii) $V_n(\theta_0, \ell_n)$ is asymptotically determined solely by the equality and inequality constraints. Thus, the construction of a suitable sample analogue for $V_n(\theta_0, \ell_n)$ intuitively only requires estimating the impact on the local parameter space that is induced by the maps Υ_F and Υ_G – a goal we accomplish by examining the impact such constraints have on the local parameter space of a corresponding $\hat{\theta}_n \in \hat{\Theta}_n \cap R$.

In order to account for the role inequality constraints play in determining the local parameter space, we conservatively estimate "binding" sets in analogy to what is done in the partially identified literature.[13] Specifically, for a sequence $\{r_n\}_{n=1}^\infty$ we define

$$G_n(\theta) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \Upsilon_G(\theta + \frac{h}{\sqrt{n}}) \leq (\Upsilon_G(\theta) - K_g r_n \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \mathbf{1_G}) \vee (-r_n \mathbf{1_G})\} , \tag{68}$$

---

[13]See Chernozhukov et al. (2007), Galichon and Henry (2009), Linton et al. (2010), and Andrews and Soares (2010).

Figure 2: Approximating Impact of Inequality Constraints
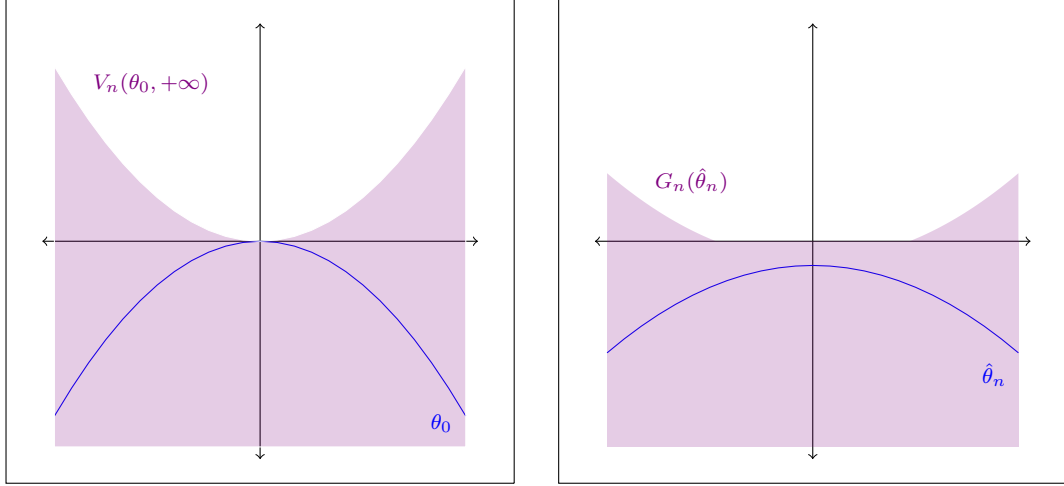
where recall $\mathbf{1_G}$ is the order unit in the AM space $\mathbf{G}$, $g_1 \vee g_2$ represents the (lattice) supremum of any two elements $g_1, g_2 \in \mathbf{G}$, and K_g is as in Assumption 6.2. Figure 2 illustrates the construction in the case in which $X_i \in \mathbf{R}$, $\mathbf{B}$ is the set of continuous functions of X_i, and we aim to test whether $\theta_0(x) \leq 0$ for all $x \in \mathbf{R}$. In this setting, assuming no equality constraints for simplicity, the local parameter space for θ_0 corresponds to the set of perturbations $h/\sqrt{n}$ such that $\theta_0 + h/\sqrt{n}$ remains negative – i.e. any function $h/\sqrt{n} \in \mathbf{B}_n$ in the shaded region of the left panel of Figure 2.[14] For an estimator $\hat{\theta}_n$ of θ_0, the set $G_n(\hat{\theta}_n)$ in turn consists of perturbations $h/\sqrt{n}$ to $\hat{\theta}_n$ such that $\hat{\theta}_n + h/\sqrt{n}$ is not "too close" to the zero function to accommodate the estimation uncertainty in $\hat{\theta}_n$ – i.e. any function $h/\sqrt{n} \in \mathbf{B}_n$ in the shaded region of the right panel of Figure 2. Intuitively, as $\hat{\theta}_n$ converges to θ_0 the set $G_n(\hat{\theta}_n)$ is thus asymptotically contained in, i.e. smaller than, the local parameter space of θ_0 which delivers size control. Unlike Figure 2, however, in settings for which Υ_G is nonlinear we must further account for the curvature of Υ_G which motivates the presence of the term $K_g r_n \|h/\sqrt{n}\|_{\mathbf{B}} \mathbf{1_G}$ in (68).

While employing $G_n(\theta)$ allows us to address the role inequality constraints play on the local parameter space of a $\theta_0 \in \Theta_{0n}(P) \cap R$, we account for equality constraints by examining their impact on the local parameter space of a corresponding $\hat{\theta}_n \in \hat{\Theta}_n \cap R$. Specifically, for a researcher chosen $\ell_n \downarrow 0$ we define $\hat{V}_n(\theta, \ell_n)$ (as utilized in (60)) by

$$\hat{V}_n(\theta, \ell_n) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \frac{h}{\sqrt{n}} \in G_n(\theta), \ \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\} . \quad (69)$$

Thus, in contrast to $V_n(\theta, \ell_n)$ (as in (52)), the set $\hat{V}_n(\theta, \ell_n)$: (i) Replaces the requirement $\Upsilon_G(\theta + h/\sqrt{n}) \leq 0$ by $h/\sqrt{n} \in G_n(\theta)$, (ii) Retains the constraint $\Upsilon_F(\theta + h/\sqrt{n}) = 0$,

---

[14]Mathematically, $\mathbf{B} = \mathbf{G}$, Υ_G is the identity map, $K_g = 0$ since Υ_G is linear, the order unit $\mathbf{1_G}$ is the function with constant value 1, and $\theta_1 \vee \theta_2$ is the pointwise (in x) maximum of the functions θ_1, θ_2.

28

Figure 3: Approximating Impact of Equality Constraints
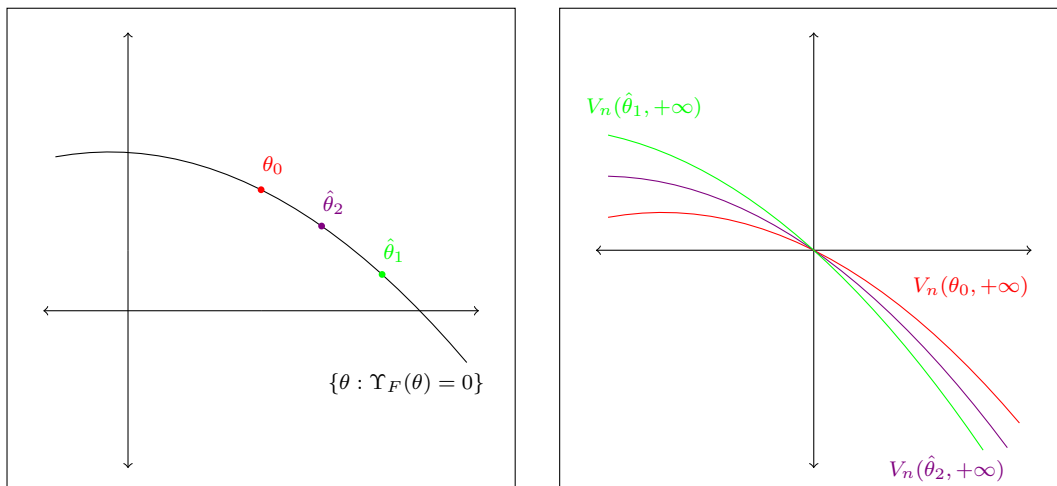
and (iii) Substitutes the $\| \cdot \|_{\mathbf{E}}$ norm constraint by $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$. Figure 3 illustrates how (ii) and (iii) allow us to account for the impact of equality constraints in the special case of no inequality constraints, $\mathbf{B} = \mathbf{R}^2$, and $\mathbf{F} = \mathbf{R}$. In this instance, the constraint $\Upsilon_F(\theta) = 0$ corresponds to a curve on $\mathbf{R}^2$ (left panel), and similarly so does the local parameter space $V_n(\theta, +\infty)$ for any $\theta \in \mathbf{R}^2$ (right panel). Since all curves $V_n(\theta, +\infty)$ pass through zero, we note that all local parameter spaces are "similar" in a neighborhood of the origin. However, for nonlinear Υ_F the size of the neighborhood of the origin in which $V_n(\hat{\theta}_n, +\infty)$ is "close" to $V_n(\theta_0, +\infty)$ crucially depends on both the distance of $\hat{\theta}_n$ to θ_0 and the curvature of Υ_F (compare $V_n(\hat{\theta}_1, +\infty)$ and $V_n(\hat{\theta}_2, +\infty)$ in Figure 3). Heuristically, the set $\hat{V}_n(\hat{\theta}_n, \ell_n)$ thus estimates the role equality constraints play on the local parameter space of θ_0 by restricting attention to the expanding neighborhood of the origin in which the local parameter space of $\hat{\theta}_n$ resembles that of θ_0. In this regard, it is crucial that the neighborhood be defined with respect to the norm under which Υ_F is smooth ($\| \cdot \|_{\mathbf{B}}$) rather than the potentially weaker norms $\| \cdot \|_{\mathbf{E}}$ or $\| \cdot \|_{\mathbf{L}}$.

Remark 6.4. In instances where the constraints $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ and $\Upsilon_G : \mathbf{B} \to \mathbf{G}$ are both linear, controlling the norm $\| \cdot \|_{\mathbf{B}}$ is no longer necessary as the "curvature" of Υ_F and Υ_G is known. As a result, it is possible to instead set $\hat{V}_n(\theta, \ell_n)$ to equal

$$\hat{V}_n(\theta, \ell_n) \equiv \{ \frac{h}{\sqrt{n}} \in \mathbf{B}_n : \frac{h}{\sqrt{n}} \in G_n(\theta), \ \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{E}} \leq \ell_n \} ; \quad (70)$$

i.e. to weaken the constraint $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ in (69) to $\|h/\sqrt{n}\|_{\mathbf{E}} \leq \ell_n$. Controlling the norm $\| \cdot \|_{\mathbf{E}}$, however, may still be necessary in order to ensure that $\hat{\mathbb{D}}_n(\theta)[h]$ is a consistent estimator for $\mathbb{D}_{n,P}(\theta)[h]$ uniformly in $h/\sqrt{n} \in \hat{V}_n(\theta, \ell_n)$. ∎

6.2 Bootstrap Approximation

Having introduced the substitutes for the isonormal process $\mathbb{W}_{n,P}$, the derivative $\mathbb{D}_{n,P}(\theta)$, and the local parameter space $V_n(\theta, \ell_n)$, we next study the bootstrap statistic $\hat{U}_n(R)$ (as in (60)). To this end, we impose the following additional Assumptions:

Assumption 6.5. *(i)* $\sup_{f \in \mathcal{F}_n} \|\hat{\mathbb{W}}_n f q_n^{k_n} - \mathbb{W}_{n,P}^\star f q_n^{k_n}\|_r = o_p(a_n)$ *uniformly in* $P \in \mathbf{P}$ *for* $\mathbb{W}_{n,P}^\star$ *an isonormal Gaussian process that is independent of* $\{V_i\}_{i=1}^n$.

Assumption 6.6. *(i) For any* $\epsilon > 0$, $\tau_n = o(S_n(\epsilon))$; *(ii) The sequences* ℓ_n, τ_n *satisfy* $k_n^{1/r}\sqrt{\log(k_n)} B_n \times \sup_{P \in \mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho} \vee (\nu_n \tau_n)^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$, $K_m \ell_n(\ell_n + \mathcal{R}_n + \nu_n \tau_n) \times S_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$, *and* $\ell_n(\ell_n + \{\mathcal{R}_n + \nu_n \tau_n\} \times S_n(\mathbf{B}, \mathbf{E}))1\{K_f > 0\} = o(a_n n^{-\frac{1}{2}})$; *(iii) The sequence* r_n *satisfies* $\limsup_{n \to \infty} 1\{K_g > 0\}\ell_n/r_n < 1/2$ *and* $(\mathcal{R}_n + \nu_n \tau_n) \times S_n(\mathbf{B}, \mathbf{E}) = o(r_n)$; *(iv) Either* $K_f = K_g = 0$ *or* $(\mathcal{R}_n + \nu_n \tau_n) \times S_n(\mathbf{B}, \mathbf{E}) = o(1)$.

Assumption 6.5 demands that the multiplier bootstrap process $\hat{\mathbb{W}}_n$ be coupled with an isonormal process $\mathbb{W}_{n,P}^\star$ that is independent of the data $\{V_i\}_{i=1}^n$. Intuitively, this condition requires that the multiplier bootstrap, which is automatically consistent for Donsker classes, still be valid in the present non-Donsker setting. Moreover, in accord with our requirements on the empirical process, Assumption 6.5 demands a coupling rate faster than a_n (see Assumption 5.1). We provide sufficient conditions for Assumption 6.5 in Appendix H that may be of independent interest; see Theorem H.1. In turn, Assumption 6.6 collects the necessary bandwidth rate requirements, which we discuss in more detail in Section 6.2.2. Assumption 6.6(i) in particular demands that $\tau_n \downarrow 0$ sufficiently fast to guarantee the one sided Hausdorff convergence of $\hat{\Theta}_n \cap R$. We note this condition is satisfied by setting $\tau_n = 0$, which we recommend unless partial identification is of particular concern. Similarly, Assumption 6.6(ii) requires $\ell_n \downarrow 0$ sufficiently fast to ensure that $\hat{\mathbb{D}}_n(\theta)[h]$ is uniformly consistent over $\theta \in \hat{\Theta}_n \cap R$ and $h/\sqrt{n} \in \hat{V}_n(\theta, \ell_n)$, and that both the intuitions behind Figures 1 and 3 are indeed valid. The latter two requirements on ℓ_n are respectively automatically satisfied by linear models ($K_m = 0$) or linear restrictions ($K_f = 0$). Assumption 6.6(iii) specifies the requirements on r_n, which amount to r_n not decreasing to zero faster than the $\|\cdot\|_{\mathbf{B}}$-rate of convergence. Finally, Assumption 6.6(iv) guarantees the directed Hausdorff consistency of $\hat{\Theta}_n \cap R$ under $\|\cdot\|_{\mathbf{B}}$ in nonlinear problems, thus allowing $\hat{V}_n(\hat{\theta}_n, \ell_n)$ to properly account for the impact of the curvatures of Υ_F and Υ_G on the local parameter space; recall Figure 3.

Given the stated assumptions, the following theorem establishes an unconditional coupling of $\hat{U}_n(R)$ that provides the basis for our subsequent inference results.

Theorem 6.1. *Let Assumptions 2.1(i), 2.2(i), 3.1, 3.2, 3.3, 3.4, 4.1, 5.1, 5.2, 5.3(i), 5.3(iii), 5.4, 6.1, 6.2, 6.3, 6.4, 6.5, and 6.6 hold. Then, uniformly in* $P \in \mathbf{P}_0$

$$\hat{U}_n(R) \geq \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, 2K_b \ell_n)} \|\mathbb{W}_{n,P}^\star \rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P), r} + o_p(a_n) .$$

Theorem 6.1 shows that with unconditional probability tending to one uniformly on $P \in \mathbf{P}_0$ our bootstrap statistic is bounded from below by a random variable that is independent of the data. The significance of this result lies in that the lower bound is equal in distribution to the upper bound for $I_n(R)$ derived in Theorem 5.1(i), and moreover that the rate of both couplings are controlled by a_n. Thus, Theorems 5.1(i) and 6.1 provide the basis for establishing that comparing $I_n(R)$ to the quantiles of $\hat{U}_n(R)$ conditional on the data provides asymptotic size control – a claim we formalize in Section 6.3. Before establishing such a result, however, we first examine whether the conclusion of Theorem 6.1 can be strengthened to hold with equality rather than inequality – i.e. whether an analogue to Theorem 5.1(ii) is available. Unfortunately, as is well understood from the moment inequalities literature, such a uniform coupling is not possible when inequality constraints are present. As we next show, however, Theorem 6.1 can be strengthened to hold with equality under conditions similar to those of Theorem 5.1(ii) in the important case of hypotheses concerning only equality restrictions.

6.2.1 Special Case: No Inequality Constraints

In this section we focus on the special yet important case in which the hypothesis of interest concerns solely equality restrictions. Such a setting encompasses, for example, the construction of confidence regions for functionals of the parameter θ_0 without imposing shape restrictions; see e.g. Horowitz (2007), Gagliardini and Scaillet (2012), and Chen and Pouzo (2015) among others. Formally, we temporarily assume R equals

$$R = \{\theta \in \mathbf{B} : \Upsilon_F(\theta) = 0\} . \tag{71}$$

Under this extra structure the formulation of the test and bootstrap statistics remain largely unchanged, with the exception that the set $\hat{V}_n(\theta, \ell_n)$ simplifies to

$$\hat{V}_n(\theta, \ell_n) = \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\} \tag{72}$$

(compare to (69)). Since these specifications are a special case of our general framework, Theorem 6.1 continues to apply.[15] In fact, as the following Theorem shows, the conclusion of Theorem 6.1 can be strengthened under the additional structure afforded by (71) and conditions analogous to those imposed in Theorem 5.1(ii).

Theorem 6.2. *Let Assumptions 2.1(i), 2.2(i), 3.1, 3.2, 3.3, 3.4, 4.1, 5.1, 5.2, 5.3, 5.4, 6.1, 6.3, 6.5, 6.6(i)-(ii) hold, the set R satisfy (71), and $(\mathcal{R}_n + \nu_n \tau_n) \times \mathcal{S}_n(\mathbf{B}, \mathbf{E}) = o(\ell_n)$.*

---

[15]To see (71) and (72) are a special case of (4) and (69) respectively, let $\mathbf{G} = \mathbf{R}$, $\Upsilon_G(\theta) = -1$ for all $\theta \in \mathbf{B}$, and then note $\{\theta \in \mathbf{B} : \Upsilon_G(\theta) \leq 0\} = \mathbf{B}$ and $G_n(\theta) = \mathbf{B}_n$ for all θ and r_n.

(i) If τ_n satisfies $(k_n^{1/r}\sqrt{\log(k_n)}J_n B_n/\sqrt{n} + \zeta_n) = o(\tau_n)$, then uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R) = \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, 2K_b\ell_n)} \|\mathbb{W}_{n,P}^\star \rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r} + o_p(a_n) \ .$$

(ii) If $\Theta_{0n}(P) \cap R = \{\theta_{0n}(P)\}$ and $\Sigma_n(P) = \{Var_P\{\rho(X_i, \theta_{0n}(P))q_n^{k_n}(Z_i)\}\}^{-\frac{1}{2}}$ for every $P \in \mathbf{P}_0$ and in addition $r = 2$, then for $c_n \equiv \dim\{\mathbf{B}_n \cap \mathcal{N}(\nabla\Upsilon_F(\theta_{0n}(P)))\}$ we have

$$\hat{U}_n(R) = \{\mathcal{X}_{k_n - c_n}^2\}^{\frac{1}{2}} + o_p(a_n) \ ,$$

uniformly in $P \in \mathbf{P}_0$, where $\mathcal{X}_d^2$ is a d-degrees of freedom chi-squared random variable.

Besides assuming a lack of inequality constraints, Theorem 6.2 demands that the rate of convergence $\mathcal{R}_n$ satisfy $\mathcal{R}_n \mathcal{S}_n(\mathbf{B}, \mathbf{E}) = o(\ell_n)$. In view of Assumption 6.6(ii) the latter requirement can be understood as imposing that either Υ_F and $\rho(X_i, \cdot)$ are linear in θ ($K_f = K_m = 0$), or the rate of convergence $\mathcal{R}_n$ is sufficiently fast – conditions that may rule out severely ill-posed nonlinear problems as also demanded in Theorem 5.1(ii). Given these requirements, and provided $\tau_n \downarrow 0$ slowly enough to ensure the Hausdorff convergence of $\hat{\Theta}_n \cap R$, Theorem 6.2(i) strengthens the conclusion of Theorem 6.1 to hold with equality rather than inequality. Moreover, the random variable to which $\hat{U}_n(R)$ is coupled by Theorem 6.2(i) shares the same distribution as the random variable to which $I_n(R)$ is coupled by Theorem 5.1(ii). Thus, Theorems 5.1(ii) and 6.2(i) together provide us with the basis for establishing that the asymptotic size of the proposed test can equal its significance level. In turn, Theorem 6.2(ii) shows that whenever $\Theta_{0n}(P) \cap R$ is a singleton, r and $\Sigma_n(P)$ may be chosen so that the coupled random variable has a pivotal distribution – a result that enables the use of analytical critical values.

Remark 6.5. Under suitable conditions Theorem 6.2(ii) can be generalized to show

$$\hat{U}_n(R) = \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{v \in \mathbb{V}_{n,P}(\theta)} \|\mathbb{W}_{n,P}^\star \rho(\cdot, \theta) * q_n^{k_n} - v\|_{\Sigma_n(P),r} + o_p(a_n) \ , \tag{73}$$

uniformly in $P \in \mathbf{P}_0$ for $\mathbb{V}_{n,P}(\theta)$ a vector subspace of $\mathbf{R}^{k_n}$ possibly depending on P and $\theta \in \Theta_{0n}(P) \cap R$. Theorem 6.2(ii) can then be seen to follow from (73) by setting $\Theta_{0n}(P) \cap R = \{\theta_{0n}(P)\}$ and $r = 2$. However, in general the characterization in (73) is not pivotal and thus does not offer an advantage over Theorem 6.2(i). In this regard, we note that setting $r = 2$ is important to ensure pivotality as projections onto linear subspaces may not admit linear selectors otherwise (Deutsch, 1982). ∎

6.2.2 Discussion: Bandwidths

In constructing our bootstrap approximation we have introduced three bandwidth parameters: τ_n, r_n, and ℓ_n. While these bandwidths are necessary for a successful boot-

strap approximation in the most general setting, there are fortunately a number of applications in which not all three bandwidths are required. With the aim of providing guidance on their selection, we therefore next revisit the role of τ_n, r_n, and ℓ_n and discuss instances in which these bandwidths may be ignored in computation.

The bandwidth τ_n was first introduced in Section 4 in the construction of the set estimator $\hat{\Theta}_n \cap R$. Its principal requirement is that it converge to zero sufficiently fast in order to guarantee the directed Hausdorff consistency of $\hat{\Theta}_n \cap R$. Since directed Hausdorff consistency is equivalent to Hausdorff consistency when $\Theta_{0n}(P) \cap R$ is a singleton, τ_n should therefore always be set to zero in models that are known to be identified; e.g. in Examples 2.1, 2.2, and 2.3. In settings where $\Theta_{0n}(P) \cap R$ is not a singleton, however, τ_n must also decrease to zero sufficiently slowly if we additionally desire $\hat{\Theta}_n \cap R$ to be Hausdorff consistent for $\Theta_{0n}(P) \cap R$. The latter stronger form of consistency can lead to a more powerful test when $\Theta_{0n}(P) \cap R$ is not a singleton, as illustrated by a comparison of Theorems 6.1 and 6.2(i). Nonetheless, even in partially identified settings it may be preferable to set τ_n to zero to simplify implementation – this is the approach implicitly pursued by Bugni et al. (2014), for example, in a related problem.

Allowing for inequality restrictions lead us to introduce the bandwidth r_n in the construction of the sample analogue to the local parameter space. Specifically, the role of r_n is to account for the impact of inequality constraints on the local parameter space and is thus unnecessary in settings where only equality restrictions are present – e.g. in Section 6.2.1. In this regard, the bandwidth r_n may be viewed as analogous to the inequality selection approach pursued in the moment inequalities literature. In particular, its principal requirement is that it decrease to zero sufficiently slowly with overly "aggressive" choices of r_n potentially causing size distortions. As in the moment inequalities literature, however, we may always set $r_n = +\infty$ which corresponds to the "least favorable" local parameter space of an element $\theta \in \Theta_n \cap R$ satisfying $\Upsilon_G(\theta) = 0$ – i.e. all inequalities bind.

The final bandwidth ℓ_n, which to the best of our knowledge does not have a precedent in the literature, plays three distinct roles. First, it ensures that the estimated derivative $\hat{\mathbb{D}}_n(\theta)[h]$ is consistent for $\mathbb{D}_{n,P}(\theta)[h]$ uniformly in $h/\sqrt{n} \in \hat{V}_n(\theta, \ell_n)$. Second, ℓ_n restricts the local parameter space to the regions where a linear approximation to the drift of the Gaussian process is indeed warranted – recall Theorem 5.1 and Figure 1. Third, it accounts for the potential nonlinearity of Υ_F and Υ_G by limiting the estimated local parameter space to areas where it asymptotically resembles the true local parameter space – recall Figures 2 and 3. As a result, the requirements on ℓ_n weaken in applications where the challenges it is meant to address are not present – for instance, when the generalized residual $\rho(X_i, \cdot)$ and/or the constraints Υ_F and Υ_G are linear, as can be seen by evaluating Assumption 6.6(ii)-(iii) when K_m, K_f, or K_g are zero.

In certain applications, it is moreover possible to show the bandwidth ℓ_n is unneces-

sary by arguing that the constraint $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ (as in (69)) is asymptotically slack. The following Lemma, for example, provides sufficient conditions for this occurrence.

Lemma 6.1. *Suppose for some $\epsilon > 0$ it follows that $\|h\|_{\mathbf{E}} \leq \nu_n\|\mathbb{D}_{n,P}(\theta)[h]\|_r$ for all $\theta \in (\Theta_{0n}(P) \cap R)^\epsilon$, $P \in \mathbf{P}_0$, and $h \in \sqrt{n}\{\mathbf{B}_n \cap R - \theta\}$. If in addition*

$$\sup_{\theta \in (\Theta_{0n}(P) \cap R)^\epsilon} \sup_{h \in \sqrt{n}\{\mathbf{B}_n \cap R - \theta\}: \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \geq \ell_n} \frac{\|\hat{\mathbb{D}}_n(\theta)[h] - \mathbb{D}_{n,P}(\theta)[h]\|_r}{\|h\|_{\mathbf{E}}} = o_p(\nu_n^{-1}) \qquad (74)$$

uniformly in $P \in \mathbf{P}_0$, Assumptions 3.1, 3.2(i), 3.3, 3.4, 4.1, 6.5, 6.6(i) hold, and $\mathcal{S}_n(\mathbf{B}, \mathbf{E})\mathcal{R}_n = o(\ell_n)$, then it follows that uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R) = \inf_{\theta \in \hat{\Theta}_n \cap R} \inf_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, +\infty)} \|\hat{\mathbb{W}}_n\rho(\cdot, \theta) * q_n^{k_n} + \hat{\mathbb{D}}_n(\theta)[h]\|_{\hat{\Sigma}_{n,r}} + o_p(a_n) . \qquad (75)$$

Heuristically, Lemma 6.1 establishes the constraint $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ is asymptotically not binding provided $\ell_n \downarrow 0$ sufficiently slowly ($\mathcal{S}_n(\mathbf{B}, \mathbf{E})\mathcal{R}_n = o(\ell_n)$).[16] In order for ℓ_n to simultaneous satisfy such a requirement and Assumption 6.6(ii)-(iii), however, it must be that either the rate of convergence $\mathcal{R}_n$ is adequately fast, or that both the generalized residual and the equality constraint are linear. Thus, while it may be possible to set ℓ_n to be infinite in applications such as Examples 2.1-2.4, the bandwidth ℓ_n can remain necessary in severely ill-posed nonlinear problems; see Appendix F.

6.3 Critical Values

The conclusions of Theorem 5.1 and Theorems 6.1 and 6.2 respectively provide us with an approximation and an estimator for the distribution of our test statistic. In this section, we conclude our main results by formally establishing the properties of a test that rejects the null hypothesis whenever $I_n(R)$ is larger than the appropriate quantile of our bootstrap approximation. To this end, we therefore define

$$c_{n,1-\alpha}(P) \equiv \inf\{u : P(I_n(R) \leq u) \geq 1 - \alpha\} \qquad (76)$$

$$\hat{c}_{n,1-\alpha} \equiv \inf\{u : P(\hat{U}_n(R) \leq u \,|\{V_i\}_{i=1}^n) \geq 1 - \alpha\} ; \qquad (77)$$

i.e. $c_{n,1-\alpha}$ denotes the $1 - \alpha$ quantile of $I_n(R)$, while $\hat{c}_{n,1-\alpha}$ denotes the corresponding quantile of the bootstrap statistic conditional on the sample.

We additionally impose the following two final Assumptions:

Assumption 6.7. *There exists a $\delta > 0$ such that for all $\epsilon > 0$ and all $\tilde{\alpha} \in [\alpha - \delta, \alpha + \delta]$ it follows that $\sup_{P \in \mathbf{P}_0} P(c_{n,1-\tilde{\alpha}}(P) - \epsilon \leq I_n(R) \leq c_{n,1-\tilde{\alpha}}(P) + \epsilon) \leq \varrho_n(\epsilon \wedge 1) + o(1)$, where*

---

[16]Following Remark 6.4, if the constraints Υ_F and Υ_G are linear, then it suffices that $\mathcal{R}_n = o(\ell_n)$.

the concentration parameter ϱ_n is smaller than the coupling rate parameter, namely $\varrho_n \le a_n^{-1}$.

Assumption 6.8. *(i) There exists a $\gamma_z > 0$ and maps $\pi_{n,P,\jmath} : \Theta_n \cap R \to \mathbf{R}^{k_{n,\jmath}}$ such that $\sup_{P \in \mathbf{P}} \sup_{\theta \in \Theta_n \cap R} \{E_P[(E_P[\rho_\jmath(X_i, \theta)|Z_{i,\jmath}] - q_{n,\jmath}^{k_{n,\jmath}}(Z_{i,\jmath})'\pi_{n,P,\jmath}(\theta))^2]\}^{\frac{1}{2}} = O(k_n^{-\gamma_z})$ for all $1 \le \jmath \le \mathcal{J}$; (ii) The eigenvalues of $E_P[q_{n,\jmath}^{k_{n,\jmath}}(Z_{i,\jmath})q_{n,\jmath}^{k_{n,\jmath}}(Z_{i,\jmath})']$ are bounded away from zero uniformly in $1 \le \jmath \le \mathcal{J}$, $n \in \mathbb{N}$, and $P \in \mathbf{P}$.*

It is well known that uniform consistent estimation of an approximating distribution is not sufficient for establishing asymptotic size control; see, e.g. Romano and Shaikh (2012). Intuitively, in order to get good size control, when critical values are estimated with noise, the approximate distribution must be suitably continuous at the quantile of interest uniformly in $P \in \mathbf{P}_0$. Assumption 6.7 imposes precisely this requirement, allowing the modulus of continuity, captured here by the concentration parameter ϱ_n, to deteriorate with the sample size provided that $\varrho_n \le a_n^{-1}$ – that is the loss of continuity must occurs at a rate slower than the coupling rate $o(a_n)$ of Theorems 5.1, 6.1, and 6.2. We refer the reader to Chernozhukov et al. (2013, 2014) for further discussion and motivation of conditions of this type, called *anti-concentration* conditions there.[17] Note that in some typical cases, the rate of concentration is $\varrho_n = 1$ with $r = 2$ and $\varrho_n \sim \sqrt{\log k_n}$ with $r = \infty$, which means that the condition on the coupling rate $o(a_n)$ arising from imposing Assumption 6.7 are mild in these cases and are expected to be mild in others. In turn, Assumption 6.8 imposes sufficient conditions for studying the power of the proposed test. In particular, Assumption 6.8(i) demands that the transformations $\{q_{k,n,\jmath}\}_{k=1}^{k_{n,\jmath}}$ be able to approximate conditional moments given $Z_{i,\jmath}$ and thus be capable of detecting violations of the null hypothesis. Finally, Assumption 6.8(ii) enables us to characterize the set of local distributions against which the test is consistent.

Theorem 6.3 exploits our previous results and the introduced assumptions to characterize the asymptotic size and power properties of our test.

Theorem 6.3. *Let the conditions imposed in Theorem 5.1(i) and Theorem 6.1 hold. (i) If in addition Assumption 6.7 is satisfied, then we can conclude that*

$$\limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(I_n(R) > \hat{c}_{n,1-\alpha}) \le \alpha .$$

(ii) If Assumption 6.7 and the conditions of Theorem 6.2(i) hold and $\mathcal{R}_n = o(\ell_n)$, then

$$\limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} |P(I_n(R) > \hat{c}_{n,1-\alpha}) - \alpha| = 0 .$$

(iii) Let $\mathbf{P}_{1,n}(M) \equiv \{P \in \mathbf{P} : \inf_{\theta \in \Theta \cap R}\{\sum_{\jmath=1}^{\mathcal{J}} \|E_P[\rho_\jmath(X_i, \theta)|Z_{i,\jmath}]\|_{L_P^2}\} \ge M\gamma_n\}$ for

---

[17]Alternatively, Assumption 6.7 can be dispensed by adding a fixed constant $\eta > 0$ to the critical value, i.e. using $\hat{c}_{n,1-\alpha} + \eta$ as the critical value; this approach is not satisfactory, since η is arbitrary and there is no adequate theory for setting this.

35

$\gamma_n \equiv \sqrt{k_n \log(k_n)} B_n J_n / \sqrt{n} + k_n^{-\gamma_z}$. *If in addition Assumption 6.8 holds, then*

$$\liminf_{M \uparrow \infty} \liminf_{n \to \infty} \inf_{P \in \mathbf{P}_{1,n}(M)} P(I_n(R) > \hat{c}_{n,1-\alpha}) = 1 \ .$$

The first claim of Theorem 6.3 exploits Theorems 5.1(i) and 6.1 to show that the proposed test delivers asymptotic size control. In turn, Theorem 6.3(ii) leverages Theorems 5.1(ii) and 6.2(ii) to conclude that the asymptotic size of the proposed test can equal its significance level when no inequality constraints are present and either the model is linear or the rate of convergence $\mathcal{R}_n$ is sufficiently fast. Under the latter structure it is also possible to obtain the same conclusion employing analytical critical by exploiting Theorem 6.2(ii); see Remark 6.6. Finally, Theorem 6.3(iii) characterizes local sequences $P_n \in \mathbf{P} \setminus \mathbf{P}_0$ for which our test has nontrivial local power.

Remark 6.6. When $\Theta_{0n}(P) \cap R$ is a singleton $\{\theta_{0n}(P)\}$ for all $P \in \mathbf{P}_0$, Theorem 6.2(ii) provides conditions under which the bootstrap statistic is in fact coupled to the square root of a chi-squared random variable. For $\chi^2_{1-\alpha}(d)$ the $1 - \alpha$ quantile of a chi-squared random variable with d degrees of freedom it is then possible to show that

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} |P(I_n^2(R) > \chi^2_{1-\alpha}(k_n - c_n)) - \alpha| = 0 \tag{78}$$

where recall $c_n \equiv \dim\{\mathbf{B}_n \cap \mathcal{N}(\nabla \Upsilon_F(\theta_{0n}(P)))\}$. As in Theorem 6.3(ii), however, we emphasize such a conclusion does not apply to nonlinear problems in which the rate of convergence is not sufficiently fast, or to hypotheses involving inequality restrictions. ∎

Remark 6.7. In the conditional moment inequalities literature, certain test statistics have been shown to converge in probability to zero when all inequalities are "slack" (Linton et al., 2010). It is worth noting that an analogous problem, which could potentially conflict with Assumption 6.7, is not automatically present in our setting. In particular, we observe that since $V_n(\theta_0, \ell_n) \subseteq \mathbf{B}_n$, Lemma 5.1 implies

$$I_n(R) \geq \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{h \in \mathbf{B}_n} \|\mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \sqrt{n} P \rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n}\|_{\Sigma_n(P),r} + o_p(a_n) \tag{79}$$

and that under regularity conditions the right hand side of (79) is non-degenerate when $\dim\{\mathbf{B}_n\} < k_n$; see also our simulations of a test of monotonicity in Section 7. ∎

7 Simulation Evidence

We examine the finite sample performance of the proposed test through a simulation study based on the nonparametric instrumental variable model

$$Y_i = \theta_0(X_i) + \epsilon_i \ , \tag{80}$$

where θ_0 is an unknown function and $E_P[\epsilon_i|Z_i] = 0$ for an observable instrument Z_i. In order to illustrate the different applications of our framework, we study both a test of a shape restriction and a test on a functional of θ_0 that imposes a shape restriction to sharpen inference. Specifically, we examine the performance of a test of whether θ_0 is monotone, and of a test that imposes monotonicity to conduct inference on the value of θ_0 at a point. These applications are closely related to Examples 2.1 and 2.2 and we refer the reader to their discussion in Appendix F for implementation details.

7.1 Design

We consider a design in which random variables $(X_i^*, Z_i^*, \epsilon_i) \in \mathbf{R}^3$ follow the distribution

$$\begin{pmatrix} X_i^* \\ Z_i^* \\ \epsilon_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & 0.5 & 0.3 \\ 0.5 & 1 & 0 \\ 0.3 & 0 & 1 \end{bmatrix} \right), \tag{81}$$

and $(X_i, Z_i) \in \mathbf{R}^2$ are generated according to $X_i = \Phi(X_i^*)$ and $Z_i = \Phi(Z_i^*)$ for Φ the c.d.f. of a standard normal random variable. The dependent variable Y_i is in turn created according to (80) with the structural function θ_0 following the specification

$$\theta_0(x) = \sigma\{1 - 2\Phi(\frac{x - 0.5}{\sigma})\} \tag{82}$$

for different choices of σ. For all positive values of σ, the function θ_0 is monotonically decreasing and satisfies $\theta_0(0.5) = 0$. Moreover, we also note $\theta_0(x) \approx 0$ for values of σ close to zero and $\theta_0(x) \approx \phi(0)(1 - 2x)$ for values of σ close to one and ϕ the derivative of Φ.[18] Thus, by varying σ in (82) we can examine the performance of our tests under different "strengths" of monotonicity. All the reported results are based on five thousand replications of samples $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ consisting of five hundred observations each.

As a sieve we employ b-Splines $\{p_{j,n}\}_{j=1}^{j_n}$ of order three with continuous derivatives and either one or no knots, which results in a dimension j_n equal to four or three respectively. Since b-Splines of order three have piecewise linear derivatives, monotonicity constraints are simple to implement as we only require to check the value of the derivative at $j_n - 1$ points. The instrument transformations $\{q_{k,n}\}_{k=1}^{k_n}$ are also chosen to be b-Splines of order three with continuous derivatives and either three, five, or ten knots placed at the population quantiles. These parameter choices correspond to a total number of moments k_n equal to six, eight, or thirteen. The test statistic $I_n(R)$ is then implemented with $r = 2$, and $\hat{\Sigma}_n$ equal to the optimal GMM weighting matrix computed with a two stage least squares estimator constrained to satisfy the null hypothesis as

---

[18]Formally, $\theta_0(x)$ converges to the 0 and $\phi(0)(1 - 2x)$ as σ approaches 0 and ∞ respectively. We find numerically, however, that $\theta_0(x)$ is very close to $\phi(0)(1 - 2x)$ for all $x \in [0, 1]$ for σ as small as one.

a first stage. Under these specifications, calculating $I_n(R)$ simply requires solving two quadratic programming problems with linear constraints.

Obtaining critical values further requires us to compute the quantiles of $\hat{U}_n(R)$ conditional on the data, which we simulate employing two hundred bootstrap samples in each replication. For the bandwidth choices, we set τ_n to zero which, despite being potentially conservative under partial identification, is both sufficient for size control and computationally simpler to implement. In turn, we explore data driven choices for the parameters r_n and ℓ_n. Specifically, setting $p_n^{j_n}(x) \equiv (p_{1,n}(x), \ldots, p_{j,n}(x))'$ and letting $\mathbb{Z}_r \sim N(0, (\hat{\Delta}_n \hat{\Sigma}_n \hat{\Delta}_n')^{-1})$ with $\hat{\Delta}_n = \frac{1}{n} \sum_i p_n^{j_n}(X_i) q_n^{k_n}(Z_i)'$, we select r_n by solving

$$q_r = P(\|p_n^{j_n \prime} \mathbb{Z}_r\|_{1,\infty} \leq r_n) \tag{83}$$

for different choices of $q_r \in \{0.05, 0.95\}$. Heuristically, r_n is thus the q_r^{th} quantile of an estimate of the asymptotic distribution of the $\|\cdot\|_{1,\infty}$ norm of the unconstrained minimizer of Q_n under fixed values for j_n and k_n. We therefore interpret $q_r = 0.05$ as an "aggressive" choice for r_n and $q_r = 0.95$ as a "conservative" one. Finally, for $\mathbb{Z}_l$ a $k_n \times j_n$ random matrix drawn from an estimate of the asymptotic distribution of $\hat{\Delta}_n$ under fixed j_n and k_n asymptotics, we select ℓ_n by solving[19]

$$q_\ell = P(\sup_{\beta \in \mathbf{R}^{j_n} : \|\beta\|_\infty \leq \ell_n} \|\mathbb{Z}_\ell \beta\|_{\hat{\Sigma}_n, 2} \leq 1) \tag{84}$$

for different values of $q_\ell \in \{0.05, 0.95\}$; see Remark 7.1 for the rationale behinds this choice. In concordance to the choice of r_n, here $q_\ell = 0.05$ also corresponds to the "aggressive" choice of ℓ_n and $q_\ell = 0.95$ to the "conservative" one.

Remark 7.1. In the hypothesis testing problems of this section, the norm constraint $\|p_n^{j_n \prime} \beta / \sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ in the definition of $\hat{V}_n(\theta, \ell_n)$ can be replaced by $\|\beta / \sqrt{n}\|_2 \leq \ell_n$; see Remark 6.4 and the discussion of Example 2.2 in Appendix F. The latter constraint, however, is in turn implied by $\|\beta / \sqrt{n}\|_\infty \leq \ell_n / \sqrt{j_n}$, which is computationally simpler to implement as it is equivalent to $2j_n$ linear constraints on β. Moreover, when Υ_F and Υ_G are linear, the sole role of ℓ_n is to ensure $\hat{\mathbb{D}}_n(\theta)[h]$ is uniformly consistent for $\mathbb{D}_{n,P}(\theta)[h]$ – recall Section 6.2.2. In the present context, such convergence is implied by

$$\sup_{\beta \in \mathbf{R}^{j_n} : \|\frac{\beta}{\sqrt{n}}\|_\infty \leq \ell_n} \|\frac{1}{n} \sum_{i=1}^n q_n^{k_n}(Z_i) p_n^{j_n}(Z_i)' \beta - E_P[q_n^{k_n}(Z_i) p_n^{j_n}(Z_i)' \beta]\|_{\hat{\Sigma}_n, 2} = o_p(a_n) , \tag{85}$$

which motivates using (84) to study the sensitivity of our tests to the choice of ℓ_n. ∎

Table 1: Monotonicity Test - Empirical Size

			$\sigma = 1$								
			$k_n = 6$			$k_n = 8$			$k_n = 13$		
j_n	q_ℓ	q_r	10%	5%	1%	10%	5%	1%	10%	5%	1%
3	5%	5%	0.084	0.044	0.010	0.087	0.042	0.009	0.095	0.046	0.009
3	5%	95%	0.060	0.030	0.006	0.068	0.033	0.007	0.079	0.037	0.008
3	95%	5%	0.083	0.043	0.010	0.086	0.042	0.009	0.095	0.046	0.009
3	95%	95%	0.060	0.030	0.006	0.068	0.033	0.007	0.079	0.037	0.008
4	5%	5%	0.048	0.023	0.004	0.055	0.028	0.005	0.070	0.032	0.006
4	5%	95%	0.047	0.023	0.004	0.055	0.027	0.005	0.069	0.031	0.006
4	95%	5%	0.047	0.023	0.004	0.055	0.027	0.005	0.070	0.032	0.006
4	95%	95%	0.047	0.023	0.004	0.055	0.027	0.005	0.069	0.031	0.006

			$\sigma = 0.1$								
			$k_n = 6$			$k_n = 8$			$k_n = 13$		
j_n	q_ℓ	q_r	10%	5%	1%	10%	5%	1%	10%	5%	1%
3	5%	5%	0.081	0.041	0.010	0.087	0.041	0.009	0.095	0.043	0.010
3	5%	95%	0.075	0.036	0.008	0.081	0.038	0.009	0.090	0.042	0.010
3	95%	5%	0.081	0.041	0.010	0.087	0.041	0.009	0.095	0.043	0.010
3	95%	95%	0.075	0.036	0.008	0.081	0.038	0.009	0.090	0.042	0.010
4	5%	5%	0.068	0.034	0.007	0.076	0.037	0.009	0.086	0.040	0.009
4	5%	95%	0.068	0.034	0.007	0.076	0.037	0.009	0.086	0.039	0.009
4	95%	5%	0.067	0.033	0.007	0.076	0.037	0.009	0.086	0.040	0.009
4	95%	95%	0.067	0.033	0.007	0.075	0.037	0.009	0.086	0.039	0.009

|			$\sigma = 0.01$								
			$k_n = 6$			$k_n = 8$			$k_n = 13$		
j_n	q_ℓ	q_r	10%	5%	1%	10%	5%	1%	10%	5%	1%
3	5%	5%	0.102	0.050	0.012	0.102	0.052	0.012	0.109	0.053	0.011
3	5%	95%	0.100	0.049	0.012	0.100	0.050	0.012	0.107	0.052	0.011
3	95%	5%	0.102	0.050	0.012	0.102	0.052	0.012	0.109	0.053	0.011
3	95%	95%	0.100	0.049	0.012	0.100	0.050	0.012	0.107	0.052	0.011
4	5%	5%	0.099	0.049	0.011	0.100	0.049	0.013	0.103	0.052	0.011
4	5%	95%	0.099	0.049	0.011	0.100	0.049	0.013	0.103	0.052	0.011
4	95%	5%	0.099	0.048	0.011	0.100	0.049	0.013	0.103	0.052	0.011
4	95%	95%	0.098	0.048	0.011	0.100	0.049	0.013	0.103	0.052	0.011

7.2 Results

We begin by first examining the performance of our inferential framework when applied to test whether the structural function θ_0 is monotonically decreasing. Table 1 reports the empirical size control of the resulting test under the different parameter choices. The test delivers good size control across specifications, though as expected can be undersized when the θ_0 is "strongly" monotonic ($\sigma = 1$). The empirical rejection rates are insensitive to the value of q_ℓ, which suggests the asymptotics of Lemma 6.1 are applicable and the bandwidth ℓ_n is not needed to ensure size control. In contrast, the empirical rejection rates are more responsive to the value of q_r, though the "aggressive" choice of $q_r = 0.05$ is still able to deliver adequate size control even in the least favorable

---

[19]The ℓ_n solving (84) is simply the reciprocal of the q_ℓ^{th} quantile of $\sup_{\beta \in \mathbf{R}^{j_n}:\|\beta\|_\infty \leq 1} \|\mathbb{Z}_\ell \beta\|_{\hat{\Sigma}_n,2}$, which we approximate using a sample of two hundred draws of $\mathbb{Z}_\ell$.

Figure 4: Monotonicity Test - Empirical Power
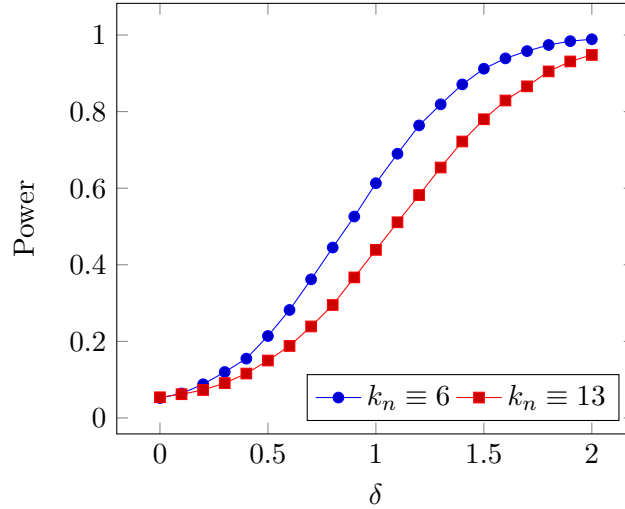
configuration ($\sigma = 0.01$). Finally, we note increasing the dimension of the sieve (j_n) can lead the test to be undersized, while increasing the number of moments (k_n) brings the empirical size of the test closer to its nominal level.

In order to study the power of the test that θ_0 is monotonically decreasing, we consider deviations from the constant zero function ($\sigma = 0$). Specifically, we examine the rejection probabilities of the test when the data is generated according to

$$Y_i = \delta X_i + \epsilon_i \tag{86}$$

for different positive values of δ. Figure 4 depicts the power function of a 5% nominal level test implemented with $j_n = 3$, $q_l = q_r = 0.05$, and different number of moments k_n. For the violation of decreasing monotonicity considered in (86), the test with fewer moments appears to be more powerful indicating the first few moments are the ones detecting the deviation from monotonicity. More generally, however, we expect the power ranking for the choices of k_n to depend on the alternative under consideration.

Table 2: Level Test - Empirical Size

σ	j_n	$k_n = 6$			$k_n = 8$			$k_n = 13$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
1	3	0.106	0.051	0.010	0.105	0.054	0.012	0.107	0.056	0.012
1	4	0.072	0.034	0.006	0.074	0.036	0.008	0.078	0.038	0.008
0.1	3	0.106	0.052	0.010	0.106	0.055	0.013	0.107	0.055	0.011
0.1	4	0.073	0.034	0.006	0.075	0.035	0.008	0.076	0.038	0.008
0.01	3	0.106	0.052	0.010	0.105	0.054	0.012	0.107	0.056	0.011
0.01	4	0.073	0.034	0.006	0.074	0.036	0.008	0.077	0.038	0.008

Table 3: Level Test Imposing Monotonicity - Empirical Size

			$\sigma = 1$								
			$k_n = 6$			$k_n = 8$			$k_n = 13$		
j_n	q_ℓ	q_r	10%	5%	1%	10%	5%	1%	10%	5%	1%
3	5%	5%	0.077	0.037	0.008	0.082	0.041	0.007	0.092	0.043	0.008
3	5%	95%	0.053	0.026	0.005	0.061	0.030	0.005	0.075	0.033	0.008
3	95%	5%	0.077	0.037	0.008	0.082	0.041	0.007	0.092	0.043	0.008
3	95%	95%	0.053	0.026	0.005	0.061	0.030	0.005	0.075	0.033	0.008
4	5%	5%	0.055	0.026	0.006	0.063	0.029	0.006	0.073	0.033	0.008
4	5%	95%	0.055	0.026	0.006	0.063	0.029	0.006	0.073	0.033	0.008
4	95%	5%	0.055	0.026	0.006	0.063	0.029	0.006	0.073	0.033	0.008
4	95%	95%	0.055	0.026	0.006	0.063	0.029	0.006	0.073	0.033	0.008
			$\sigma = 0.1$								
			$k_n = 6$			$k_n = 8$			$k_n = 13$		
j_n	q_ℓ	q_r	10%	5%	1%	10%	5%	1%	10%	5%	1%
3	5%	5%	0.078	0.038	0.008	0.084	0.042	0.009	0.090	0.044	0.009
3	5%	95%	0.072	0.035	0.007	0.079	0.037	0.008	0.085	0.040	0.009
3	95%	5%	0.078	0.038	0.008	0.084	0.042	0.008	0.090	0.044	0.009
3	95%	95%	0.072	0.034	0.007	0.079	0.037	0.008	0.085	0.040	0.009
4	5%	5%	0.068	0.034	0.008	0.075	0.037	0.008	0.084	0.039	0.009
4	5%	95%	0.068	0.034	0.008	0.075	0.037	0.008	0.084	0.039	0.009
4	95%	5%	0.067	0.033	0.008	0.075	0.037	0.008	0.084	0.039	0.009
4	95%	95%	0.067	0.033	0.008	0.075	0.037	0.008	0.084	0.039	0.009
			$\sigma = 0.01$								
			$k_n = 6$			$k_n = 8$			$k_n = 13$		
j_n	q_ℓ	q_r	10%	5%	1%	10%	5%	1%	10%	5%	1%
3	5%	5%	0.102	0.053	0.012	0.106	0.054	0.013	0.109	0.054	0.011
3	5%	95%	0.100	0.051	0.011	0.104	0.053	0.012	0.107	0.053	0.011
3	95%	5%	0.102	0.053	0.012	0.106	0.054	0.013	0.109	0.054	0.011
3	95%	95%	0.100	0.051	0.011	0.104	0.053	0.012	0.107	0.053	0.011
4	5%	5%	0.101	0.051	0.011	0.103	0.049	0.012	0.106	0.052	0.011
4	5%	95%	0.101	0.051	0.011	0.103	0.049	0.012	0.106	0.052	0.011
4	95%	5%	0.101	0.051	0.011	0.102	0.049	0.012	0.106	0.052	0.011
4	95%	95%	0.101	0.051	0.011	0.102	0.049	0.012	0.106	0.052	0.011

Next, we apply our inferential framework to conduct inference on the value of the structural function θ_0 at the point $x = 0.5$ – recall that for all values of σ in (82), $\theta_0(0.5) = 0$. First, we examine the size control of a test that does not impose monotonicity, so that the set R consists of all functions θ satisfying $\theta(0.5) = 0$. For such a hypothesis, r_n is unnecessary and we therefore examine the quality of the chi-squared approximation of Theorem 6.2(ii) and Remark 6.6. The empirical size of the corresponding test are summarized in Table 2, which shows adequate size control and an insensitivity to the "degree" of monotonicity (σ) of the structural function.

In addition, we also examine the size of a test that conducts inference on the level of θ_0 at the point $x = 0.5$ while imposing the monotonicity of θ_0 – i.e. the set R consists of monotonically decreasing functions satisfying $\theta(0.5) = 0$. Table 3 reports the

Figure 5: Level Test - Empirical Power
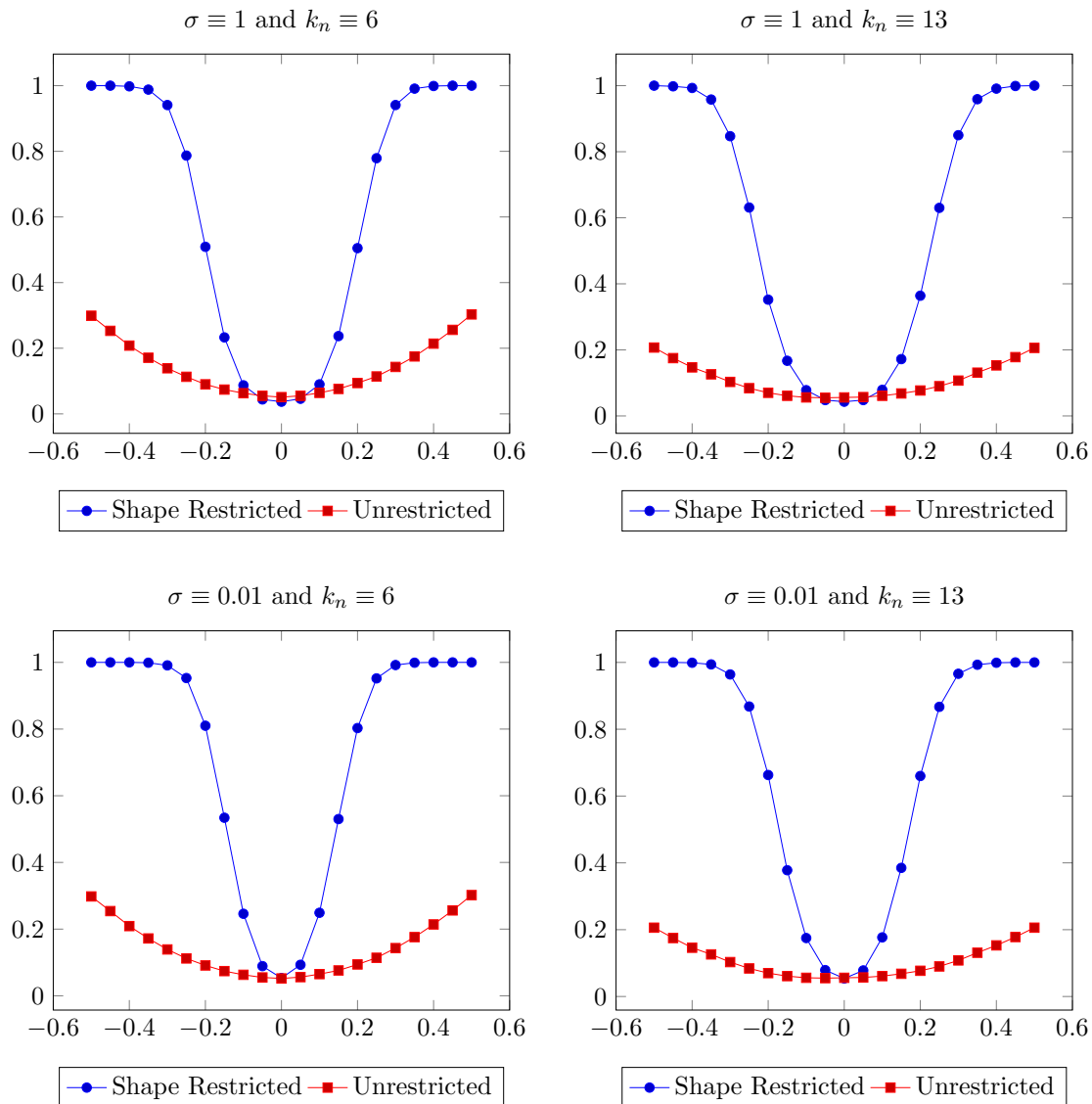
empirical size of the corresponding test under different parameter values. The results are qualitatively similar to those corresponding to the test of monotonicity summarized in Table 1. Namely, (i) All parameter choices yield adequate size control; (ii) The test can be undersized in the strongly monotonic specifications ($\sigma = 1$); (iii) Empirical rejection rates are insensitive to the bandwidth ℓ_n; and (iv) Both the "conservative" ($q_r = 0.95$) and "aggressive" ($q_r = 0.05$) choices for r_n yield good size control.

Finally, we compare the power of the test that imposes monotonicity with the power of the test that does not. To this end, we consider data generated according to

$$Y_i = \sigma\{1 - 2\Phi(\frac{x - 0.5}{\sigma})\} + \delta + \epsilon_i \,, \tag{87}$$

so that the structural function is still monotonically decreasing but satisfies $\theta_0(0.5) = \delta$ instead of the tested null hypothesis $\theta_0(0.5) = 0$. We implement the tests with $j_n = 3$, and $q_\ell = q_r = 0.05$ since such specification yields empirical size closest to the nominal level of the test (rather than being undersized). The corresponding power curves are depicted for different values of instruments (k_n) and degree of monotonicity (σ) in Figure 5. The power gains of imposing monotonicity are substantial, even when the true structural function is "strongly" monotonic $(\sigma = 1)$. This evidence is consistent with our earlier claims of our framework being able to capture the strong finite-sample gains from imposing monotonicity. At the nearly constant specification $(\sigma = 0.01)$, the power of the test that imposes monotonicity improves while the power of the test that does not remains constant. As a result, the power differences between both tests are further accentuated at $\sigma = 0.01$.

8 Conclusion

In this paper, we have developed an inferential framework for testing "equality" and "inequality" constraints in models defined by conditional moment restrictions. Notably, the obtained results are sufficiently general to enable us to test for shape restrictions or to impose them when conducting inference. While our results focus on conditional moment restriction models, the insights developed for accounting for nonlinear local parameter spaces are more generally applicable to other settings. As such, we believe our theoretical analysis will be useful in the study of nonparametric constraints in complementary contexts such as likelihood based models.

Appendix A - Notation and AM Spaces

For ease of reference, in this Appendix we collect the notation employed throughout the paper, briefly review AM spaces and their basic properties, and provide a description of the organization of the remaining Appendices. We begin with Table 4 below, which contains the norms and mathematical notation used. In turn, Table 5 presents the sequences utilized in the text as well as the location of their introduction.

Table 4: List of norms, spaces, and notation.

$a \lesssim b$	$a \leq Mb$ for some constant M that is universal in the proof.
$\|\cdot\|_{L_P^q}$	For a measure P and function f, $\|f\|_{L_P^q}^q \equiv \int |f|^q dP$.
$\|\cdot\|_r$	For a vector $a = (a^{(1)}, \ldots, a^{(k)})'$, $\|a\|_r^r = \sum_{i=1}^k |a^{(i)}|^r$.
$\|\cdot\|_{o,r}$	For a $k \times k$ matrix A, $\|A\|_{o,r} \equiv \sup_{\|a\|_r=1} \|Aa\|_r$.
$\|\cdot\|_{\mathcal{G}}$	For a set $\mathcal{G}$ and a map $f : \mathcal{G} \to \mathbf{R}$, the norm $\|f\|_{\mathcal{G}} \equiv \sup_{g \in \mathcal{G}} |f(g)|$.
$\vec{d}_H(\cdot, \cdot, \|\cdot\|)$	For sets A, B, $\vec{d}_H(A, B, \|\cdot\|) \equiv \sup_{a \in A} \inf_{b \in B} \|a - b\|$.
$d_H(\cdot, \cdot, \|\cdot\|)$	For sets A, B, $d_H(A, B, \|\cdot\|) \equiv \max\{\vec{d}_H(A, B, \|\cdot\|), \vec{d}_H(B, A, \|\cdot\|)\}$.
$N_{[]}(\epsilon, \mathcal{G}, \|\cdot\|)$	The ϵ bracketing numbers for a class $\mathcal{G}$ under $\|\cdot\|$.
$J_{[]}(\delta, \mathcal{G}, \|\cdot\|)$	The entropy integral $J_{[]}(\delta, \mathcal{G}, \|\cdot\|) \equiv \int_0^\delta \{1 + \log N_{[]}(\epsilon, \mathcal{G}, \|\cdot\|)\}^{1/2} d\epsilon$.
$\mathcal{S}_n(A, B)$	The modulus of continuity of norms on normed spaces A and B.

Table 5: List of sequences.

a_n	A bound on the rate of convergence of the coupling results.
B_n	A bound on the sup norm of $\{q_{k,n,j}\}$. Introduced in Assumption 3.2(i).
$S_n(\epsilon)$	How "well separated" the minimum is. Introduced in Assumption 4.1(ii).
J_n	A bound on the entropy of $\mathcal{F}_n$. Introduced in Assumption 3.3(iii).
k_n	The number of moments employed.
$\mathcal{R}_n$	Convergence rate of $\hat{\Theta}_n \cap R$ with $\tau_n = o(n^{-\frac{1}{2}})$. Introduced in Theorem 4.1.
τ_n	A sequence defining $\hat{\Theta}_n \cap R$. Introduced in equation (36).
ν_n	Controls the strength of identification. Introduced in Assumption 4.2.
ζ_n	A bound on the population minimum. Introduced in Assumption 4.1(i).

Since AM spaces are not often employed in econometrics, we next provide a brief introduction that highlights the properties we need for our analysis. The definitions and results presented here can be found in Chapters 8 and 9 of Aliprantis and Border (2006), and we refer the reader to said reference for a more detailed exposition. Before proceeding, we first recall the definitions of a partially ordered set and a lattice:

Definition A.1. A *partially ordered set* $(\mathbf{G}, \geq)$ is a set $\mathbf{G}$ with a partial order relationship $\geq$ defined on it – i.e. $\geq$ is a transitive ($x \geq y$ and $y \geq z$ implies $x \geq z$), reflexive ($x \geq x$), and antisymmetric ($x \geq y$ implies the negation of $y \geq x$) relation. ∎

Definition A.2. A *lattice* is a partially ordered set $(\mathbf{G}, \geq)$ such that any pair $x, y \in \mathbf{G}$ has a least upper bound (denoted $x \vee y$) and a greatest lower bound (denoted $x \wedge y$). ∎

Whenever $\mathbf{G}$ is both a vector space and a lattice, it is possible to define objects that depend on both the vector space and lattice operations. In particular, for $x \in \mathbf{G}$ the positive part x^+, the negative part x^-, and the absolute value $|x|$ are defined by

$$x^+ \equiv x \vee 0 \qquad\qquad x^- \equiv (-x) \vee 0 \qquad\qquad |x| \equiv x \vee (-x) . \qquad \text{(A.1)}$$

In addition, it is natural to demand that the order relation $\geq$ interact with the algebraic operations of the vector space in a manner analogous to that of $\mathbf{R}$ – i.e. to expect

$$x \geq y \text{ implies } x + z \geq y + z \text{ for each } z \in \mathbf{G} \qquad\qquad \text{(A.2)}$$

$$x \geq y \text{ implies } \alpha x \geq \alpha y \text{ for each } 0 \leq \alpha \in \mathbf{R} . \qquad\qquad \text{(A.3)}$$

A complete normed vector space that shares these familiar properties of $\mathbf{R}$ under a given order relation $\geq$ is referred to as a *Banach lattice*. Formally, we define:

Definition A.3. A Banach space $\mathbf{G}$ with norm $\| \cdot \|_{\mathbf{G}}$ is a *Banach lattice* if (i) $\mathbf{G}$ is a lattice under $\geq$, (ii) $\|x\|_{\mathbf{G}} \leq \|y\|_{\mathbf{G}}$ when $|x| \leq |y|$, (iii) (A.2) and (A.3) hold. ∎

An AM space, is then simply a Banach lattice in which the norm $\| \cdot \|_{\mathbf{G}}$ is such that the maximum of the norms of two positive elements is equal to the norm of the maximums of the two elements – e.g. L_P^∞ under pointwise ordering. The norm having such property is called the M-norm.

Definition A.4. A Banach lattice $\mathbf{G}$ is called an AM space if for any elements $0 \leq x, y \in \mathbf{G}$ it follows that $\|x \vee y\|_{\mathbf{G}} = \max\{\|x\|_{\mathbf{G}}, \|y\|_{\mathbf{G}}\}$. ∎

In certain Banach lattices there may exist an element $\mathbf{1_G} > 0$ called an *order unit* such that for any $x \in \mathbf{G}$ there exists a $0 < \lambda \in \mathbf{R}$ for which $|x| \leq \lambda \mathbf{1_G}$ – for example, in $\mathbf{R}^d$ the vector $(1, \ldots, 1)'$ is an order unit. The order unit $\mathbf{1_G}$ can be used to define

$$\|x\|_\infty \equiv \{\inf \lambda > 0 : |x| \leq \lambda \mathbf{1_G}\} , \qquad\qquad \text{(A.4)}$$

which is easy to see constitutes a norm on the original Banach lattice $\mathbf{G}$. In principle, the norm $\| \cdot \|_\infty$ need not be related to the original norm $\| \cdot \|_{\mathbf{G}}$ with which $\mathbf{G}$ was endowed. Fortunately, however, if $\mathbf{G}$ is an AM space, then the original norm $\| \cdot \|_{\mathbf{G}}$ and the norm $\| \cdot \|_\infty$ are equivalent in the sense that they generate the same topology (Aliprantis and Border, 2006, page 358). Hence, without loss of generality we refer to $\mathbf{G}$ as an *AM space with unit* $\mathbf{1_G}$ if these conditions are satisfied: (i) $\mathbf{G}$ is an AM space, (ii) $\mathbf{1_G}$ is an order unit in $\mathbf{G}$, and (iii) The norm of $\mathbf{G}$ equals $\| \cdot \|_\infty$ (as in (A.4)).

We conclude Appendix A by outlining the organization of the remaining Appendices.

APPENDIX B: Contains the proofs of the results in Section 4 concerning consistency of the set estimator (Lemma 4.1) and its rates of convergence (Theorem 4.1). ∎

APPENDIX C: Develops the proofs for the results in Section 5, including the preliminary local approximation (Lemma 5.1) and the final drift linearization (Theorem 5.1). ∎

APPENDIX D: Contains the proofs for all results in Section 6, including the lower bound for the bootstrap statistic (Theorem 6.1), conditions under which the lower bound coupling is "sharp" (Theorem 6.2), and the analysis of the test that compares the proposed test statistic to the quantiles of the bootstrap distribution (Theorem 6.3). ∎

APPENDIX E: Develops the auxiliary results concerning the approximation of the local parameter space. These results depend on the characterization of R only, and thus may be of independent interest as they are broadly applicable to hypotheses testing problems similarly concerned with examining equality and inequality restrictions. ∎

APPENDIX F: Provides additional details concerning the implementation of our test and the implications of our Assumptions in the context of the motivating examples introduced in Section 2.2. ∎

APPENDIX G: Derives primitive conditions for verifying the coupling requirements of Assumption 5.1. The results employ the Hungarian construction in Koltchinskii (1994) and may be of independent interest. ∎

APPENDIX H: Provides primitive conditions for the validity of the Gaussian multiplier bootstrap as imposed in Assumption 6.5. These results more generally provide sufficient conditions for the Gaussian multiplier bootstrap to be consistent for the law of the empirical process over expanding classes $\mathcal{F}_n$, and may be of independent interest. The arguments in this Appendix can also be employed to obtain alternative sufficient conditions for Assumption 5.1 that complement those in Appendix G. ∎

APPENDIX B - Proofs for Section 4

PROOF OF LEMMA 4.1: First fix $\epsilon > 0$ and notice that by definition of $\hat{\Theta}_n \cap R$ we have

$$P(\overrightarrow{d}_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{B}}) > \epsilon)$$
$$\leq P(\inf_{\theta \in (\Theta_n \cap R) \setminus (\Theta_{0n}(P) \cap R)^\epsilon} \frac{1}{\sqrt{n}} Q_n(\theta) \leq \inf_{\theta \in \hat{\Theta}_n \cap R} \frac{1}{\sqrt{n}} Q_n(\theta) + \tau_n) \quad \text{(B.1)}$$

for all n and all $P \in \mathbf{P}_0$. Moreover, setting $\bar{Q}_{n,P}(\theta) \equiv \|\sqrt{n} E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_{\hat{\Sigma}_n, r}$,

it then follows from Lemmas B.2 and B.3, and Markov's inequality that

$$\inf_{\theta \in (\Theta_n \cap R) \setminus (\Theta_{0n}(P) \cap R)^\epsilon} \frac{1}{\sqrt{n}} \bar{Q}_{n,P}(\theta)$$

$$\leq \inf_{\theta \in (\Theta_n \cap R) \setminus (\Theta_{0n}(P) \cap R)^\epsilon} \frac{1}{\sqrt{n}} Q_n(\theta) + O_p \left(\frac{k_n^{1/r} \sqrt{\log(k_n)} J_n B_n}{\sqrt{n}} \right) \quad \text{(B.2)}$$

uniformly in $P \in \mathbf{P}_0$. In addition, by similar arguments we obtain uniformly in $P \in \mathbf{P}_0$

$$\inf_{\theta \in \Theta_n \cap R} \frac{1}{\sqrt{n}} Q_n(\theta) \leq \inf_{\theta \in \Theta_n \cap R} \frac{1}{\sqrt{n}} \bar{Q}_{n,P}(\theta) + O_p \left(\frac{k_n^{1/r} \sqrt{\log(k_n)} J_n B_n}{\sqrt{n}} \right)$$

$$\leq \|\hat{\Sigma}_n\|_{o,r} \times \zeta_n + O_p \left(\frac{k_n^{1/r} \sqrt{\log(k_n)} J_n B_n}{\sqrt{n}} \right) = O_p \left(\zeta_n + \frac{k_n^{1/r} \sqrt{\log(k_n)} J_n B_n}{\sqrt{n}} \right) , \quad \text{(B.3)}$$

where the second inequality results from Assumption 4.1(i) and the equality follows from Lemma B.3. For conciseness set $\eta_n \equiv (\zeta_n + \tau_n + k_n^{1/r} \sqrt{\log(k_n)} J_n B_n / \sqrt{n})$. Then note that combining results (B.1), (B.2), and (B.3) we can conclude that

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(\overrightarrow{d}_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{B}}) > \epsilon)$$

$$\leq \limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P \left(\inf_{\theta \in (\Theta_n \cap R) \setminus (\Theta_{0n}(P) \cap R)^\epsilon} \frac{1}{\sqrt{n}} \bar{Q}_{n,P}(\theta) \leq M \eta_n \right) . \quad \text{(B.4)}$$

Next note that for any $a \in \mathbf{R}^{k_n}$ we have $\|a\|_r = \|\hat{\Sigma}_n^{-1} \hat{\Sigma}_n a\|_r \leq \|\hat{\Sigma}_n^{-1}\|_{o,r} \|a\|_{\hat{\Sigma}_n,r}$ (provided $\hat{\Sigma}_n^{-1}$ exists). Thus, by Assumption 4.1(ii) and Lemma B.3 we obtain for any $M < \infty$

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P \left(\inf_{\theta \in (\Theta_n \cap R) \setminus (\Theta_{0n}(P) \cap R)^\epsilon} \frac{1}{\sqrt{n}} \bar{Q}_{n,P}(\theta) \leq M \eta_n \right)$$

$$\leq \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(S_n(\epsilon) \leq \|\hat{\Sigma}_n^{-1}\|_{o,r} M \eta_n) = 0 \quad \text{(B.5)}$$

which together with (B.4) establishes the first claim of the Lemma.

In order to establish (44), we employ the definition of $\hat{\Theta}_n \cap R$ to obtain for all $P \in \mathbf{P}_0$

$$P(\Theta_{0n}(P) \cap R \subseteq \hat{\Theta}_n \cap R) \geq P \left(\sup_{\theta \in \Theta_{0n}(P) \cap R} \frac{1}{\sqrt{n}} Q_n(\theta) \leq \tau_n \right) . \quad \text{(B.6)}$$

Therefore, setting $\delta_n \equiv k_n^{1/r} \sqrt{\log(k_n)} J_n B_n / \sqrt{n}$, exploiting Lemmas B.2 and B.3, and the definition of $\Theta_{0n}(P) \cap R$ we then obtain uniformly in $P \in \mathbf{P}_0$ that

$$\sup_{\theta \in \Theta_{0n}(P) \cap R} \frac{1}{\sqrt{n}} Q_n(\theta) \leq \sup_{\theta \in \Theta_{0n}(P) \cap R} \frac{1}{\sqrt{n}} \bar{Q}_{n,P}(\theta) + O_p(\delta_n)$$

$$\leq \|\hat{\Sigma}_n\|_{o,r} \times \inf_{\theta \in \Theta_n \cap R} \|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_r + O_p(\delta_n) = O_p(\zeta_n + \delta_n) . \quad \text{(B.7)}$$

Hence, (44) follows from results (B.6), (B.7) and $\tau_n/(\delta_n + \zeta_n) \to \infty$. ∎

PROOF OF THEOREM 4.1: To begin, we first define the event $A_n \equiv A_{n1} \cap A_{n2}$ where

$$A_{n1} \equiv \{\hat{\Theta}_n \cap R \subseteq (\Theta_{0n}(P) \cap R)^\epsilon\}$$
$$A_{n2} \equiv \{\hat{\Sigma}_n^{-1} \text{ exists and } \max\{\|\hat{\Sigma}_n^{-1}\|_{o,r}, \|\hat{\Sigma}_n\|_{o,r}\} < B\} , \qquad (B.8)$$

where recall $(\Theta_{0n}(P) \cap R)^\epsilon \equiv \{\theta \in \Theta_n \cap R : \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{B}}) < \epsilon\}$. Moreover, note that for any $\epsilon > 0$ and B sufficiently large, Lemmas 4.1 and B.3 imply

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(A_n^c) = 0 . \qquad (B.9)$$

Hence, for $\eta_n^{-1} \equiv \nu_n\{k_n^{1/r}\sqrt{\log(k_n)}J_n B_n/\sqrt{n} + \tau_n + \zeta_n\}$ we obtain for any M that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\eta_n \overrightarrow{d}_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{E}}) > 2^M)$$
$$= \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\eta_n \overrightarrow{d}_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{E}}) > 2^M; \ A_n) \quad (B.10)$$

by result (B.9). Next, for each $P \in \mathbf{P}_0$, partition $(\Theta_{0n}(P) \cap R)^\epsilon \setminus (\Theta_{0n}(P) \cap R)$ into

$$S_{n,j}(P) \equiv \{\theta \in (\Theta_{0n}(P) \cap R)^\epsilon : 2^{j-1} < \eta_n \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{E}}) \leq 2^j\} . \quad (B.11)$$

Since $\hat{\Theta}_n \cap R \subseteq (\Theta_{0n}(P) \cap R)^\epsilon$ with probability tending to one uniformly in $P \in \mathbf{P}_0$ by (B.9), it follows from the definition of $\hat{\Theta}_n \cap R$ and result (B.10) that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\eta_n \overrightarrow{d}_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{E}}) > 2^M)$$
$$\leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} \sum_{j\geq M}^{\infty} P(\inf_{\theta\in S_{n,j}(P)} \frac{1}{\sqrt{n}}Q_n(\theta) \leq \inf_{\theta\in\Theta_n\cap R} \frac{1}{\sqrt{n}}Q_n(\theta) + \tau_n; \ A_n) . \quad (B.12)$$

In addition, letting $\bar{Q}_{n,P}(\theta) \equiv \|\sqrt{n}E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_{\hat{\Sigma}_n, r}$, we obtain from (B.8), Lemma B.2, and the definition of ζ_n in Assumption 4.1(i) that under the event A_n

$$\inf_{\theta\in\Theta_n\cap R} \frac{1}{\sqrt{n}}Q_n(\theta) \leq \inf_{\theta\in\Theta_n\cap R} \frac{1}{\sqrt{n}}\bar{Q}_{n,P}(\theta) + \|\hat{\Sigma}_n\|_{o,r} \times \mathcal{Z}_{n,P} \leq B\{\mathcal{Z}_{n,P} + \zeta_n\} . \quad (B.13)$$

Therefore, exploiting result (B.12) and that (B.13) holds under A_n we can conclude

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\eta_n \overrightarrow{d}_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{E}}) > 2^M)$$
$$\leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} \sum_{j\geq M}^{\infty} P(\inf_{\theta\in S_{n,j}(P)} \frac{1}{\sqrt{n}}Q_n(\theta) \leq B\{\mathcal{Z}_{n,P} + \zeta_n\} + \tau_n; \ A_n) . \quad (B.14)$$

Further note that for any $a \in \mathbf{R}^{k_n}$, $\|a\|_r = \|\hat{\Sigma}_n^{-1}\hat{\Sigma}_n a\|_r \leq \|\hat{\Sigma}_n^{-1}\|_r \|a\|_{\hat{\Sigma}_n, r} \leq B\|a\|_{\hat{\Sigma}_n, r}$

48

under the event A_n. Therefore, Lemma B.2 implies that under the event A_n

$$\inf_{\theta \in S_{n,j}(P)} \frac{1}{\sqrt{n}} Q_n(\theta) \geq \inf_{\theta \in S_{n,j}(P)} \frac{1}{\sqrt{n}} \bar{Q}_{n,P}(\theta) - \|\hat{\Sigma}_n\|_{o,r} \times \mathcal{Z}_{n,P}$$
$$\geq B^{-1} \times \inf_{\theta \in S_{n,j}(P)} \|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_r - B \times \mathcal{Z}_{n,P} . \quad \text{(B.15)}$$

Moreover, since $\zeta_n \leq (\eta_n \nu_n)^{-1}$, definition (B.11) implies that for j sufficiently large

$$\nu_n^{-1} \times \inf_{\theta \in S_{n,j}(P)} \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}(P) \cap R, \| \cdot \|_{\mathbf{E}}) - O(\zeta_n) \geq \frac{2^{(j-1)}}{\eta_n \nu_n} - O(\zeta_n) \geq \frac{2^{(j-2)}}{\eta_n \nu_n} . \quad \text{(B.16)}$$

Thus, $S_{n,j}(P) \subseteq (\Theta_{0n}(P) \cap R)^{\epsilon}$, Assumption 4.2, and (B.16) imply for j large that

$$\inf_{\theta \in S_{n,j}(P)} \|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_r \geq \frac{2^{(j-2)}}{\eta_n \nu_n} . \quad \text{(B.17)}$$

Hence, we can conclude from results (B.14), (B.15), and (B.17) that we must have

$$\limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(\eta_n \overrightarrow{d}_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \| \cdot \|_{\mathbf{E}}) > 2^M)$$
$$\leq \limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} \sum_{j \geq M}^{\infty} P(\frac{1}{B}(\frac{2^{(j-2)}}{\eta_n \nu_n}) \leq 2B\mathcal{Z}_{n,P} + B\zeta_n + \tau_n)$$
$$\leq \limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} \sum_{j \geq M}^{\infty} P(\frac{1}{4B}(\frac{2^{(j-2)}}{\eta_n \nu_n}) \leq 2B\mathcal{Z}_{n,P}) , \quad \text{(B.18)}$$

where in the final inequality we exploited that the definition of η_n implies $(\eta_n \nu_n)^{-1} \geq \tau_n$ and $(\eta_n \nu_n)^{-1} \geq \zeta_n$. Therefore, $\mathcal{Z}_{n,P} \in \mathbf{R}_+$, Lemma B.2, and Markov's inequality yield

$$\limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} \sum_{j \geq M} P(\frac{1}{8B^2}(\frac{2^{(j-2)}}{\eta_n \nu_n}) \leq \mathcal{Z}_{n,P})$$
$$\lesssim \limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sum_{j \geq M} 2^{-j} \times \frac{\eta_n \nu_n k_n^{1/r} \sqrt{\log(k_n)} J_n B_n}{\sqrt{n}} = 0 , \quad \text{(B.19)}$$

where in the final result we used $\eta_n \nu_n \leq \sqrt{n}/k_n^{1/r} \sqrt{\log(k_n)} J_n B_n$, and that $\sum_{j=1}^{\infty} 2^{-j} < \infty$. Hence, the first claim of the Theorem follows from (B.18) and (B.19).

To establish the second claim, define the event $A_{n3} \equiv \{\Theta_{0n}(P) \cap R \subseteq \hat{\Theta}_n \cap R\}$. Since $\overrightarrow{d}_H(\Theta_{0n}(P) \cap R, \hat{\Theta}_n \cap R, \| \cdot \|_{\mathbf{E}}) = 0$ whenever A_{n3} occurs, we obtain from Lemma 4.1

$$\limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(\eta_n d_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \| \cdot \|_{\mathbf{E}}) > 2^M)$$
$$= \limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(\eta_n \overrightarrow{d}_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \| \cdot \|_{\mathbf{E}}) > 2^M) = 0 \quad \text{(B.20)}$$

due to result (48). Therefore, the second claim of the Theorem follows from (B.20). ∎

Lemma B.1. *Let Assumption 3.2(i) hold, and define the class of functions*

$$\mathcal{G}_n \equiv \{f(x)q_{k,n,\jmath}(z_\jmath) : f \in \mathcal{F}_n, \ 1 \leq \jmath \leq \mathcal{J}, \ and \ 1 \leq k \leq k_{n,\jmath}\} \ . \tag{B.21}$$

Then, it follows that $N_{[]}(\epsilon, \mathcal{G}_n, \|\cdot\|_{L^2_P}) \leq k_n \times N_{[]}(\epsilon/B_n, \mathcal{F}_n, \|\cdot\|_{L^2_P})$ for all $P \in \mathbf{P}$.

PROOF: Note that by Assumption 3.2(i) we have $\sup_{P\in\mathbf{P}} \|q_{k,n,\jmath}\|_{L^\infty_P} \leq B_n$ for all $1 \leq \jmath \leq \mathcal{J}$ and $1 \leq k \leq k_{n,\jmath}$, and define $q^+_{k,n,\jmath}(z_\jmath) \equiv q_{k,n,\jmath}(z_\jmath)\mathbf{1}\{q_{k,n,\jmath}(z_\jmath) \geq 0\}$ and $q^-_{k,n,\jmath}(z_\jmath) \equiv q_{k,n,\jmath}(z_\jmath)\mathbf{1}\{q_{k,n,\jmath}(z_\jmath) \leq 0\}$. If $\{[f_{i,l,P}, f_{i,u,P}]\}_i$ is a collection of brackets for $\mathcal{F}_n$ with

$$\int (f_{i,u,P} - f_{i,l,P})^2 dP \leq \epsilon^2 \tag{B.22}$$

for all i, then it follows that the following collection of brackets covers the class $\mathcal{G}_n$

$$\{[q^+_{k,n,\jmath}f_{i,l,P} + q^-_{k,n,\jmath}f_{i,u,P}, \ q^-_{k,n,\jmath}f_{i,l,P} + q^+_{k,n,\jmath}f_{i,u,P}]\}_{i,k,\jmath} \ . \tag{B.23}$$

Moreover, since $|q_{k,n,\jmath}| = q^+_{k,n,\jmath} - q^-_{k,n,\jmath}$ by construction, we also obtain from (B.22) that

$$\int (q^+_{k,n,\jmath}f_{i,u,P} + q^-_{k,n,\jmath}f_{i,l,P} - q^+_{k,n,\jmath}f_{i,l,P} - q^-_{k,n,\jmath}f_{i,u,P})^2 dP$$
$$= \int (f_{i,u,P} - f_{i,l,P})^2 |q_{k,n,\jmath}|^2 dP \leq \epsilon^2 B_n^2 \ . \tag{B.24}$$

Since there are $k_n \times N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{L^2_P})$ brackets in (B.23), we can conclude from (B.24)

$$N_{[]}(\epsilon, \mathcal{G}_n, \|\cdot\|_{L^2_P}) \leq k_n \times N_{[]}(\frac{\epsilon}{B_n}, \mathcal{F}_n, \|\cdot\|_{L^2_P}) \ , \tag{B.25}$$

for all $P \in \mathbf{P}$, which establishes the claim of the Lemma. ∎

Lemma B.2. *Let $\bar{Q}_{n,P}(\theta) \equiv \|\sqrt{n}E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_{\hat{\Sigma}_n, r}$, and Assumptions 3.1, 3.2(i), 3.3(ii)-(iii) hold. Then, for each $P \in \mathbf{P}$ there are random $\mathcal{Z}_{n,P} \in \mathbf{R}_+$ with*

$$\frac{1}{\sqrt{n}}|Q_n(\theta) - \bar{Q}_{n,P}(\theta)| \leq \|\hat{\Sigma}_n\|_{o,r} \times \mathcal{Z}_{n,P} \ , \tag{B.26}$$

for all $\theta \in \Theta_n \cap R$ and in addition $\sup_{P\in\mathbf{P}} E_P[\mathcal{Z}_{n,P}] = O(k_n^{1/r}\sqrt{\log(k_n)}J_nB_n/\sqrt{n})$.

PROOF: Let $\mathcal{G}_n \equiv \{f(x)q_{k,n,\jmath}(z_\jmath) : f \in \mathcal{F}_n, \ 1 \leq \jmath \leq \mathcal{J} \ and \ 1 \leq k \leq k_{n,\jmath}\}$. Note that by Assumption 3.2(i), $\sup_{P\in\mathbf{P}} \|q_{k,n,\jmath}\|_{L^\infty_P} \leq B_n$ for all $1 \leq \jmath \leq \mathcal{J}$ and $1 \leq k \leq k_{n,\jmath}$. Hence, letting F_n be the envelope for $\mathcal{F}_n$, as in Assumption 3.3(ii), it follows that $G_n(v) \equiv B_nF_n(v)$ is an envelope for $\mathcal{G}_n$ satisfying $\sup_{P\in\mathbf{P}} E_P[G_n^2(V_i)] < \infty$. Thus,

$$\sup_{P\in\mathbf{P}} E_P[\sup_{g\in\mathcal{G}_n} |\mathbb{G}_{n,P}g|] \lesssim \sup_{P\in\mathbf{P}} J_{[]}(\|G_n\|_{L^2_P}, \mathcal{G}_n, \|\cdot\|_{L^2_P}) \ . \tag{B.27}$$

by Theorem 2.14.2 in van der Vaart and Wellner (1996). Moreover, also notice that Lemma B.1, the change of variables $u = \epsilon/B_n$ and $B_n \geq 1$ imply

$$\sup_{P \in \mathbf{P}} J_{[]}(\|G_n\|_{L_P^2}, \mathcal{G}_n, \|\cdot\|_{L_P^2}) \leq \sup_{P \in \mathbf{P}} \int_0^{\|G_n\|_{L_P^2}} \sqrt{1 + \log(k_n N_{[]}(\epsilon/B_n, \mathcal{F}_n, \|\cdot\|_{L_P^2}))} d\epsilon$$
$$\leq (1 + \sqrt{\log(k_n)})B_n \times \sup_{P \in \mathbf{P}} J_{[]}(\|F_n\|_{L_P^2}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = O(\sqrt{\log(k_n)}B_n J_n) , \quad \text{(B.28)}$$

where the final equality follows from Assumption 3.3(iii). Next define $\mathcal{Z}_{n,P} \in \mathbf{R}_+$ by

$$\mathcal{Z}_{n,P} \equiv \frac{k_n^{1/r}}{\sqrt{n}} \times \sup_{g \in \mathcal{G}_n} |\mathbb{G}_{n,P}g| \quad \text{(B.29)}$$

and note (B.27) and (B.28) imply $\sup_{P \in \mathbf{P}} E_P[\mathcal{Z}_{n,P}] = O(k_n^{1/r}\sqrt{\log(k_n)}B_n J_n/\sqrt{n})$ as desired. Since we also have that $\|\mathbb{G}_{n,P}\rho(\cdot,\theta) * q_n^{k_n}\|_r \leq k_n^{1/r} \times \sup_{g \in \mathcal{G}_n} |\mathbb{G}_{n,P}g|$ for all $\theta \in \Theta_n \cap R$ by definition of $\mathcal{G}_n$, we can in turn conclude by direct calculation

$$\frac{1}{\sqrt{n}}|Q_n(\theta) - \bar{Q}_{n,P}(\theta)| \leq \frac{\|\hat{\Sigma}_n\|_{o,r}}{\sqrt{n}} \times \|\mathbb{G}_{n,P}\rho(\cdot,\theta) * q_n^{k_n}\|_r \leq \|\hat{\Sigma}_n\|_{o,r} \times \mathcal{Z}_{n,P} , \quad \text{(B.30)}$$

which establishes the claim of the Lemma. ∎

Lemma B.3. *If Assumption 3.4 holds, then there exists a constant $B < \infty$ such that*

$$\liminf_{n \to \infty} \inf_{P \in \mathbf{P}} P(\hat{\Sigma}_n^{-1} \text{ exists and } \max\{\|\hat{\Sigma}_n\|_{o,r}, \|\hat{\Sigma}_n^{-1}\|_{o,r}\} < B) = 1 . \quad \text{(B.31)}$$

PROOF: First note that by Assumption 3.4(iii) there exists a $B < \infty$ such that

$$\sup_{n \geq 1} \sup_{P \in \mathbf{P}} \max\{\|\Sigma_n(P)\|_{o,r}, \|\Sigma_n(P)^{-1}\|_{o,r}\} < \frac{B}{2} . \quad \text{(B.32)}$$

Next, let I_n denote the $k_n \times k_n$ identity matrix and for each $P \in \mathbf{P}$ rewrite $\hat{\Sigma}_n$ as

$$\hat{\Sigma}_n = \Sigma_n(P)\{I_n - \Sigma_n(P)^{-1}(\Sigma_n(P) - \hat{\Sigma}_n)\} . \quad \text{(B.33)}$$

By Theorem 2.9 in Kress (1999), the matrix $\{I_n - \Sigma_n(P)^{-1}(\Sigma_n(P) - \hat{\Sigma}_n)\}$ is invertible and the operator norm of its inverse is bounded by two when $\Sigma_n(P)^{-1}(\Sigma_n(P) - \hat{\Sigma}_n)\} < 1/2$. Since by Assumption 3.4(ii) and the equality in (B.33) it follows that $\hat{\Sigma}_n$ is invertible if and only if $\{I_n - \Sigma_n(P)^{-1}(\Sigma_n(P) - \hat{\Sigma}_n)\}$ is invertible, we obtain that

$$P(\hat{\Sigma}_n^{-1} \text{ exists and } \|\{I_n - \Sigma_n(P)^{-1}(\Sigma_n(P) - \hat{\Sigma}_n)\}^{-1}\|_{o,r} < 2)$$
$$\geq P(\|\Sigma_n(P)^{-1}(\hat{\Sigma}_n - \Sigma_n(P))\|_{o,r} < \frac{1}{2}) \geq P(\|\hat{\Sigma}_n - \Sigma_n(P)\|_{o,r} < \frac{1}{B}) , \quad \text{(B.34)}$$

where we exploited $\|\Sigma_n(P)^{-1}(\hat{\Sigma}_n - \Sigma_n(P))\|_{o,r} \leq \|\Sigma_n(P)^{-1}\|_{o,r}\|\hat{\Sigma}_n - \Sigma_n(P)\|_{o,r}$ and

(B.32). Hence, since $\|\Sigma_n(P)\|_{o,r} < B/2$ for all $P \in \mathbf{P}$ and n, (B.33) and (B.34) yield

$$P(\hat{\Sigma}_n^{-1} \text{ exists and } \|\hat{\Sigma}_n^{-1}\|_{o,r} < B) \geq P(\|\hat{\Sigma}_n - \Sigma_n(P)\|_{o,r} < \frac{1}{B}) . \tag{B.35}$$

Finally, since $\|\hat{\Sigma}_n\|_{o,r} \leq B/2 + \|\hat{\Sigma}_n - \Sigma_n(P)\|_{o,r}$ by (B.32), result (B.35) implies that

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P(\hat{\Sigma}_n^{-1} \text{ exists and } \max\{\|\hat{\Sigma}_n\|_{o,r}, \|\hat{\Sigma}_n^{-1}\|_{o,r}\} < B)$$

$$\geq \liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P(\|\hat{\Sigma}_n - \Sigma_n(P)\|_{o,r} < \min\{\frac{B}{2}, \frac{1}{B}\}) = 1 , \quad \text{(B.36)}$$

where the equality, and hence the Lemma, follows from Assumption 3.4(i). ∎

Lemma B.4. *If $a \in \mathbf{R}^d$, then $\|a\|_{\tilde{r}} \leq d^{(\frac{1}{\tilde{r}} - \frac{1}{r})+} \|a\|_r$ for any $\tilde{r}, r \in [2, \infty]$.*

PROOF: The case $r \leq \tilde{r}$ trivially follows from $\|a\|_{\tilde{r}} \leq \|a\|_r$ for all $a \in \mathbf{R}^d$. For the case $r > \tilde{r}$, let $a = (a^{(1)}, \dots, a^{(d)})'$ and note that by Hölder's inequality we can obtain that

$$\|a\|_{\tilde{r}}^{\tilde{r}} = \sum_{i=1}^{d} |a^{(i)}|^{\tilde{r}} = \sum_{i=1}^{d} \{|a^{(i)}|^{\tilde{r}} \times 1\}$$

$$\leq \{\sum_{i=1}^{d} (|a^{(i)}|^{\tilde{r}})^{\frac{r}{\tilde{r}}}\}^{\frac{\tilde{r}}{r}} \{\sum_{i=1}^{d} 1^{\frac{r}{r-\tilde{r}}}\}^{1-\frac{\tilde{r}}{r}} = \{\sum_{i=1}^{d} |a^{(i)}|^r\}^{\frac{\tilde{r}}{r}} d^{1-\frac{\tilde{r}}{r}} . \quad \text{(B.37)}$$

Thus, the claim of the Lemma for $r > \tilde{r}$ follows from taking the $1/\tilde{r}$ power in (B.37). ∎

APPENDIX C - Proofs for Section 5

PROOF OF LEMMA 5.1: First note that the existence of the required sequence $\{\ell_n\}$ is guaranteed by Assumption 5.3(i). Next, for $\eta_n = o(a_n)$ let $\hat{\theta}_n \in \Theta_n \cap R$ satisfy

$$Q_n(\hat{\theta}_n) \leq \inf_{\theta\in\Theta_n\cap R} Q_n(\theta) + \eta_n . \tag{C.1}$$

Applying Theorem 4.1 with $\tau_n \equiv \eta_n/\sqrt{n}$ and noting $\tau_n = o(k_n^{1/r}/\sqrt{n})$, then yields that

$$\overrightarrow{d}_H(\{\hat{\theta}_n\}, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{E}}) = O_p(\mathcal{R}_n) \tag{C.2}$$

uniformly in $P \in \mathbf{P}_0$. Hence, defining for each $P \in \mathbf{P}_0$ the shrinking neighborhood $(\Theta_{0n}(P) \cap R)^{\ell_n} \equiv \{\theta \in \Theta_n \cap R : \overrightarrow{d}_H(\theta, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{E}}) \leq \ell_n\}$, we obtain

$$I_n(R) = \inf_{\theta\in(\Theta_{0n}(P)\cap R)^{\ell_n}} Q_n(\theta) + o_p(a_n) \tag{C.3}$$

uniformly in $P \in \mathbf{P}_0$ due to $\mathcal{R}_n = o(\ell_n)$, $\eta_n = o(a_n)$, and results (C.1) and (C.2). Defining

$$Q_{n,P}^0(\theta) \equiv \|\mathbb{W}_{n,P}\rho(\cdot,\theta) * q_n^{k_n} + \sqrt{n}P\rho(\cdot,\theta) * q_n^{k_n}\|_{\hat{\Sigma}_n,r} \tag{C.4}$$

we also obtain from Assumption 5.1 and Lemmas B.3 and C.1 that uniformly in $P \in \mathbf{P}_0$

$$\left| \inf_{\theta \in (\Theta_{0n}(P) \cap R)^{\ell_n}} Q_n(\theta) - \inf_{\theta \in (\Theta_{0n}(P) \cap R)^{\ell_n}} Q_{n,P}^0(\theta) \right|$$
$$\leq \mathcal{J}\|\hat{\Sigma}_n\|_{o,r} \times \sup_{f \in \mathcal{F}_n} \|\mathbb{G}_{n,P}fq_n^{k_n} - \mathbb{W}_{n,P}fq_n^{k_n}\|_r = o_p(a_n) . \tag{C.5}$$

Similarly, exploiting Lemmas B.3 and C.1 together with Lemma C.2 and ℓ_n satisfying $k_n^{1/r}\sqrt{\log(k_n)}B_n \times \sup_{P \in \mathbf{P}} J_{[]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$ by hypothesis yields

$$\inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \|\mathbb{W}_{n,P}\rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n} + \sqrt{n}P\rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n}\|_{\hat{\Sigma}_n,r}$$

$$= \inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \|\mathbb{W}_{n,P}\rho(\cdot, \theta_0) * q_n^{k_n} + \sqrt{n}P\rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n}\|_{\hat{\Sigma}_n,r} + o_p(a_n)$$
$$\tag{C.6}$$

uniformly in $P \in \mathbf{P}_0$. Thus, the Lemma follows from results (C.3), (C.5), and (C.6) together with Lemma C.3. ∎

PROOF OF THEOREM 5.1: First we note that Assumption 4.1(i) implies

$$\sup_{P \in \mathbf{P}_0} \sup_{\theta_0 \in \Theta_{0n}(P) \cap R} \|\sqrt{n}P\rho(\cdot, \theta_0) * q_n^{k_n}\|_r \leq \sqrt{n}\zeta_n . \tag{C.7}$$

Therefore, Lemma B.4, result (C.7), and the law of iterated expectations yield that for all $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}(P) \cap R$, and $h/\sqrt{n} \in V_n(\theta_0, \ell_n)$ we must have

$$\|\sqrt{n}P\rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n} - \mathbb{D}_{n,P}(\theta_0)[h]\|_r$$
$$\leq \|\sqrt{n}\{P\rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n} - P\rho(\cdot, \theta_0) * q_n^{k_n}\} - \mathbb{D}_{n,P}(\theta_0)[h]\|_2 + \sqrt{n}\zeta_n . \tag{C.8}$$

Moreover, Lemma C.5 and the maps $m_{P,\jmath} : \mathbf{B}_n \to L_P^2$ satisfying Assumption 5.4(i) imply

$$\sum_{\jmath=1}^{\mathcal{J}} \sum_{k=1}^{k_{n,\jmath}} \langle \sqrt{n}\{m_{P,\jmath}(\theta_0 + \frac{h}{\sqrt{n}}) - m_{P,\jmath}(\theta_0)\} - \nabla m_{P,\jmath}(\theta_0)[h], q_{k,n,\jmath}\rangle_{L_P^2}^2$$

$$\leq \sum_{\jmath=1}^{\mathcal{J}} C_0\|\sqrt{n}\{m_{P,\jmath}(\theta_0 + \frac{h}{\sqrt{n}}) - m_{P,\jmath}(\theta_0) - \nabla m_{P,\jmath}(\theta_0)[\frac{h}{\sqrt{n}}]\}\|_{L_P^2}^2$$

$$\leq \sum_{\jmath=1}^{\mathcal{J}} C_0 K_m^2 \times n \times \|\frac{h}{\sqrt{n}}\|_{\mathbf{L}}^2 \times \|\frac{h}{\sqrt{n}}\|_{\mathbf{E}}^2 \tag{C.9}$$

53

for some constant $C_0 < \infty$ and all $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}(P) \cap R$, and $h/\sqrt{n} \in V_n(\theta_0, \ell_n)$. Therefore, by results (C.8) and (C.9), and the definition of $\mathcal{S}_n(\mathbf{L}, \mathbf{E})$ in (35) we get

$$\sup_{P \in \mathbf{P}_0} \sup_{\theta_0 \in \Theta_{0n}(P) \cap R} \sup_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \| \sqrt{n} P \rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n} - \mathbb{D}_{n,P}(\theta_0)[h] \|_r$$

$$\leq \sqrt{\mathcal{J} C_0} K_m \times \sqrt{n} \times \ell_n^2 \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) + \sqrt{n} \zeta_n = o(a_n) \quad \text{(C.10)}$$

due to $K_m \ell_n^2 \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$ by hypothesis and $\sqrt{n} \zeta_n = o(a_n)$ by Assumption 5.3(ii). Next, note that since $k_n^{1/r} \sqrt{\log(k_n)} B_n \times \sup_{P \in \mathbf{P}} J_{[]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \| \cdot \|_{L_P^2}) = o(a_n)$, Assumption 5.3(i) implies there is a sequence $\tilde{\ell}_n$ satisfying the conditions of Lemma 5.1 and $\ell_n = o(\tilde{\ell}_n)$. Therefore, applying Lemma 5.1 we obtain that

$$I_n(R) = \inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \tilde{\ell}_n)} \| \mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \sqrt{n} P \rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n} \|_{\Sigma_n(P), r} + o_p(a_n)$$

$$\text{(C.11)}$$

Moreover, since $\ell_n = o(\tilde{\ell}_n)$ implies $V_n(\theta, \tilde{\ell}_n) \subseteq V_n(\theta, \ell_n)$ for all $\theta \in \Theta_n \cap R$, we have

$$\inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \tilde{\ell}_n)} \| \mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \sqrt{n} P \rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n} \|_{\Sigma_n(P), r}$$

$$\leq \inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \| \mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \sqrt{n} P \rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n} \|_{\Sigma_n(P), r}$$

$$= \inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \| \mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h] \|_{\Sigma_n(P), r} + o_p(a_n) \quad \text{(C.12)}$$

uniformly in $P \in \mathbf{P}_0$, with the final equality following from (C.10), Assumption 3.4(iii) and Lemma C.1. Thus, the first claim of the Theorem follows from (C.11) and (C.12), while the second follows by noting that if $K_m \mathcal{R}_n^2 \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$, then we may set ℓ_n to simultaneously satisfy the conditions of Lemma 5.1 and $K_m \ell_n^2 \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$, which obviates the need to introduce $\tilde{\ell}_n$ in (C.11) and (C.12). ∎

Lemma C.1. *If Λ is a set, $A : \Lambda \to \mathbf{R}^k$, $B : \Lambda \to \mathbf{R}^k$, and W is a $k \times k$ matrix, then*

$$| \inf_{\lambda \in \Lambda} \| W A(\lambda) \|_r - \inf_{\lambda \in \Lambda} \| W B(\lambda) \|_r | \leq \| W \|_{o,r} \times \sup_{\lambda \in \Lambda} \| A(\lambda) - B(\lambda) \|_r .$$

PROOF: Fix $\eta > 0$, and let $\lambda_a \in \Lambda$ satisfy $\| W A(\lambda_a) \|_r \leq \inf_{\lambda \in \Lambda} \| W A(\lambda) \|_r + \eta$. Then,

$$\inf_{\lambda \in \Lambda} \| W B(\lambda) \|_r - \inf_{\lambda \in \Lambda} \| W A(\lambda) \|_r \leq \| W B(\lambda_a) \|_r - \| W A(\lambda_a) \|_r + \eta$$

$$\leq \| W \{ B(\lambda_a) - A(\lambda_a) \} \|_r + \eta \leq \| W \|_{o,r} \times \sup_{\lambda \in \Lambda} \| A(\lambda) - B(\lambda) \|_r + \eta \quad \text{(C.13)}$$

where the second result follows from the triangle inequality, and the final result from

$\|Wv\|_r \leq \|W\|_{o,r}\|v\|_r$ for any $v \in \mathbf{R}^k$. In turn, by identical manipulations we also have

$$\inf_{\lambda \in \Lambda} \|WA(\lambda)\|_r - \inf_{\lambda \in \Lambda} \|WB(\lambda)\|_r \leq \|W\|_{o,r} \times \sup_{\lambda \in \Lambda} \|A(\lambda) - B(\lambda)\|_r + \eta . \qquad (C.14)$$

Thus, since η was arbitrary, the Lemma follows from results (C.13) and (C.14). ∎

Lemma C.2. *Let Assumptions 3.2(i), 3.4, and 5.2(i) hold. If $\delta_n \downarrow 0$ is such that $k_n^{1/r}\sqrt{\log(k_n)}B_n \times \sup_{P \in \mathbf{P}} J_{[]}(\delta_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$, then uniformly in $P \in \mathbf{P}$:*

$$\sup_{\theta_0 \in \Theta_{0n}(P) \cap R} \sup_{\frac{h}{\sqrt{n}} \in V_n(\theta_0,\delta_n)} \|\mathbb{W}_{n,P}\rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n} - \mathbb{W}_{n,P}\rho(\cdot, \theta_0) * q_n^{k_n}\|_{\hat{\Sigma}_n, r} = o_p(a_n) .$$

PROOF: Since $\|q_{k,n,j}\|_{L_P^\infty} \leq B_n$ for all $1 \leq j \leq \mathcal{J}$ and $1 \leq k \leq k_{n,j}$ by Assumption 3.2(i), Assumption 5.2(i) yields for any $P \in \mathbf{P}$, $\theta \in \Theta_n \cap R$, and $h/\sqrt{n} \in V_n(\theta, \delta_n)$ that

$$E_P[\|\rho(X_i, \theta + \frac{h}{\sqrt{n}}) - \rho(X_i, \theta)\|_2^2 q_{k,n,j}^2(Z_i)] \leq K_\rho^2 B_n^2 \|\frac{h}{\sqrt{n}}\|_{\mathbf{E}}^{2\kappa_\rho} \leq K_\rho^2 B_n^2 \delta_n^{2\kappa_\rho} . \qquad (C.15)$$

Next, define the class of functions $\mathcal{G}_n \equiv \{f(x)q_{k,n,j}(z)$ for some $f \in \mathcal{F}_n$, $1 \leq j \leq \mathcal{J}$ and $1 \leq k \leq k_{n,j}\}$, and note that (C.15) implies that for all $1 \leq j \leq \mathcal{J}$ and $P \in \mathbf{P}$

$$\sup_{\theta_0 \in \Theta_{0n}(P) \cap R} \sup_{\frac{h}{\sqrt{n}} \in V_n(\theta_0,\delta_n)} \max_{1 \leq k \leq k_{n,j}} |\mathbb{W}_{n,P}\rho_j(\cdot, \theta_0 + \frac{h}{\sqrt{n}})q_{k,n,j} - \mathbb{W}_{n,P}\rho_j(\cdot, \theta_0)q_{k,n,j}|$$

$$\leq \sup_{g_1, g_2 \in \mathcal{G}_n : \|g_1 - g_2\|_{L_P^2} \leq K_\rho B_n \delta_n^{\kappa_\rho}} |\mathbb{W}_{n,P}g_1 - \mathbb{W}_{n,P}g_2| . \qquad (C.16)$$

Hence, since $\|\hat{\Sigma}_n a\|_r \leq \|\hat{\Sigma}_n\|_{o,r}\|a\|_r \leq \|\hat{\Sigma}_n\|_{o,r}k_n^{1/r}\|a\|_\infty$ for any $a \in \mathbf{R}^{k_n}$, (C.16) yields

$$\sup_{\theta_0 \in \Theta_{0n}(P) \cap R} \sup_{\frac{h}{\sqrt{n}} \in V_n(\theta_0,\delta_n)} \|\mathbb{W}_{n,P}\rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n} - \mathbb{W}_{n,P}\rho(\cdot, \theta_0) * q_n^{k_n}\|_{\hat{\Sigma}_n, r}$$

$$\leq \|\hat{\Sigma}_n\|_{o,r}k_n^{1/r} \times \sup_{g_1, g_2 \in \mathcal{G}_n : \|g_1 - g_2\|_{L_P^2} \leq K_\rho B_n \delta_n^{\kappa_\rho}} |\mathbb{W}_{n,P}g_1 - \mathbb{W}_{n,P}g_2| . \qquad (C.17)$$

Moreover, Corollary 2.2.8 in van der Vaart and Wellner (1996) implies that

$$\sup_{P \in \mathbf{P}} E_P[\sup_{g_1, g_2 \in \mathcal{G}_n : \|g_1 - g_2\|_{L_P^2} \leq K_\rho B_n \delta_n^{\kappa_\rho}} |\mathbb{W}_{n,P}g_1 - \mathbb{W}_{n,P}g_2|]$$

$$\leq \sup_{P \in \mathbf{P}} C_0 \int_0^{K_\rho B_n \delta_n^{\kappa_\rho}} \sqrt{\log N_{[]}(\epsilon/2, \mathcal{G}_n, \|\cdot\|_{L_P^2})}d\epsilon \qquad (C.18)$$

for some $C_0 < \infty$. In turn, Lemma B.1 and the change of variables $u = \epsilon/2B_n$ yields

$$\sup_{P \in \mathbf{P}} \int_0^{K_\rho B_n \delta_n^{\kappa\rho}} \sqrt{\log N_{[]}(\epsilon/2, \mathcal{G}_n, \|\cdot\|_{L_P^2})} d\epsilon$$

$$\leq \sup_{P \in \mathbf{P}} \sqrt{\log(k_n)} \int_0^{K_\rho B_n \delta_n^{\kappa\rho}} \sqrt{1 + \log N_{[]}(\epsilon/2B_n, \mathcal{F}_n, \|\cdot\|_{L_P^2})} d\epsilon$$

$$\leq \sup_{P \in \mathbf{P}} 2\sqrt{\log(k_n)} B_n \int_0^{K_\rho \delta_n^{\kappa\rho}/2} \sqrt{1 + \log N_{[]}(u, \mathcal{F}_n, \|\cdot\|_{L_P^2})} du \ . \quad \text{(C.19)}$$

However, since $N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{L_P^2})$ is a decreasing function of ϵ, we can also conclude

$$\sup_{P \in \mathbf{P}} \int_0^{K_\rho \delta_n^{\kappa\rho}/2} \sqrt{1 + \log N_{[]}(u, \mathcal{F}_n, \|\cdot\|_{L_P^2})} du$$

$$\leq \max\{\frac{K_\rho}{2}, 1\} \times \sup_{P \in \mathbf{P}} J_{[]}(\delta_n^{\kappa\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) \quad \text{(C.20)}$$

by definition of $J_{[]}(\delta, \mathcal{F}_n, \|\cdot\|_{L_P^2})$. Therefore, the Lemma follows from (C.17), $\|\hat{\Sigma}_n\|_{o,r} = O_p(1)$ by Lemma B.3, and Markov's inequality combined with results (C.18), (C.19), (C.20), and $k_n^{1/r} \sqrt{\log(k_n)} B_n \times \sup_{P \in \mathbf{P}} J_{[]}(\delta_n^{\kappa\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$ by hypothesis. ∎

Lemma C.3. *Let Assumptions 3.2(i), 3.3(ii), 3.4, 4.1(i), and 5.3(ii)-(iii) hold. For any sequence $\delta_n \downarrow 0$ it then follows that uniformly in $P \in \mathbf{P}_0$ we have*

$$\inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \delta_n)} \|\mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \sqrt{n} P \rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n}\|_{\Sigma_n(P), r}$$

$$= \inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \delta_n)} \|\mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \sqrt{n} P \rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n}\|_{\hat{\Sigma}_n, r} + o_p(a_n)$$

PROOF: First note that by Assumptions 3.4(ii) there exists a constant $C_0 < \infty$ such that $\max\{\|\Sigma_n(P)\|_{o,r}, \|\Sigma_n(P)^{-1}\|_{o,r}\} \leq C_0$ for all n and $P \in \mathbf{P}$. Thus, we obtain

$$\|\mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \sqrt{n} P \rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n}\|_{\hat{\Sigma}_n, r}$$

$$\leq \{C_0 \|\hat{\Sigma}_n - \Sigma_n(P)\|_{o,r} + 1\} \|\mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \sqrt{n} P \rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n}\|_{\Sigma_n(P), r}$$

$$\text{(C.21)}$$

by the triangle inequality. Moreover, since $0 \in V_n(\theta_0, \delta_n)$ for all θ_0, we also have that

$$\inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \delta_n)} \|\mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \sqrt{n} P \rho(\cdot, \theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n}\|_{\Sigma_n(P), r}$$

$$\leq C_0 \times \inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \|\mathbb{W}_{n,P} \rho(\cdot, \theta_0) * q_n^{k_n} + \sqrt{n} P \rho(\cdot, \theta_0) * q_n^{k_n}\|_r \ . \quad \text{(C.22)}$$

Hence, Lemma C.4, Markov's inequality, and Assumptions 4.1(i) and 5.3(ii) establish

$$
\inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \|\mathbb{W}_{n,P}\rho(\cdot,\theta_0) * q_n^{k_n} + \sqrt{n}P\rho(\cdot,\theta_0) * q_n^{k_n}\|_r
$$

$$
\leq \sup_{\theta \in \Theta_n \cap R} \|\mathbb{W}_{n,P}\rho(\cdot,\theta) * q_n^{k_n}\|_r + o(1) = O_p(k_n^{1/r}\sqrt{\log(k_n)}B_n J_n) \quad \text{(C.23)}
$$

uniformly in $P \in \mathbf{P}_0$. Therefore, (C.21), (C.22), (C.23), and Assumption 5.3(iii) imply

$$
\inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0,\delta_n)} \|\mathbb{W}_{n,P}\rho(\cdot,\theta_0) * q_n^{k_n} + \sqrt{n}P\rho(\cdot,\theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n}\|_{\hat{\Sigma}_n,r}
$$

$$
\leq \inf_{\theta_0 \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0,\delta_n)} \|\mathbb{W}_{n,P}\rho(\cdot,\theta_0) * q_n^{k_n} + \sqrt{n}P\rho(\cdot,\theta_0 + \frac{h}{\sqrt{n}}) * q_n^{k_n}\|_{\Sigma_n(P),r} + o_p(a_n)
$$

$$
\text{(C.24)}
$$

uniformly in $P \in \mathbf{P}_0$. The reverse inequality to (C.24) follows by identical arguments but relying on Lemma B.3 implying $\|\hat{\Sigma}_n\|_{o,r} = O_p(1)$ and $\|\hat{\Sigma}_n^{-1}\|_{o,r} = O_p(1)$ uniformly in $P \in \mathbf{P}$ rather than on $\max\{\|\Sigma_n(P)\|_{o,r}, \|\Sigma_n(P)^{-1}\|_{o,r}\} \leq C_0$. ∎

Lemma C.4. *If Assumptions 3.2(i) and 3.3(ii)-(iii) hold, then for some $K_0 > 0$,*

$$
\sup_{P \in \mathbf{P}} E_P[\sup_{\theta \in \Theta_n \cap R} \|\mathbb{W}_{n,P}\rho(\cdot,\theta) * q_n^{k_n}\|_r] \leq K_0 k_n^{1/r}\sqrt{\log(k_n)}B_n J_n \ .
$$

PROOF: Define the class $\mathcal{G}_n \equiv \{f(x)q_{k,n,\jmath}(z) : f \in \mathcal{F}_n, \ 1 \leq \jmath \leq \mathcal{J}, \text{ and } 1 \leq k \leq k_{n,\jmath}\}$, and note $\|a\|_r \leq d^{1/r}\|a\|_\infty$ for any $a \in \mathbf{R}^d$ implies that for any $P \in \mathbf{P}$ we have

$$
E_P[\sup_{\theta \in \Theta_n \cap R} \|\mathbb{W}_{n,P}\rho(\cdot,\theta) * q_n^{k_n}\|_r] \leq k_n^{1/r} E_P[\sup_{g \in \mathcal{G}_n} |\mathbb{W}_{n,P}g|]
$$

$$
\leq k_n^{1/r}\{E_P[|\mathbb{W}_{n,P}g_0|] + C_1 \int_0^\infty \sqrt{\log N_{[\,]}(\epsilon/2, \mathcal{G}_n, \|\cdot\|_{L_P^2})}d\epsilon\} \ , \quad \text{(C.25)}
$$

where the final inequality holds for any $g_0 \in \mathcal{G}_n$ and some $C_1 < \infty$ by Corollary 2.2.8 in van der Vaart and Wellner (1996). Next, let $G_n(v) \equiv B_n F_n(v)$ for F_n as in Assumption 3.3(ii) and note Assumption 3.2(i) implies G_n is an envelope for $\mathcal{G}_n$. Thus $[-G_n, G_n]$ is a bracket of size $2\|G_n\|_{L_P^2}$ covering $\mathcal{G}_n$, and hence the change of variables $u = \epsilon/2$ yields

$$
\int_0^\infty \sqrt{\log N_{[\,]}(\epsilon/2, \mathcal{G}_n, \|\cdot\|_{L_P^2})}d\epsilon
$$

$$
= 2 \int_0^{2\|G_n\|_{L_P^2}} \sqrt{1 + \log N_{[\,]}(u, \mathcal{G}_n, \|\cdot\|_{L_P^2})}du \leq C_2\sqrt{\log(k_n)}B_n J_n \ , \quad \text{(C.26)}
$$

where the final inequality holds for some $C_2 < \infty$ by result (B.28) and $N_{[\,]}(u, \mathcal{G}_n, \|\cdot\|_{L_P^2})$

being decreasing in u. Furthermore, since $E_P[|\mathbb{W}_{n,P}g_0|] \leq \|g_0\|_{L^2_P} \leq \|G_n\|_{L^2_P}$ we have

$$E_P[|\mathbb{W}_{n,P}g_0|] \leq \|G_n\|_{L^2_P} \leq \int_0^{\|G_n\|_{L^2_P}} \sqrt{1 + \log N_{[\,]}(u, \mathcal{G}_n, \|\cdot\|_{L^2_P})} du . \qquad (C.27)$$

Thus, the claim of the Lemma follows from (C.25), (C.26), and (C.27). ∎

Lemma C.5. *Let Assumption 3.2(ii) hold. It then follows that there exists a constant $C < \infty$ such that for all $P \in \mathbf{P}$, $n \geq 1$, $1 \leq \jmath \leq \mathcal{J}$, and functions $f \in L^2_P$ we have*

$$\sum_{k=1}^{k_{n,\jmath}} \langle f, q_{k,n,\jmath} \rangle_{L^2_P}^2 \leq C E_P[(E_P[f(V_i)|Z_{i,\jmath}])^2] . \qquad (C.28)$$

PROOF: Let $L^2_P(Z_{i,\jmath})$ denote the subspace of L^2_P consisting of functions depending on $Z_{i,\jmath}$ only, and set $\ell^2(\mathbb{N}) \equiv \{\{c_k\}_{k=1}^\infty : c_k \in \mathbf{R} \text{ and } \|\{c_k\}\|_{\ell^2(\mathbb{N})} < \infty\}$, where $\|\{c_k\}\|_{\ell^2(\mathbb{N})}^2 \equiv \sum_k c_k^2$. For any sequence $\{c_k\} \in \ell^2(\mathbb{N})$, then define the map $J_{P,n,\jmath} : \ell^2(\mathbb{N}) \to L^2_P(Z_{i,\jmath})$ by

$$J_{P,n,\jmath}\{c_k\}(z) = \sum_{k=1}^{k_{n,\jmath}} c_k q_{k,n,\jmath}(z) . \qquad (C.29)$$

Clearly, the maps $J_{P,n,\jmath} : \ell^2(\mathbb{N}) \to L^2_P(Z_{i,\jmath})$ are linear, and moreover we note that by Assumption 3.2(ii) there exists a constant $C < \infty$ such that the largest eigenvalue of $E_P[q_{n,\jmath}^{k_{n,\jmath}}(Z_{i,\jmath})q_{n,\jmath}^{k_{n,\jmath}}(Z_{i,\jmath})']$ is bounded by C for all $n \geq 1$ and $P \in \mathbf{P}$. Therefore, we obtain

$$\sup_{P \in \mathbf{P}} \sup_{n \geq 1} \|J_{P,n,\jmath}\|_o^2 = \sup_{P \in \mathbf{P}} \sup_{n \geq 1} \sup_{\{c_k\}:\sum_k c_k^2 = 1} \|J_{P,n,\jmath}\{c_k\}\|_{L^2_P(Z_{i,\jmath})}^2$$

$$= \sup_{P \in \mathbf{P}} \sup_{n \geq 1} \sup_{\{c_k\}:\sum_k c_k^2 = 1} E_P[(\sum_{k=1}^{k_{n,\jmath}} c_k q_{k,n,\jmath}(Z_{i,\jmath}))^2] \leq \sup_{\{c_k\}:\sum_k c_k^2 = 1} C \sum_{k=1}^\infty c_k^2 = C \quad (C.30)$$

which implies $J_{P,n,\jmath}$ is continuous. Next, define $J_{P,n,\jmath}^* : L^2_P(Z_{i,\jmath}) \to \ell^2(\mathbb{N})$ to be given by

$$J_{P,n,\jmath}^* g = \{a_k(g)\}_{k=1}^\infty \qquad a_k(g) \equiv \begin{cases} \langle g, q_{k,n,\jmath} \rangle_{L^2_P(Z_{i,\jmath})} & \text{if } k \leq k_{n,\jmath} \\ 0 & \text{if } k > k_{n,\jmath} \end{cases} , \qquad (C.31)$$

and note $J_{P,n,\jmath}^*$ is the adjoint of $J_{P,n,\jmath}$. Therefore, since $\|J_{P,n,\jmath}\|_o = \|J_{P,n,\jmath}^*\|_o$ by Theorem 6.5.1 in Luenberger (1969), we obtain for any $P \in \mathbf{P}$, $n \geq 1$, and $g \in L^2_P(Z_{i,\jmath})$

$$\sum_{k=1}^{k_{n,\jmath}} \langle g, q_{k,n,\jmath} \rangle_{L^2_P(Z_{i,\jmath})}^2 = \|J_{P,n,\jmath}^* g\|_{\ell^2(\mathbb{N})}^2 \leq \|J_{P,n,\jmath}^*\|_o^2 \|g\|_{L^2_P(Z_{i,\jmath})}^2 = \|J_{P,n,\jmath}\|_o^2 \|g\|_{L^2_P(Z_{i,\jmath})}^2 .$$

$$(C.32)$$

Therefore, since $E_P[f(V_i)q_{k,n,\jmath}(Z_{i,\jmath})] = E_P[E_P[f(V_i)|Z_{i,\jmath}]q_{k,n,\jmath}(Z_{i,\jmath})]$ for any $f \in L^2_P$, setting $g(Z_{i,\jmath}) = E_P[f(V_i)|Z_{i,\jmath}]$ in (C.32) and exploiting (C.30) yields the Lemma. ∎

APPENDIX D - Proofs for Section 6

PROOF OF THEOREM 6.1: First note that Lemma D.1 implies that uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R) = \inf_{\theta \in \hat{\Theta}_n \cap R} \inf_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, \ell_n)} \|\mathbb{W}_{n,P}^{\star}\rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P), r} + o_p(a_n) . \quad \text{(D.1)}$$

Thus, we may select $\hat{\theta}_n \in \hat{\Theta}_n \cap R$ and $\hat{h}_n/\sqrt{n} \in \hat{V}_n(\hat{\theta}_n, \ell_n)$ so that uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R) = \|\mathbb{W}_{n,P}^{\star}\rho(\cdot, \hat{\theta}_n) * q_n^{k_n} + \mathbb{D}_{n,P}(\hat{\theta}_n)[\hat{h}_n]\|_{\Sigma_n(P), r} + o_p(a_n) . \quad \text{(D.2)}$$

To proceed, note that by Assumptions 5.3(i) and 6.6(ii)-(iv) we may select a δ_n so that $\delta_n \mathcal{S}_n(\mathbf{B}, \mathbf{E}) = o(r_n)$, $1\{K_f \vee K_g > 0\}\delta_n \mathcal{S}_n(\mathbf{B}, \mathbf{E}) = o(1)$, $\mathcal{R}_n + \nu_n\tau_n = o(\delta_n)$, and

$$\ell_n \delta_n \times \mathcal{S}_n(\mathbf{B}, \mathbf{E})1\{K_f > 0\} = o(a_n n^{-\frac{1}{2}}) \quad \text{(D.3)}$$

$$K_m \delta_n \ell_n \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}}) \quad \text{(D.4)}$$

$$k_n^{1/r}\sqrt{\log(k_n)}B_n \times \sup_{P \in \mathbf{P}} J_{[]}(\delta_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n) . \quad \text{(D.5)}$$

Next, notice that Theorem 4.1 implies that there exist $\theta_{0n} \in \Theta_{0n}(P) \cap R$ such that

$$\|\hat{\theta}_n - \theta_{0n}\|_{\mathbf{E}} = O_p(\mathcal{R}_n + \nu_n\tau_n) \quad \text{(D.6)}$$

uniformly in $P \in \mathbf{P}_0$. Further note that since $\|q_{k,n,\jmath}\|_{L_P^\infty} \leq B_n$ for all $1 \leq k \leq k_{n,\jmath}$ by Assumption 3.2(i), we obtain from Assumption 5.2(i), result (D.6) and $\mathcal{R}_n + \nu_n\tau_n = o(\delta_n)$ that with probability tending to one uniformly in $P \in \mathbf{P}_0$ we have

$$E_P[\|\rho(X_i, \hat{\theta}_n) - \rho(X_i, \theta_{0n})\|_2^2 q_{k,n,\jmath}^2(Z_{i,\jmath})] \leq B_n^2 K_\rho^2 \delta_n^{2\kappa_\rho} . \quad \text{(D.7)}$$

Hence, letting $\mathcal{G}_n \equiv \{f(x)q_{k,n,\jmath}(z_\jmath) : f \in \mathcal{F}_n, \ 1 \leq \jmath \leq \mathcal{J}, \text{ and } 1 \leq k \leq k_{n,\jmath}\}$, we obtain from $\|\Sigma_n(P)\|_{o,r}$ being uniformly bounded by Assumption 3.4(iii), results (C.18)-(C.20), Markov's inequality, and δ_n satisfying (D.5) that uniformly in $P \in \mathbf{P}$

$$\|\mathbb{W}_{n,P}^{\star}\rho(\cdot, \hat{\theta}_n) * q_n^{k_n} - \mathbb{W}_{n,P}^{\star}\rho(\cdot, \theta_{0n}) * q_n^{k_n}\|_{\Sigma_n(P), r}$$
$$\leq \|\Sigma_n(P)\|_{o,r}\mathcal{J}k_n^{1/r} \sup_{g_1, g_2 \in \mathcal{G}_n : \|g_1 - g_2\|_{L_P^2} \leq B_n K_\rho \delta_n^{\kappa_\rho}} |\mathbb{W}_{n,P}^{\star}g_1 - \mathbb{W}_{n,P}^{\star}g_2| = o_p(a_n) . \quad \text{(D.8)}$$

Similarly, since $\hat{\theta}_n \in (\Theta_{0n}(P) \cap R)^\epsilon$ with probability tending to one uniformly in $P \in \mathbf{P}_0$ by Lemma 4.1, we can exploit Lemma D.3 to obtain for some $C < \infty$ that

$$\|\mathbb{D}_{n,P}(\theta_{0n})[\hat{h}_n] - \mathbb{D}_{n,P}(\hat{\theta}_n)[\hat{h}_n]\|_{\Sigma_n(P), r} \leq \|\Sigma_n(P)\|_{o,r} \times CK_m\|\hat{\theta}_n - \theta_{0n}\|_{\mathbf{L}}\|\hat{h}\|_{\mathbf{E}} + o_p(a_n)$$
$$\leq \|\Sigma_n(P)\|_{o,r} \times CK_m K_b \mathcal{S}_n(\mathbf{L}, \mathbf{E})\delta_n \ell_n\sqrt{n} + o_p(a_n) = o_p(a_n) \quad \text{(D.9)}$$

59

where the second inequality follows from $\|\hat{h}_n/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ due to $\hat{h}_n/\sqrt{n} \in \hat{V}_n(\hat{\theta}_n, \ell_n)$, $\|\hat{h}_n\|_{\mathbf{E}} \leq K_b\|\hat{h}_n\|_{\mathbf{B}}$ by Assumption 6.1(i), and $\mathcal{R}_n + \nu_n\tau_n = o(\delta_n)$. In turn, the final result in (D.9) is implied by (D.4) and $\|\Sigma_n(P)\|_{o,r}$ being uniformly bounded due to Assumption 3.4(iii). Next, we note that (D.6) and $\mathcal{R}_n + \nu_n\tau_n = o(\delta_n)$ imply

$$\|\hat{\theta}_n - \theta_{0n}\|_{\mathbf{B}} = o_p(\delta_n \times \mathcal{S}_n(\mathbf{B}, \mathbf{E})) \tag{D.10}$$

uniformly in $P \in \mathbf{P}_0$. Thus, since $\delta_n\mathcal{S}_n(\mathbf{B}, \mathbf{E})1\{K_f \vee K_g > 0\} = o(1)$, $\delta_n\mathcal{S}_n(\mathbf{B}, \mathbf{E}) = o(r_n)$, and $\limsup_{n\to\infty} \ell_n/r_n 1\{K_g > 0\} < 1/2$ by Assumption 6.6(iii), we obtain

$$r_n \geq (M_g\delta_n\mathcal{S}_n(\mathbf{B}, \mathbf{E}) + K_g\delta_n^2\mathcal{S}_n^2(\mathbf{B}, \mathbf{E})) \vee 2(\ell_n + \delta_n\mathcal{S}_n(\mathbf{B}, \mathbf{E}))1\{K_g > 0\} \tag{D.11}$$

for n sufficiently large. Hence, applying Theorem E.1 and exploiting Assumption 6.1(ii), and $\|h\|_{\mathbf{E}} \leq K_b\|h\|_{\mathbf{B}}$ for all $h \in \mathbf{B}_n$ and $P \in \mathbf{P}$ by Assumption 6.1(i), we obtain that there is an $M < \infty$ for which with probability tending to one uniformly in $P \in \mathbf{P}_0$

$$\inf_{\frac{h_0}{\sqrt{n}} \in V_n(\theta_{0n}, 2K_b\ell_n)} \|\frac{\hat{h}_n}{\sqrt{n}} - \frac{h_0}{\sqrt{n}}\|_{\mathbf{B}} \leq M\ell_n(\ell_n + \delta_n\mathcal{S}_n(\mathbf{B}, \mathbf{E}))1\{K_f > 0\} . \tag{D.12}$$

In particular, it follows from Assumption 6.6(ii) and (D.3) that we may find a $h_{0n}/\sqrt{n} \in V_n(\theta_{0n}, 2K_b\ell_n)$ such that $\|h_{0n} - \hat{h}_n\|_{\mathbf{B}} = o_p(a_n)$ uniformly in $P \in \mathbf{P}_0$, and hence Assumption 3.4(iii), Lemma D.3, and $\|h\|_{\mathbf{E}} \leq K_b\|h\|_{\mathbf{B}}$ by Assumption 6.1(i) yield

$$\|\mathbb{D}_{n,P}(\theta_{0n})[\hat{h}_n] - \mathbb{D}_{n,P}(\theta_{0n})[h_{0n}]\|_{\Sigma_n(P),r} \leq \|\Sigma_n(P)\|_{o,r} \times CM_m\|\hat{h} - h_{0n}\|_{\mathbf{E}} = o_p(a_n) \tag{D.13}$$

uniformly in $P \in \mathbf{P}_0$. Therefore, combining results (D.2), (D.8), (D.9), and (D.13) together with $\theta_{0n} \in \Theta_{0n}(P) \cap R$ and $h_{0n}/\sqrt{n} \in V_n(\theta_{0n}, 2K_b\ell_n)$ imply

$$\hat{U}_n(R) = \|\mathbb{W}_{n,P}^\star\rho(\cdot, \theta_{0n}) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n})[h_{0n}]\|_{\Sigma_n(P),r} + o_p(a_n)$$
$$\geq \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, 2K_b\ell_n)} \|\mathbb{W}_{n,P}^\star\rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r} + o_p(a_n) \tag{D.14}$$

uniformly in $P \in \mathbf{P}_0$, thus establishing the claim of the Theorem. $\blacksquare$

PROOF OF THEOREM 6.2: First set $\mathbf{G} \equiv \mathbf{R}$, $\Upsilon_G(\theta) = -1$ for all $\theta \in \mathbf{B}$, and note that

$$R = \{\theta \in \mathbf{B} : \Upsilon_F(\theta) = 0\} = \{\theta \in \mathbf{B} : \Upsilon_F(\theta) = 0 \text{ and } \Upsilon_G(\theta) \leq 0\} . \tag{D.15}$$

Moreover, also note Assumption 6.2 is automatically satisfied with $K_g = M_g = 0$ and $\nabla\Upsilon_G(\theta)[h] = 0$ for all $\theta, h \in \mathbf{B}$, while Assumption 6.4 holds with $h_0 = 0$ and $\epsilon = -1$.

Similarly, since $K_g = 0$, definition (68) implies $G_n(\theta) = \mathbf{B}_n$ for all $\theta \in \mathbf{B}$, and hence

$$\{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \frac{h}{\sqrt{n}} \in G_n(\theta), \ \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\}$$
$$= \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\} \ . \quad \text{(D.16)}$$

Furthermore, since (D.16) holds for any r_n, we may set r_n so Assumption 6.6(iii) holds. Thus, it follows that we may apply Theorem 6.1 to obtain uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R) \geq \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, 2K_b \ell_n)} \|\mathbb{W}_{n,P}^{\star} \rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r} + o_p(a_n) \ . \quad \text{(D.17)}$$

Next, note that since $(k_n^{1/r} \sqrt{\log(k_n)} J_n B_n / \sqrt{n} + \zeta_n) = o(\tau_n)$ by hypothesis, we have

$$\liminf_{n \to \infty} \inf_{P \in \mathbf{P}_0} P(\Theta_{0n}(P) \cap R \subseteq \hat{\Theta}_n \cap R) = 1 \quad \text{(D.18)}$$

by Lemma 4.1(ii). For notational simplicity define $\eta_n^{-1} \equiv \mathcal{S}_n(\mathbf{B}, \mathbf{E})$, and then note that $\|h\|_{\mathbf{B}} \leq \ell_n$ for any $h \in \mathbf{B}_n$ satisfying $\|h\|_{\mathbf{E}} \leq \eta_n \ell_n$. Thus, we obtain by definitions of $V_n(\theta, \ell)$ and $\hat{V}_n(\theta, \ell)$ that for any $P \in \mathbf{P}$ and $\theta \in \Theta_n \cap R$ we have

$$V_n(\theta, \eta_n \ell_n) = \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \theta + \frac{h}{\sqrt{n}} \in \Theta_n \cap R \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{E}} \leq \eta_n \ell_n\}$$
$$\subseteq \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\} = \hat{V}_n(\theta, \ell_n) \ . \quad \text{(D.19)}$$

Therefore, Lemma D.1 and results (D.18) and (D.19) imply that uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R) \leq \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, \eta_n \ell_n)} \|\mathbb{W}_{n,P}^{\star} \rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r} + o_p(a_n) \ . \quad \text{(D.20)}$$

Furthermore, we also note that the definition of $\mathcal{S}_n(\mathbf{B}, \mathbf{E})$ and Assumption 6.1(i) yield

$$\|h\|_{\mathbf{B}} \leq \mathcal{S}_n(\mathbf{B}, \mathbf{E}) \times \|h\|_{\mathbf{E}} \leq \mathcal{S}_n(\mathbf{B}, \mathbf{E}) \times K_b \|h\|_{\mathbf{B}} \ , \quad \text{(D.21)}$$

for any $h \in \mathbf{B}_n$, which implies $\mathcal{S}_n(\mathbf{B}, \mathbf{E}) \geq 1/K_b$, and thus $\eta_n = O(1)$. Hence, since $\mathcal{R}_n \mathcal{S}_n(\mathbf{B}, \mathbf{E}) = o(\ell_n)$, we have $\mathcal{R}_n = o(\ell_n \eta_n \wedge \ell_n)$. Similarly, Assumption 6.6(ii) implies $k_n^{1/r} \sqrt{\log(k_n)} B_n \sup_{P \in \mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho} \vee (\ell_n \eta_n)^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$ and $K_m(\ell_n^2 \vee \ell_n^2 \eta_n^2) \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$. Thus, applying Lemma D.4 with $\ell_n = \ell_n$ and $\tilde{\ell}_n = \ell_n \eta_n$ yields

$$\inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, \eta_n \ell_n)} \|\mathbb{W}_{n,P}^{\star} \rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r}$$
$$= \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, 2K_b \ell_n)} \|\mathbb{W}_{n,P}^{\star} \rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r} + o_p(a_n) \quad \text{(D.22)}$$

uniformly in $P \in \mathbf{P}_0$. Hence, results (D.17), (D.20), and (D.22) allow us to conclude

$$\hat{U}_n(R) = \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, 2K_b \ell_n)} \|\mathbb{W}_{n,P}^{\star} \rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r} + o_p(a_n)$$

(D.23)

uniformly in $P \in \mathbf{P}_0$, which establishes the first claim of the Theorem.

In order to establish the second claim of the Theorem, we first define the set

$$N_n(\theta, \ell) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \nabla \Upsilon_F(\theta)[h] = 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \le \ell\} .$$

(D.24)

Next, note that since $\Theta_{0n}(P) \cap R = \{\theta_{0n}(P)\}$, Theorem 4.1 yields uniformly in $P \in \mathbf{P}_0$

$$d_H(\hat{\Theta}_n \cap R, \theta_{0n}(P), \|\cdot\|_{\mathbf{E}}) = \vec{d}_H(\hat{\Theta}_n \cap R, \theta_{0n}(P), \|\cdot\|_{\mathbf{E}}) = O_p(\mathcal{R}_n + \nu_n \tau_n) . \quad (D.25)$$

Furthermore, since $(\mathcal{R}_n + \nu_n \tau_n) \times \mathcal{S}_n(\mathbf{B}, \mathbf{E}) = o(\ell_n)$, Assumptions 5.3(i) and 6.6(ii) imply there is a $\delta_n \downarrow 0$ satisfying (D.3)-(D.5), $\mathcal{R}_n + \nu_n \tau_n = o(\delta_n)$, and

$$\delta_n \times \mathcal{S}_n(\mathbf{B}, \mathbf{E}) = o(\ell_n) .$$

(D.26)

Moreover, identical arguments to those employed in (D.8), (D.9), and (D.10) yield

$$\sup_{\theta \in \hat{\Theta}_n \cap R} \|\mathbb{W}_{n,P}^{\star} \rho(\cdot, \theta) * q_n^{k_n} - \mathbb{W}_{n,P}^{\star} \rho(\cdot, \theta_{0n}(P)) * q_n^{k_n}\|_{\Sigma_n(P),r} = o_p(a_n)$$

(D.27)

$$\sup_{\theta \in \hat{\Theta}_n \cap R} \sup_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, \ell_n)} \|\mathbb{D}_{n,P}(\theta)[h] - \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P),r} = o_p(a_n)$$

(D.28)

$$\sup_{\theta \in \hat{\Theta}_n \cap R} \|\theta - \theta_{0n}(P)\|_{\mathbf{B}} \times \{\mathcal{S}_n(\mathbf{B}, \mathbf{E})\}^{-1} = o_p(\delta_n)$$

(D.29)

uniformly in $P \in \mathbf{P}_0$. Therefore, we can conclude from Lemma D.1 and results (D.27) and (D.28) that we may select a $\hat{\theta}_n \in \hat{\Theta}_n \cap R$ so that uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R) = \inf_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\hat{\theta}_n, \ell_n)} \|\mathbb{W}_{n,P}^{\star} \rho(\cdot, \hat{\theta}_n) * q_n^{k_n} + \mathbb{D}_{n,P}(\hat{\theta}_n)[h]\|_{\Sigma_n(P),r} + o_p(a_n)$$

$$= \inf_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\hat{\theta}_n, \ell_n)} \|\mathbb{W}_{n,P}^{\star} \rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P),r} + o_p(a_n) . \quad (D.30)$$

We next proceed by establishing upper and lower bounds for the right hand side of (D.30). To this end, note that by (D.30), we may select a $\hat{h}_{ln}/\sqrt{n} \in \hat{V}_n(\hat{\theta}_n, \ell_n)$ so that

$$\hat{U}_n(R) = \|\mathbb{W}_{n,P}^{\star} \rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[\hat{h}_{ln}]\|_{\Sigma_n(P),r} + o_p(a_n)$$

(D.31)

uniformly in $P \in \mathbf{P}_0$. Also observe that the final equality in (D.19), result (D.29), and Lemma E.1 (see (E.29)) imply that there exist $M < \infty$ and $\tilde{h}_{ln}/\sqrt{n}$ in $N_n(\theta_{0n}(P), 2\ell_n)$

such that with probability tending to one uniformly in $P \in \mathbf{P}_0$ we have

$$\|\frac{\hat{h}_{ln}}{\sqrt{n}} - \frac{\tilde{h}_{ln}}{\sqrt{n}}\|_{\mathbf{B}} \leq M \times \ell_n(\ell_n + \delta_n \times \mathcal{S}_n(\mathbf{B}, \mathbf{E}))1\{K_f > 0\} . \qquad \text{(D.32)}$$

Thus, we obtain from (D.32), $\|\hat{h}_{ln} - \tilde{h}_{ln}\|_{\mathbf{E}} \leq K_b \|\hat{h}_{ln} - \tilde{h}_{ln}\|_{\mathbf{B}}$, $\|\Sigma_n(P)\|_{o,r}$ being uniformly bounded by Assumption 3.4(iii), and Lemma D.3 that uniformly in $P \in \mathbf{P}_0$

$$\|\mathbb{D}_{n,P}(\theta_{0n}(P))[\hat{h}_{ln}] - \mathbb{D}_{n,P}(\theta_{0n}(P))[\tilde{h}_{ln}]\|_{\Sigma_n(P),r} \leq \|\Sigma_n(P)\|_{o,r} \times CM_m \|\hat{h}_{ln} - \tilde{h}_{ln}\|_{\mathbf{E}}$$
$$\lesssim \ell_n(\ell_n + \delta_n \times \mathcal{S}_n(\mathbf{B}, \mathbf{E}))\sqrt{n}1\{K_f > 0\} + o_p(a_n) = o_p(a_n) , \quad \text{(D.33)}$$

where the final equality follows by (D.26) and $\ell_n^2 \sqrt{n}1\{K_f > 0\} = o(a_n)$ by Assumption 6.6(ii). Hence, (D.31), (D.33), and $\tilde{h}_{ln}/\sqrt{n} \in N_n(\theta_{0n}(P), 2\ell_n)$ yield uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R) = \|\mathbb{W}_{n,P}^{\star}\rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[\tilde{h}_{ln}]\|_{\Sigma_n(P),r} + o_p(a_n)$$
$$\geq \inf_{\frac{h}{\sqrt{n}} \in N_n(\theta_{0n}(P), 2\ell_n)} \|\mathbb{W}_{n,P}^{\star}\rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P),r} + o_p(a_n) .$$
$$\text{(D.34)}$$

To establish an upper bound for (D.30), let $\hat{h}_{un}/\sqrt{n} \in N_n(\theta_{0n}(P), \ell_n/2)$ satisfy

$$\inf_{\frac{h}{\sqrt{n}} \in N_n(\theta_{0n}(P), \frac{\ell_n}{2})} \|\mathbb{W}_{n,P}^{\star}\rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P),r}$$
$$= \|\mathbb{W}_{n,P}^{\star}\rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[\hat{h}_{un}]\|_{\Sigma_n(P),r} + o_p(a_n) \quad \text{(D.35)}$$

uniformly in $P \in \mathbf{P}_0$. Next note that by Lemma E.1 (see (E.30)), the final equality in (D.19), and (D.29) we may pick a $\tilde{h}_{un}/\sqrt{n} \in \hat{V}_n(\hat{\theta}_n, \ell_n)$ such that for some $M < \infty$

$$\|\frac{\hat{h}_{un}}{\sqrt{n}} - \frac{\tilde{h}_{un}}{\sqrt{n}}\|_{\mathbf{B}} \leq M \times \ell_n(\ell_n + \delta_n \times \mathcal{S}_n(\mathbf{B}, \mathbf{E}))1\{K_f > 0\} \qquad \text{(D.36)}$$

with probability tending to one uniformly in $P \in \mathbf{P}_0$. Therefore, exploiting Lemma D.3, $\|\Sigma_n(P)\|_{o,r}$ being uniformly bounded by Assumption 3.4(iii), result (D.36), and $\|\hat{h}_{un} - \tilde{h}_{un}\|_{\mathbf{E}} \leq K_b \|\hat{h}_{un} - \tilde{h}_{un}\|_{\mathbf{B}}$ implies that uniformly in $P \in \mathbf{P}_0$

$$\|\mathbb{D}_{n,P}(\theta_{0n}(P))[\hat{h}_{un}] - \mathbb{D}_{n,P}(\theta_{0n}(P))[\tilde{h}_{un}]\|_{\Sigma_n(P),r} \leq \|\Sigma_n(P)\|_{o,r} \times CM_m \|\hat{h}_{un} - \tilde{h}_{un}\|_{\mathbf{E}}$$
$$\lesssim \ell_n(\ell_n + \delta_n \times \mathcal{S}_n(\mathbf{B}, \mathbf{E}))\sqrt{n}1\{K_f > 0\} + o_p(a_n) = o_p(a_n) , \quad \text{(D.37)}$$

where in the final equality we exploited $\ell_n^2 1\{K_f > 0\} = o(a_n n^{-1/2})$ by Assumption

6.6(ii) and δ_n satisfying (D.26). Hence, we conclude uniformly in $P \in \mathbf{P}_0$ that

$$
\begin{aligned}
\hat{U}_n(R) &\leq \|\mathbb{W}^\star_{n,P}\rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[\tilde{h}_{un}]\|_{\Sigma_n(P),r} + o_p(a_n) \\
&= \inf_{\frac{h}{\sqrt{n}} \in N_n(\theta_{0n}(P), \frac{\ell_n}{2})} \|\mathbb{W}^\star_{n,P}\rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P),r} + o_p(a_n) ,
\end{aligned}
\tag{D.38}
$$

where the inequality follows from (D.30) and $\tilde{h}_{un}/\sqrt{n} \in \hat{V}_n(\hat{\theta}_n, \ell_n)$, while the equality is implied by results (D.35) and (D.37).

Finally, we obtain from results (D.34) and (D.38) together with Lemma D.5(i) that

$$
\hat{U}_n(R) = \inf_{h \in \mathbf{B}_n \cap \mathcal{N}(\nabla \Upsilon_F(\theta_{0n}(P)))} \|\mathbb{W}^\star_{n,P}\rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P),r} + o_p(a_n)
\tag{D.39}
$$

uniformly in $P \in \mathbf{P}_0$. Setting $\mathbb{V}_{n,P} \equiv \{v = \Sigma_n(P)\mathbb{D}_{n,P}(\theta_{0n}(P))[h] \text{ for some } h \in \mathbf{B}_n \cap \mathcal{N}(\nabla \Upsilon_F(\theta_{0n}(P)))\}$, then note that $\mathbb{V}_{n,P}$ is a vector subspace of $\mathbf{R}^{k_n}$ by linearity of $\mathbb{D}_{n,P}(\theta_{0n}(P))$ and its dimension for n sufficiently large is equal to $c_n \equiv \dim\{\mathbf{B}_n \cap \mathcal{N}(\nabla \Upsilon_F(\theta_{0n}(P)))\}$ by Lemma D.5(ii) and $\Sigma_n(P)$ being full rank by Assumption 3.4(iii). Letting $\mathbb{Z}_n \in \mathbf{R}^{k_n}$ denote a standard normal random variable, we then obtain from $r = 2$, $\Sigma_n(P) = \{\text{Var}_P\{\rho(X_i, \theta_{0n}(P))q_n^{k_n}(Z_i)\}\}^{-\frac{1}{2}}$, and (D.39) that uniformly in $P \in \mathbf{P}_0$

$$
\hat{U}_n(R) = \inf_{v \in \mathbb{V}_{n,P}} \|\mathbb{Z}_n - v\|_2 + o_p(a_n) = \{\mathcal{X}^2_{k_n - c_n}\}^{\frac{1}{2}} + o_p(a_n) ,
\tag{D.40}
$$

where the final equality follows by observing that the projection of $\mathbb{Z}_n$ onto $\mathbb{V}_{n,P}$ can be written as $\Pi_P \mathbb{Z}_n$ for some $k_n \times k_n$ idempotent matrix Π_P of rank c_n. ∎

PROOF OF LEMMA 6.1: First, let $\hat{\theta}_n \in \hat{\Theta}_n \cap R$ and $\hat{h}_n \in \hat{V}_n(\hat{\theta}_n, +\infty)$ be such that

$$
\inf_{\theta \in \hat{\Theta}_n \cap R} \inf_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, +\infty)} \|\hat{\mathbb{W}}_n \rho(\cdot, \theta) * q_n^{k_n} + \hat{\mathbb{D}}_n(\theta)[h]\|_{\hat{\Sigma}_n, r}
$$
$$
= \|\hat{\mathbb{W}}_n \rho(\cdot, \hat{\theta}_n) * q_n^{k_n} + \hat{\mathbb{D}}_n(\hat{\theta}_n)[\hat{h}_n]\|_{\hat{\Sigma}_n, r} + o(a_n) . \tag{D.41}
$$

Then note that in order to establish the claim of the Lemma it suffices to show that

$$
\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(\|\frac{\hat{h}_n}{\sqrt{n}}\|_{\mathbf{B}} \geq \ell_n) = 0 .
\tag{D.42}
$$

To this end, observe that since $0 \in \hat{V}_n(\theta, +\infty)$ for all $\theta \in \Theta_n \cap R$, we obtain from the

triangle inequality, $\|\hat{\Sigma}_n\|_{o,r} = O_p(1)$ by Lemma B.3 and Assumption 6.5 that

$$\|\hat{\mathbb{D}}_n(\hat{\theta}_n)[\hat{h}_n]\|_{\hat{\Sigma}_n,r} \le \|\hat{\mathbb{W}}_n\rho(\cdot,\hat{\theta}_n) * q_n^{k_n} + \hat{\mathbb{D}}_n(\hat{\theta}_n)[\hat{h}_n]\|_{\hat{\Sigma}_n,r} + \|\hat{\mathbb{W}}_n\rho(\cdot,\hat{\theta}_n) * q_n^{k_n}\|_{\hat{\Sigma}_n,r}$$
$$\le 2\|\hat{\Sigma}_n\|_{o,r}\|\hat{\mathbb{W}}_n\rho(\cdot,\hat{\theta}_n) * q_n^{k_n}\|_r + o(a_n) \le 2\|\hat{\Sigma}_n\|_{o,r}\|\mathbb{W}_{n,P}^\star\rho(\cdot,\hat{\theta}_n) * q_n^{k_n}\|_r + o_p(a_n)$$
(D.43)

uniformly in $P \in \mathbf{P}$. Hence, since $\hat{\Theta}_n \cap R \subseteq \Theta_n \cap R$ almost surely, we obtain from result (D.43), $\|\hat{\Sigma}_n\|_{o,r} = O_p(1)$, and Lemma C.4 together with Markov's inequality

$$\|\hat{\mathbb{D}}_n(\hat{\theta}_n)[\hat{h}_n]\|_{\hat{\Sigma}_n,r}$$
$$\le 2\|\hat{\Sigma}_n\|_{o,r} \times \sup_{\theta\in\Theta_n\cap R}\|\mathbb{W}_{n,P}^\star\rho(\cdot,\theta) * q_n^{k_n}\|_r + o_p(a_n) = O_p(k_n^{1/r}\sqrt{\log(k_n)}B_nJ_n) \quad \text{(D.44)}$$

uniformly in $P \in \mathbf{P}$. Moreover, note that since $\hat{\Theta}_n \cap R \subseteq (\Theta_{0n}(P) \cap R)^\epsilon$ with probability tending to one uniformly in $\mathbf{P}_0$ by Lemma 4.1, and $\hat{h}_n/\sqrt{n} \in \hat{V}_n(\hat{\theta}_n, +\infty)$ implies $\hat{h}_n \in \sqrt{n}\{\mathbf{B}_n \cap R - \hat{\theta}_n\}$ we obtain from the first hypothesis of the Lemma that

$$\limsup_{n\to\infty}\sup_{P\in\mathbf{P}_0} P(\ell_n \le \|\frac{\hat{h}_n}{\sqrt{n}}\|_\mathbf{B})$$
$$= \limsup_{n\to\infty}\sup_{P\in\mathbf{P}_0} P(\ell_n \le \|\frac{\hat{h}_n}{\sqrt{n}}\|_\mathbf{B} \text{ and } \|\hat{h}_n\|_\mathbf{E} \le \nu_n\|\mathbb{D}_{n,P}(\hat{\theta}_n)[\hat{h}_n]\|_r)$$
$$\le \limsup_{n\to\infty}\sup_{P\in\mathbf{P}_0} P(\ell_n \le \|\frac{\hat{h}_n}{\sqrt{n}}\|_\mathbf{B} \text{ and } \|\hat{h}_n\|_\mathbf{E} \le 2\nu_n\|\hat{\mathbb{D}}_n(\hat{\theta}_n)[\hat{h}_n]\|_r) , \quad \text{(D.45)}$$

where the inequality follows from (74). Hence, results (D.44) and (D.45), the definitions of $\mathcal{S}_n(\mathbf{B},\mathbf{E})$ and $\mathcal{R}_n$, and $\mathcal{S}_n(\mathbf{B},\mathbf{E})\mathcal{R}_n = o(\ell_n)$ by hypothesis yield

$$\limsup_{n\to\infty}\sup_{P\in\mathbf{P}_0} P(\ell_n \le \|\frac{\hat{h}_n}{\sqrt{n}}\|_\mathbf{B})$$
$$\le \limsup_{n\to\infty}\sup_{P\in\mathbf{P}_0} P(\ell_n \le 2\frac{\nu_n}{\sqrt{n}}\mathcal{S}_n(\mathbf{B},\mathbf{E})\|\hat{\mathbb{D}}_n(\hat{\theta}_n)[\hat{h}_n]\|_r) = 0 , \quad \text{(D.46)}$$

which establishes (D.42) and hence the claim of the Lemma. ∎

PROOF OF THEOREM 6.3: We establish the first claim by appealing to the first claim of Lemma D.6. To this end, we note condition (D.97) holds by Assumption 6.7 and define

$$U_{n,P}^\star(R) \equiv \inf_{\theta\in\Theta_{0n}(P)\cap R} \inf_{\frac{h}{\sqrt{n}}\in V_n(\theta,2K_b\ell_n)} \|\mathbb{W}_{n,P}^\star\rho(\cdot,\theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r} , \quad \text{(D.47)}$$

which is independent of $\{V_i\}_{i=1}^n$ by Assumption 6.5. Moreover, Theorem 6.1 yields

$$\hat{U}_n(R) \ge U_{n,P}^\star(R) + o_p(a_n) \quad \text{(D.48)}$$

uniformly in $P \in \mathbf{P}_0$, while Assumption 6.6(ii) implies $K_m(2K_b\ell_n)^2 \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$ and $k_n^{1/r}\sqrt{\log(k_n)}B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}((2K_b\ell_n)^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$, and hence

$$I_n(R) \leq \inf_{\theta\in\Theta_{0n}(P)\cap R} \inf_{\frac{h}{\sqrt{n}}\in V_n(\theta, 2K_b\ell_n)} \|\mathbb{W}_{n,P}\rho(\cdot,\theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r} + o_p(a_n) \quad \text{(D.49)}$$

uniformly in $P \in \mathbf{P}_0$ by Theorem 5.1(i). Since the right hand side of (D.49) shares the same distribution as $U_{n,P}^\star(R)$, the first claim of the Theorem holds by Lemma D.6(i).

For the second claim of the Theorem we first note that Theorem 6.2(i) yields that

$$\hat{U}_n(R) = U_{n,P}^\star(R) + o_p(a_n) \quad \text{(D.50)}$$

uniformly in $P \in \mathbf{P}_0$. Furthermore, as already argued $2K_b\ell_n$ satisfies the conditions of Theorem 5.1(i) by Assumption 6.6(ii) while $\mathcal{R}_n = o(\ell_n)$ in addition implies that

$$I_n(R) = \inf_{\theta\in\Theta_{0n}(P)\cap R} \inf_{\frac{h}{\sqrt{n}}\in V_n(\theta, 2K_b\ell_n)} \|\mathbb{W}_{n,P}\rho(\cdot,\theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r} + o_p(a_n) \quad \text{(D.51)}$$

uniformly in $P \in \mathbf{P}_0$ (see (C.12) and subsequent discussion). Hence, since the right hand side of (D.51) shares the same distribution as $U_{n,P}^\star(R)$ and condition (D.97) of Lemma D.6 holds by Assumption 6.7, the second claim of the Theorem follows from results (D.50) and (D.51), and Lemma D.6(ii).

In order to establish the final claim of the Theorem, we next note that since $\hat{\Theta}_n \cap R \subseteq \Theta_n \cap R$ and $0 \in \hat{V}_n(\theta, \ell_n)$ for all $\theta \in \Theta_n \cap R$ it follows from Assumption 6.5 and $\|\hat{\Sigma}_n\|_{o,r} = O_p(1)$ uniformly in $P \in \mathbf{P}$ by Lemma B.3 that we must have

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} P(\hat{U}_n(R) > Mk_n^{1/r}\sqrt{\log(k_n)}B_n J_n)$$
$$\leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}} P(\sup_{\theta\in\Theta_n\cap R} \|\hat{\mathbb{W}}_{n,P}\rho(\cdot,\theta) * q_n^{k_n}\|_{\hat{\Sigma}_n,r} > Mk_n^{1/r}\sqrt{\log(k_n)}B_n J_n)$$
$$= \limsup_{n\to\infty} \sup_{P\in\mathbf{P}} P(\sup_{\theta\in\Theta_n\cap R} \|\mathbb{W}_{n,P}^\star\rho(\cdot,\theta) * q_n^{k_n}\|_{\hat{\Sigma}_n,r} > Mk_n^{1/r}\sqrt{\log(k_n)}B_n J_n) . \quad \text{(D.52)}$$

Therefore, (D.52), Markov's inequality, and Lemmas B.3 and C.4 allow us to conclude

$$\limsup_{M\uparrow\infty} \limsup_{n\to\infty} \sup_{P\in\mathbf{P}} P(\hat{U}_n(R) > Mk_n^{1/r}\sqrt{\log(k_n)}B_n J_n) = 0 . \quad \text{(D.53)}$$

We thus obtain from the definition of $\hat{c}_{n,1-\alpha}$, result (D.53), and Markov's inequality

$$\limsup_{M\uparrow\infty} \limsup_{n\to\infty} \sup_{P\in\mathbf{P}} P(\hat{c}_{n,1-\alpha} > Mk_n^{1/r}\sqrt{\log(k_n)}B_n J_n)$$
$$= \limsup_{M\uparrow\infty} \limsup_{n\to\infty} \sup_{P\in\mathbf{P}} P(P(\hat{U}_n(R) > Mk_n\sqrt{\log(k_n)}B_n J_n|\{V_i\}_{i=1}^n) > \alpha) = 0 . \quad \text{(D.54)}$$

Next observe that $\|a\|_r \leq \|\hat{\Sigma}_n^{-1}\|_{o,r}\|a\|_{\hat{\Sigma}_n,r}$ for any $a \in \mathbf{R}^{k_n}$, and hence by Lemma B.2 we

obtain for some $\mathcal{Z}_{n,P} \in \mathbf{R}_+$ satisfying $\sup_{P \in \mathbf{P}} E_P[\mathcal{Z}_{n,P}] = O(k_n^{1/r} \sqrt{\log(k_n)} J_n B_n / \sqrt{n})$

$$I_n(R) \geq \sqrt{n} \|\hat{\Sigma}_n^{-1}\|_{o,r}^{-1} \times \inf_{\theta \in \Theta_n \cap R} \|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_r - \sqrt{n} \|\hat{\Sigma}_n\|_{o,r} \mathcal{Z}_{n,P} . \quad \text{(D.55)}$$

Moreover, assuming without loss of generality that $\pi_{n,P,\jmath}(\theta)' q_{n,\jmath}^{k_{n,\jmath}}$ is the $\|\cdot\|_{L_P^2}$ projection of $E_P[\rho_\jmath(X_i, \theta)|Z_{i,\jmath}]$ onto the span of $q_{n,\jmath}^{k_{n,\jmath}}(Z_{i,\jmath})$, we obtain by Lemma B.4 that

$$\inf_{\theta \in \Theta_n \cap R} \|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_r$$
$$\geq \mathcal{J}^{-\frac{1}{2}} k_n^{\frac{1}{r} - \frac{1}{2}} \inf_{\theta \in \Theta_n \cap R} \{\sum_{\jmath=1}^{\mathcal{J}} \|E_P[q_{n,\jmath}^{k_{n,\jmath}}(Z_{i,\jmath}) q_{n,\jmath}^{k_{n,\jmath}}(Z_{i,\jmath})'] \pi_{n,P,\jmath}(\theta)\|_2\} . \quad \text{(D.56)}$$

Thus, since the eigenvalues of $E_P[q_{n,\jmath}^{k_{n,\jmath}}(Z_{i,\jmath}) q_{n,\jmath}^{k_{n,\jmath}}(Z_{i,\jmath})']$ are uniformly bounded away from zero by Assumption 6.8(ii), we can conclude from result (D.56) that

$$\inf_{\theta \in \Theta_n \cap R} \|E_P[\rho(X_i, \theta) * q_n^{k_n}(Z_i)]\|_r \gtrsim k_n^{\frac{1}{r} - \frac{1}{2}} \inf_{\theta \in \Theta_n \cap R} \{\sum_{\jmath=1}^{\mathcal{J}} \|q_{n,\jmath}^{k_{n,\jmath}'} \pi_{n,P,\jmath}(\theta)\|_{L_P^2}\}$$
$$\geq k_n^{\frac{1}{r} - \frac{1}{2}} \inf_{\theta \in \Theta \cap R} \{\sum_{\jmath=1}^{\mathcal{J}} \|E_P[\rho_\jmath(X_i, \theta)|Z_{i,\jmath}]\|_{L_P^2}\} - k_n^{\frac{1}{r} - \frac{1}{2}} K_0 k_n^{-\gamma_z} \quad \text{(D.57)}$$

where the second inequality must hold for some $K_0 < \infty$ by Assumption 6.8(i) and $\Theta_n \cap R \subseteq \Theta \cap R$. Hence, results (D.55) and (D.57) imply that for $M > K_0$ and some $\epsilon_0 > 0$ it follows that for any $P \in \mathbf{P}_{1,n}(M)$ we must have

$$I_n(R) \geq \|\hat{\Sigma}_n^{-1}\|_{o,r}^{-1} \epsilon_0 M k_n^{1/r} \sqrt{\log(k_n)} J_n B_n - \sqrt{n} \|\hat{\Sigma}_n\|_{o,r} \mathcal{Z}_{n,P} . \quad \text{(D.58)}$$

Thus, since (D.58) holds for all $P \in \mathbf{P}_{1,n}(M)$ with $M > K_0$, $\mathbf{P}_{1,n}(M) \subseteq \mathbf{P}$ implies that

$$\inf_{P \in \mathbf{P}_{1,n}(M)} P(I_n(R) > \hat{c}_{n,1-\alpha})$$
$$\geq \inf_{P \in \mathbf{P}_{1,n}(M)} P(\epsilon_0 M \|\hat{\Sigma}_n^{-1}\|_{o,r}^{-1} k_n^{1/r} \sqrt{\log(k_n)} J_n B_n > \hat{c}_{n,1-\alpha} + \sqrt{n} \|\hat{\Sigma}_n\|_{o,r} \mathcal{Z}_{n,P})$$
$$\geq \inf_{P \in \mathbf{P}} P(\epsilon_0 M \|\hat{\Sigma}_n^{-1}\|_{o,r}^{-1} k_n^{1/r} \sqrt{\log(k_n)} J_n B_n > \hat{c}_{n,1-\alpha} + \sqrt{n} \|\hat{\Sigma}_n\|_{o,r} \mathcal{Z}_{n,P}) . \quad \text{(D.59)}$$

In particular, since: (i) $\max\{\|\hat{\Sigma}_n\|_{o,r}, \|\hat{\Sigma}_n^{-1}\|_{o,r}\} = O_p(1)$ uniformly in $P \in \mathbf{P}$ by Lemma B.3, (ii) $\mathcal{Z}_{n,P} = O_p(k_n^{1/r} \sqrt{\log(k_n)} J_n B_n / \sqrt{n})$ uniformly in $P \in \mathbf{P}$ by Markov's inequality and $\sup_{P \in \mathbf{P}} E_P[\mathcal{Z}_{n,P}] = O(k_n^{1/r} \sqrt{\log(k_n)} B_n J_n / \sqrt{n})$ by Lemma B.2, and (iii) $\hat{c}_{n,1-\alpha} = O_p(k_n^{1/r} \sqrt{\log(k_n)} J_n B_n)$ uniformly in $P \in \mathbf{P}$ by result (D.53), it follows from (D.59) that

$$\liminf_{M \uparrow \infty} \liminf_{n \to \infty} \inf_{P \in \mathbf{P}_{1,n}(M)} P(I_n(R) > \hat{c}_{n,1-\alpha}) = 1 , \quad \text{(D.60)}$$

which establishes the final claim of the Theorem. ∎

Lemma D.1. *Let Assumptions 3.1, 3.2(i)-(ii), 3.3, 3.4, 4.1, 5.1, 5.2(i), 5.3(iii), 5.4(i), 6.1, 6.5, and 6.6(i)-(ii) hold. It then follows that uniformly in $P \in \mathbf{P}_0$ we have*

$$\hat{U}_n(R) = \inf_{\theta \in \hat{\Theta}_n \cap R} \inf_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, \ell_n)} \|\mathbb{W}_{n,P}^\star \rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P), r} + o_p(a_n) . \quad \text{(D.61)}$$

PROOF: For an arbitrary $\epsilon > 0$, observe that Lemma 4.1 and Assumption 6.6(i) imply

$$\liminf_{n \to \infty} \inf_{P \in \mathbf{P}_0} P(\hat{\Theta}_n \cap R \subseteq (\Theta_{0n}(P) \cap R)^{\frac{\epsilon}{2}}) = 1 . \quad \text{(D.62)}$$

Furthermore, for any $\theta \in \hat{\Theta}_n \cap R$ and $h/\sqrt{n} \in \hat{V}_n(\theta, \ell_n)$ note that $\Upsilon_G(\theta + h/\sqrt{n}) \leq 0$ and $\Upsilon_F(\theta + h/\sqrt{n}) = 0$ by definition of $\hat{V}_n(\theta, \ell_n)$. Thus, $\theta + h/\sqrt{n} \in R$ for any $\theta \in \hat{\Theta}_n \cap R$ and $h/\sqrt{n} \in \hat{V}_n(\theta, \ell_n)$, and hence result (D.62) and Assumption 6.1(ii) yield

$$\liminf_{n \to \infty} \inf_{P \in \mathbf{P}_0} P(\theta + \frac{h}{\sqrt{n}} \in \Theta_n \cap R \text{ for all } \theta \in \hat{\Theta}_n \cap R \text{ and } \frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, \ell_n))$$

$$= \liminf_{n \to \infty} \inf_{P \in \mathbf{P}_0} P(\theta + \frac{h}{\sqrt{n}} \in \Theta_n \text{ for all } \theta \in \hat{\Theta}_n \cap R \text{ and } \frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, \ell_n)) = 1 \quad \text{(D.63)}$$

due to $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n \downarrow 0$ for any $h/\sqrt{n} \in \hat{V}_n(\theta, \ell_n)$. Therefore, results (D.62) and (D.63) together with Assumption 6.6(ii) and Lemma D.2 yield that uniformly in $P \in \mathbf{P}_0$

$$\sup_{\theta \in \hat{\Theta}_n \cap R} \sup_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, \ell_n)} \|\hat{\mathbb{D}}_n(\theta)[h] - \mathbb{D}_{n,P}(\theta)[h]\|_r = o_p(a_n) . \quad \text{(D.64)}$$

Moreover, since $\hat{\Theta}_n \cap R \subseteq \Theta_n \cap R$ almost surely, we also have from Assumption 6.5 that

$$\sup_{\theta \in \hat{\Theta}_n \cap R} \|\hat{\mathbb{W}}_n \rho(\cdot, \theta) * q_n^{k_n} - \mathbb{W}_{n,P}^\star \rho(\cdot, \theta) * q_n^{k_n}\|_r$$

$$\leq \mathcal{J} \times \sup_{f \in \mathcal{F}_n} \|\hat{\mathbb{W}}_n f q_n^{k_n} - \mathbb{W}_{n,P}^\star f q_n^{k_n}\|_r = o_p(a_n) \quad \text{(D.65)}$$

uniformly in $P \in \mathbf{P}$. Therefore, since $\|\hat{\Sigma}_n\|_{o,r} = O_p(1)$ uniformly in $P \in \mathbf{P}$ by Lemma B.3, we obtain from results (D.64) and (D.65) and Lemma C.1 that uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R) = \inf_{\theta \in \hat{\Theta}_n \cap R} \inf_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, \ell_n)} \|\mathbb{W}_{n,P}^\star \rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\hat{\Sigma}_n, r} + o_p(a_n) . \quad \text{(D.66)}$$

Next, note that by Assumption 3.4(iii) there exists a constant $C_0 < \infty$ such that $\|\Sigma_n(P)^{-1}\|_{o,r} \leq C_0$ for all n and $P \in \mathbf{P}$. Thus, we obtain that

$$\|\mathbb{W}_{n,P}^\star \rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\hat{\Sigma}_n, r}$$

$$\leq \{C_0 \|\hat{\Sigma}_n - \Sigma_n(P)\|_{o,r} + 1\} \|\mathbb{W}_{n,P}^\star \rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P), r} \quad \text{(D.67)}$$

for any $\theta \in \Theta_n \cap R$, $h \in \mathbf{B}_n$, and $P \in \mathbf{P}$. In particular, since $0 \in \hat{V}_n(\theta, \ell_n)$ for any $\theta \in \Theta_n \cap R$, Assumptions 3.4(iii), 5.3(iii), Markov's inequality, and Lemma C.4 yield

$$\|\hat{\Sigma}_n - \Sigma_n(P)\|_{o,r} \times \inf_{\theta \in \hat{\Theta}_n \cap R} \inf_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, \ell_n)} \|\mathbb{W}_{n,P}^{\star}\rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\hat{\Sigma}_n, r}$$

$$\leq \|\hat{\Sigma}_n - \Sigma_n(P)\|_{o,r} \times \sup_{\theta \in \Theta_n \cap R} \|\mathbb{W}_{n,P}^{\star}\rho(\cdot, \theta) * q_n^{k_n}\|_{\Sigma_n(P), r} = o_p(a_n) \quad \text{(D.68)}$$

uniformly in $P \in \mathbf{P}$. It then follows from (D.67) and (D.68) that uniformly in $P \in \mathbf{P}$

$$\inf_{\theta \in \hat{\Theta}_n \cap R} \inf_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, \ell_n)} \|\mathbb{W}_{n,P}^{\star}\rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\hat{\Sigma}_n, r}$$

$$\leq \inf_{\theta \in \hat{\Theta}_n \cap R} \inf_{\frac{h}{\sqrt{n}} \in \hat{V}_n(\theta, \ell_n)} \|\mathbb{W}_{n,P}^{\star}\rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P), r} + o_p(a_n) . \quad \text{(D.69)}$$

The reverse inequality to (D.69) can be obtained by identical arguments and exploiting $\max\{\|\hat{\Sigma}_n\|_{o,r}, \|\hat{\Sigma}_n^{-1}\|_{o,r}\} = O_p(1)$ uniformly in $P \in \mathbf{P}$ by Lemma B.3. The claim of the Lemma then follows from (D.66) and (D.69) (and its reverse inequality). ∎

Lemma D.2. *Let Assumptions 3.2(i)-(ii), 5.1, 5.2(i), 5.4(i), 6.1(i) hold, and define*

$$D_n(\theta) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \theta + \frac{h}{\sqrt{n}} \in \Theta_n \cap R \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\} . \quad \text{(D.70)}$$

If $\ell_n \downarrow 0$ satisfies $k_n^{1/r} \sqrt{\log(k_n)} B_n \times \sup_{P \in \mathbf{P}} J_{[]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$ and $K_m \ell_n^2 \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$, then there is an $\epsilon > 0$ such that uniformly in $P \in \mathbf{P}$

$$\sup_{\theta \in (\Theta_{0n}(P) \cap R)^{\epsilon}} \sup_{\frac{h}{\sqrt{n}} \in D_n(\theta)} \|\hat{\mathbb{D}}_n(\theta)[h] - \mathbb{D}_{n,P}(\theta)[h]\|_r = o_p(a_n) . \quad \text{(D.71)}$$

PROOF: By definition of the set $D_n(\theta)$, we have $\theta + h/\sqrt{n} \in \Theta_n \cap R$ for any $\theta \in \Theta_n \cap R$, $h/\sqrt{n} \in D_n(\theta)$. Therefore, since $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ for all $h/\sqrt{n} \in D_n(\theta)$ we obtain that

$$\sup_{\theta \in \Theta_n \cap R} \sup_{\frac{h}{\sqrt{n}} \in D_n(\theta)} \|\hat{\mathbb{D}}_n(\theta)[h] - \sqrt{n}\{P\rho(\cdot, \theta + \frac{h}{\sqrt{n}}) * q_n^{k_n} - P\rho(\cdot, \theta) * q_n^{k_n}\}\|_r$$

$$\leq \sup_{\theta_1, \theta_2 \in \Theta_n \cap R : \|\theta_1 - \theta_2\|_{\mathbf{B}} \leq \ell_n} \|\mathbb{G}_{n,P}\rho(\cdot, \theta_1) * q_n^{k_n} - \mathbb{G}_{n,P}\rho(\cdot, \theta_2) * q_n^{k_n}\|_r . \quad \text{(D.72)}$$

Further note that Assumptions 3.2(i), 5.2(i), and 6.1(i) additionally imply that

$$\sup_{P \in \mathbf{P}} \sup_{\theta_1, \theta_2 \in \Theta_n \cap R : \|\theta_1 - \theta_2\|_{\mathbf{B}} \leq \ell_n} E_P[\|\rho(X_i, \theta_1) - \rho(X_i, \theta_2)\|_2^2 q_{k,n,\jmath}^2(Z_{i,\jmath})] \leq B_n^2 K_\rho^2 K_b^{2\kappa_\rho} \ell_n^{2\kappa_\rho} .$$

(D.73)

Next, let $\mathcal{G}_n \equiv \{f(x)q_{k,n,\jmath}(z) : f \in \mathcal{F}_n, 1 \leq \jmath \leq \mathcal{J} \text{ and } 1 \leq k \leq k_{n,\jmath}\}$, and then observe

that Assumption 5.1, result (D.73) and $\|v\|_r \le k_n^{1/r}\|v\|_\infty$ for any $v \in \mathbf{R}^{k_n}$ yield

$$\sup_{\theta_1,\theta_2\in\Theta_n\cap R:\|\theta_1-\theta_2\|_{\mathbf{B}}\le\ell_n} \|\mathbb{G}_{n,P}\rho(\cdot,\theta_1)*q_n^{k_n} - \mathbb{G}_{n,P}\rho(\cdot,\theta_2)*q_n^{k_n}\|_r$$

$$\le 2\mathcal{J}k_n^{1/r} \times \sup_{g_1,g_2\in\mathcal{G}_n:\|g_1-g_2\|_{L_P^2}\le B_n K_\rho K_b^{\kappa_\rho}\ell_n^{\kappa_\rho}} |\mathbb{W}_{n,P}g_1 - \mathbb{W}_{n,P}g_2| + o_p(a_n) \quad (\text{D.74})$$

uniformly in $P \in \mathbf{P}$. Therefore, from results (C.18)-(C.20), Markov's inequality, and $k_n^{1/r}\sqrt{\log(k_n)}B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$ by hypothesis, we conclude

$$\sup_{\theta\in\Theta_n\cap R} \sup_{\frac{h}{\sqrt{n}}\in D_n(\theta)} \|\hat{\mathbb{D}}_n(\theta)[h] - \sqrt{n}\{P\rho(\cdot,\theta+\frac{h}{\sqrt{n}})*q_n^{k_n} - P\rho(\cdot,\theta)*q_n^{k_n}\}\|_r = o_p(a_n) \quad (\text{D.75})$$

uniformly in $P \in \mathbf{P}$. Moreover, setting $\epsilon > 0$ sufficiently small for Assumption 5.4(i) to hold, we then conclude from Lemmas B.4 and C.5, and Assumption 5.4(i) that

$$\sup_{\theta\in(\Theta_{0n}(P)\cap R)^\epsilon} \sup_{\frac{h}{\sqrt{n}}\in D_n(\theta)} \|\sqrt{n}\{P\rho(\cdot,\theta+\frac{h}{\sqrt{n}})*q_n^{k_n} - P\rho(\cdot,\theta)*q_n^{k_n}\} - \mathbb{D}_{n,P}(\theta)[h]\|_r$$

$$\le \sup_{\theta\in(\Theta_{0n}(P)\cap R)^\epsilon} \sup_{\frac{h}{\sqrt{n}}\in D_n(\theta)} \{\sqrt{C\mathcal{J}}K_m \times \sqrt{n} \times \|\frac{h}{\sqrt{n}}\|_{\mathbf{E}} \times \|\frac{h}{\sqrt{n}}\|_{\mathbf{L}}\} \quad (\text{D.76})$$

for some $C < \infty$. Therefore, since $\|h\|_{\mathbf{E}} \le K_b\|h\|_{\mathbf{B}}$ for all $h \in \mathbf{B}_n$ and $P \in \mathbf{P}$ by Assumption 6.1(i), we conclude from $K_m\ell_n^2 \times \mathcal{S}_n(\mathbf{L},\mathbf{E}) = o(a_n n^{-\frac{1}{2}})$ that

$$\sup_{\theta\in(\Theta_{0n}(P)\cap R)^\epsilon} \sup_{\frac{h}{\sqrt{n}}\in D_n(\theta)} \|\sqrt{n}\{P\rho(\cdot,\theta+\frac{h}{\sqrt{n}})*q_n^{k_n} - P\rho(\cdot,\theta)*q_n^{k_n}\} - \mathbb{D}_{n,P}(\theta)[h]\|_r = o(a_n)$$

$$(\text{D.77})$$

uniformly in $P \in \mathbf{P}$. Hence, the Lemma follows from results (D.75) and (D.77). ∎

Lemma D.3. *Let Assumptions 3.2(ii) and 5.4 hold. Then there is an $\epsilon > 0$ and $C < \infty$ such that for all n, $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, $\theta_1 \in (\Theta_{0n}(P) \cap R)^\epsilon$, and $h_0, h_1 \in \mathbf{B}_n$*

$$\|\mathbb{D}_{n,P}(\theta_0)[h_0] - \mathbb{D}_{n,P}(\theta_1)[h_1]\|_r \le C\{M_m\|h_0 - h_1\|_{\mathbf{E}} + K_m\|\theta_0 - \theta_1\|_{\mathbf{L}}\|h_1\|_{\mathbf{E}}\} .$$

PROOF: We first note that by Lemmas B.4 and C.5 there is a constant $C_0 < \infty$ with

$$\|\mathbb{D}_{n,P}(\theta_0)[h_0] - \mathbb{D}_{n,P}(\theta_1)[h_1]\|_r \le \{\sum_{j=1}^{\mathcal{J}} C_0\|\nabla m_{P,j}(\theta_0)[h_0] - \nabla m_{P,j}(\theta_1)[h_1]\|_{L_P^2}^2\}^{\frac{1}{2}} . \quad (\text{D.78})$$

Moreover, since $(h_0 - h_1) \in \mathbf{B}_n$, we can also conclude from Assumption 5.4(iii) that

$$\|\nabla m_{P,j}(\theta_0)[h_0 - h_1]\|_{L_P^2} \le M_m \times \|h_0 - h_1\|_{\mathbf{E}} . \quad (\text{D.79})$$

Similarly, letting $\epsilon > 0$ be such that Assumption 5.4(ii) holds, we further obtain

$$\|\nabla m_{P,\jmath}(\theta_0)[h_1] - \nabla m_{P,\jmath}(\theta_1)[h_1]\|_{L_P^2} \leq K_m\|\theta_1 - \theta_0\|_{\mathbf{L}}\|h_1\|_{\mathbf{E}} \tag{D.80}$$

due $\theta_1 \in (\Theta_{0n}(P) \cap R)^\epsilon$. Thus, the Lemma follows from (D.78)-(D.80). ∎

Lemma D.4. *Let Assumptions 3.1, 3.2, 3.3, 3.4, 4.1, 4.2, 5.1, 5.2, 5.3, 5.4(i), 6.5 hold. If ℓ_n, $\tilde{\ell}_n$ satisfy $k_n^{1/r}\sqrt{\log(k_n)}B_n \sup_{P \in \mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho} \vee \tilde{\ell}_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$, $\mathcal{R}_n = o(\ell_n \wedge \tilde{\ell}_n)$, and $K_m(\ell_n^2 \vee \tilde{\ell}_n^2)\mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$, then uniformly in $P \in \mathbf{P}_0$*

$$\inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, \ell_n)} \|\mathbb{W}_{n,P}^\star\rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r}$$

$$= \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, \tilde{\ell}_n)} \|\mathbb{W}_{n,P}^\star\rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r} + o_p(a_n) . \tag{D.81}$$

PROOF: For notational simplicity, for any $f : \Theta_n \cap R \to \mathbf{R}^{k_n}$, and $\ell \in \mathbf{R}_+$ define

$$T_{n,P}(f, \ell) \equiv \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, \ell)} \|f(\theta) + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r} . \tag{D.82}$$

Next, note that since $\mathbb{W}_{n,P}$ and $\mathbb{W}_{n,P}^\star$ have the same law for every P, it follows that

$$P(|T_{n,P}(\mathbb{W}_{n,P}^\star\rho * q_n^{k_n}, \ell_n) - T_{n,P}(\mathbb{W}_{n,P}^\star\rho * q_n^{k_n}, \tilde{\ell}_n)| > \epsilon)$$
$$= P(|T_{n,P}(\mathbb{W}_{n,P}\rho * q_n^{k_n}, \ell_n) - T_{n,P}(\mathbb{W}_{n,P}\rho * q_n^{k_n}, \tilde{\ell}_n)| > \epsilon) \tag{D.83}$$

for any $\epsilon > 0$. However, by Lemma 5.1 we also have uniformly in $P \in \mathbf{P}_0$ that

$$I_n(R) = \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, \ell_n)} \|\mathbb{W}_{n,P}\rho(\cdot, \theta) * q_n^{k_n} + \sqrt{n}P\rho(\cdot, \theta + \frac{h}{\sqrt{n}}) * q_n^{k_n}\|_{\Sigma_n(P),r} + o_p(a_n)$$

$$= \inf_{\theta \in \Theta_{0n}(P) \cap R} \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta, \ell_n)} \|\mathbb{W}_{n,P}\rho(\cdot, \theta) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta)[h]\|_{\Sigma_n(P),r} + o_p(a_n) , \tag{D.84}$$

where the second equality follows from (C.10), $K_m\ell_n^2 \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$ by hypothesis, and Lemma C.1. Hence, since the same arguments in (D.84) apply if we employ $\tilde{\ell}_n$ in place of ℓ_n, it follows from (D.82) and (D.84) that uniformly in $P \in \mathbf{P}_0$

$$T_{n,P}(\mathbb{W}_{n,P}\rho * q_n^{k_n}, \ell_n) = I_n(R) + o_p(1) = T_{n,P}(\mathbb{W}_{n,P}\rho * q_n^{k_n}, \tilde{\ell}_n) + o_p(a_n) . \tag{D.85}$$

Thus, the claim of the Lemma follows from (D.83) and (D.85). ∎

Lemma D.5. *Let Assumptions 2.1, 2.2(i), 3.2, 3.3(ii)-(iii), 3.4(ii)-(iii), 4.2, 5.4(i)-(ii), 6.1, and 6.3 hold, $\Theta_{0n}(P) \cap R = \{\theta_{0n}(P)\}$, R satisfy (71), and define the set*

$$N_n(\theta, \ell) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \nabla\Upsilon_F(\theta)[\frac{h}{\sqrt{n}}] = 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell\} .$$

Further assume that $\ell_n \downarrow 0$ satisfies $K_m \ell_n^2 \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(1)$, $\ell_n^2 1\{K_f > 0\} = o(n^{-\frac{1}{2}})$, and $\mathcal{R}_n \mathcal{S}_n(\mathbf{B}, \mathbf{E}) = o(\ell_n)$. (i) It then follows that uniformly in $P \in \mathbf{P}_0$ we have

$$\inf_{\frac{h}{\sqrt{n}} \in N_n(\theta_{0n}(P), \ell_n)} \|\mathbb{W}_{n,P}^\star \rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P), r}$$

$$= \inf_{h \in \mathbf{B}_n \cap \mathcal{N}(\nabla \Upsilon_F(\theta_{0n}(P)))} \|\mathbb{W}_{n,P}^\star \rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P), r} + o_p(a_n) .$$

(ii) For n large, $\mathbb{D}_{n,P}(\theta_{0n}(P)) : \mathbf{B}_n \cap \mathcal{N}(\nabla \Upsilon_F(\theta_{0n}(P))) \to \mathbf{R}^{k_n}$ is injective for all $P \in \mathbf{P}_0$.

PROOF: To begin, select $\hat{h}_{n,P}/\sqrt{n} \in N_n(\theta_{0n}(P), \ell_n)$ so that uniformly in $P \in \mathbf{P}_0$

$$\inf_{\frac{h}{\sqrt{n}} \in N_n(\theta_{0n}(P), \ell_n)} \|\mathbb{W}_{n,P}^\star \rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P), r}$$

$$= \|\mathbb{W}_{n,P}^\star \rho(\cdot, \theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[\hat{h}_{n,P}]\|_{\Sigma_n(P), r} + o_p(1) . \quad \text{(D.86)}$$

Further note that for $L_n(\theta_{0n}(P), 2\ell_n)$ as defined in (E.27), Lemma E.1 (see (E.30)) implies that there exists a $\tilde{h}_{n,P}/\sqrt{n} \in L_n(\theta_{0n}(P), 2\ell_n)$ for which for n sufficiently large

$$\|\frac{\hat{h}_{n,P}}{\sqrt{n}} - \frac{\tilde{h}_{n,P}}{\sqrt{n}}\|_{\mathbf{B}} \le M \times \ell_n^2 1\{K_f > 0\} \quad \text{(D.87)}$$

for some $M < \infty$. Moreover, since $\Upsilon_F(\theta_{0n}(P) + \tilde{h}_{n,P}/\sqrt{n}) = 0$ and R satisfies (71), it follows that $\theta_{0n}(P) + \tilde{h}_{n,P}/\sqrt{n} \in \mathbf{B}_n \cap R$. Thus, we obtain from Assumption 6.1 and $\|\tilde{h}_{n,P}/\sqrt{n}\|_{\mathbf{B}} \le 2\ell_n$ that for n sufficiently large we have for all $P \in \mathbf{P}_0$ that

$$\frac{\tilde{h}_{n,P}}{\sqrt{n}} \in V_n(\theta_{0n}(P), 2K_b \ell_n) . \quad \text{(D.88)}$$

In particular, we obtain $\theta_{0n}(P) + \tilde{h}_{n,P}/\sqrt{n} \in (\Theta_{0n}(P) \cap R)^\epsilon$ for n sufficiently large, and thus from Assumption 4.2 and $\Theta_{0n}(P) \cap R = \{\theta_{0n}(P)\}$ we can conclude that

$$\|\frac{\tilde{h}_{n,P}}{\sqrt{n}}\|_{\mathbf{E}} \le \nu_n \{\|E_P[\rho(X_i, \theta_{0n}(P) + \frac{\tilde{h}_{n,P}}{\sqrt{n}}) * q_n^{k_n}(Z_i)]\|_r + O(\zeta_n)\}$$

$$\lesssim \nu_n \{\|\mathbb{D}_{n,P}(\theta_{0n}(P))[\frac{\tilde{h}_{n,P}}{\sqrt{n}}]\|_r + K_m \frac{\ell_n^2}{\sqrt{n}} \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) + O(\zeta_n)\} , \quad \text{(D.89)}$$

where the second equality in (D.89) holds by result (C.10). Moreover, also note that

$$\|\mathbb{D}_{n,P}(\theta_{0n}(P))[\frac{\tilde{h}_{n,P}}{\sqrt{n}}] - \mathbb{D}_{n,P}(\theta_{0n}(P))[\frac{\hat{h}_{n,P}}{\sqrt{n}}]\|_r \lesssim \ell_n^2 1\{K_f > 0\} \quad \text{(D.90)}$$

by Lemma D.3, Assumption 6.1(i), and result (D.87). Thus, combining results (D.87),

(D.89), and (D.90) together with $\nu_n^{-1} = O(1)$ by Assumption 4.2 we obtain that

$$\|\frac{\hat{h}_{n,P}}{\sqrt{n}}\|_{\mathbf{E}} \lesssim \nu_n\{\|\mathbb{D}_{n,P}(\theta_{0n}(P))[\frac{\hat{h}_{n,P}}{\sqrt{n}}]\|_r + \ell_n^2(\frac{K_m}{\sqrt{n}}\mathcal{S}_n(\mathbf{L},\mathbf{E}) + 1\{K_f > 0\}) + \zeta_n\} \quad (D.91)$$

for all $P \in \mathbf{P}_0$ for n sufficiently large. Also note that $\hat{h}_{n,P}$ satisfying (D.86) and $0 \in N_n(\theta_{0n}(P), \ell_n)$ imply together with $\|\Sigma_n(P)\|_{o,r}$ being bounded uniformly in P by Assumption 3.4(ii), Lemma C.4, and Markov's inequality that uniformly in $P \in \mathbf{P}_0$

$$\|\mathbb{D}_{n,P}(\theta_{0n}(P))[\hat{h}_{n,P}]\|_{\Sigma_n(P),r}$$
$$\leq 2\|\mathbb{W}_{n,P}\rho(\cdot,\theta_{0n}(P)) * q_n^{k_n}\|_{\Sigma_n(P),r} + o_p(1) = O_p(k_n^{1/r}\sqrt{\log(k_n)}B_n J_n) . \quad (D.92)$$

Hence, employing the definition of $\mathcal{R}_n$ (see (47)) and that $\nu_n/\sqrt{n} \leq \mathcal{R}_n$, we can conclude from Assumption 3.4(iii) and results (D.91) and (D.92) that uniformly in $P \in \mathbf{P}_0$

$$\|\frac{\hat{h}_{n,P}}{\sqrt{n}}\|_{\mathbf{E}} = O_p(\mathcal{R}_n\{1 + K_m\ell_n^2\mathcal{S}_n(\mathbf{L},\mathbf{E}) + \sqrt{n}\ell_n^2 1\{K_f > 0\}\}) = O_p(\mathcal{R}_n) , \quad (D.93)$$

where the second equality follows from $K_m\ell_n^2\mathcal{S}_n(\mathbf{L},\mathbf{E}) = o(1)$ and $\ell_n^2 1\{K_f > 0\} = o(n^{-\frac{1}{2}})$ by hypothesis. Therefore, since $\mathcal{R}_n\mathcal{S}_n(\mathbf{B},\mathbf{E}) = o(\ell_n)$ we obtain $\|\hat{h}_{n,P}/\sqrt{n}\|_{\mathbf{B}} = o_p(\ell_n)$ and hence with probability tending to one uniformly in $P \in \mathbf{P}_0$

$$\inf_{\frac{h}{\sqrt{n}} \in N_n(\theta_{0n}(P),\ell_n)} \|\mathbb{W}_{n,P}^\star\rho(\cdot,\theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P),r}$$
$$= \inf_{\frac{h}{\sqrt{n}} \in N_n(\theta_{0n}(P),\ell_n/2)} \|\mathbb{W}_{n,P}^\star\rho(\cdot,\theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P),r} . \quad (D.94)$$

However, since $\mathbb{D}_{n,P}(\theta_{0n}(P)) : \mathbf{B}_n \to \mathbf{R}^{k_n}$ is linear, the function being minimized in (D.94) is convex. As a result, it follows that whenever (D.94) holds, we must also have

$$\inf_{\frac{h}{\sqrt{n}} \in N_n(\theta_{0n}(P),\ell_n)} \|\mathbb{W}_{n,P}^\star\rho(\cdot,\theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P),r}$$
$$= \inf_{\frac{h}{\sqrt{n}} \in N_n(\theta_{0n}(P),+\infty)} \|\mathbb{W}_{n,P}^\star\rho(\cdot,\theta_{0n}(P)) * q_n^{k_n} + \mathbb{D}_{n,P}(\theta_{0n}(P))[h]\|_{\Sigma_n(P),r} , \quad (D.95)$$

and hence the first claim of the Lemma follows from the definition of $N_n(\theta,\ell_n)$.

We establish the second claim of the Lemma by contradiction. Suppose there exists a subsequence $\{n_j\}_{j=1}^\infty$ of $\{n\}_{n=1}^\infty$ and sequence $\{P_j\}_{j=1}^\infty \subseteq \mathbf{P}_0$ such that $\mathbb{D}_{n_j,P_j}(\theta_{0n_j}(P_j)) : \mathbf{B}_{n_j} \cap \mathcal{N}(\nabla\Upsilon_F(\theta_{0n_j}(P_j))) \to \mathbf{R}^{k_{n_j}}$ is not injective, and then note the linearity of the map $\mathbb{D}_{n_j,P_j}(\theta_{0n_j}(P_j))$ implies there exists a $h_{n_j}^c \in \mathbf{B}_{n_j} \cap \mathcal{N}(\nabla\Upsilon_F(\theta_{0n_j}(P_j)))$ such that $\|h_{n_j}^c/\sqrt{n_j}\|_{\mathbf{B}} = 1$ and $\mathbb{D}_{n_j,P_j}(\theta_{0n_j}(P_j))[h_{n_j}^c] = 0$. Then observe that $\ell_{n_j}h_{n_j}^c/\sqrt{n_j} \in N_{n_j}(\theta_{0n_j}(P_j), \ell_{n_j})$, and that as a result (D.91) also holds with $\ell_{n_j}h_{n_j}^c/\sqrt{n_j}$ in place of

$\hat{h}_{n,P}/\sqrt{n}$. Therefore, since $\mathbb{D}_{n_j,P_j}(\theta_{0n_j}(P_j))[h^c_{n_j}] = 0$ we can conclude that

$$\ell_{n_j}\|\frac{h^c_{n_j}}{\sqrt{n_j}}\|_{\mathbf{E}} \lesssim \nu_{n_j}\{\ell_{n_j}^2(\frac{K_m}{\sqrt{n_j}}\mathcal{S}_{n_j}(\mathbf{L},\mathbf{E}) + 1\{K_f > 0\}) + \zeta_{n_j}\} = O(\mathcal{R}_{n_j}) \qquad \text{(D.96)}$$

where the final equality follows by exploiting the definition of $\mathcal{R}_n$ and the fact that $K_m\ell_n^2\mathcal{S}_n(\mathbf{L},\mathbf{E}) = o(1)$ and $\ell_n^2 1\{K_f > 0\} = o(n^{-\frac{1}{2}})$. However, result (D.96) and $\|h^c_n/\sqrt{n}\|_{\mathbf{E}} \leq K_b\|h^c_n/\sqrt{n}\|_{\mathbf{B}} = 1$ by Assumption 6.1(i) imply $\ell_{n_j} = O(\mathcal{R}_{n_j})$, which contradicts $\mathcal{R}_n\mathcal{S}_n(\mathbf{B},\mathbf{E}) = o(\ell_n)$ due to $\{\mathcal{S}_n(\mathbf{B},\mathbf{E})\}^{-1} \geq 1/K_b$ by (D.21), and hence the second claim of the Lemma follows. ∎

Lemma D.6. *Suppose there exists a $\delta > 0$ such that for all $\epsilon > 0$ and $\tilde{\alpha} \in [\alpha - \delta, \alpha + \delta]$*

$$\sup_{P \in \mathbf{P}_0} P(c_{n,1-\tilde{\alpha}}(P) - \epsilon \leq I_n(R) \leq c_{n,1-\tilde{\alpha}}(P) + \epsilon) \leq a_n^{-1}(\epsilon \wedge 1) + o(1) . \qquad \text{(D.97)}$$

(i) If $I_n(R) \leq U_{n,P}(R) + o_p(a_n)$ and $\hat{U}_n(R) \geq U_{n,P}^\star(R) + o_p(a_n)$ uniformly in $P \in \mathbf{P}_0$ for some $U_{n,P}^\star(R)$ independent to $\{V_i\}_{i=1}^n$ and equal in distribution to $U_{n,P}(R)$, then

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(I_n(R) > \hat{c}_{n,1-\alpha}) \leq \alpha . \qquad \text{(D.98)}$$

(ii) If $I_n(R) = U_{n,P}(R) + o_p(a_n^{-1})$ and $\hat{U}_n(R) = U_{n,P}^\star(R) + o_p(a_n^{-1})$ uniformly in $P \in \mathbf{P}_0$ for some $U_{n,P}^\star(R)$ independent to $\{V_i\}_{i=1}^n$ and equal in distribution to $U_{n,P}(R)$, then

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} |P(I_n(R) > \hat{c}_{n,1-\alpha}) - \alpha| = 0 . \qquad \text{(D.99)}$$

PROOF: For the first claim, note that by hypothesis there exists a positive sequence b_n such that $b_n = o(a_n)$ and in addition we have uniformly in $P \in \mathbf{P}_0$ that

$$I_n(R) \leq U_{n,P}(R) + o_p(b_n) \qquad \hat{U}_n(R) \geq U_{n,P}^\star(R) + o_p(b_n) . \qquad \text{(D.100)}$$

Next, observe that by Markov's inequality and result (D.100) we can conclude that

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(P(U_{n,P}^\star(R) > \hat{U}_n(R) + b_n|\{V_i\}_{i=1}^n) > \epsilon)$$

$$\leq \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} \frac{1}{\epsilon}P(U_{n,P}^\star(R) > \hat{U}_n(R) + b_n) = 0 . \qquad \text{(D.101)}$$

Thus, it follows from (D.101) that there exists some sequence $\eta_n \downarrow 0$ such that the event

$$\Omega_n(P) \equiv \{\{V_i\}_{i=1}^n | P(U_{n,P}^\star(R) > \hat{U}_n(R) + b_n|\{V_i\}_{i=1}^n) \leq \eta_n\} \qquad \text{(D.102)}$$

satisfies $P(\Omega_n(P)^c) = o(1)$ uniformly in $P \in \mathbf{P}_0$. Hence, for any $t \in \mathbf{R}$ we obtain that

$$
\begin{aligned}
P(\hat{U}_n(R) \leq t | \{V_i\}_{i=1}^n) & 1\{\{V_i\}_{i=1}^n \in \Omega_n(P)\} \\
& \leq P(\hat{U}_n(R) \leq t \text{ and } U_{n,P}^\star(R) \leq \hat{U}_n(R) + b_n | \{V_i\}_{i=1}^n) + \eta_n \\
& \leq P(U_{n,P}^\star(R) \leq t + b_n) + \eta_n \;,
\end{aligned}
\tag{D.103}
$$

where the final inequality exploited that $U_{n,P}^\star(R)$ is independent of $\{V_i\}_{i=1}^n$. Next, define

$$
q_{n,1-\alpha}(P) \equiv \inf\{u : P(U_{n,P}(R) \leq u) \geq 1 - \alpha\}
\tag{D.104}
$$

and note that by evaluating (D.103) at $t = \hat{c}_{n,1-\alpha}$ we obtain that $\Omega_n(P)$ implies $\hat{c}_{n,1-\alpha} + b_n \geq q_{n,1-\alpha-\eta_n}(P)$. Therefore, $P(\Omega_n(P)^c) = o(1)$ uniformly in $P \in \mathbf{P}_0$ yields

$$
\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(q_{n,1-\alpha-\eta_n}(P) \leq \hat{c}_{n,1-\alpha}(P) + b_n) \geq \liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(\{V_i\}_{i=1}^n \in \Omega_n(P)) = 1 \;.
\tag{D.105}
$$

Furthermore, arguing as in (D.103) it follows that for some sequence $\tilde{\eta}_n = o(1)$ we have

$$
c_{n,1-\alpha-\tilde{\eta}_n}(P) \leq q_{n,1-\alpha-\eta_n}(P) + b_n \;.
\tag{D.106}
$$

Thus, exploiting (D.105), (D.106), condition (D.97), and $b_n = o(a_n)$, we conclude that

$$
\begin{aligned}
\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(I_n(R) > \hat{c}_{n,1-\alpha}) & \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(I_n(R) > q_{n,1-\alpha-\eta_n}(P) - b_n) \\
& \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(I_n(R) > c_{n,1-\alpha-\tilde{\eta}_n}(P) - 2b_n) = 1 - \alpha \;.
\end{aligned}
\tag{D.107}
$$

The proof of the second claim follows arguments similar to those already employed and hence we keep the exposition more concise. Moreover, we further note that since the first part of the Lemma implies (D.97) holds, it suffices to show that

$$
\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(I_n(R) > \hat{c}_{n,1-\alpha}) \geq \alpha \;.
\tag{D.108}
$$

First, note that we may now set the sequence b_n so that $b_n = o(a_n)$ and in addition

$$
I_n(R) = U_{n,P}(R) + o_p(b_n) \qquad \hat{U}_n(R) = U_{n,P}^\star(R) + o_p(b_n)
\tag{D.109}
$$

uniformly in $P \in \mathbf{P}_0$. Moreover, arguing as in (D.101) implies that $P(|\hat{U}_n(R) - U_{n,P}^\star(R)| > b_n | \{V_i\}_{i=1}^n) = o_p(\eta_n)$ uniformly in $P \in \mathbf{P}_0$ for some $\eta_n \downarrow 0$, and therefore

$$
\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(\hat{c}_{n,1-\alpha} \leq q_{n,1-\alpha+\eta_n}(P) + b_n) = 1
\tag{D.110}
$$

by Lemma 11 in Chernozhukov et al. (2013). Furthermore, by analogous arguments and

again relying on Lemma 11 in Chernozhukov et al. (2013) we can also conclude that

$$q_{n,1-\alpha+\eta_n}(P) \leq c_{n,1-\alpha+\tilde\eta_n}(P) + b_n \tag{D.111}$$

for some $\tilde\eta_n = o(1)$. Therefore, combining results (D.110) and (D.111) we obtain

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(I_n(R) > \hat{c}_{n,1-\alpha}) \geq \liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(I_n(R) > q_{n,1-\alpha+\eta_n}(P) + b_n)$$
$$\geq \liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(I_n(R) > c_{n,1-\alpha+\tilde\eta_n}(P) + 2b_n) = 1 - \alpha , \quad \text{(D.112)}$$

where the final equality follows from condition (D.97). ∎

APPENDIX E - Local Parameter Space

In this Appendix, we develop analytical results analyzing the approximation rate of the local parameter spaces. The main result of this Appendix is Theorem E.1, which plays an instrumental role in the proof of the results of Section 6.

Theorem E.1. *Let Assumptions 2.1(i), 2.2(i), 6.2, 6.3, and 6.4 hold, $\{\ell_n, \delta_n, r_n\}_{n=1}^{\infty}$ satisfy $\ell_n \downarrow 0$, $\delta_n 1\{K_f > 0\} \downarrow 0$, $r_n \geq (M_g \delta_n + K_g \delta_n^2) \vee 2(\ell_n + \delta_n)1\{K_g > 0\}$, and define*

$$G_n(\theta) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \Upsilon_G(\theta + \frac{h}{\sqrt{n}}) \leq (\Upsilon_G(\theta) - K_g r_n \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \mathbf{1_G}) \vee (-r_n \mathbf{1_G})\} \tag{E.1}$$

$$A_n(\theta) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \frac{h}{\sqrt{n}} \in G_n(\theta), \ \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0 \ and \ \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\} \tag{E.2}$$

$$T_n(\theta) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0, \ \Upsilon_G(\theta + \frac{h}{\sqrt{n}}) \leq 0 \ and \ \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq 2\ell_n\} \tag{E.3}$$

Then it follows that there exists a $M < \infty$, $\epsilon > 0$, and $n_0 < \infty$ such that for all $n > n_0$, $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, and $\theta_1 \in (\Theta_{0n}(P) \cap R)^\epsilon$ satisfying $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$ we have

$$\sup_{\frac{h_1}{\sqrt{n}} \in A_n(\theta_1)} \inf_{\frac{h_0}{\sqrt{n}} \in T_n(\theta_0)} \|\frac{h_1}{\sqrt{n}} - \frac{h_0}{\sqrt{n}}\|_{\mathbf{B}} \leq M \times \ell_n(\ell_n + \delta_n)1\{K_f > 0\} . \tag{E.4}$$

PROOF: Throughout, let $\tilde\epsilon$ be such that Assumptions 6.2 and 6.3 hold, set $\epsilon = \tilde\epsilon/2$, and for any $\delta > 0$ let $N_{n,P}(\delta) \equiv \{\theta \in \mathbf{B}_n : \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{B}}) < \epsilon\}$. For ease of exposition, we next break up the proof intro four distinct steps.

<u>STEP 1:</u> (Decompose $h/\sqrt{n}$). For any $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, and $h \in \mathbf{B}_n$ set

$$h^{\perp\theta_0} \equiv \nabla\Upsilon_F(\theta_0)^- \nabla\Upsilon_F(\theta_0)[h] \qquad h^{\mathcal{N}_{\theta_0}} \equiv h - h^{\perp\theta_0} , \tag{E.5}$$

where recall $\nabla\Upsilon_F(\theta_0)^- : \mathbf{F}_n \to \mathbf{B}_n$ denotes the right inverse of $\nabla\Upsilon_F(\theta_0) : \mathbf{B}_n \to \mathbf{F}_n$.

Further note that $h^{\mathcal{N}_{\theta_0}} \in \mathcal{N}(\nabla\Upsilon_F(\theta_0))$ since $\nabla\Upsilon_F(\theta_0)\nabla\Upsilon_F(\theta_0)^- = I$ implies that

$$\nabla\Upsilon_F(\theta_0)[h^{\mathcal{N}_{\theta_0}}] = \nabla\Upsilon_F(\theta_0)[h] - \nabla\Upsilon_F(\theta_0)\nabla\Upsilon_F(\theta_0)^-\nabla\Upsilon_F(\theta_0)[h] = 0 , \qquad \text{(E.6)}$$

by definition of $h^{\perp_{\theta_0}}$ in (E.5). Next, observe that if $\theta_1 \in (\Theta_{0n}(P) \cap R)^{\epsilon}$ and $h/\sqrt{n} \in \mathbf{B}_n$ satisfies $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ and $\Upsilon_F(\theta_1 + h/\sqrt{n}) = 0$, then $\theta_1 + h/\sqrt{n} \in N_{n,P}(\tilde{\epsilon})$ for n sufficiently large, and hence by Assumption 6.3(i) and $\Upsilon_F(\theta_1) = 0$ due to $\theta_1 \in \Theta_n \cap R$

$$\|\nabla\Upsilon_F(\theta_1)[\frac{h}{\sqrt{n}}]\|_{\mathbf{F}} = \|\Upsilon_F(\theta_1 + \frac{h}{\sqrt{n}}) - \Upsilon_F(\theta_1) - \nabla\Upsilon_F(\theta_1)[\frac{h}{\sqrt{n}}]\|_{\mathbf{F}} \leq K_f\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}}^2 . \text{ (E.7)}$$

Therefore, Assumption 6.3(ii), result (E.7), $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$, and $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ imply

$$\|\nabla\Upsilon_F(\theta_0)[\frac{h}{\sqrt{n}}]\|_{\mathbf{F}}$$
$$\leq \|\nabla\Upsilon_F(\theta_0) - \nabla\Upsilon_F(\theta_1)\|_o\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} + K_f\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}}^2 \leq K_f\ell_n(\delta_n + \ell_n) . \quad \text{(E.8)}$$

Moreover, since $\nabla\Upsilon_F(\theta_0) : \mathbf{F}_n \to \mathbf{B}_n$ satisfies Assumption 6.3(iv), we also have that

$$K_f\|h^{\perp_{\theta_0}}\|_{\mathbf{B}} = K_f\|\nabla\Upsilon_F(\theta_0)^-\nabla\Upsilon(\theta_0)[h]\|_{\mathbf{B}}$$
$$\leq K_f\|\nabla\Upsilon_F(\theta_0)^-\|_o\|\nabla\Upsilon_F(\theta_0)[h]\|_{\mathbf{F}} \leq M_f\|\nabla\Upsilon_F(\theta_0)[h]\|_{\mathbf{F}} . \quad \text{(E.9)}$$

Further note that if $K_f = 0$, then (E.5) and (E.8) imply that $h^{\perp_{\theta_0}} = 0$. Thus, combining results (E.8) and (E.9) to handle the case $K_f > 0$ we conclude that for any $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, $\theta_1 \in (\Theta_{0n}(P) \cap R)^{\epsilon}$ satisfying $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$ and any $h/\sqrt{n} \in \mathbf{B}_n$ such that $\Upsilon_F(\theta_1 + h/\sqrt{n}) = 0$ and $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ we have the norm bound

$$\|\frac{h^{\perp_{\theta_0}}}{\sqrt{n}}\|_{\mathbf{B}} \leq M_f\ell_n(\delta_n + \ell_n)1\{K_f > 0\} . \qquad \text{(E.10)}$$

STEP 2: (Inequality Constraints). In what follows, it is convenient to define the set

$$S_n(\theta_0, \theta_1) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) \leq 0, \ \Upsilon_F(\theta_1 + \frac{h}{\sqrt{n}}) = 0, \ \text{and} \ \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\} .$$
$$\text{(E.11)}$$

Then note $r_n \geq (M_g\delta_n + K_g\delta_n^2) \vee 2(\ell_n + \delta_n)1\{K_g > 0\}$ and Lemma E.2 imply that

$$A_n(\theta_1) \subseteq S_n(\theta_0, \theta_1) \qquad \text{(E.12)}$$

for n sufficiently large, all $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, and $\theta_1 \in (\Theta_{0n}(P) \cap R)^{\epsilon}$ satisfying $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$. The proof will proceed by verifying (E.4) holds with $S_n(\theta_0, \theta_1)$ in place of $A_n(\theta_1)$. In particular, if $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ is linear, then $\Upsilon_F(\theta_0) = \Upsilon_F(\theta_1)$ and (E.12) implies $A_n(\theta_1) \subseteq S_n(\theta_0, \theta_1) \subseteq T_n(\theta_0)$, which establishes (E.4) for the case $K_f = 0$.

For the rest of the proof we therefore assume $K_f > 0$. We further note that Lemma E.3 implies that for any $\eta_n \downarrow 0$, there is a sufficiently large n and constant $1 \leq C < \infty$ (independent of η_n) such that for all $P \in \mathbf{P}$ and $\theta_0 \in \Theta_{0n}(P) \cap R$ there exists a $h_{\theta_0,n}/\sqrt{n} \in \mathbf{B}_n \cap \mathcal{N}(\nabla\Upsilon_F(\theta_0))$ such that for any $\tilde{h}/\sqrt{n} \in \mathbf{B}_n$ for which there exists a $h/\sqrt{n} \in S_n(\theta_0, \theta_1)$ satisfying $\|(\tilde{h} - h)/\sqrt{n}\|_{\mathbf{B}} \leq \eta_n$ the following inequalities hold

$$\Upsilon_G(\theta_0 + \frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{\tilde{h}}{\sqrt{n}}) \leq 0 \qquad \|\frac{h_{\theta_0,n}}{\sqrt{n}}\|_{\mathbf{B}} \leq C\eta_n . \qquad (\text{E.13})$$

STEP 3: (Equality Constraints). The results in this step allow us to address the challenge that $h/\sqrt{n} \in S_n(\theta_0, \theta_1)$ satisfies $\Upsilon_F(\theta_1 + h/\sqrt{n}) = 0$ but not necessarily $\Upsilon_F(\theta_0 + h/\sqrt{n}) = 0$. To this end, let $\mathcal{R}(\nabla\Upsilon_F(\theta_0)^-\nabla\Upsilon_F(\theta_0))$ denote the range of the operator $\nabla\Upsilon_F(\theta_0)^-\nabla\Upsilon_F(\theta_0) : \mathbf{B}_n \to \mathbf{B}_n$ and define the vector subspaces

$$\mathbf{B}_n^{\mathcal{N}_{\theta_0}} \equiv \mathbf{B}_n \cap \mathcal{N}(\nabla\Upsilon_F(\theta_0)) \qquad \mathbf{B}_n^{\perp_{\theta_0}} \equiv \mathcal{R}(\nabla\Upsilon_F(\theta_0)^-\nabla\Upsilon_F(\theta_0)) , \qquad (\text{E.14})$$

which note are closed due to $\mathbf{B}_n$ being finite dimensional by Assumption 3.2(iii). Moreover, since $h^{\mathcal{N}_{\theta_0}} \in \mathbf{B}_n^{\mathcal{N}_{\theta_0}}$ by (E.6), the definitions in (E.5) and (E.14) imply that $\mathbf{B}_n = \mathbf{B}_n^{\mathcal{N}_{\theta_0}} + \mathbf{B}_n^{\perp_{\theta_0}}$. Furthermore, since $\nabla\Upsilon_F(\theta_0)\nabla\Upsilon_F(\theta_0)^- = I$, we also have

$$\nabla\Upsilon_F(\theta_0)^-\nabla\Upsilon_F(\theta_0)[h] = h \qquad (\text{E.15})$$

for any $h \in \mathbf{B}_n^{\perp_{\theta_0}}$, and thus that $\mathbf{B}_n^{\mathcal{N}_{\theta_0}} \cap \mathbf{B}_n^{\perp_{\theta_0}} = \{0\}$. Since $\mathbf{B}_n = \mathbf{B}_n^{\mathcal{N}_{\theta_0}} + \mathbf{B}_n^{\perp_{\theta_0}}$, it then follows that $\mathbf{B}_n = \mathbf{B}_n^{\mathcal{N}_{\theta_0}} \oplus \mathbf{B}_n^{\perp_{\theta_0}}$ – i.e. the decomposition in (E.5) is unique. Moreover, we observe that $\mathbf{B}_n^{\mathcal{N}_{\theta_0}} \cap \mathbf{B}_n^{\perp_{\theta_0}} = \{0\}$ further implies the restricted map $\nabla\Upsilon_F(\theta_0) : \mathbf{B}_n^{\perp_{\theta_0}} \to \mathbf{F}_n$ is in fact bijective, and by (E.15) its inverse is $\nabla\Upsilon_F(\theta_0)^- : \mathbf{F}_n \to \mathbf{B}_n^{\perp_{\theta_0}}$.

We next note Assumption 6.3(i) implies that for all n and $P \in \mathbf{P}$, Υ_F is Fréchet differentiable at all $\theta \in \mathbf{B}_n$ such that $\|\theta - \theta_0\|_{\mathbf{B}} \leq \tilde{\epsilon}$ for some $\theta_0 \in \Theta_{0n}(P) \cap R$. Therefore, applying Lemma E.5 with $\mathbf{A}_1 = \mathbf{B}_n^{\mathcal{N}_{\theta_0}}$, $\mathbf{A}_2 = \mathbf{B}_n^{\perp_{\theta_0}}$ and $K_0 \equiv K_f \vee M_f \vee M_f/K_f$ yields that for any $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$ and $h^{\mathcal{N}_{\theta_0}} \in \mathbf{B}_n^{\mathcal{N}_{\theta_0}}$ satisfying $\|h^{\mathcal{N}_{\theta_0}}\|_{\mathbf{B}} \leq \{\tilde{\epsilon}/2 \wedge (2K_0)^{-2} \wedge 1\}^2$, there exists a $h^\star(h^{\mathcal{N}_{\theta_0}}) \in \mathbf{B}_n^{\perp_{\theta_0}}$ such that

$$\Upsilon_F(\theta_0 + h^{\mathcal{N}_{\theta_0}} + h^\star(h^{\mathcal{N}_{\theta_0}})) = 0 \qquad \|h^\star(h^{\mathcal{N}_{\theta_0}})\|_{\mathbf{B}} \leq 2K_0^2\|h^{\mathcal{N}_{\theta_0}}\|_{\mathbf{B}}^2 . \qquad (\text{E.16})$$

In addition, note that for any $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, $\theta_1 \in (\Theta_{0n}(P) \cap R)^\epsilon$ and any $h/\sqrt{n} \in \mathbf{B}_n$ such that $\Upsilon_F(\theta_1 + h/\sqrt{n}) = 0$ and $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$, result (E.10), the decomposition in (E.5), and $\delta_n \downarrow 0$ (since $K_f > 0$), $\ell_n \downarrow 0$ imply that for n large

$$\|\frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}}\|_{\mathbf{B}} \leq \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} + \|\frac{h^{\perp_{\theta_0}}}{\sqrt{n}}\|_{\mathbf{B}} \leq 2\ell_n . \qquad (\text{E.17})$$

Thus, for $h_{\theta_0,n} \in \mathbf{B}_n^{\mathcal{N}_{\theta_0}}$ as in (E.13), $C \geq 1$, and results (E.16) and (E.17) imply that

for n sufficiently large we must have for all $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, $\theta_1 \in \Theta_n \cap R$ with $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$ and $h/\sqrt{n} \in \mathbf{B}_n$ satisfying $\Upsilon_F(\theta_1 + h/\sqrt{n}) = 0$ that

$$\Upsilon_F(\theta_0 + \frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}} + h^\star(\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}})) = 0 \tag{E.18}$$

$$\|h^\star(\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}})\|_{\mathbf{B}} - 16K_0^2 C^2(\ell_n^2 + \eta_n^2) \leq 0 . \tag{E.19}$$

Step 4: (Build Approximation). In order to exploit Steps 2 and 3, we now set η_n to

$$\eta_n = 32(M_f + C^2 K_0^2)\ell_n(\ell_n + \delta_n) . \tag{E.20}$$

In addition, for any $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, $\theta_1 \in \Theta_n \cap R$ satisfying $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$, and any $h/\sqrt{n} \in S_n(\theta_0, \theta_1)$, we let $h^{\mathcal{N}_{\theta_0}}$ be as in (E.5) and define

$$\frac{\hat{h}}{\sqrt{n}} \equiv \frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}} + h^\star(\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}}) . \tag{E.21}$$

From Steps 2 and 3 it then follows that for n sufficiently large (independent of $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, $\theta_1 \in \Theta_n \cap R$ with $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$ or $h/\sqrt{n} \in S_n(\theta_0, \theta_1)$) we have

$$\Upsilon_F(\theta_0 + \frac{\hat{h}}{\sqrt{n}}) = 0 . \tag{E.22}$$

Moreover, from results (E.19) and (E.20) we also obtain that for n sufficiently large

$$\|h^\star(\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}})\|_{\mathbf{B}} \leq 16C^2 K_0^2(\ell_n^2 + \eta_n^2) \leq \frac{\eta_n}{2} + 16C^2 K_0^2 \eta_n^2 \leq \frac{3}{4}\eta_n . \tag{E.23}$$

Thus, $h = h^{\mathcal{N}_{\theta_0}} + h^{\perp_{\theta_0}}$, (E.10), (E.20), (E.21) and (E.23) imply $\|(\hat{h} - h - h_{\theta_0,n})/\sqrt{n}\|_{\mathbf{B}} \leq \eta_n$ for n sufficiently large, and exploiting (E.13) with $\tilde{h} = (\hat{h} - h_{\theta_0,n})/\sqrt{n}$) yields

$$\Upsilon_G(\theta_0 + \frac{\hat{h}}{\sqrt{n}}) \leq 0 . \tag{E.24}$$

Furthermore, since $\|h_{\theta_0,n}/\sqrt{n}\|_{\mathbf{B}} \leq C\eta_n$ by (E.13), results (E.10), (E.19), and $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ for any $h/\sqrt{n} \in S_n(\theta_0, \theta_1)$ imply by (E.20) and $\ell_n \downarrow 0$, $\delta_n \downarrow 0$ that

$$\|\frac{\hat{h}}{\sqrt{n}}\|_{\mathbf{B}} \leq \|\frac{h_{\theta_0,n}}{\sqrt{n}}\|_{\mathbf{B}} + \|h^\star(\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}})\|_{\mathbf{B}} + \|\frac{h^{\perp_{\theta_0}}}{\sqrt{n}}\|_{\mathbf{B}} + \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}}$$
$$\leq C\eta_n + 16C^2 K_0^2(\ell_n^2 + \eta_n^2) + M_f \ell_n(\delta_n + \ell_n) + \ell_n \leq 2\ell_n \tag{E.25}$$

for n sufficiently large. Therefore, we conclude from (E.22), (E.24), and (E.25) that

$\hat{h}/\sqrt{n} \in T_n(\theta_0)$. Similarly, (E.10), (E.13), (E.19), and (E.20) yield for some $M < \infty$

$$\|\frac{\hat{h}}{\sqrt{n}} - \frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \|\frac{h_{\theta_0,n}}{\sqrt{n}}\|_{\mathbf{B}} + \|h^\star(\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}})\|_{\mathbf{B}} + \|\frac{h^{\perp_{\theta_0}}}{\sqrt{n}}\|_{\mathbf{B}}$$
$$\leq C\eta_n + 16C^2 K_0^2(\ell_n^2 + \eta_n^2) + M_f \ell_n(\ell_n + \delta_n) \leq M\ell_n(\ell_n + \delta_n) , \quad \text{(E.26)}$$

which establishes the (E.4) for the case $K_f > 0$. ∎

Lemma E.1. *Let Assumptions 2.1(i), 2.2(i), 6.3 hold, $\{\ell_n, \delta_n\}_{n=1}^{\infty}$ be given and define*

$$L_n(\theta, \ell) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0 \ and \ \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell\} \quad \text{(E.27)}$$

$$N_n(\theta, \ell) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \nabla\Upsilon_F(\theta)[\frac{h}{\sqrt{n}}] = 0 \ and \ \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell\} . \quad \text{(E.28)}$$

If $\ell_n \downarrow 0$, $\delta_n 1\{K_f > 0\} \downarrow 0$, then there are $M < \infty$, $n_0 < \infty$, and $\epsilon > 0$ such that for all $n > n_0$, $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$ and $\theta_1 \in (\Theta_{0n}(P) \cap R)^\epsilon$ with $\|\theta_1 - \theta_0\|_{\mathbf{B}} \leq \delta_n$, we have

$$\sup_{\frac{h}{\sqrt{n}} \in L_n(\theta_1, \ell_n)} \inf_{\frac{\tilde{h}}{\sqrt{n}} \in N_n(\theta_0, 2\ell_n)} \|\frac{h}{\sqrt{n}} - \frac{\tilde{h}}{\sqrt{n}}\|_{\mathbf{B}} \leq M \times \ell_n(\ell_n + \delta_n)1\{K_f > 0\} \quad \text{(E.29)}$$

$$\sup_{\frac{h}{\sqrt{n}} \in N_n(\theta_0, \ell_n)} \inf_{\frac{\tilde{h}}{\sqrt{n}} \in L_n(\theta_1, 2\ell_n)} \|\frac{h}{\sqrt{n}} - \frac{\tilde{h}}{\sqrt{n}}\|_{\mathbf{B}} \leq M \times \ell_n(\ell_n + \delta_n)1\{K_f > 0\} . \quad \text{(E.30)}$$

PROOF: The proof exploits manipulations similar to those employed in Theorem E.1. First, let $\tilde{\epsilon}$ be such that Assumption 6.3 holds, set $\epsilon = \tilde{\epsilon}/2$ and note that for any $\theta_1 \in (\Theta_{0n}(P) \cap R)^\epsilon$ and $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ we have $\overrightarrow{d}_H(\{\theta_1 + h/\sqrt{n}\}, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{B}}) < \tilde{\epsilon}$ for n sufficiently large. In particular, if $K_f = 0$, then Assumptions 6.3(i)-(ii) yield

$$\Upsilon_F(\theta_1 + \frac{h}{\sqrt{n}}) = \nabla\Upsilon_F(\theta_1)[\frac{h}{\sqrt{n}}] = \nabla\Upsilon_F(\theta_0)[\frac{h}{\sqrt{n}}] = \Upsilon_F(\theta_0 + \frac{h}{\sqrt{n}}) . \quad \text{(E.31)}$$

Thus, $L_n(\theta_1, \ell) = L_n(\theta_0, \ell) = N_n(\theta_1, \ell) = N_n(\theta_0, \ell)$ and hence both (E.29) and (E.30) automatically hold. In what follows, we therefore assume $K_f > 0$.

Next, for each $h \in \mathbf{B}_n$, $P \in \mathbf{P}$, and $\theta \in (\Theta_{0n}(P) \cap R)^\epsilon$ we decompose h according to

$$h^{\perp_\theta} \equiv \nabla\Upsilon_F(\theta)^- \nabla\Upsilon_F(\theta)[h] \qquad\qquad h^{\mathcal{N}_\theta} \equiv h - h^{\perp_\theta} , \quad \text{(E.32)}$$

where recall $\nabla\Upsilon_F(\theta)^- : \mathbf{F}_n \to \mathbf{B}_n$ denotes the right inverse of $\nabla\Upsilon_F(\theta) : \mathbf{B}_n \to \mathbf{F}_n$. Next, note that result (E.10) implies that for n sufficiently large, we have for any $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, $\theta_1 \in (\Theta_{0n}(P) \cap R)^\epsilon$ satisfying $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$ and $h/\sqrt{n} \in L_n(\theta_1, \ell_n)$

$$\|\frac{h^{\perp_{\theta_0}}}{\sqrt{n}}\|_{\mathbf{B}} \leq M_f \ell_n(\ell_n + \delta_n)1\{K_f > 0\} . \quad \text{(E.33)}$$

Furthermore, note that for any $h/\sqrt{n} \in L_n(\theta_1, \ell_n)$ and n sufficiently large $h^{\mathcal{N}_{\theta_0}}/\sqrt{n}$ satisfies $\nabla \Upsilon_F(\theta_0)[h^{\mathcal{N}_{\theta_0}}] = 0$ by (E.6) and $\|h^{\mathcal{N}_{\theta_0}}/\sqrt{n}\|_{\mathbf{B}} \leq 2\ell_n$ by (E.17), and thus $h^{\mathcal{N}_{\theta_0}}/\sqrt{n} \in N_n(\theta_0, 2\ell_n)$. In particular, it follows that for any $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, and $\theta_1 \in (\Theta_{0n}(P) \cap R)^\epsilon$ with $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$ we must have that

$$
\sup_{\frac{h}{\sqrt{n}} \in L_n(\theta_1, \ell_n)} \inf_{\frac{\tilde{h}}{\sqrt{n}} \in N_n(\theta_0, 2\ell_n)} \|\frac{h}{\sqrt{n}} - \frac{\tilde{h}}{\sqrt{n}}\|_{\mathbf{B}} \leq \sup_{\frac{h}{\sqrt{n}} \in L_n(\theta_1, \ell_n)} \|\frac{h}{\sqrt{n}} - \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}}\|_{\mathbf{B}}
$$
$$
= \sup_{\frac{h}{\sqrt{n}} \in L_n(\theta_1, \ell_n)} \|\frac{h^{\perp_{\theta_0}}}{\sqrt{n}}\|_{\mathbf{B}} \leq M_f \ell_n (\ell_n + \delta_n) 1\{K_f > 0\} , \quad \text{(E.34)}
$$

where the first inequality follows from $h^{\mathcal{N}_{\theta_0}}/\sqrt{n} \in N_n(\theta_0, 2\ell_n)$, the equality from (E.32), and the second inequality is implied by (E.33). Thus, (E.29) follows.

In order to establish (E.30) when $K_f > 0$ note that for any $h/\sqrt{n} \in N_n(\theta_0, \ell_n)$, $\nabla \Upsilon_F(\theta_0)[h/\sqrt{n}] = 0$, $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$ and Assumption 6.3(ii) imply that

$$
\|\nabla \Upsilon_F(\theta_1)[\frac{h}{\sqrt{n}}]\|_{\mathbf{F}} = \|\nabla \Upsilon_F(\theta_1)[\frac{h}{\sqrt{n}}] - \nabla \Upsilon_F(\theta_0)[\frac{h}{\sqrt{n}}]\|_{\mathbf{F}} \leq K_f \delta_n \ell_n . \quad \text{(E.35)}
$$

Therefore, from definition (E.32), Assumption 6.3(iv) and result (E.35) we can conclude

$$
\|\frac{h^{\perp_{\theta_1}}}{\sqrt{n}}\|_{\mathbf{B}} = \|\nabla \Upsilon_F(\theta_1)^- \nabla \Upsilon_F(\theta_1)[\frac{h}{\sqrt{n}}]\|_{\mathbf{B}} \leq M_f \delta_n \ell_n . \quad \text{(E.36)}
$$

Moreover, identical arguments to those employed in establishing (E.16) (with θ_1 in place of θ_0) imply that for sufficiently large n it follows that for all $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, and $h/\sqrt{n} \in N_n(\theta_0, \ell_n)$ there is a $h^\star(h^{\mathcal{N}_{\theta_1}}/\sqrt{n})$ such that for some $K_0 < \infty$

$$
\Upsilon_F(\theta_1 + \frac{h^{\mathcal{N}_{\theta_1}}}{\sqrt{n}} + h^\star(\frac{h^{\mathcal{N}_{\theta_1}}}{\sqrt{n}})) = 0 \qquad \|h^\star(\frac{h^{\mathcal{N}_{\theta_1}}}{\sqrt{n}})\|_{\mathbf{B}} \leq 2K_0^2 \|\frac{h^{\mathcal{N}_{\theta_1}}}{\sqrt{n}}\|_{\mathbf{B}}^2 . \quad \text{(E.37)}
$$

Since $\ell_n, \delta_n \downarrow 0$, it follows that $M_f \delta_n \ell_n + 2K_0^2 (\ell_n + M_f \delta_n \ell_n)^2 \leq \ell_n$ for n sufficiently large. Therefore, $\|h^{\mathcal{N}_{\theta_1}}/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n + \|h^{\perp_{\theta_1}}/\sqrt{n}\|_{\mathbf{B}}$, (E.36) and (E.37) imply that for n sufficiently large, we have for all $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$ and $h/\sqrt{n} \in N_n(\theta_0, \ell_n)$ that

$$
\frac{h^{\mathcal{N}_{\theta_1}}}{\sqrt{n}} + h^\star(\frac{h^{\mathcal{N}_{\theta_1}}}{\sqrt{n}}) \in L_n(\theta_1, 2\ell_n) . \quad \text{(E.38)}
$$

Hence, for n sufficiently large we can conclude from result (E.38) that for all $P \in \mathbf{P}$,

$\theta_0 \in \Theta_{0n}(P) \cap R$, and $\theta_1 \in \Theta_n \cap R$ with $\|\theta_0 - \theta_1\|_{\mathbf{B}} \le \delta_n$ we have that

$$\sup_{\frac{h}{\sqrt{n}} \in N_n(\theta_0, \ell_n)} \inf_{\frac{\tilde{h}}{\sqrt{n}} \in L_n(\theta_1, 2\ell_n)} \|\frac{h}{\sqrt{n}} - \frac{\tilde{h}}{\sqrt{n}}\|_{\mathbf{B}} \le \sup_{\frac{h}{\sqrt{n}} \in N_n(\theta_0, \ell_n)} \|\frac{h}{\sqrt{n}} - \{\frac{h^{\mathcal{N}_{\theta_1}}}{\sqrt{n}} + h^\star(\frac{h^{\mathcal{N}_{\theta_1}}}{\sqrt{n}})\}\|_{\mathbf{B}}$$

$$\le \sup_{\frac{h}{\sqrt{n}} \in N_n(\theta_0, \ell_n)} \{\|\frac{h^{\perp_{\theta_1}}}{\sqrt{n}}\|_{\mathbf{B}} + \|h^\star(\frac{h^{\mathcal{N}_{\theta_1}}}{\sqrt{n}})\|_{\mathbf{B}}\} \le M_f \ell_n \delta_n + 2K_0^2(\ell_n + M_f \delta_n \ell_n)^2 \quad (\text{E.39})$$

where the final inequality holds by (E.36), (E.37) and $\|h^{\mathcal{N}_{\theta_1}}/\sqrt{n}\|_{\mathbf{B}} \le \ell_n + M_f \delta_n \ell_n$. Thus, (E.30) follows from (E.39), which establishes the claim of the Lemma. ∎

Lemma E.2. *Let Assumptions 2.1(i), 2.2(i), and 6.2 hold, and $\ell_n \downarrow 0$ be given. Then, there exist $n_0 < \infty$ and $\epsilon > 0$ such that for all $n > n_0$, $P \in \mathbf{P}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, and $\theta_1 \in \mathbf{B}_n$ satisfying $\overrightarrow{d}_H(\{\theta_1\}, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{B}}) < \epsilon$ it follows that*

$$\{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) \le (\Upsilon_G(\theta_1) - K_g r\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \mathbf{1_G}) \vee (-r \mathbf{1_G}), \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \le \ell_n\}$$

$$\subseteq \{\frac{h}{\sqrt{n}} \in \mathbf{B}_n : \Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) \le 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \le \ell_n\} . \quad (\text{E.40})$$

for any $r \ge \{M_g\|\theta_0 - \theta_1\|_{\mathbf{B}} + K_g\|\theta_0 - \theta_1\|_{\mathbf{B}}^2\} \vee 2\{\ell_n + \|\theta_0 - \theta_1\|_{\mathbf{B}}\} \mathbf{1}\{K_g > 0\}$.

PROOF: Let $\tilde{\epsilon} > 0$ be such that Assumption 6.2 holds, set $\epsilon = \tilde{\epsilon}/2$, and for notational simplicity let $N_{n,P}(\delta) \equiv \{\theta \in \mathbf{B}_n : \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}(P) \cap R, \|\cdot\|_{\mathbf{B}}) < \delta\}$ for any $\delta > 0$. Then note that for any $\theta_1 \in N_{n,P}(\epsilon)$ and $\|h/\sqrt{n}\|_{\mathbf{B}} \le \ell_n$ we have $\theta_1 + h/\sqrt{n} \in N_{n,P}(\tilde{\epsilon})$ for n sufficiently large. Therefore, by Assumption 6.2(ii) we obtain that

$$\|\Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) - \Upsilon_G(\theta_1) - \nabla \Upsilon_G(\theta_1)[\frac{h}{\sqrt{n}}]\|_{\mathbf{G}} \le K_g \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}}^2 . \quad (\text{E.41})$$

Similarly, Assumption 6.2(ii) implies that if $\theta_0 \in \Theta_{0n}(P) \cap R$ and $\theta_1 \in N_{n,P}(\epsilon)$, then

$$\|\nabla \Upsilon_G(\theta_0)[\frac{h}{\sqrt{n}}] - \nabla \Upsilon_G(\theta_1)[\frac{h}{\sqrt{n}}]\|_{\mathbf{G}}$$

$$\le \|\nabla \Upsilon_G(\theta_0) - \nabla \Upsilon_G(\theta_1)\|_o \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \le K_g \|\theta_0 - \theta_1\|_{\mathbf{B}} \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \quad (\text{E.42})$$

for any $h/\sqrt{n} \in \mathbf{B}_n$. Hence, since $\Upsilon_G(\theta_0) \le 0$ due to $\theta_0 \in \Theta_n \cap R$ we can conclude that

$$\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) + \{\Upsilon_G(\theta_1) - \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}})\}$$

$$\le \{\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) - \Upsilon_G(\theta_0)\} + \{\Upsilon_G(\theta_1) - \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}})\}$$

$$\le K_g \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \{2\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} + \|\theta_0 - \theta_1\|_{\mathbf{B}}\} \mathbf{1_G} , \quad (\text{E.43})$$

by (E.41), (E.42), and Lemma E.4. Also note for any $\theta_0 \in \Theta_{0n}(P) \cap R$, $\theta_1 \in N_{n,P}(\epsilon)$ and

82

$h/\sqrt{n} \in \mathbf{B}_n$ with $\|h/\sqrt{n}\|_\mathbf{B} \leq \ell_n$ we have $\theta_0 + h/\sqrt{n} \in N_{n,P}(\tilde{\epsilon})$ and $\theta_1 + h/\sqrt{n} \in N_{n,P}(\tilde{\epsilon})$ for n sufficiently large. Therefore, Assumptions 6.2(i), 6.2(iii), and Lemma E.4 yield

$$\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) - \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) \leq \nabla\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}})[\theta_0 - \theta_1] + K_g\|\theta_0 - \theta_1\|_\mathbf{B}^2 \mathbf{1_G}$$
$$\leq \{M_g\|\theta_0 - \theta_1\|_\mathbf{B} + K_g\|\theta_0 - \theta_1\|_\mathbf{B}^2\}\mathbf{1_G} . \qquad \text{(E.44)}$$

Hence, (E.43) and (E.44) yield for $r \geq \{M_g\|\theta_0 - \theta_1\|_\mathbf{B} + K_g\|\theta_0 - \theta_1\|_\mathbf{B}^2\} \vee 2\{\ell_n + \|\theta_0 - \theta_1\|_\mathbf{B}\}1\{K_g > 0\}$, $\theta_0 \in \Theta_{0n}(P) \cap R$, $\theta_1 \in N_{n,P}(\epsilon)$, $\|h/\sqrt{n}\|_\mathbf{B} \leq \ell_n$, and n large

$$\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) \leq \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) + (K_g r\|\frac{h}{\sqrt{n}}\|_\mathbf{B} - \Upsilon_G(\theta_1))\mathbf{1_G} \wedge r\mathbf{1_G}$$
$$= \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) - (\Upsilon_G(\theta_1) - K_g r\|\frac{h}{\sqrt{n}}\|_\mathbf{B})\mathbf{1_G} \vee (-r\mathbf{1_G}) \qquad \text{(E.45)}$$

where the equality follows from $(-a) \vee (-b) = -(a \wedge b)$ by Theorem 8.6 in Aliprantis and Border (2006). Thus, since $a_1 \leq a_2$ and $b_1 \leq b_2$ implies $a_1 \wedge b_1 \leq a_2 \wedge b_2$ in $\mathbf{G}$ by Corollary 8.7 in Aliprantis and Border (2006), (E.45) implies that for n sufficiently large and any $\theta_0 \in \Theta_{0n}(P) \cap R$, $\theta_1 \in N_{n,P}(\epsilon)$ and $h/\sqrt{n} \in \mathbf{B}_n$ satisfying $\|h/\sqrt{n}\|_\mathbf{B} \leq \ell_n$ and

$$\Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) \leq (\Upsilon_G(\theta_1) - K_g r\|\frac{h}{\sqrt{n}}\|_\mathbf{B}\mathbf{1_G}) \vee (-r\mathbf{1_G}) \qquad \text{(E.46)}$$

we must have $\Upsilon_G(\theta_0 + h/\sqrt{n}) \leq 0$, which verifies (E.40) indeed holds. ∎

Lemma E.3. *If Assumptions 2.1(i), 2.2(i), 6.2, 6.4(ii) hold, and $\eta_n \downarrow 0$, $\ell_n \downarrow 0$, then there is a n_0 (depending on η_n, ℓ_n) and a $C < \infty$ (independent of η_n, ℓ_n) such that for all $n > n_0$, $P \in \mathbf{P}$, and $\theta_0 \in \Theta_{0n}(P) \cap R$ there is $h_{\theta_0,n}/\sqrt{n} \in \mathbf{B}_n \cap \mathcal{N}(\nabla\Upsilon_F(\theta_0))$ with*

$$\Upsilon_G(\theta_0 + \frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{\tilde{h}}{\sqrt{n}}) \leq 0 \qquad \qquad \|\frac{h_{\theta_0,n}}{\sqrt{n}}\|_\mathbf{B} \leq C\eta_n \qquad \text{(E.47)}$$

for all $\tilde{h}/\sqrt{n} \in \mathbf{B}_n$ for which there is a $h/\sqrt{n} \in \mathbf{B}_n$ satisfying $\|\tilde{h}/\sqrt{n} - h/\sqrt{n}\|_\mathbf{B} \leq \eta_n$, $\|h/\sqrt{n}\|_\mathbf{B} \leq \ell_n$ and the inequality $\Upsilon_G(\theta_0 + h/\sqrt{n}) \leq 0$.

PROOF: By Assumption 6.4(ii) there are $\epsilon > 0$ and $K_d < \infty$ such that for every $P \in \mathbf{P}$, n, and $\theta_0 \in \Theta_{0n}(P) \cap R$ there exists a $\bar{h}_{\theta_0,n} \in \mathbf{B}_n \cap \mathcal{N}(\nabla\Upsilon_F(\theta_0))$ satisfying

$$\Upsilon_G(\theta_0) + \nabla\Upsilon_G(\theta_0)[\bar{h}_{\theta_0,n}] \leq -\epsilon\mathbf{1_G} \qquad \text{(E.48)}$$

and $\|\bar{h}_{\theta_0,n}\|_\mathbf{B} \leq K_d$. Moreover, for any $h/\sqrt{n} \in \mathbf{B}_n$ such that $\|h/\sqrt{n}\|_\mathbf{B} \leq \ell_n$, Assump-

tion 6.2(i), Lemma E.4 and $K_g \ell_n^2 \leq M_g \ell_n$ for n sufficiently large yield

$$\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) \leq \Upsilon_G(\theta_0) + \nabla\Upsilon_G(\theta_0)[\frac{h}{\sqrt{n}}] + K_g\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}}^2 \mathbf{1_G}$$

$$\leq \Upsilon_G(\theta_0) + \{\|\nabla\Upsilon_G(\theta_0)\|_o \ell_n + K_g \ell_n^2\}\mathbf{1_G} \leq \Upsilon_G(\theta_0) + 2M_g\ell_n\mathbf{1_G} . \quad \text{(E.49)}$$

Hence, (E.48) and (E.49) imply for any $h/\sqrt{n} \in \mathbf{B}_n$ with $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ we must have

$$\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) + \nabla\Upsilon_G(\theta_0)[\bar{h}_{\theta_0,n}] \leq \{2M_g\ell_n - \epsilon\}\mathbf{1_G} . \quad \text{(E.50)}$$

Next, we let $C_0 > 8M_g/\epsilon$ and aim to show (E.47) holds with $C = C_0 K_d$ by setting

$$\frac{h_{\theta_0,n}}{\sqrt{n}} \equiv C_0\eta_n\bar{h}_{\theta_0,n} . \quad \text{(E.51)}$$

To this end, we first note that if $\theta_0 \in \Theta_{0n}(P) \cap R$, $h/\sqrt{n} \in \mathbf{B}_n$ satisfies $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ and $\Upsilon_G(\theta_0 + h/\sqrt{n}) \leq 0$, and $\tilde{h}/\sqrt{n} \in \mathbf{B}_n$ is such that $\|h/\sqrt{n} - \tilde{h}/\sqrt{n}\|_{\mathbf{B}} \leq \eta_n$, then definition (E.51) implies that $\|\theta_0 + (h_{\theta_0,n} + \tilde{h})/\sqrt{n} - \theta_0\|_{\mathbf{B}} = o(1)$. Therefore, Assumption 6.2(i), Lemma E.4, and $\|(\tilde{h} - h)/\sqrt{n}\|_{\mathbf{B}} \leq \eta_n$ together allow us to conclude

$$\Upsilon_G(\theta_0 + \frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{\tilde{h}}{\sqrt{n}})$$

$$\leq \Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) + \nabla\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}})[\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{(\tilde{h} - h)}{\sqrt{n}}] + 2K_g(\|\frac{h_{\theta_0,n}}{\sqrt{n}}\|_{\mathbf{B}}^2 + \eta_n^2)\mathbf{1_G}$$

$$\leq \Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) + \nabla\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}})[\frac{h_{\theta_0,n}}{\sqrt{n}}] + \{2K_g\|\frac{h_{\theta_0,n}}{\sqrt{n}}\|_{\mathbf{B}}^2 + 2M_g\eta_n\}\mathbf{1_G} , \quad \text{(E.52)}$$

where the final result follows from Assumption 6.2(iii) and $2K_g\eta_n^2 \leq M_g\eta_n$ for n sufficiently large. Similarly, Assumption 6.2(ii) and Lemma E.4 yield

$$\nabla\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}})[\frac{h_{\theta_0,n}}{\sqrt{n}}] \leq \nabla\Upsilon_G(\theta_0)[\frac{h_{\theta_0,n}}{\sqrt{n}}] + \|\nabla\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) - \nabla\Upsilon_G(\theta_0)\|_o\|\frac{h_{\theta_0,n}}{\sqrt{n}}\|_{\mathbf{B}}\mathbf{1_G}$$

$$\leq \nabla\Upsilon_G(\theta_0)[\frac{h_{\theta_0,n}}{\sqrt{n}}] + K_g\ell_n\|\frac{h_{\theta_0,n}}{\sqrt{n}}\|_{\mathbf{B}}\mathbf{1_G} . \quad \text{(E.53)}$$

Hence, combining results (E.52) and (E.53), $\|h_{\theta_0,n}/\sqrt{n}\|_{\mathbf{B}} \leq C_0 K_d\eta_n$ due to $\|\bar{h}_{\theta_0,n}\|_{\mathbf{B}} \leq K_d$, and $\eta_n \downarrow 0$, $\ell_n \downarrow 0$, we obtain that for n sufficiently large we have

$$\Upsilon_G(\theta_0 + \frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{\tilde{h}}{\sqrt{n}}) \leq \Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) + \nabla\Upsilon_G(\theta_0)[\frac{h_{\theta_0,n}}{\sqrt{n}}] + 4M_g\eta_n\mathbf{1_G} . \quad \text{(E.54)}$$

In addition, since $C_0\eta_n \downarrow 0$, we have $C_0\eta_n \leq 1$ eventually, and hence $\Upsilon_G(\theta_0 + h/\sqrt{n}) \leq 0$,

84

$2M_g\ell_n \leq \epsilon/2$ for n sufficiently large due to $\ell_n \downarrow 0$ and result (E.50) imply that

$$\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) + C_0\eta_n\nabla\Upsilon_G(\theta_0)[\bar{h}_{\theta_0,n}]$$

$$\leq C_0\eta_n\{\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) + \nabla\Upsilon_G(\theta_0)[\bar{h}_{\theta_0,n}]\} \leq C_0\eta_n\{2M_g\ell_n - \epsilon\}\mathbf{1_G} \leq -\frac{C_0\eta_n\epsilon}{2}\mathbf{1_G} \ .$$

$$\text{(E.55)}$$

Thus, we can conclude from results (E.51),(E.54) and (E.55), and $C_0 > 8M_g/\epsilon$ that

$$\Upsilon_G(\theta_0 + \frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{\tilde{h}}{\sqrt{n}}) \leq \{4M_g - \frac{C_0\epsilon}{2}\}\eta_n\mathbf{1_G} \leq 0 \ , \tag{E.56}$$

for n sufficiently large, which establishes the claim of the Lemma. ∎

Lemma E.4. *If $\mathbf{A}$ is an AM space with norm $\|\cdot\|_{\mathbf{A}}$ and unit $\mathbf{1_A}$, and $a_1, a_2 \in \mathbf{A}$, then it follows that $a_1 \leq a_2 + C\mathbf{1_A}$ for any $a_1, a_2 \in \mathbf{A}$ satisfying $\|a_1 - a_2\|_{\mathbf{A}} \leq C$.*

PROOF: Since $\mathbf{A}$ is an AM space with unit $\mathbf{1_A}$ we have that $\|a_1 - a_2\|_{\mathbf{A}} \leq C$ implies $|a_1 - a_2| \leq C\mathbf{1_A}$, and hence the claim follows trivially from $a_1 - a_2 \leq |a_1 - a_2|$. ∎

Lemma E.5. *Let $\mathbf{A}$ and $\mathbf{C}$ be Banach spaces with norms $\|\cdot\|_{\mathbf{A}}$ and $\|\cdot\|_{\mathbf{C}}$, $\mathbf{A} = \mathbf{A}_1 \oplus \mathbf{A}_2$ and $F : \mathbf{A} \to \mathbf{C}$. Suppose $F(a_0) = 0$ and that there are $\epsilon_0 > 0$ and $K_0 < \infty$ such that:*

(i) $F : \mathbf{A} \to \mathbf{C}$ is Fréchet differentiable at all $a \in \mathcal{B}_{\epsilon_0}(a_0) \equiv \{a \in \mathbf{A} : \|a - a_0\|_{\mathbf{A}} \leq \epsilon_0\}$.
(ii) $\|F(a + h) - F(a) - \nabla F(a)[h]\|_{\mathbf{C}} \leq K_0\|h\|_{\mathbf{A}}^2$ for all $a, a + h \in \mathcal{B}_{\epsilon_0}(a_0)$.
(iii) $\|\nabla F(a_1) - \nabla F(a_2)\|_o \leq K_0\|a_1 - a_2\|_{\mathbf{A}}$ for all $a_1, a_2 \in \mathcal{B}_{\epsilon_0}(a_0)$.
(iv) $\nabla F(a_0) : \mathbf{A} \to \mathbf{C}$ has $\|\nabla F(a_0)\|_o \leq K_0$.
(v) $\nabla F(a_0) : \mathbf{A}_2 \to \mathbf{C}$ is bijective and $\|\nabla F(a_0)^{-1}\|_o \leq K_0$.

Then, for all $h_1 \in \mathbf{A}_1$ with $\|h_1\|_{\mathbf{A}} \leq \{\frac{\epsilon_0}{2} \wedge (4K_0^2)^{-1} \wedge 1\}^2$ there is a unique $h_2^\star(h_1) \in \mathbf{A}_2$ with $F(a_0 + h_1 + h_2^\star(h_1)) = 0$. In addition, $h_2^\star(h_1)$ satisfies $\|h^\star(h_1)\|_{\mathbf{A}} \leq 4K_0^2\|h_1\|_{\mathbf{A}}$ for arbitrary $\mathbf{A}_1$, and $\|h^\star(h_1)\|_{\mathbf{A}} \leq 2K_0^2\|h_1\|_{\mathbf{A}}^2$ when $\mathbf{A}_1 = \mathcal{N}(\nabla F(a_0))$.

PROOF: We closely follow the arguments in the proof of Theorems 4.B in Zeidler (1985). First, we define $g : \mathbf{A}_1 \times \mathbf{A}_2 \to \mathbf{C}$ pointwise for any $h_1 \in \mathbf{A}_1$ and $h_2 \in \mathbf{A}_2$ by

$$g(h_1, h_2) \equiv \nabla F(a_0)[h_2] - F(a_0 + h_1 + h_2) \ . \tag{E.57}$$

Since $\nabla F(a_0) : \mathbf{A}_2 \to \mathbf{C}$ is bijective by hypothesis, $F(a_0 + h_1 + h_2) = 0$ if and only if

$$h_2 = \nabla F(a_0)^{-1}[g(h_1, h_2)] \ . \tag{E.58}$$

Letting $T_{h_1} : \mathbf{A}_2 \to \mathbf{A}_2$ be given by $T_{h_1}(h_2) = \nabla F(a_0)^{-1}[g(h_1, h_2)]$, we see from (E.58) that the desired $h_2^\star(h_1)$ must be a fixed point of T_{h_1}. Next, define the set

$$M_0 \equiv \{h_2 \in \mathbf{A}_2 : \|h_2\|_{\mathbf{A}} \leq \delta_0\} \tag{E.59}$$

for $\delta_0 \equiv \frac{\epsilon_0}{2} \wedge (4K_0^2)^{-1} \wedge 1$, and consider an arbitrary $h_1 \in \mathbf{A}_1$ with $\|h_1\|_{\mathbf{A}} \le \delta_0^2$. Notice that then $a_0 + h_1 + h_2 \in \mathcal{B}_{\epsilon_0}(a_0)$ for any $h_2 \in M_0$ and hence g is differentiable with respect to h_2 with derivative $\nabla_2 g(h_1, h_2) \equiv \nabla F(a_0) - \nabla F(a_0 + h_1 + h_2)$. Thus, if $h_2, \tilde{h}_2 \in M_0$, then Proposition 7.3.2 in Luenberger (1969) implies that

$$
\begin{aligned}
\|g(h_1, h_2) - g(h_1, \tilde{h}_2)\|_{\mathbf{C}} &\le \sup_{0 < \tau < 1} \|\nabla_2 g(h_1, h_2 + \tau(\tilde{h}_2 - h_2))\|_o \|h_2 - \tilde{h}_2\|_{\mathbf{A}} \\
&= \sup_{0 < \tau < 1} \|\nabla F(a_0) - \nabla F(a_0 + h_1 + h_2 + \tau(\tilde{h}_2 - h_2))\|_o \|h_2 - \tilde{h}_2\|_{\mathbf{A}} \\
&\le \frac{1}{2K_0} \|h_2 - \tilde{h}_2\|_{\mathbf{A}} \ ,
\end{aligned} \tag{E.60}
$$

where the final inequality follows by Condition (iii) and $\delta_0^2 \le \delta_0 \le (4K_0^2)^{-1}$. Moreover,

$$
\begin{aligned}
\|\nabla F(a_0)[h_2] &- \nabla F(a_0 + h_1)[h_2]\|_{\mathbf{C}} \\
&\le \|\nabla F(a_0) - \nabla F(a_0 + h_1)\|_o \|h_2\|_{\mathbf{A}} \le K_0 \|h_1\|_{\mathbf{A}} \|h_2\|_{\mathbf{A}} \le \frac{\|h_2\|_{\mathbf{A}}}{4K_0} \quad (E.61)
\end{aligned}
$$

by Condition (iv) and $\|h_1\|_{\mathbf{A}} \le \delta_0 \le (4K_0^2)^{-1}$. Similarly, for any $h_2 \in M_0$ we have

$$
\|F(a_0 + h_1 + h_2) - F(a_0 + h_1) - \nabla F(a_0 + h_1)[h_2]\|_{\mathbf{C}} \le K_0 \|h_2\|_{\mathbf{A}}^2 \le \frac{\|h_2\|_{\mathbf{A}}}{4K_0} \quad (E.62)
$$

due to $a_0 + h_1 \in \mathcal{B}_{\epsilon_0}(a_0)$ and Condition (ii). In turn, since $F(a_0) = 0$ by hypothesis, Condition (iii), $\|h_1\|_{\mathbf{A}} \le \delta_0^2$ and $\delta_0 \le (4K_0^2)^{-1}$ yield that

$$
\|F(a_0 + h_1)\|_{\mathbf{C}} = \|F(a_0 + h_1) - F(a_0)\|_{\mathbf{C}} \le K_0 \|h_1\|_{\mathbf{A}}^2 + \|\nabla F(a_0)\|_o \|h_1\|_{\mathbf{A}} \le \frac{\delta_0}{2K_0} \ . \quad (E.63)
$$

Hence, by (E.57) and (E.61)-(E.63) we obtain for any $h_2 \in M_0$ and h_1 with $\|h_1\|_{\mathbf{A}} \le \delta_0^2$

$$
\|g(h_1, h_2)\|_{\mathbf{C}} \le \frac{\|h_2\|_{\mathbf{A}}}{2K_0} + \frac{\delta_0}{2K_0} \le \frac{\delta_0}{K_0} \ . \tag{E.64}
$$

Thus, since $\|\nabla F(a_0)^{-1}\|_o \le K_0$ by Condition (v), result (E.64) implies $T_{h_1} : M_0 \to M_0$, and (E.60) yields $\|T_{h_1}(h_2) - T_{h_1}(\tilde{h}_2)\|_{\mathbf{A}} \le 2^{-1} \|h_2 - \tilde{h}_2\|_{\mathbf{A}}$ for any $h_2, \tilde{h}_2 \in M_0$. By Theorem 1.1.1.A in Zeidler (1985) we then conclude T_{h_1} has a unique fixed point $h_2^\star(h_1) \in M_0$, and the first claim of the Lemma follows from (E.57) and (E.58).

Next, we note that since $h_2^\star(h_1)$ is a fixed point of T_{h_1}, we can conclude that

$$
\|h_2^\star(h_1)\|_{\mathbf{A}} = \|T_{h_1}(h_2^\star(h_1))\|_{\mathbf{A}} \le \|T_{h_1}(h_2^\star(h_1)) - T_{h_1}(0)\|_{\mathbf{A}} + \|T_{h_1}(0)\|_{\mathbf{A}} \ . \tag{E.65}
$$

Thus, since (E.60) and $\|\nabla F(a_0)^{-1}\|_o \le K_0$ imply that $\|T_{h_1}(h_2^\star(h_1)) - T_{h_1}(0)\|_{\mathbf{A}} \le$

$2^{-1}\|h_2^\star(h_1)\|_{\mathbf{A}}$, it follows from result (E.65) and $T_{h_1}(0) \equiv -\nabla F(a_0)^{-1}F(a_0 + h_1)$ that

$$\frac{1}{2}\|h_2^\star(h_1)\|_{\mathbf{A}} \leq \|T_{h_1}(0)\|_{\mathbf{A}} \leq K_0\|F(a_0 + h_1)\|_{\mathbf{C}}$$
$$\leq K_0\{K_0\|h_1\|_{\mathbf{A}}^2 + \|\nabla F(a_0)\|_o\|h_1\|_{\mathbf{A}}\} \leq 2K_0^2\|h_1\|_{\mathbf{A}} \ , \quad \text{(E.66)}$$

where in the second inequality we exploited $\|\nabla F(a_0)^{-1}\|_o \leq K_0$, in the third inequality we used (E.63) and in the final inequality we exploited $\|h_1\|_{\mathbf{A}} \leq 1$. While the estimate in (E.66) applies for generic $\mathbf{A}_1$, we note that if in addition $\mathbf{A}_1 = \mathcal{N}(\nabla F(a_0))$, then

$$\frac{1}{2}\|h_2^\star(h_1)\|_{\mathbf{A}} \leq \|T_{h_1}(0)\|_{\mathbf{A}} \leq K_0\|F(a_0 + h_1)\|_{\mathbf{C}} \leq K_0^2\|h_1\|_{\mathbf{A}}^2 \ , \quad \text{(E.67)}$$

due to $F(a_0) = 0$ and $\nabla F(a_0)[h_1] = 0$, and thus the final claim of the Lemma follows. ∎

Appendix F - Motivating Examples Details

In this Appendix, we revisit Examples 2.2, 2.1, 2.3, and 2.4 in order to illustrate our results. We focus in particular in deriving explicit expressions for the test and bootstrap statistics $I_n(R)$ and $\hat{U}_n(R)$, clarifying the role of the norms $\|\cdot\|_{\mathbf{E}}$, $\|\cdot\|_{\mathbf{L}}$, and $\|\cdot\|_{\mathbf{B}}$, as well as computing the rate requirements imposed by our Assumptions.

Discussion of Example 2.1

Since in this example we require g_0 to be continuously differentiable to evaluate the Slutsky restriction, it is natural to set the Banach space $\mathbf{B}$ to equal $\mathbf{B} = C^1(\mathbf{R}_+^2) \times \mathbf{R}^{d_w}$. Further recall that in this instance $Z_i = (P_i, Y_i, W_i)$, $X_i = (Q_i, Z_i)$, and

$$\rho(X_i, \theta) = Q_i - g(P_i, Y_i) - W_i'\gamma \quad \text{(F.1)}$$

for any $(g, \gamma) = \theta \in \mathbf{B}$. For simplicity, we assume the support of (P_i, Y_i) under P is bounded uniformly in $P \in \mathbf{P}$ and for some $C_0 < \infty$ set the parameter space Θ to be

$$\Theta \equiv \{(g, \theta) \in \mathbf{B} : \|g\|_{2,\infty} \leq C_0 \text{ and } \|\gamma\|_2 \leq C_0\} \ , \quad \text{(F.2)}$$

which is compact under the norm $\|\theta\|_{\mathbf{B}} = \|g\|_{1,\infty} \vee \|\gamma\|_2$ – for calculations with non-compact Θ see Examples 2.3 and 2.4 below. In order to approximate the function g_0 we utilize linear sieves $\{p_{j,n}\}_{j=1}^{j_n}$ and let $p_n^{j_n}(P_i, Y_i) \equiv (p_{1,n}(P_i, Y_i), \ldots, p_{j_n,n}(P_i, Y_i))'$. For $T : \mathbf{R}^{d_w} \to \mathbf{R}^{d_w}$ a bounded transformation we then set $q_n^{k_n}(Z_i) = (T(W_i)', p_n^{j_n}(P_i, Y_i)')'$

as our instruments so that $k_n = j_n + d_w$. Therefore, $I_n(R)$ is here equivalent to

$$I_n(R) = \inf_{(\beta,\gamma)} \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{Q_i - p_n^{j_n}(P_i, Y_i)'\beta - W_i'\gamma\} q_n^{k_n}(Z_i)\|_{\hat{\Sigma}_n, r}$$

$$\text{s.t. (i)} \ \|\gamma\|_2 \vee \|p_n^{j_n'}\beta\|_{2,\infty} \le C_0, \ \text{(ii)} \ p_n^{j_n}(p_0, y_0)'\beta = c_0,$$

$$\text{(iii)} \ \frac{\partial}{\partial p} p_n^{j_n}(p,y)'\beta + p_n^{j_n}(p,y)'\beta \frac{\partial}{\partial y} p_n^{j_n}(p,y)'\beta \le 0 , \tag{F.3}$$

where constraint (i) imposes that $(p_n^{j_n'}\beta, \gamma) \in \Theta$, restriction (ii) corresponds to $\Upsilon_F(\theta) = 0$, and (iii) enforces the Slutsky constraint $\Upsilon_G(\theta) \le 0$.

Whenever $\{p_{j,n}\}_{j=1}^{j_n}$ are chosen to be a tensor product of b-Splines or local polynomials it follows that $\sup_{(p,y)} \|p_n^{j_n}(p,y)\|_2 \lesssim \sqrt{j_n}$ and hence Assumption 3.2(i) holds with $B_n \asymp \sqrt{j_n}$ since $T(W_i)$ was assumed bounded (Belloni et al., 2015). Moreover, we note Assumption 3.3(ii) holds if $\sup_{P \in \mathbf{P}} E_P[\|W_i\|_2^2 + Q_i^2] < \infty$, while Assumption 3.3(iii) is satisfied with $J_n = O(1)$ by Theorem 2.7.1 in van der Vaart and Wellner (1996). By Remark 4.2, it also follows that if the eigenvalues of the matrix

$$E_P[q_n^{k_n}(Z_i)(W_i', p_n^{j_n}(P_i, Y_i)')] \tag{F.4}$$

are bounded from above and away from zero uniformly in $P \in \mathbf{P}$, then Assumption 4.2(ii) is satisfied with $\nu_n \asymp j_n^{1/2-1/r}$ and $\|\theta\|_{\mathbf{E}} = \sup_{P \in \mathbf{P}} \|g\|_{L_P^2} + \|\gamma\|_2$. Since we expect (g_0, γ_0) to be identified, we set $\tau_n = 0$ and thus under the no-bias condition of Assumption 5.3(ii) Theorem 4.1 yields a rate of convergence under $\|\cdot\|_{\mathbf{E}}$ equal to

$$\mathcal{R}_n = \frac{j_n \sqrt{\log(j_n)}}{\sqrt{n}} . \tag{F.5}$$

We refer to Corollary G.1 for verifying Assumption 5.1 and also note that Assumption 5.2 is satisfied with $\kappa_\rho = 1$ and $K_\rho^2 = 2(1 + \sup_{P \in \mathbf{P}} E_P[\|W_i\|_2^2])$ by (F.1) and definition of $\|\cdot\|_{\mathbf{E}}$. In turn, we note that since $\mathcal{F}_n$ is an Euclidean class, we also have

$$\sup_{P \in \mathbf{P}} J_{[\ ]}(\mathcal{R}_n, \mathcal{F}_n, \|\cdot\|_{L_P^2}) \lesssim \sqrt{j_n} \mathcal{R}_n \log(n) \tag{F.6}$$

and thus $B_n \lesssim \sqrt{j_n}$ and result (F.5) imply that Assumption 5.3(i) holds provided that $j_n^{2+1/r} \log^2(n) = o(a_n\sqrt{n})$. Since equation (F.1) implies that in this model

$$m_P(\theta)(Z_i) \equiv E_P[Q_i|Z_i] - g(P_i, Y_i) - W_i'\gamma , \tag{F.7}$$

we also observe that Assumption 5.4 holds with $\nabla m_P(\theta)[h](Z_i) = -g(P_i, Y_i) - W_i'\gamma$ for any $(g, \gamma) = h \in \mathbf{B}$, $K_m = 0$, and $M_m = 1 + \sup_{P \in \mathbf{P}} E_P[\|W_i\|_2]$.

With regards to the bootstrap statistic, we let $(\hat{\beta}, \hat{\gamma})$ be a minimizer of (F.3) and

$$
\hat{\mathbb{W}}_n \rho(\cdot, \hat{\theta}) * q_n^{k_n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_i \{ (Q_i - p_n^{j_n}(P_i, Y_i)'\hat{\beta} - W_i'\hat{\gamma}) q_n^{k_n}(Z_i)
$$
$$
- \frac{1}{n} \sum_{j=1}^{n} (Q_j - p_n^{j_n}(P_j, Y_j)'\hat{\beta} - W_j'\hat{\gamma}) q_n^{k_n}(Z_j) \} , \quad \text{(F.8)}
$$

where recall $\{\omega_i\}_{i=1}^{n}$ is an i.i.d. sample of standard normal random variables that are independent of the data $\{Q_i, P_i, Y_i, W_i\}_{i=1}^{n}$. Since in this model the moment conditions are linear in θ, in this case the numerical derivative in (63) simply reduces to

$$
\hat{\mathbb{D}}_n(\hat{\theta})[h] = -\frac{1}{n} \sum_{i=1}^{n} (p_n^{j_n}(Y_i, P_i)'\beta + W_i'\gamma) q_n^{k_n}(Z_i) \quad \text{(F.9)}
$$

for any $h = (p_n^{j_n'}\beta, \gamma) \in \mathbf{B}_n$. We also note that result (F.7) similarly implies that

$$
\mathbb{D}_{n,P}(\theta_0)[h] = -E_P[(p_n^{j_n}(Y_i, P_i)'\beta + W_i'\gamma) q_n^{k_n}(Z_i)] \quad \text{(F.10)}
$$

for $h = (p_n^{j_n'}\beta, \gamma) \in \mathbf{B}_n$. Moreover, since the requirement that the eigenvalues of (F.4) be bounded away from zero and infinity implies that similarly the eigenvalues of

$$
E_P[p_n^{j_n}(Y_i, P_i) p_n^{j_n}(Y_i, P_i)'] \quad \text{(F.11)}
$$

are bounded away from zero and infinity uniformly in $P \in \mathbf{P}$, it follows that $\|h\|_{\mathbf{E}} \asymp \|\gamma\|_2 + \|\beta\|_2$ for any $(p_n^{j_n'}\beta, \gamma) \in \mathbf{B}_n$. We therefore obtain from (F.9) and (F.10) that

$$
\sup_{\|h\|_{\mathbf{E}} \leq 1} \|\hat{\mathbb{D}}_n(\hat{\theta})[h] - \mathbb{D}_{n,P}(\theta_0)[h]\|_r
$$
$$
\lesssim \| \frac{1}{n} \sum_{i=1}^{n} q_n^{k_n}(Z_i)(W_i', p_n^{j_n}(P_i, Y_i)') - E_P[q_n^{k_n}(Z_i)(W_i', p_n^{j_n}(P_i, Y_i)')] \|_{o,2} . \quad \text{(F.12)}
$$

Thus, since $\sup_{(p,y)} \|p_n^{j_n}(p,y)\|_2 \lesssim \sqrt{j_n}$ we conclude by standard arguments and Theorem 6.1 in Tropp (2012) (Bernstein's inequality for matrices) that uniformly in $P \in \mathbf{P}$

$$
\sup_{\|h\|_{\mathbf{E}} \leq 1} \|\hat{\mathbb{D}}_n(\hat{\theta})[h] - \mathbb{D}_{n,P}(\theta_0)[h]\|_r = O_p(\frac{j_n \sqrt{\log(j_n)}}{\sqrt{n}}) . \quad \text{(F.13)}
$$

We next note that given the definitions of $\| \cdot \|_{\mathbf{E}}$ and $\| \cdot \|_{\mathbf{B}}$, Assumption 6.1(i) is trivially satisfied with $K_b = 2$. Furthermore, since in this example $\mathbf{G} = C(\mathbf{R}_+^2)$, we obtain by defining for any $(g_1, \theta_1) \in \mathbf{B}$ the map $\nabla \Upsilon(\theta_1) : \mathbf{B} \to \mathbf{G}$ according to

$$
\nabla \Upsilon_G(\theta_1)[h](p, y) = \frac{\partial}{\partial p} g(p, y) + g_1(p, y) \frac{\partial}{\partial y} g(p, y) + g(p, y) \frac{\partial}{\partial y} g_1(p, y) \quad \text{(F.14)}
$$

for any $(g, \gamma) = h \in \mathbf{B}$, that Assumptions 6.2(i)-(ii) hold with $K_g = 2$. Similarly, exploiting (F.14) and the definition of Θ in (F.2) it also follows that Assumption 6.2(iii) is satisfied with $M_g = 1 + 2C_0$. In turn, we observe that since $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ is linear (with $\mathbf{F} = \mathbf{R}$), Assumption 6.4 is automatically satisfied, while Assumption 6.3 holds with $K_f = 0$ and $M_f = 1$. Sufficient conditions for Assumption 6.5 are given by Theorem H.1, and we note the preceding discussion implies Assumption 6.6(ii) imposes $j_n^{1+1/r} \log^2(n) \ell_n = o(a_n)$ while Assumptions 6.6(iii)-(iv) respectively demand $j_n^5 \log(j_n) = o(nr_n^2)$ and $j_n^5 \log(j_n) = o(n)$ because $\mathcal{S}_n(\mathbf{B}, \mathbf{E}) \lesssim j_n^{3/2}$ (Newey, 1997). These rate requirements are compatible with setting ℓ_n to satisfy $\mathcal{R}_n \mathcal{S}_n(\mathbf{B}, \mathbf{E}) = o(\ell_n)$, and hence result (F.13), $\nu_n \asymp j_n^{1/2-1/r}$, and the eigenvalues of (F.4) being bounded away from zero imply that the conditions of Lemma 6.1 are satisfied. Therefore, the bandwidth ℓ_n is unnecessary – i.e. we may set $\ell_n = +\infty$ – and hence $\hat{U}_n(R)$ becomes

$$\hat{U}_n(R) = \inf_{(\pi, \beta)} \|\hat{\mathbb{W}}_n \rho(\cdot, \hat{\theta}) * q_n^{k_n} + \hat{\mathbb{D}}_n(\hat{\theta})[(\pi, p_n^{j_n'}\beta)]\|_{\hat{\Sigma}_n, r} \quad \text{s.t.} \quad \text{(i)} \ p_n^{j_n}(p_0, y_0)'\beta = 0,$$

$$\text{(ii)} \ \Upsilon_G(\hat{\theta} + \frac{p_n^{j_n'}\beta}{\sqrt{n}}) \leq (\Upsilon_G(\hat{\theta}) - 2r_n \|\frac{p_n^{j_n'}\beta}{\sqrt{n}}\|_{1,\infty}) \vee (-r_n \mathbf{1_G}) , \quad \text{(F.15)}$$

where constraints (i) and (ii) correspond to $\Upsilon_F(\hat{\theta} + h/\sqrt{n}) = 0$ and $h/\sqrt{n} \in G_n(\hat{\theta})$ in definition (69). It is worth noting that if constraints (i) and (ii) in (F.15) are replaced by more demanding restrictions, then the test would continue to control size. For computational simplicity, it may hence be preferable to replace constraint (ii) with

$$\Upsilon_G(\hat{\theta} + \frac{p_n^{j_n'}\beta}{\sqrt{n}}) \leq (\Upsilon_G(\hat{\theta}) - 2\frac{r_n j_n^{3/2}}{\sqrt{n}} \|\beta\|_2) \vee (-r_n \mathbf{1_G}) \quad \text{(F.16)}$$

where we have exploited that $\|p_n^{j_n'}\beta\|_{1,\infty} \lesssim j_n^{1/3} \|\beta\|_2$ by $\sup_{P \in \mathbf{P}} \|p_n^{j_n'}\beta\|_{L_P^2} \asymp \|\beta\|_2$ and $\mathcal{S}_n(\mathbf{B}, \mathbf{E}) \lesssim j_n^{3/2}$. Finally, we observe that in a model with endogeneity the arguments would remain similar, except the rate of convergence $\mathcal{R}_n$ would be slower leading to different requirements on r_n; see Chen and Christensen (2013) for the optimal $\mathcal{R}_n$. ∎

Discussion of Example 2.2

In the monotonic regression discontinuity example the Banach space $\mathbf{B}$ was set to equal $\mathbf{B} = C^1([-1, 0]) \times C^1([0, 1])$ while $(g_-, g_+) = \theta_0 \in \mathbf{B}$ satisfied the restriction

$$E[Y_i - g_-(R_i)(1 - D_i) - g_+(R_i)D_i | R_i, D_i] = 0 . \quad \text{(F.17)}$$

For the parameter space Θ we may for instance set Θ to be a C_0-ball in $\mathbf{B}$, so that

$$\Theta \equiv \{(g_1, g_2) \in \mathbf{B} : \|g_1\|_{1,\infty} \vee \|g_2\|_{1,\infty} \leq C_0\} \quad \text{(F.18)}$$

for C_0 sufficiently large to ensure $(g_-, g_+) \in \Theta$. In turn, we employ linear sieves

$\{p_{-,j,n}\}_{j=1}^{j_n}$ and $\{p_{+,j,n}\}_{j=1}^{j_n}$ for $C^1([-1,0])$ and $C^1([0,1])$ respectively and set the vector of instruments $q_n^{k_n}(R_i, D_i) \equiv ((1 - D_i)p_{-,n}^{j_n}(R_i)', D_i p_{-,n}^{j_n}(R_i)')$ where $k_n = 2j_n$, $p_{-,n}^{j_n} = (p_{-,1,n}, \ldots, p_{-,j_n,n})'$, and $p_{+,n}^{j_n} = (p_{+,1,n}, \ldots, p_{+,j_n,n})'$. Thus, $I_n(R)$ becomes

$$I_n(R) = \inf_{(\beta_1, \beta_2)} \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{Y_i - (1 - D_i)p_{-,n}^{j_n}(R_i)'\beta_1 - D_i p_{+,n}^{j_n}(R_i)'\beta_2\} q_n^{k_n}(R_i, D_i)\|_{\hat{\Sigma}_{n,r}}$$

$$\text{s.t. (i) } p_{-,n}^{j_n}(0)'\beta_1 - p_{+,n}^{j_n}(0)'\beta_2 = 0, \text{ (ii) } \nabla p_{-,n}^{j_n\prime}\beta_1 \geq 0, \text{ (iii) } \nabla p_{+,n}^{j_n\prime}\beta_2 \geq 0 \quad \text{(F.19)}$$

where constraint (i) corresponds to $\Upsilon_F(\theta) = 0$, constraints (ii) and (iii) impose $\Upsilon_G(\theta) \leq 0$, and the restriction $(p_{-,n}^{j_n\prime}\beta_1, p_{+,n}^{j_n\prime}\beta_2) \in \Theta$ can be ignored by Remark 6.3.

For concreteness, suppose $\{p_{-,j,n}\}_{j=1}^{j_n}$ and $\{p_{+,j,n}\}_{j=1}^{j_n}$ are orthonormalized b-splines, in which case the constant B_n of Assumption 3.2(i) satisfies $B_n \lesssim \sqrt{j_n}$. Moreover, we note that Theorem 2.7.1 in van der Vaart and Wellner (1996) implies the sequence J_n of Assumption 3.3(iii) satisfies $J_n = O(1)$. In turn, provided that the eigenvalues of

$$E_P[q_n^{k_n}(R_i, D_i) q_n^{k_n}(R_i, D_i)'] \quad \text{(F.20)}$$

are bounded from above and away from zero uniformly in $P \in \mathbf{P}$, Remark 4.2 implies Assumption 4.2 holds with $\nu_n \asymp j_n^{1/2-1/r}$ when using for any $(g_1, g_2) = \theta \in \mathbf{B}$ the norm $\|\theta\|_{\mathbf{E}} \equiv \sup_{P \in \mathbf{P}} \|g_1\|_{L_P^2} + \sup_{P \in \mathbf{P}} \|g_2\|_{L_P^2}$. We further note that since (g_-, g_+) is identified we may set $\tau_n = 0$ and hence, under the no bias condition of Assumption 5.3(ii), we obtain from Theorem 4.1 a rate of convergence under $\| \cdot \|_{\mathbf{E}}$ equal to

$$\mathcal{R}_n = \frac{j_n \sqrt{\log(j_n)}}{\sqrt{n}} . \quad \text{(F.21)}$$

We conjecture the above rate is suboptimal in that it does not exploit the linearity of the moment condition in θ, and employing the arguments in Belloni et al. (2015) it should be possible to derive the refined rate $\mathcal{R}_n = \sqrt{j_n \log(j_n)}/\sqrt{n}$ at least for the case $r = 2$.

Sufficient conditions for Assumption 5.1 are provided by Corollary G.1, while we observe Assumption 5.2 holds with $\kappa_\rho = 1$ by linearity of the moment condition and definition of $\| \cdot \|_{\mathbf{E}}$. Furthermore, since $\mathcal{F}_n$ is an Euclidean class, we further obtain

$$\sup_{P \in \mathbf{P}} J_{[]}(\mathcal{R}_n, \mathcal{F}_n, \| \cdot \|_{L_P^2}) \lesssim \sqrt{j_n} \mathcal{R}_n \log(n) , \quad \text{(F.22)}$$

and hence under the bound $B_n \lesssim \sqrt{j_n}$ and $\mathcal{R}_n \lesssim j_n \sqrt{\log(j_n)}/\sqrt{n}$ by (F.21), Assumption 5.3(i) reduces to $j_n^{2+1/r} \log^2(n) = o(a_n \sqrt{n})$. In turn, we note that

$$m_P(\theta)(Z_i) = E_P[Y_i - g_1(R_i)(1 - D_i) - g_2(R_i)D_i | R_i, D_i] \quad \text{(F.23)}$$

is linear for any $(g_1, g_2) \in \mathbf{B}$, and hence Assumptions 5.4(i)-(ii) hold with $K_m = 0$,

while Assumption 5.4(iii) is satisfied with $M_m = 1$. Similarly, if we metrize the product topology on $\mathbf{B} = C^1([-1,0]) \times C^1([0,1])$ by $\|\theta\|_{\mathbf{B}} = \|g_1\|_{1,\infty} \vee \|g_2\|_{1,\infty}$ for any $(g_1, g_2) = \theta \in \mathbf{B}$, then Assumption 6.1(i) holds with $K_b = 2$, Assumptions 6.2 and 6.3 are satisfied with $K_g = K_f = 0$ and $M_g = 1$ and $M_f = 2$, and Assumption 6.4 holds by linearity of Υ_F. Therefore, for any $(g_1, g_2) = \theta \in \mathbf{B}$ the set $G_n(\theta)$ becomes

$$G_n(\theta) \equiv \left\{ \left(\frac{p_{-,n}^{j_n'}\beta_1}{\sqrt{n}}, \frac{p_{+,n}^{j_n'}\beta_2}{\sqrt{n}} \right) : \begin{array}{ll} \nabla g_1(a) + \frac{\nabla p_{-,n}^{j_n}(a)'\beta_1}{\sqrt{n}} & \geq \nabla g_1(a) \wedge r_n \; \forall a \in [-1,0] \\ \nabla g_2(a) + \frac{\nabla p_{+,n}^{j_n}(a)'\beta_2}{\sqrt{n}} & \geq \nabla g_2(a) \wedge r_n \; \forall a \in [0,1] \end{array} \right\}$$
(F.24)

where we exploited $\mathbf{1_G} = (\mathbf{1}_{C([-1,0])}, \mathbf{1}_{C([0,1])})$ for $\mathbf{1}_{C([-1,0])}$ and $\mathbf{1}_{C([0,1])}$ respectively the constant functions equal to one on $[-1,0]$ and $[0,1]$.

With regards to other elements needed to construct the bootstrap statistic, we next let $(\hat{\beta}_1, \hat{\beta}_2)$ be a minimizer of (F.19) and for $\hat{\theta} = (p_{-,n}^{j_n'}\hat{\beta}_1, p_{+,n}^{j_n'}\hat{\beta}_2)$ note that

$$\hat{\mathbb{W}}_n \rho(\cdot, \hat{\theta}) * q_n^{k_n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_i \Big\{ (Y_i - (1 - D_i)p_{-,n}^{j_n}(R_i)'\hat{\beta}_1 - D_i p_{+,n}^{j_n}(R_i)'\hat{\beta}_2) q_n^{k_n}(R_i, D_i)$$
$$- \frac{1}{n} \sum_{j=1}^{n} (Y_j - (1 - D_j)p_{-,n}^{j_n}(R_j)'\hat{\beta}_1 - D_j p_{+,n}^{j_n}(R_j)'\hat{\beta}_2) q_n^{k_n}(R_j, D_j) \Big\} . \quad \text{(F.25)}$$

Since the moment condition is linear in θ, there is no need to employ a numerical derivative to estimate $\mathbb{D}_{n,P}(\theta_0)$ and hence following Remark 6.1 we just set

$$\hat{\mathbb{D}}_n(\hat{\theta})[h] = -\frac{1}{n} \sum_{i=1}^{n} \{(1 - D_i)p_{-,n}^{j_n}(R_i)'\beta_1 + D_i p_{+,n}^{j_n}(R_i)'\beta_2\} q_n^{k_n}(R_i, D_i) \quad \text{(F.26)}$$

for any $h = (p_{-,n}^{j_n'}\beta_1, p_{+,n}^{j_n'}\beta_2)$. Analogously, in this instance $\mathbb{D}_{n,P}(\theta_0) : \mathbf{B}_n \to \mathbf{R}^{k_n}$ equals

$$\mathbb{D}_{n,P}(\theta_0)[h] = -E_P[\{(1 - D_i)p_{-,n}^{j_n}(R_i)'\beta_1 + D_i p_{+,n}^{j_n}(R_i)'\beta_2\} q_n^{k_n}(R_i, D_i)] \quad \text{(F.27)}$$

for any $h = (p_{-,n}^{j_n'}\beta_1, p_{+,n}^{j_n'}\beta_2)$. Provided the eigenvalues of (F.20) are bounded from above and away from zero, and recalling $\nu_n \asymp j_n^{1/2-1/r}$, it is then straightforward to show $\|h\|_{\mathbf{E}} \leq \nu_n \|\mathbb{D}_{n,P}(\theta)[h]\|_r$ for any $h \in \mathbf{B}_n$. Moreover, by direct calculation

$$\sup_{\|h\|_{\mathbf{E}} \leq 1} \|\hat{\mathbb{D}}_n(\hat{\theta})[h] - \mathbb{D}_{n,P}(\theta_0)[h]\|_r$$
$$\lesssim \|\frac{1}{n} \sum_{i=1}^{n} q_n^{k_n}(R_i, D_i) q_n^{k_n}(R_i, D_i)' - E_P[q_n^{k_n}(R_i, D_i) q_n^{k_n}(R_i, D_i)']\|_{o,2} \quad \text{(F.28)}$$

and hence from Theorem 6.1 in Tropp (2012) we can conclude that uniformly in $P \in \mathbf{P}_0$

$$\sup_{\|h\|_{\mathbf{E}} \leq 1} \|\hat{\mathbb{D}}_n(\hat{\theta})[h] - \mathbb{D}_{n,P}(\theta_0)[h]\|_r = O_p\left(\frac{j_n \sqrt{\log(j_n)}}{\sqrt{n}}\right) . \quad \text{(F.29)}$$

In particular, (74) holds provided $j_n^{3/2-1/r}\sqrt{\log(j_n)} = o(\sqrt{n})$, and by Lemma 6.1 and Remark 6.4 it follows that the bandwidth ℓ_n can be ignored if it is possible to set $\ell_n \downarrow 0$ such that $\mathcal{R}_n = o(\ell_n)$. The additional requirements on ℓ_n are dictated by Assumption 6.6, which here become $j_n^{1/r}\sqrt{\log(j_n)}\sqrt{j_n}\sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$. Since we have shown $j_n^{1/r}\sqrt{\log(j_n)}\sqrt{j_n}\sup_{P\in\mathbf{P}} J_{[\,]}(\mathcal{R}_n, \mathcal{F}_n, \|\cdot\|_{L_P^2}) = o(a_n)$, we conclude $\mathcal{R}_n = o(\ell_n)$ is feasible, and hence in this example we may employ

$$\hat{V}_n(\theta, +\infty) \equiv \left\{ \left(\frac{p_{-,n}^{j_n\prime}\beta_1}{\sqrt{n}}, \frac{p_{+,n}^{j_n\prime}\beta_2}{\sqrt{n}}\right) \in G_n(\theta) : p_{-,n}^{j_n}(0)'\beta_1 - p_{+,n}^{j_n}(0)'\beta_2 = 0 \right\}. \qquad \text{(F.30)}$$

Thus, combining (F.24), (F.25), (F.26), and (F.30), the bootstrap statistic becomes

$$\hat{U}_n(R) = \inf_{(\beta_1,\beta_2)} \|\hat{\mathbb{W}}_n\rho(\cdot,\hat{\theta}) * q_n^{k_n} + \hat{\mathbb{D}}_n(\hat{\theta})[(p_{-,n}^{j_n\prime}\beta_1, p_{+,n}^{j_n\prime}\beta_2)]\|_{\hat{\Sigma}_n,r}$$

$$\text{s.t.} \quad \text{(i) } \left(\frac{p_{-,n}^{j_n\prime}\beta_1}{\sqrt{n}}, \frac{p_{+,n}^{j_n\prime}\beta_2}{\sqrt{n}}\right) \in G_n(\hat{\theta}), \text{ (ii) } p_{-,n}^{j_n}(0)'\beta_1 - p_{-,n}^{j_n}(0)'\beta_2 = 0. \quad \text{(F.31)}$$

Finally, we note that Theorem H.1 provides sufficient conditions for Assumption 6.5, while using the bound $\mathcal{S}_n(\mathbf{B},\mathbf{E}) \lesssim j_n^{3/2}$ from Newey (1997) and (F.21) implies Assumption 6.6(iii) is satisfied provided $j_n^5 \log(n) = o(nr_n^2)$. $\blacksquare$

Discussion of Example 2.3

Recall that in this application the parameter θ consists of a finite dimensional component $(\gamma_1, \gamma_2, \alpha) \in \mathbf{R}^{2d_\gamma + d_\alpha}$ and a nonparametric function $\delta \in C(\mathbf{R}^{d_y})$. For notational simplicity, we let $(\gamma_1, \gamma_2, \alpha) = \pi \in \mathbf{R}^{d_\pi}$ with $d_\pi = 2d_\gamma + d_\alpha$, and define the function

$$M_1(Z_i, \theta)$$
$$\equiv \int 1\{W_i'\gamma_1 + \epsilon_1 \geq 0,\ W_i'\gamma_2 + \delta(Y_i) + \epsilon_2 \leq 0,\ W_i'\gamma_1 + \epsilon_1 \geq W_i'\gamma_2 + \epsilon_2\}dG(\epsilon|\alpha),$$

which constitutes the part of (16) that depends on θ. Similarly, we further define

$$M_2(Z_i, \theta) \equiv \int 1\{\epsilon_1 \leq -W_i'\gamma_1,\ \epsilon_2 \leq -W_i'\gamma_2\}dG(\epsilon|\alpha) \qquad \text{(F.32)}$$

$$M_3(Z_i, \theta) \equiv \int 1\{\epsilon_1 + \delta(Y_i) \geq -W_i'\gamma_1,\ \epsilon_2 + \delta(Y_i) \geq -W_i'\gamma_2\}dG(\epsilon|\alpha), \qquad \text{(F.33)}$$

which correspond to the moment conditions in (17) and (18). For A_i the observed bundle purchased by agent i let $X_i = (A_i, Z_i)$, and then note that the generalized residuals are

$$\rho_1(X_i, \theta) \equiv 1\{A_i = (1,0)\} - M_1(Z_i, \theta)$$
$$\rho_2(X_i, \theta) \equiv 1\{A_i = (0,0)\} - M_2(Z_i, \theta)$$
$$\rho_3(X_i, \theta) \equiv 1\{A_i = (1,1)\} - M_3(Z_i, \theta). \qquad \text{(F.34)}$$

so that $\rho(X_i, \theta) = (\rho_1(X_i, \theta), \rho_2(X_i, \theta), \rho_3(X_i, \theta))'$ for a total of $\mathcal{J} = 3$ restrictions.

In this instance, $\mathbf{B} = \mathbf{R}^{d_\pi} \times C(\mathbf{R}^{d_y})$ and for illustrative purposes we select a non-compact parameter space by setting $\Theta = \mathbf{B}$. For $\{p_{j,n}\}_{j=1}^{j_n}$ a sequence of linear sieves in $C(\mathbf{R}^{d_y})$ and $p_n^{j_n}(y) = (p_{1,n}(y), \ldots, p_{j_n,n}(y))'$ we then let Θ_n be given by

$$\Theta_n \equiv \{(\pi, \delta) \in \mathbf{R}^{d_\pi} \times C(\mathbf{R}^{d_y}) : \|\pi\|_2 \leq C_0 \text{ and } \delta = p_n^{j_n\prime}\beta \text{ for some } \|\beta\|_2 \leq C_n\} \quad \text{(F.35)}$$

for some constant $C_0 < \infty$ and sequence $\{C_n\}_{n=1}^\infty$ satisfying $C_n \uparrow \infty$. In turn, for $1 \leq \jmath \leq \mathcal{J}$ we let $\{q_{k,n,\jmath}\}_{k=1}^{k_{n,\jmath}}$ denote a sequence of transformations of $Z_i = (W_i, Y_i)$, and recall $q_n^{k_n}(z) = (q_{n,1}^{k_{n,1}}(z)', q_{n,2}^{k_{n,2}}(z)', q_{n,3}^{k_{n,3}}(z)')'$ where $q_{n,\jmath}^{k_{n,\jmath}}(z) = (q_{1,n,\jmath}(z), \ldots, q_{k_{n,\jmath},n,\jmath}(z))'$ and $k_n = k_{n,1} + k_{n,2} + k_{n,3}$. We note, however, that since the conditioning variable is the same for all three moment conditions, in this instance we may in fact let $q_{n,1}^{k_{n,1}} = q_{n,2}^{k_{n,2}} = q_{n,3}^{k_{n,3}}$ – i.e. employ the same transformation of Z_i for all three moment conditions. Thus, given the above specifications, the test statistic $I_n(R)$ is equivalent to

$$I_n(R) = \inf_{(\pi,\beta)} \|\frac{1}{\sqrt{n}} \sum_{i=1}^n \rho(X_i, (\pi, p_n^{j_n\prime}\beta)) * q_n^{k_n}\|_{\hat{\Sigma}_n, r}$$

$$\text{s.t. (i) } p_n^{j_n\prime}\beta \leq 0, \text{ (ii) } \|\pi\|_2 \leq C_0, \text{ (iii) } \|\beta\|_2 \leq C_n , \quad \text{(F.36)}$$

where constraint (i) corresponds to $\Upsilon_G(\theta) \leq 0$, while constraints (ii) and (iii) impose that $(\pi, p_n^{j_n\prime}\beta) = \theta \in \Theta_n$. The latter two restrictions are standard sieve compactness conditions imposed in (even parametric) nonconvex estimation problems.

Next, let $M(Z_i, \theta) = (M_1(Z_i, \theta), M_2(Z_i, \theta), M_3(Z_i, \theta))'$ which we will assume to be differentiable, with $\nabla_\pi M(Z_i, \theta)$ denoting the derivative with respect to π, and

$$\nabla_\delta M(Z_i, \theta) \equiv \frac{\partial}{\partial \tau} M(Z_i, \theta + \tau e_\delta)\Big|_{\tau=0} \quad \text{(F.37)}$$

for $e_\delta \in \mathbf{B}$ equal to $(0, \mathbf{1}_{C(\mathbf{R}^{d_y})})$ and $\mathbf{1}_{C(\mathbf{R}^{d_y})}$ the constant function that equals one everywhere. For $\| \cdot \|_F$ the Frobenius norm of a matrix, we further define

$$F_\pi(z) \equiv \sup_{\theta \in \Theta} \|\nabla_\pi M(z, \theta)\|_F \qquad F_\delta(z) \equiv \sup_{\theta \in \Theta} \|\nabla_\delta M(z, \theta)\|_2 , \quad \text{(F.38)}$$

and then observe that by the mean value theorem and the Cauchy-Schwarz inequality

$$|M_\jmath(Z_i, (\pi_1, p_n^{j_n\prime}\beta_1)) - M_\jmath(Z_i, (\pi_2, p_n^{j_n\prime}\beta_2))|$$
$$\leq F_\pi(Z_i)\|\pi_1 - \pi_2\|_2 + F_\delta(Z_i)\|p_n^{j_n}(Y_i)'(\beta_1 - \beta_2)\|_2$$
$$\leq \{F_\pi(Z_i) + \|p_n^{j_n}(Y_i)\|_2 F_\delta(Z_i)\}\{\|\pi_1 - \pi_2\|_2 + \|\beta_1 - \beta_2\|_2\} \quad \text{(F.39)}$$

for any $1 \leq \jmath \leq \mathcal{J}$. Defining $F_{\theta,n}(z) \equiv \{F_\pi(Z_i) + \|p_n^{j_n}(Y_i)\|_2 F_\delta(Z_i)\}$ and assuming that

$\sup_{P \in \mathbf{P}} E_P[F_\pi(Z_i)^2] < \infty$, $F_\delta(z)$ is bounded uniformly in z, and that the matrix

$$E_P[p_n^{j_n}(Y_i)p_n^{j_n}(Y_i)'] \tag{F.40}$$

has eigenvalues bounded away from zero and infinity uniformly in n and $P \in \mathbf{P}$, it then follows that $\sup_{P \in \mathbf{P}} \|F_{\theta,n}\|_{L_P^2} \lesssim \sqrt{j_n}$. Therefore, by result (F.39) and Theorem 2.7.11 in van der Vaart and Wellner (1996) we can conclude that

$$\sup_{P \in \mathbf{P}} N_{[\,]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{L_P^2}) \lesssim \left(\frac{C_n \sqrt{j_n}}{\epsilon}\right)^{j_n}. \tag{F.41}$$

Hence, since $\int_0^a \log(M/u)du = a\log(M/a) + a$ and $C_n\sqrt{j_n} \uparrow \infty$, result (F.41) yields

$$\sup_{P \in \mathbf{P}} J_{[\,]}(\eta, \mathcal{F}_n, \|\cdot\|_{L_P^2}) \lesssim \int_0^\eta \{1 + j_n \log(\frac{C_n\sqrt{j_n}}{\epsilon})\}^{1/2} d\epsilon$$

$$\lesssim \sqrt{j_n} \int_0^\eta \log(\frac{C_n\sqrt{j_n}}{\epsilon})d\epsilon = \sqrt{j_n} \times \{\eta \log(\frac{C_n\sqrt{j_n}}{\eta}) + \eta\}. \tag{F.42}$$

In particular, since $F(X_i) = 1$ is an envelope for $\mathcal{F}_n$, we obtain by setting $\eta = 1$ in (F.42) that Assumption 3.3(iii) holds with $J_n \asymp \sqrt{j_n}\log(C_n j_n)$.

We next study the rate of convergence under the assumption that the model is identified and hence set $\tau_n = 0$ – sufficient conditions for identification are provided by Fox and Lazzati (2014). For any $(\pi, \delta) = \theta \in \mathbf{B}$, we then define the norm $\|\theta\|_{\mathbf{E}} = \|\pi\|_2 + \sup_{P \in \mathbf{P}} \|\delta\|_{L_P^2}$ and note that since the eigenvalues of $E_P[p_n^{j_n}(Y_i)p_n^{j_n}(Y_i)']$ were assumed to be bounded away from zero and infinity, Remark 4.2 implies that Assumption 4.2 holds with $\nu_n \asymp k_n^{1/2-1/r}$ provided that the smallest singular value of the matrix

$$E_P\left[\begin{pmatrix} q_n^{k_n}(Z_i) * \nabla_\pi M(Z_i, \theta) \\ q_n^{k_n}(Z_i) * \nabla_\delta M(Z_i, \theta)p_n^{j_n}(Y_i)' \end{pmatrix}\right] \tag{F.43}$$

is bounded away from zero uniformly in $\theta \in (\Theta_{0n}(P) \cap R)^\epsilon$, n, and $P \in \mathbf{P}_0$. Therefore, assuming $\|q_{k,n,\jmath}\|_{L_P^\infty}$ is uniformly bounded in k, n, $\jmath$, and $P \in \mathbf{P}$ for simplicity, we obtain that under Assumption 5.3(ii) the rate $\mathcal{R}_n$ delivered by Theorem 4.1 becomes

$$\mathcal{R}_n = \frac{\sqrt{k_n j_n \log(k_n)} \times \log(C_n j_n)}{\sqrt{n}}, \tag{F.44}$$

where we exploited $\nu_n \asymp k_n^{1/2-1/r}$ and that as previously argued $J_n \asymp \sqrt{j_n}\log(C_n j_n)$.

Corollary G.1 provides sufficient conditions for verifying Assumption 5.1, while the definition of $M(Z_i, \theta)$, equation (F.38), the mean value theorem, and the Cauchy Schwarz

inequality imply for any $\theta_1 = (\pi_1, p_n^{j_n\prime}\beta_1) \in \Theta_n$ and $\theta_2 = (\pi_2, p_n^{j_n\prime}\beta_2) \in \Theta_n$ that

$$
\begin{aligned}
E_P[\|\rho(X_i,\theta_1) - \rho(X_i,\theta_2)\|_2^2] &= E_P[\|M(Z_i,\theta_1) - M(Z_i,\theta_2)\|_2^2] \\
&\leq 6E_P[F_\pi(Z_i)^2\|\pi_1 - \pi_2\|_2^2 + F_\delta(Z_i)^2(p_n^{j_n}(Y_i)'(\beta_1 - \beta_2))^2] \lesssim \|\theta_1 - \theta_2\|_{\mathbf{E}}^2 , \quad \text{(F.45)}
\end{aligned}
$$

where in the final inequality we exploited that $\sup_{P\in\mathbf{P}} E_P[F_\pi(Z_i)^2] < \infty$ and $F_\delta(z)$ is uniformly bounded by hypothesis. Hence, we conclude from (F.45) that Assumption 5.2 holds with $\kappa_\rho = 1$, and combining (F.42) and (F.44) we obtain that

$$
\frac{j_n k_n^{1/r+1/2}\log(k_n)\log^2(C_n j_n n)}{\sqrt{n}} = o(a_n) \quad \text{(F.46)}
$$

implies Assumption 5.3(i) holds. Unlike in Examples 2.1 and 2.2, however, $\rho(X_i,\theta)$ is nonlinear in θ and hence Assumption 5.4 is harder to verify. To this end, recall $m_{P,J}(\theta) \equiv E_P[\rho_J(X_i,\theta)|Z_i]$, and for any $(\pi,\delta) = h \in \mathbf{B}$ define

$$
\nabla m_{P,J}(\theta)[h] = \nabla_\pi M_J(Z_i,\theta)\pi + \nabla_\delta M_J(Z_i,\theta)\delta(Y_i) , \quad \text{(F.47)}
$$

which we note satisfies Assumption 5.4(iii) with $M_m = \sup_{P\in\mathbf{P}}\|F_\pi\|_{L_P^2} + \|F_\delta\|_\infty$. Next, we suppose that for any $\theta_1,\theta_2 \in (\Theta_{0n}(P) \cap R)^\epsilon$ with $\theta_1 = (\pi_1,\delta_1)$ and $\theta_2 = (\pi_2,\delta_2)$

$$
\begin{aligned}
\|\nabla_\pi M(Z_i,\theta_1) - \nabla_\pi M(Z_i,\theta_2)\|_F &\leq G_\pi(Z_i)\|\pi_1 - \pi_2\|_2 + G_\delta\|\delta_1 - \delta_2\|_\infty \\
\|\nabla_\delta M(Z_i,\theta_1) - \nabla_\delta M(Z_i,\theta_2)\|_2 &\leq G_\delta\{\|\pi_1 - \pi_2\|_2 + \|\delta_1 - \delta_2\|_\infty\} \quad \text{(F.48)}
\end{aligned}
$$

for some functions G_π satisfying $\sup_{P\in\mathbf{P}} E_P[G_\pi(Z_i)^2] < \infty$ and a constant $G_\delta < \infty$ – a sufficient conditions is that $M(Z_i,\theta)$ be twice continuously differentiable with respect to θ and that such derivatives be uniformly bounded. Exploiting results (F.47) and (F.48), we then obtain by the mean value theorem that for any $(\pi,\delta) = h \in \mathbf{B}$

$$
\begin{aligned}
\|m_{P,J}(\theta + h) &- m_{P,J}(\theta) - \nabla m_{P,J}(\theta)[h]\|_{L_P^2} \\
&\leq \{\|\pi\|_2 + \sup_{P\in\mathbf{P}}\|\delta\|_{L_P^2}\} \times \{(G_\delta + \sup_{P\in\mathbf{P}}\|G_\pi\|_{L_P^2})\|\pi\|_2 + G_\delta \times \|\delta\|_\infty\} . \quad \text{(F.49)}
\end{aligned}
$$

Therefore, setting $K_m = G_\delta + \sup_{P\in\mathbf{P}}\|G_\pi\|_{L_P^2}$ we conclude that Assumption 5.4(i) holds with the norm $\|\theta\|_{\mathbf{L}} = \|\pi\|_2 + \|\delta\|_\infty$ for any $(\pi,\delta) = \theta \in \mathbf{B}$. Identical arguments as in (F.49) further verify Assumption 5.4(ii) for the same choice of K_m and $\|\cdot\|_{\mathbf{L}}$. Because $\|\cdot\|_{\mathbf{L}}$ in fact metrizes the product topology in $\mathbf{B}$, in this example we actually have $\mathbf{B} = \mathbf{L}$ and $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{\mathbf{L}}$. Moreover, since the smallest eigenvalue of $E_P[p_n^{j_n}(Y_i)p_n^{j_n}(Y_i)']$ was assumed to be bounded away from zero uniformly in $P \in \mathbf{P}$, we also obtain that

$$
\mathcal{S}_n(\mathbf{B},\mathbf{E}) = \mathcal{S}_n(\mathbf{L},\mathbf{E}) \lesssim \sup_{\beta\in\mathbf{R}^{j_n}} \frac{\|p_n^{j_n\prime}\beta\|_\infty}{\|\beta\|_2} \lesssim \sqrt{j_n} , \quad \text{(F.50)}
$$

where the final inequality applies when $\{p_{j,n}\}_{j=1}^{j_n}$ are Fourier, Spline, or Wavelet series since then $\sup_y \|p_n^{j_n}(y)\|_2 \lesssim \sqrt{j_n}$ (Belloni et al., 2015; Chen and Christensen, 2013). If $\{p_{j,n}\}_{j=1}^{j_n}$ are polynomial series instead, then (F.49) holds with j_n in place of $\sqrt{j_n}$.

Now turning to the construction of the bootstrap statistic, we let $(\hat{\pi}, \hat{\beta})$ be a minimizer of (F.36), and setting $\hat{\theta} = (\hat{\pi}, p_n^{j_n \prime} \hat{\beta})$ we then define

$$\hat{\mathbb{W}}_n \rho(\cdot, \hat{\theta}) * q_n^{k_n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \omega_i \{\rho(X_i, \hat{\theta}) * q_n^{k_n}(Z_i) - \frac{1}{n} \sum_{j=1}^n \rho(X_j, \hat{\theta}) * q_n^{k_n}(Z_j)\} \qquad \text{(F.51)}$$

for $\rho(X_i, \theta)$ as defined in (F.34). Since $\rho(X_i, \theta)$ is differentiable in θ we do not employ a numerical derivative, but instead follow Remark 6.1 and set for any $(\pi, \delta) = h \in \mathbf{B}_n$

$$\hat{\mathbb{D}}_n(\hat{\theta})[h] \equiv \frac{1}{n} \sum_{i=1}^n \{\nabla_\pi M(Z_i, \hat{\theta})\pi + \nabla_\delta M(Z_i, \hat{\theta})\delta(Y_i)\} * q_n^{k_n}(Z_i) . \qquad \text{(F.52)}$$

Next, we note that Assumption 6.1(i) holds with $K_b = 1$, while linearity of Υ_G implies Assumptions 6.2(i)-(ii) hold with $K_g = 0$, while Assumption 6.2(iii) is satisfied with $M_g = 1$ by direct calculation. In turn, since no equality restrictions are present in this problem, Assumptions 6.3 and 6.4 are not needed – formally they are automatically satisfied by setting $\mathbf{F} = \mathbf{R}$ and letting $\Upsilon_F(\theta) = 0$ for all $\theta \in \mathbf{B}$. Thus, here we have

$$G_n(\hat{\theta}) = \left\{ (\frac{\pi}{\sqrt{n}}, \frac{p_n^{j_n \prime} \beta}{\sqrt{n}}) \in \mathbf{B}_n : \frac{p_n^{j_n}(y)' \beta}{\sqrt{n}} \leq \max\{0, -p_n^{j_n}(y)' \hat{\beta} - r_n\} \text{ for all } y \right\} . \qquad \text{(F.53)}$$

Hence, since $K_g = K_f = 0$, according to Remark 6.4 we may set $\hat{V}_n(\hat{\theta}, \ell_n)$ to equal

$$\hat{V}_n(\hat{\theta}, \ell_n) = \{ \frac{h}{\sqrt{n}} \in \mathbf{B}_n : \frac{h}{\sqrt{n}} \in G_n(\hat{\theta}_n) \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{E}} \leq \ell_n \} . \qquad \text{(F.54)}$$

Finally, we observe that Theorem H.1 provides sufficient conditions for Assumption 6.5, while Assumption 6.6(i) is automatically satisfied since $\tau_n = 0$. Moreover, exploiting results (F.42), (F.44), and (F.50), it follows that Assumption 6.6(ii) reduces to

$$\ell_n \times \{(j_n n)^{1/4} \vee \sqrt{j_n k_n \log(k_n)} \log(\frac{C_n j_n}{\ell_n})\} = o(a_n) \qquad \text{(F.55)}$$

and Assumption 6.6(iii)-(iv) are satisfied whenever $j_n \sqrt{k_n \log(k_n)} \log(C_n j_n) = o(\sqrt{n} r_n)$. Furthermore, since the eigenvalues of (F.40) have been assumed to be bounded away from zero and infinity, it follows that $\|h\|_{\mathbf{E}} \asymp \|\pi\|_2 \vee \|\beta\|_2$ uniformly in $(\pi, p_n^{j_n \prime} \beta) = h \in$

$\mathbf{B}_n$ and n. Therefore, from results (F.53) and (F.54), the bootstrap statistic equals

$$\hat{U}_n(R) = \inf_{(\pi,\beta)} \|\hat{\mathbb{W}}_n \rho(\cdot,\hat{\theta}) * q_n^{k_n} + \hat{\mathbb{D}}_n(\hat{\theta})[(\pi, p_n^{j_n\prime}\beta)]\|_{\hat{\Sigma}_n, r}$$

$$\text{s.t. (i)} \ \frac{p_n^{j_n}(y)'\beta}{\sqrt{n}} \leq \max\{0, -p_n^{j_n}(y)'\hat{\beta} - r_n\} \ \forall y, \ \text{(ii)} \ \|\pi\|_2 \vee \|\beta\|_2 \leq \sqrt{n}\ell_n \ . \quad \text{(F.56)}$$

Alternatively, under slightly stronger requirements, it is possible to appeal to Lemma 6.1 to conclude that the bandwidth ℓ_n is unnecessary – i.e. the second constraint in (F.56) can be ignored. To this end, we note that for any $(\pi, p_n^{j_n\prime}\beta) = h \in \mathbf{B}_n$, we have

$$\mathbb{D}_{n,P}(\theta)[h] = E_P[\{\nabla_\pi M(Z_i, \theta)\pi + \nabla_\delta M(Z_i, \theta)p_n^{j_n}(Y_i)'\beta\} * q_n^{k_n}(Z_i)] \ . \quad \text{(F.57)}$$

Since $\nu_n \asymp k_n^{\frac{1}{2}-\frac{1}{r}}$, $\|a\|_2 \leq k_n^{\frac{1}{2}-\frac{1}{r}}\|a\|_r$ for any $a \in \mathbf{R}^{k_n}$, and we assumed the smallest singular value of (F.43) and the largest eigenvalue of (F.40) are respectively bounded away from zero and infinity uniformly in $\theta \in (\Theta_{0n}(P) \cap R)^\epsilon$, n, and $P \in \mathbf{P}_0$, we obtain

$$\|h\|_{\mathbf{E}} \leq \nu_n \|\mathbb{D}_{n,P}(\theta)[h]\|_r \quad \text{(F.58)}$$

for any $\theta \in (\Theta_{0n}(P) \cap R)^\epsilon$, $P \in \mathbf{P}_0$, and $h \in \mathbf{B}_n$. In order to verify (74), we define the class $\mathcal{G}_{\delta,n} \equiv \{g : g(z) = \nabla_\delta M_\jmath(z,\theta)q_{k,n,\jmath}(z)p_n^{j_n}(y)'\beta$ for some $\theta \in \Theta_n$, $\|\beta\|_2 \leq 1$, $1 \leq j \leq j_n$, $1 \leq k \leq k_n$, and $1 \leq \jmath \leq \mathcal{J}\}$. Since $\sup_y \|p_n^{j_n}(y)\|_2 \lesssim \sqrt{j_n}$, as exploited in (F.50), and $\|q_{k,n,\jmath}\|_{L_P^\infty}$ was assumed to be uniformly bounded, it follows from (F.38) that $F_\delta \sqrt{j_n}K_0$ is an envelope for $\mathcal{G}_{\delta,n}$ for K_0 sufficiently large. Therefore, it follows that

$$E_P[\sup_{g \in \mathcal{G}_{\delta,n}} |\mathbb{G}_n g|] \lesssim J_{[]}(\|\sqrt{j_n}F_\delta K_0\|_{L_P^2}, \mathcal{G}_{n,\delta}, \|\cdot\|_{L_P^2}) \quad \text{(F.59)}$$

by Theorem 2.14.2 in van der Vaart and Wellner (1996). Furthermore, exploiting once again that $\sup_y \|p_n^{j_n}(y)\|_2 \lesssim \sqrt{j_n}$ and condition (F.48) we can conclude

$$|\nabla_\delta M_\jmath(Z_i, \theta_1)q_{k,n,\jmath}(Z_i)p_n^{j_n}(Y_i)'\tilde{\beta}_1 - \nabla_\delta M_\jmath(Z_i, \theta_2)q_{k,n,\jmath}(Z_i)p_n^{j_n}(Y_i)'\tilde{\beta}_2|$$
$$\lesssim \|F_\delta\|_\infty \|p_n^{j_n\prime}(\tilde{\beta}_1 - \tilde{\beta}_2)\|_\infty + \{\|\pi_1 - \pi_2\|_2 + \|p_n^{j_n\prime}(\beta_1 - \beta_2)\|_\infty\}\|p_n^{j_n\prime}\tilde{\beta}_2\|_\infty$$
$$\lesssim \sqrt{j_n}\|\tilde{\beta}_1 - \tilde{\beta}_2\|_2 + j_n\{\|\pi_1 - \pi_2\|_2 + \|\beta_1 - \beta_2\|_2\} \ . \quad \text{(F.60)}$$

for any $\theta_1 = (\pi_1, p_n^{j_n\prime}\beta_1)$ and $\theta_2 = (\pi_2, p_n^{j_n\prime}\beta_2)$. Thus, result (F.60), Theorem 2.7.11 in van der Vaart and Wellner (1996), and arguing as in (F.41) and (F.42) implies $\sup_{P \in \mathbf{P}} J_{[]}(\|\sqrt{j_n}F_\delta\|_{L_P^2}, \mathcal{G}_{\delta,n}, \|\cdot\|_{L_P^2}) = O(j_n \log(C_n j_n))$. Similarly, if we define $\mathcal{G}_{\pi,n} \equiv \{g : g(z) = p_{j,n}(y)q_{k,n,\jmath}(z)\nabla_\pi M_\jmath(z,\theta)\pi$ for some $1 \leq j \leq j_n$, $1 \leq k \leq k_n$, $1 \leq \jmath \leq \mathcal{J}$, $\theta \in \Theta_n$ and $\|\pi\|_2 \leq 1\}$, then by analogous arguments we can conclude

$$\sup_{P \in \mathbf{P}} E_P[\sup_{g \in \mathcal{G}_{\pi,n}} |\mathbb{G}_n g|] = O(j_n \log(C_n j_n)) \ . \quad \text{(F.61)}$$

Thus, employing the above results together with Markov's inequality finally implies that

$$\sup_{\theta \in \Theta_n} \sup_{h \in \mathbf{B}_n : \|h\|_{\mathbf{E}} = 1} \|\hat{\mathbb{D}}_n(\theta)[h] - \mathbb{D}_{n,P}(\theta)[h]\|_r$$

$$\lesssim \frac{\sqrt{k_n}}{\sqrt{n}} \times \{ \sup_{g \in \mathcal{G}_{\pi,n}} |\mathbb{G}_n g| + \sup_{g \in \mathcal{G}_{\delta,n}} |\mathbb{G}_n g| \} = O_p(\frac{\sqrt{k_n} j_n \log(C_n j_n)}{\sqrt{n}}) . \quad \text{(F.62)}$$

Exploiting (F.58) and (F.62), it then follows that the conditions of Lemma 6.1 are satisfied and constraint (ii) in (F.56) can be ignored if ℓ_n can be chosen to simultaneously satisfy Assumption 6.6 and $\ell_n = o(\mathcal{R}_n)$. These requirements are met provided that

$$\frac{k_n j_n^{3/2} \log(k_n) \log^2(C_n j_n)}{\sqrt{n}} = o(a_n) , \quad \text{(F.63)}$$

which represents a strengthening of (F.46) – note strengthening (F.46) to (F.63) potentially makes Assumption 5.3(ii) less tenable. Hence, under (F.63) we may set

$$\hat{U}_n(R) = \inf_{(\pi,\beta)} \|\hat{\mathbb{W}}_n \rho(\cdot, \hat{\theta}) * q_n^{k_n} + \hat{\mathbb{D}}_n(\hat{\theta})[(\pi, p_n^{j_{n'}}\beta)]\|_{\hat{\Sigma}_n, r}$$

$$\text{s.t.} \quad \frac{p_n^{j_n}(y)'\beta}{\sqrt{n}} \le \max\{0, -p_n^{j_n}(y)'\hat{\beta} - r_n\} \text{ for all } y , \quad \text{(F.64)}$$

i.e. the second constraint in (F.56) may be ignored when computing the bootstrap critical values. ∎

Discussion of Example 2.4

An observation i in this example was assumed to consist of an instrument $Z_i \in \mathbf{R}^{d_z}$ and a pair of individuals $j \in \{1, 2\}$ for whom we observe the hospital H_{ij} in the network $\mathcal{H}$ that they were referred to, as well as for all hospitals $h \in \mathcal{H}$ the cost of treatment $P_{ij}(h)$ and distance to hospital $D_{ij}(h)$. Ho and Pakes (2014) then derive

$$E[\sum_{j=1}^{2} \{\gamma_0(P_{ij}(H_{ij}) - P_{ij'}(H_{ij'})) + g_0(D_{ij}(H_{ij})) - g_0(D_{ij}(H_{ij'}))\}|Z_i] \le 0 \quad \text{(F.65)}$$

where $\gamma_0 \in \mathbf{R}$, $g_0 : \mathbf{R}_+ \to \mathbf{R}_+$ is an unknown monotonic function, and $j' \equiv \{1, 2\} \setminus \{j\}$. For notational simplicity, we let $X_i = (\{\{P_{ij}(h), D_{ij}(h)\}_{h \in \mathcal{H}}, H_{ij}\}_{j=1}^2, Z_i)$ and define

$$\psi(X_i, \gamma, g) \equiv \sum_{j=1}^{2} \{\gamma(P_{ij}(H_{ij}) - P_{ij'}(H_{ij'})) + g(D_{ij}(H_{ij})) - g(D_{ij}(H_{ij'}))\} . \quad \text{(F.66)}$$

In addition, we assume that the supports of $D_{ij}(H_{ij})$ and $D_{ij}(H_{ij'})$ are contained in a bounded set uniformly in $P \in \mathbf{P}$ and $j \in \{1, 2\}$, which we normalize to $[0, 1]$. Finally,

recall that in this example γ_0 is the parameter of interest, and that we aim to test

$$H_0 : \gamma_0 = c_0 \qquad H_1 : \gamma_0 \neq c_0 \tag{F.67}$$

employing the moment restriction (F.65) while imposing monotonicity of $g_0 : \mathbf{R}_+ \to \mathbf{R}_+$.

In order to map this problem into our framework, we follow the discussion of Example 2.4 in the main text – see equations (21)-(22) – and rewrite restriction (F.65) as

$$E[\psi(X_i, \gamma_0, g_0)|Z_i] + \lambda_0(Z_i) = 0 , \tag{F.68}$$

for a function λ_0 satisfying $\lambda_0(Z_i) \geq 0$. We further define the Hilbert space L_U^2 by

$$L_U^2 \equiv \{f : [0,1] \to \mathbf{R} : \|f\|_{L_U^2} < \infty \text{ for } \|f\|_{L_U^2}^2 \equiv \int_0^1 f^2(u) du\} \tag{F.69}$$

and note $C^1([0,1]) \subset L_U^2$. The parameter in this example is thus $\theta_0 = (\gamma_0, g_0, \lambda_0)$ which we view as an element of $\mathbf{B} = \mathbf{R} \times C^1([0,1]) \times \ell^\infty(\mathbf{R}^{d_z})$, and set as the residual

$$\rho(X_i, \theta) = \psi(X_i, \gamma, g) + \lambda(Z_i) \tag{F.70}$$

for any $\theta = (\gamma, g, \lambda)$. The hypothesis in (F.67) can then be tested by letting $\mathbf{F} = \mathbf{R}$ and $\Upsilon_F(\theta) = \gamma - c_0$ for any $\theta \in \mathbf{B}$, and imposing the monotonicity constraint on g and positivity restriction on λ by setting $\mathbf{G} = \ell^\infty([0,1]) \times \ell^\infty(\mathbf{R}^{d_z})$ and $\Upsilon_G(\theta) = -(g', \lambda)$ for any $\theta \in \mathbf{B}$. Finally, as in Example 2.3 we utilize a noncompact parameter space and let $\Theta = \mathbf{B}$, which together with the preceding discussion verifies the general structure of our framework imposed in Assumptions 2.1 and 2.2.

To build the test statistic $I_n(R)$, we let $\{A_{k,n}\}_{k=1}^{k_n}$ denote a triangular array of partitions of $\mathbf{R}^{d_z}$, and set $q_n^{k_n}(z) \equiv (1\{z \in A_{1,n}\}, \ldots, 1\{z \in A_{k_n,n}\})'$. For ease of computation, it is convenient to also employ $\{1\{z \in A_{k,n}\}\}_{k=1}^{k_n}$ as a sieve for $\ell^\infty(\mathbf{R}^{d_z})$ and thus for a sequence $C_n \uparrow \infty$ we approximate $\ell^\infty(\mathbf{R}^{d_z})$ by the set

$$\Lambda_n \equiv \{\lambda \in \ell^\infty(\mathbf{R}^{d_z}) : \lambda = q_n^{k_n\prime}\pi \text{ for some } \|\pi\|_2 \leq C_n\} . \tag{F.71}$$

In turn, for $\{p_{j,n}\}_{j=1}^{j_n}$ a triangular array of orthonormal functions in L_U^2 such as b-splines or wavelets, and $p_n^{j_n}(u) = (p_{1,n}(u), \ldots, p_{j_n,n}(u))'$ we let the sieve for $\mathbf{B}$ be given by

$$\Theta_n \equiv \{(\gamma, g, \lambda) \in \mathbf{B} : g = p_n^{j_n\prime}\beta, \ \|\beta\|_2 \leq C_n, \ \lambda \in \Lambda_n\} . \tag{F.72}$$

Given the stated parameter choices, the test statistic $I_n(R)$ is then equivalent to

$$I_n(R) = \inf_{(\gamma,\beta,\pi)} \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho(X_i, (\gamma, p_n^{j_n\prime}\beta, q_n^{k_n\prime}\pi)) * q_n^{k_n}(Z_i)\|_{\hat{\Sigma}_{n,r}}$$

$$\text{s.t. (i) } \|\beta\|_2 \le C_n, \text{ (ii) } \gamma = c_0, \text{ (iii) } \nabla p_n^{j_n\prime}\beta \ge 0, \text{ (iv) } \pi \ge 0 , \quad \text{(F.73)}$$

where restriction (i) imposes that $\theta \in \Theta_n$ and the requirement $\|\pi\|_2 \le C_n$ can be ignored as argued in Remark 6.3, and restrictions (ii) and (iii)-(iv) respectively enforce the equality ($\Upsilon_F(\theta) = 0$) and inequality ($\Upsilon_G(\theta) \le 0$) constraints. While we introduce (F.73) due to its link to our general formulation in (26), it is actually more convenient to work with a profiled version of $I_n(R)$. Specifically, the choice of sieve Λ_n enables us to easily profile the optimal $\pi \in \mathbf{R}^{k_n}$ in (F.73) for any given choice of (γ, β) leading to

$$I_n(R) = \inf_{(\gamma,\beta)} \|\{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i, \gamma, p_n^{j_n\prime}\beta) * q_n^{k_n}(Z_i)\} \vee 0\|_{\hat{\Sigma}_{n,r}}$$

$$\text{s.t. (i) } \|\beta\|_2 \le C_n, \text{ (ii) } \gamma = c_0, \text{ (iii) } \nabla p_n^{j_n\prime}\beta \ge 0 . \quad \text{(F.74)}$$

The ability to profile out the nuisance parameter λ grants this problem an additional structure that enables us to weaken some of our assumptions. In particular, the rate of convergence of the minimizers of $I_n(R)$ in (F.74) is better studied through direct arguments rather than a reliance on Theorem 4.1. To this end, let

$$\Gamma_{0n}(P) \cap R \equiv \{(\gamma, g) : (\gamma, g, \lambda) \in \Theta_{0n}(P) \cap R \text{ for some } \lambda \in \ell^\infty(\mathbf{R}^{d_z})\}$$

$$\Gamma_n \cap R \equiv \{(\gamma, g) : (\gamma, g, \lambda) \in \Theta_n \cap R \text{ for some } \lambda \in \ell^\infty(\mathbf{R}^{d_z})\} \quad \text{(F.75)}$$

denote the profiled set $\Theta_{0n}(P) \cap R$ and profiled sieve $\Theta_n \cap R$. For each $(\gamma, g) \in \Gamma_n \cap R$ we then denote the corresponding population and sample profiled λ by

$$\lambda_{n,P}^{(\gamma,g)}(z) \equiv -\sum_{k=1}^{k_n} 1\{z \in A_{k,n}\}\{E_P[\psi(X_i, \gamma, g)|Z_i \in A_{k,n}] \wedge 0\} \quad \text{(F.76)}$$

$$\hat{\lambda}_n^{(\gamma,g)}(z) \equiv -\sum_{k=1}^{k_n} 1\{z \in A_{k,n}\}\{\frac{\sum_{i=1}^{n} 1\{Z_i \in A_{k,n}\}\psi(X_i, \gamma, g)}{\sum_{i=1}^{n} 1\{Z_i \in A_{k,n}\}} \wedge 0\} , \quad \text{(F.77)}$$

and observe that $\Theta_{0n}(P) \cap R = \{(\gamma, g, \lambda_{n,P}^{(\gamma,g)}) : (\gamma, g) \in \Gamma_{0n}(P) \cap R\}$. Therefore, defining

$$\mathcal{P}_n(\gamma, g) \equiv \|\{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i, \gamma, g) * q_n^{k_n}(Z_i)\} \vee 0\|_{\hat{\Sigma}_{n,r}} , \quad \text{(F.78)}$$

we can then construct a set estimator for the profiled identified set $\Gamma_{0n}(P) \cap R$ by setting

$$\hat{\Gamma}_n \cap R \equiv \{(\gamma, g) \in \Gamma_n \cap R : \mathcal{P}_n(\gamma, g) \leq \inf_{\theta \in \Theta_n \cap R} Q_n(\theta) + \tau_n\} \qquad \text{(F.79)}$$

for some $\tau_n \downarrow 0$. Next, we note that the collection of transformations $\{q_{k,n}\}_{k=1}^{k_n}$ is orthogonal in L_P^2, yet not orthonormal. In order to normalize them, we suppose that

$$\frac{1}{k_n} \asymp \inf_{P \in \mathbf{P}} \inf_{1 \leq k \leq k_n} P(Z_i \in A_{k,n}) \asymp \sup_{P \in \mathbf{P}} \sup_{1 \leq k \leq k_n} P(Z_i \in A_{k,n}) \qquad \text{(F.80)}$$

which implies $\|q_{k,n}\|_{L_P^2} \asymp k_n^{-1/2}$ uniformly in $P \in \mathbf{P}$, $1 \leq k \leq k_n$, and n. Following the discussion of Examples 2.1, 2.2, and 2.3 we further impose the condition

$$\vec{d}_H((\gamma, g), \Gamma_{0n}(P) \cap R, \| \cdot \|_2 + \| \cdot \|_{L_U^2})$$
$$\lesssim k_n^{1-\frac{1}{r}} \{\|E_P[\psi(X_i, \gamma, g) * q_n^{k_n}(Z_i)] \vee 0\|_r + O(\zeta_n)\} . \quad \text{(F.81)}$$

Defining $\mathcal{G}_n \equiv \{\psi(x, \gamma, g) q_{k,n}(z) : (\gamma, g) \in \Gamma_n \cap R, \ 1 \leq k \leq k_n\}$, next suppose that

$$\|g\|_{L_U^2}^2 \asymp E_P[g^2(D_{ij}(H_{ij}))] \asymp E_P[g^2(D_{ij}(H_{ij'}))] \qquad \text{(F.82)}$$

uniformly in $g \in L_U^2$, $j \in \{1, 2\}$, and $P \in \mathbf{P}$. If in addition $E_P[P_{ij}^2(H_{ij}) + P_{ij}^2(H_{ij'})]$ is bounded uniformly in $P \in \mathbf{P}$ and $j \in \{1, 2\}$, then $\mathcal{G}_n$ has an envelope G_n satisfying $\sup_{P \in \mathbf{P}} \|G_n\|_{L_P^2} \lesssim C_n$. Arguing as in (F.42) it is then possible to show

$$\sup_{P \in \mathbf{P}} J_{[]}(C_n, \mathcal{G}_n, \| \cdot \|_{L_P^2}) \lesssim C_n \sqrt{j_n \log(k_n)} , \qquad \text{(F.83)}$$

and hence letting $\bar{\mathcal{P}}_{n,P}(\gamma, g) \equiv \|\sqrt{n} E_P[\psi(X_i, \gamma, g) * q_n^{k_n}(Z_i)] \vee 0\|_{\hat{\Sigma}_n, r}$ for any $(\gamma, g) \in \mathbf{R} \times L_U^2$ we obtain by Theorem 2.14.2 in van der Vaart and Wellner (1996) that

$$\sup_{(\gamma, g) \in \Gamma_n \cap R} |\mathcal{P}_n(\gamma, g) - \bar{\mathcal{P}}_{n,P}(\gamma, g)| = O_p(k_n^{1/r} C_n \sqrt{j_n \log(k_n)}) \qquad \text{(F.84)}$$

uniformly in $P \in \mathbf{P}$ since $\|\hat{\Sigma}_n\|_{o,r} = O_p(1)$ uniformly in $P \in \mathbf{P}$ by Lemma B.3. Under (F.81) and (F.84) the proof of Theorem 4.1 applies without changes, and therefore under the no-bias condition of Assumption 5.3(ii) we obtain uniformly in $P \in \mathbf{P}$

$$\vec{d}_H(\hat{\Gamma}_n \cap R, \Gamma_{0n}(P) \cap R, \| \cdot \|_2 + \| \cdot \|_{L_U^2}) = O_p(\frac{k_n C_n \sqrt{j_n \log(k_n)}}{\sqrt{n}} + k_n^{1-\frac{1}{r}} \tau_n) . \quad \text{(F.85)}$$

Moreover, it also follows from (F.82) that for any $(\gamma_1, g_1), (\gamma_2, g_2) \in \Gamma_n \cap R$ we have

$$\sup_{P \in \mathbf{P}} \|\lambda_{n,P}^{(\gamma_1, g_1)} - \lambda_{n,P}^{(\gamma_2, g_2)}\|_{L_P^2} \lesssim \|g_1 - g_2\|_{L_U^2} . \qquad \text{(F.86)}$$

102

In addition, by standard arguments – see e.g. Lemmas 2.2.9 and 2.2.10 in van der Vaart and Wellner (1996) – it also follows from (F.80) that uniformly in $P \in \mathbf{P}$ we have

$$\max_{1 \le k \le k_n} |\frac{1}{n} \sum_{i=1}^{n} \frac{1\{Z_i \in A_{k,n}\}}{P(Z_i \in A_{k,n})} - 1| = O_p(\frac{\sqrt{k_n \log(k_n)}}{\sqrt{n}}) , \tag{F.87}$$

while under (F.83) Theorem 2.14.2 in van der Vaart and Wellner (1996) implies that

$$\sup_{f \in \mathcal{G}_n} |\frac{1}{n} \sum_{i=1}^{n} f(X_i) - E_P[f(X_i)]| = O_p(\frac{C_n \sqrt{j_n \log(k_n)}}{\sqrt{n}}) . \tag{F.88}$$

Thus, combining (F.80), (F.87), (F.88) with the definitions in (F.76) and (F.77) yields

$$\sup_{(\gamma,g) \in \Gamma_n \cap R} \sup_{P \in \mathbf{P}} \|\hat{\lambda}_n^{(\gamma,g)} - \lambda_{n,P}^{(\gamma,g)}\|_{L_P^2} = O_p(\frac{k_n C_n \sqrt{(k_n \vee j_n) \log(k_n)}}{\sqrt{n}}) . \tag{F.89}$$

Hence, setting $\|\theta\|_{\mathbf{E}} = \|\gamma\|_2 + \|g\|_{L_U^2} + \sup_{P \in \mathbf{P}} \|\lambda\|_{L_P^2}$ for any $(\gamma, g, \lambda) = \theta \in \mathbf{B}$, and

$$\hat{\Theta}_n \cap R = \{(\gamma, g, \hat{\lambda}_n^{(\gamma,g)}) : (\gamma, g) \in \hat{\Gamma}_n \cap R\} , \tag{F.90}$$

we finally obtain from $\Theta_{0n}(P) \cap R = \{(\gamma, g, \lambda_{n,P}^{(\gamma,g)}) : (\gamma, g) \in \Gamma_{0n}(P) \cap R\}$, results (F.85), (F.86), and (F.89) that uniformly in $P \in \mathbf{P}$ we have that

$$\overrightarrow{d}_H(\hat{\Theta}_n \cap R, \Theta_{0n}(P) \cap R, \| \cdot \|_{\mathbf{E}}) = O_p(\frac{k_n C_n \sqrt{(k_n \vee j_n) \log(k_n)}}{\sqrt{n}} + k_n^{1-\frac{1}{r}} \tau_n) . \tag{F.91}$$

For the rest of the following discussion, we thus let $\nu_n = k_n^{1-\frac{1}{r}}$ and set $\mathcal{R}_n$ to equal

$$\mathcal{R}_n = \frac{k_n C_n \sqrt{(k_n \vee j_n) \log(k_n)}}{\sqrt{n}} ; \tag{F.92}$$

compare to result (48) in the main text.

With regards to Section 5, we refer to Corollary G.1 for sufficient conditions for Assumption 5.1, while Assumption 5.2 is satisfied with $\kappa_\rho = 1$ and some $K_\rho < \infty$ since $\|\theta\|_{\mathbf{E}} = \|\gamma\|_2 + \|g\|_{L_U^2} + \sup_{P \in \mathbf{P}} \|\lambda\|_{L_P^2}$ for any $(\gamma, g, \lambda) = \theta \in \mathbf{B}$ and we assumed (F.82) holds. Moreover, from (F.72) and arguing as in (F.42) we calculate

$$\sup_{P \in \mathbf{P}} J_{[]}(\eta, \mathcal{F}_n, \| \cdot \|_{L_P^2}) \lesssim \sqrt{j_n \vee k_n}(\eta \log(\frac{C_n}{\eta}) + \eta) , \tag{F.93}$$

which together with (F.92) implies that a sufficient conditions for Assumption 5.3(i) is

$$\frac{k_n^{1+\frac{1}{r}}(j_n \vee k_n)C_n \log^2(n)}{\sqrt{n}} = o(a_n) . \tag{F.94}$$

In turn, we note the definitions of $\psi(X_i, \gamma, g)$ and $\rho(X_i, \theta)$ in (F.66) and (F.70) imply

$$\nabla m_P(\theta)[h] = E_P[\psi(X_i, \gamma, g)|Z_i] + \lambda(Z_i) , \tag{F.95}$$

and hence Assumption 5.4(i)-(ii) holds with $K_m = 0$, while Assumption 5.4(iii) is satisfied for some $M_m < \infty$ since we assumed that $E_P[P_{ij}^2(H_{ij}) + P_{ij}^2(H_{ij'})]$ is bounded uniformly in $P \in \mathbf{P}$ and imposed condition (F.82).

Turning to the construction of the bootstrap statistic, we first recall that any $\theta \in \hat{\Theta}_n \cap R$ is of the form $\theta = (\gamma, g, \hat{\lambda}_n^{(\gamma,g)})$ for some $(\gamma, g) \in \hat{\Gamma}_n \cap R$ – see (F.90). Therefore,

$$\hat{\mathbb{W}}_n \rho(\cdot, \theta) * q_n^{k_n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \omega_i \Big\{ (\psi(X_i, \gamma, g) + \hat{\lambda}_n^{(\gamma,g)}(Z_i)) * q_n^{k_n}(Z_i)$$
$$- (\frac{1}{n} \sum_{j=1}^n \psi(X_j, \gamma, g) * q_n^{k_n}(Z_j)) \vee 0 \Big\} \tag{F.96}$$

for any $(\gamma, g, \lambda) = \theta \in \hat{\Theta}_n \cap R$. Next, also note that for any $\theta \in \mathbf{B}$ and $h = (\gamma, g, \lambda)$, definitions (F.66) and (F.70) imply that the estimator $\hat{\mathbb{D}}_n(\theta)[h]$ (as in (63)) is equal to

$$\hat{\mathbb{D}}_n(\theta)[h] = \frac{1}{n} \sum_{i=1}^n \{\psi(X_i, \gamma, g) + \lambda(Z_i)\} * q_n^{k_n}(Z_i) . \tag{F.97}$$

We further observe that if we metrize the topology on $\mathbf{B} = \mathbf{R} \times C^1([0,1]) \times \ell^\infty(\mathbf{R}^{d_z})$ by $\|\theta\|_{\mathbf{B}} = \|\gamma\|_2 \vee \|g\|_{1,\infty} \vee \|\lambda\|_\infty$ for any $\theta = (\gamma, g, \lambda) \in \mathbf{B}$, then Assumption 6.1(i) holds with $K_b = 3$. In turn, we also note that since $\Upsilon_G : \mathbf{B} \to \ell^\infty([0,1]) \times \ell^\infty(\mathbf{R}^{d_z})$ is linear and continuous, Assumptions 6.2(i)-(iii) are satisfied with $K_g = 0$, some finite M_g, and $\nabla \Upsilon_G(\theta)[h] = -(g', \lambda)$ for any $\theta \in \mathbf{B}$ and $h = (\gamma, g, \lambda)$. Similarly, since $\Upsilon_F : \mathbf{B} \to \mathbf{R}$ is affine and $\nabla \Upsilon_F(\theta)[h] = \gamma$ for any $\theta \in \mathbf{B}$ and $h = (\gamma, g, \lambda)$, Assumption 6.4 is automatically satisfied, while Assumption 6.3 holds with $K_f = 0$ and $M_f = 1$.

Writing each $\hat{\theta} \in \hat{\Theta}_n \cap R$ in the form $\hat{\theta} = (\hat{\gamma}, \hat{g}, q_n^{k_n'}\hat{\pi})$, we then finally obtain that

$$\hat{U}_n(R) = \inf_{\hat{\theta} \in \hat{\Theta}_n \cap R} \inf_{(\gamma,\beta,\pi)} \|\hat{\mathbb{W}}_n \rho(\cdot, \hat{\theta}) * q_n^{k_n} + \hat{\mathbb{D}}_n(\hat{\theta})[(\gamma, p_n^{j_n'}\beta, q_n^{k_n'}\pi)]\|_{\hat{\Sigma}_n, r}$$

s.t. (i) $\gamma = 0$,   (ii) $\frac{\pi}{\sqrt{n}} \geq 0 \wedge (r_n - \hat{\pi})$

     (iii) $\frac{\nabla p_n^{j_n'}\beta}{\sqrt{n}} \geq 0 \wedge (r_n - \hat{g}')$,   (iv) $\|\frac{p_n^{j_n'}\beta}{\sqrt{n}}\|_{1,\infty} \vee \|\frac{\pi}{\sqrt{n}}\|_\infty \leq \ell_n$ (F.98)

where constraint (i) corresponds to $\Upsilon_F(\hat{\theta} + h/\sqrt{n}) = 0$, the restrictions in (ii) and (iii) impose $h/\sqrt{n} \in G_n(\hat{\theta})$, and the constraint in (iv) demand $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ – compare to the definition of $\hat{V}_n(\theta, \ell_n)$ in (69). As in Section 7, constraint (iii) reduces to a finite number of linear constraints when employing orthonormalized b-Splines of Order 3 as

the basis $\{p_{j,n}\}_{j=1}^{j_n}$. Moreover, under such a choice of sieve we further have

$$\|p_n^{j_n\prime}\beta\|_{1,\infty} \lesssim j_n^{3/2}\|\beta\|_2 , \tag{F.99}$$

(see Newey (1997)) and thus exploiting $\|\beta\|_2 \leq j_n^{1/2}\|\beta\|_\infty$ it may be preferable for easy of computation to replace the constraint $\|p_n^{j_n\prime}\beta\|_{1,\infty} \leq \sqrt{n}\ell_n$ in (F.98) by the more conservative but linear constraints $\|\beta\|_\infty \leq \sqrt{n}\ell_n/j_n^2$.

We refer to Theorem H.1 for sufficient conditions for Assumption 6.5 to hold, and focus on the rate requirements imposed by Assumption 6.6. To this end, we first observe

$$\sup_{\pi\in\mathbf{R}^{k_n}} \frac{\|q_n^{k_n\prime}\pi\|_\infty}{\sup_{P\in\mathbf{P}}\|q_n^{k_n\prime}\pi\|_{L_P^2}} \lesssim \sup_{\pi\in\mathbf{R}^{k_n}} \frac{\|q_n^{k_n\prime}\pi\|_\infty}{\|\pi\|_2/\sqrt{k_n}} = \sqrt{k_n} \tag{F.100}$$

where in the second inequality we exploited (F.80) implies that $\sup_{P\in\mathbf{P}}\|q_n^{k_n\prime}\pi\|_{L_P^2} \asymp \|\pi\|_2/\sqrt{k_n}$, and the final inequality follows from $\|q_n^{k_n\prime}\pi\|_\infty = \|\pi\|_\infty$ due to $\{A_{k,n}\}_{k=1}^{k_n}$ being a partition of $\mathbf{R}^{d_z}$. Hence, results (F.99) and (F.100) together imply that

$$\mathcal{S}_n(\mathbf{B},\mathbf{E}) \lesssim j_n^{3/2} \vee k_n^{1/2} . \tag{F.101}$$

We think it advisable to set $\tau_n = 0$ which automatically satisfies Assumption 6.6 and is simpler to implement, though we note that in contrast to Examples 2.1, 2.2, and 2.3 such a choice may lead to a loss of power. We further note that the rate of convergence derived in (F.92) and the bound in (F.101) imply Assumption 6.6(iii) is satisfied provided

$$\frac{k_n C_n \sqrt{(k_n \vee j_n)\log(k_n)}(j_n^{3/2} \vee k_n^{1/2})}{\sqrt{n}} = o(r_n) . \tag{F.102}$$

Finally, we note (F.93) implies the conditions imposed on ℓ_n by Assumption 6.6 become

$$\ell_n(k_n^{1/r}\sqrt{(j_n \vee k_n)\log(k_n)}\log(C_n) = o(a_n) . \tag{F.103}$$

In parallel to Examples 2.1, 2.2, and 2.3 it may be possible to establish under additional conditions that the bandwidth ℓ_n is not necessary – i.e. that constraint (iv) may be dropped in (F.100). Unfortunately, such a conclusion cannot be reached by applying Lemma 6.1 due to a failure of the condition $\|h\|_{\mathbf{E}} \leq \nu_n\|\mathbb{D}_{n,P}(\theta_0)[h]\|_r$ for all $\theta_0 \in (\Theta_{0n}(P) \cap R)^\epsilon$ and $h \in \sqrt{n}\{\mathbf{B}_n \cap R - \theta\}$, which is required by Lemma 6.1. ∎

Appendix G - Uniform Coupling Results

In this Appendix we develop uniform coupling results for empirical processes that help verify Assumption 5.1 in specific applications. The results are based on the Hungarian construction of Massart (1989) and Koltchinskii (1994), and are stated in a notation

that abstracts from the rest of the paper due to the potential independent interest of the results. Thus, in this Appendix we consider $V \in \mathbf{R}^{d_v}$ as a generic random variable distributed according to $P \in \mathbf{P}$ and denote its support under P by $\Omega(P) \subset \mathbf{R}^{d_v}$.

The rates obtained through a Hungarian construction crucially depend on the ability of the functions inducing the empirical process to be approximated by a suitable Haar basis. Here, we follow Koltchinskii (1994) and control the relevant approximation errors through primitive conditions stated in terms of the integral modulus of continuity. Specifically, for λ the Lebesgue measure and a function $f : \mathbf{R}^{d_v} \to \mathbf{R}$, the integral modulus of continuity of f is the function $\varpi(f, \cdot, P) : \mathbf{R}_+ \to \mathbf{R}_+$ given by

$$\varpi(f, h, P) \equiv \sup_{\|s\| \leq h} \left(\int_{\Omega(P)} (f(v + s) - f(v))^2 1\{v + s \in \Omega(P)\} d\lambda(v) \right)^{\frac{1}{2}}. \tag{G.1}$$

Intuitively, the integral modulus of continuity quantifies the "smoothness" of a function f by examining the difference between f and its own translation. For Lipschitz function f, it is straightforward to show for instance that $\varpi(f, h, P) \lesssim h$. In contrast indicator functions such as $f(v) = 1\{v \leq t\}$ typically satisfy $\varpi(f, h, P) \lesssim h^{1/2}$.

The uniform coupling result will established under the following Assumptions:

Assumption G.1. *(i) For all $P \in \mathbf{P}$, $P \ll \lambda$ and $\Omega(P)$ is compact; (ii) The densities $dP/d\lambda$ satisfy $\sup_{P \in \mathbf{P}} \sup_{v \in \Omega(P)} \frac{dP}{d\lambda}(v) < \infty$ and $\inf_{P \in \mathbf{P}} \inf_{v \in \Omega(P)} \frac{dP}{d\lambda}(v) > 0$.*

Assumption G.2. *(i) For each $P \in \mathbf{P}$ there is a continuously differentiable bijection $T_P : [0, 1]^{d_v} \to \Omega(P)$; (ii) The Jacobian $JT_P : [0, 1]^{d_v} \to \mathbf{R}$ and derivative $T_P' : [0, 1]^{d_v} \to \Omega(P)$ satisfy $\inf_{P \in \mathbf{P}} \inf_{v \in [0,1]^{d_v}} |JT_P(v)| > 0$ and $\sup_{P \in \mathbf{P}} \sup_{v \in [0,1]^{d_v}} \|JT_P(v)\|_{o,2} < \infty$.*

Assumption G.3. *The classes of functions $\mathcal{F}_n$ satisfy: (i) $\sup_{P \in \mathbf{P}} \sup_{f \in \mathcal{F}_n} \varpi(f, h, P) \leq \varphi_n(h)$ for some $\varphi_n : \mathbf{R}_+ \to \mathbf{R}_+$ satisfying $\varphi_n(Ch) \leq C^\kappa \varphi_n(h)$ for all n, $C > 0$, and some $\kappa > 0$; and (ii) $\sup_{P \in \mathbf{P}} \sup_{f \in \mathcal{F}_n} \|f\|_{L_P^\infty} \leq K_n$ for some $K_n > 0$.*

In Assumption G.1 we impose that $V \sim P$ be continuously distributed for all $P \in \mathbf{P}$, with uniformly (in P) bounded supports and densities bounded from above and away from zero. Assumption G.2 requires that the support of V under each P be "smooth" in the sense that it may be seen as a differentiable transformation of the unit square. Together, Assumptions G.1 and G.2 enable us to construct partitions of $\Omega(P)$ such that the diameter of each set in the partition is controlled uniformly in P; see Lemma G.1. As a result, the approximation error by the Haar bases implied by each partition can be controlled uniformly by the integral modulus of continuity; see Lemma G.2. Together with Assumption G.3, which imposes conditions on the integral modulus of continuity of $\mathcal{F}_n$ uniformly in P, we can obtain a uniform coupling result through Koltchinskii (1994). We note that the homogeneity condition on φ_n in Assumption G.3(i) is not necessary, but imposed to simplify the expression for the bound.

The following theorem provides us with the foundation for verifying Assumption 5.1.

Theorem G.1. *Let Assumptions G.1-G.3 hold, $\{V_i\}_{i=1}^\infty$ be i.i.d. with $V_i \sim P \in \mathbf{P}$ and for any $\delta_n \downarrow 0$ let $N_n \equiv \sup_{P \in \mathbf{P}} N_{[]}(\delta_n, \mathcal{F}_n, \|\cdot\|_{L_P^2})$, $J_n \equiv \sup_{P \in \mathbf{P}} J_{[]}(\delta_n, \mathcal{F}_n, \|\cdot\|_{L_P^2})$, and*

$$S_n \equiv \left(\sum_{i=0}^{\lceil \log_2 n \rceil} 2^i \varphi_n^2 (2^{-\frac{i}{d_v}}) \right)^{\frac{1}{2}} . \tag{G.2}$$

If $N_n \uparrow \infty$, then there exist processes $\{\mathbb{W}_{n,P}\}_{n=1}^\infty$ such that uniformly in $P \in \mathbf{P}$ we have

$$\|\mathbb{G}_{n,P} - \mathbb{W}_{n,P}\|_{\mathcal{F}_n}$$
$$= O_p\left(\frac{K_n \log(nN_n)}{\sqrt{n}} + \frac{K_n \sqrt{\log(nN_n)\log(n)} S_n}{\sqrt{n}} + J_n\left(1 + \frac{J_n K_n}{\delta_n^2 \sqrt{n}}\right) \right) . \tag{G.3}$$

Theorem G.1 is a mild modification of the results in Koltchinskii (1994). Intuitively, the proof of Theorem G.1 relies on a coupling of the empirical process on a sequence of grids of cardinality N_n, and relying on equicontinuity of both the empirical and isonormal processes to obtain a coupling on $\mathcal{F}_n$. The conclusion of Theorem G.1 applies to any choice of grid accuracy δ_n. In order to obtain the best rate that Theorem G.1 can deliver, however, the sequence δ_n must be chosen to balance the terms in (G.3) and thus depends on the metric entropy of the class $\mathcal{F}_n$. The following Corollary illustrates the use of Theorem G.1 by establishing a coupling result for Euclidean classes. We emphasize, however, that different metric entropy assumptions on $\mathcal{F}_n$ lead to alternative optimal choices of δ_n in Theorem G.1 and thus also to differing coupling rates.

Corollary G.1. *Let Assumption 3.1, 3.2(i), G.1, G.2 hold, and $\sup_{f \in \mathcal{F}_n} \|f\|_{L_P^\infty}$ be bounded uniformly in n and $P \in \mathbf{P}$. If $\sup_{P \in \mathbf{P}} \sup_{f \in \mathcal{F}_n} \max_{k,j} \varpi(f q_{k,n,j}, h, P) \le A_n h^\gamma$ for some $\gamma_f \in (0,1]$, $\sup_{P \in \mathbf{P}} N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{L_P^2}) \le (D/\epsilon)^{j_n}$ for some $j_n \uparrow \infty$ and $D < \infty$, and $\log(k_n) = O(j_n)$, then it follows that uniformly in $P \in \mathbf{P}$ we have*

$$\|\mathbb{G}_{n,P} f q_n^{k_n} - \mathbb{W}_{n,P} f q_n^{k_n}\|_r = O_p\left(k_n^{1/r} B_n \log(B_n k_n n) \left\{ \frac{A_n \sqrt{j_n}}{n^{\gamma/d_v}} + \frac{j_n \log(B_n n)}{\sqrt{n}} \right\} \right) .$$

Below, we include the proofs of Theorem G.1, Corollary G.1, and auxiliary results.

PROOF OF THEOREM G.1: Let $\{\Delta_i(P)\}$ be a sequence of partitions of $\Omega(P)$ as in Lemma G.1, and $\mathcal{B}_{P,i}$ the σ-algebra generated by $\Delta_i(P)$. By Lemma G.2 and Assumption G.3,

$$\sup_{P \in \mathbf{P}} \sup_{f \in \mathcal{F}_n} \left(\sum_{i=0}^{\lceil \log_2 n \rceil} 2^i E_P[(f(V) - E_P[f(V)|\mathcal{B}_{P,i}])^2] \right)^{\frac{1}{2}}$$

$$\le C_1 \left(\sum_{i=0}^{\lceil \log_2 n \rceil} 2^i \varphi_n^2 (2^{-\frac{i}{d_v}}) \right)^{\frac{1}{2}} \equiv C_1 S_n \tag{G.4}$$

for some constant $C_1 > 0$, and for S_n as defined in (G.2). Next, let $\mathcal{F}_{P,n,\delta_n} \subseteq \mathcal{F}_n$ denote a finite δ_n-net of $\mathcal{F}_n$ with respect to $\|\cdot\|_{L_P^2}$. Since $N(\epsilon, \mathcal{F}_n, \|\cdot\|_{L_P^2}) \leq N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{L_P^2})$, it follows from the definition of N_n that we may choose $\mathcal{F}_{P,n,\delta_n}$ so that

$$\sup_{P \in \mathbf{P}} \operatorname{card}(\mathcal{F}_{P,n,\delta_n}) \leq \sup_{P \in \mathbf{P}} N_{[]}(\delta_n, \mathcal{F}_n, \|\cdot\|_{L_P^2}) \equiv N_n . \tag{G.5}$$

By Theorem 3.5 in Koltchinskii (1994), (G.4) and (G.5), it follows that for each $n \geq 1$ there exists an isonormal process $\mathbb{W}_{n,P}$, such that for all $\eta_1 > 0$, $\eta_2 > 0$

$$\sup_{P \in \mathbf{P}} P\left(\frac{\sqrt{n}}{K_n} \|\mathbb{G}_{n,P} - \mathbb{W}_{n,P}\|_{\mathcal{F}_{P,n,\delta_n}} \geq \eta_1 + \sqrt{\eta_1}\sqrt{\eta_2}(C_1 S_n + 1)\right)$$
$$\lesssim N_n e^{-C_2 \eta_1} + n e^{-C_2 \eta_2} , \tag{G.6}$$

for some $C_2 > 0$. Since $N_n \uparrow \infty$, (G.6) implies for any $\varepsilon > 0$ there are $C_3 > 0$, $C_4 > 0$ sufficiently large, such that setting $\eta_1 \equiv C_3 \log(N_n)$ and $\eta_2 \equiv C_3 \log(n)$ yields

$$\sup_{P \in \mathbf{P}} P\left(\|\mathbb{G}_{n,P} - \mathbb{W}_{n,P}\|_{\mathcal{F}_{P,n,\delta_n}} \geq C_4 K_n \times \frac{\log(nN_n) + \sqrt{\log(N_n)\log(n)}S_n}{\sqrt{n}}\right) < \varepsilon . \tag{G.7}$$

Next, note that by definition of $\mathcal{F}_{P,n,\delta_n}$, there exists a $\Gamma_{n,P} : \mathcal{F}_n \to \mathcal{F}_{P,n,\delta_n}$ such that $\sup_{P \in \mathbf{P}} \sup_{f \in \mathcal{F}_n} \|f - \Gamma_{n,P}f\|_{L_P^2} \leq \delta_n$. Let $D(\epsilon, \mathcal{F}_n, \|\cdot\|_{L_P^2})$ denote the ϵ-packing number for $\mathcal{F}_n$ under $\|\cdot\|_{L_P^2}$, and note $D(\epsilon, \mathcal{F}_n, \|\cdot\|_{L_P^2}) \leq N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{L_P^2})$. Therefore, by Corollary 2.2.8 in van der Vaart and Wellner (1996) we can conclude that

$$\sup_{P \in \mathbf{P}} E_P[\|\mathbb{W}_{n,P} - \mathbb{W}_{n,P} \circ \Gamma_{n,P}\|_{\mathcal{F}_n}]$$
$$\lesssim \sup_{P \in \mathbf{P}} \int_0^{\delta_n} \sqrt{\log D(\epsilon, \mathcal{F}_n, \|\cdot\|_{L_P^2})}d\epsilon \leq \sup_{P \in \mathbf{P}} J_{[]}(\delta_n, \mathcal{F}_n, \|\cdot\|_{L_P^2}) \equiv J_n . \tag{G.8}$$

Similarly, employing Lemma 3.4.2 in van der Vaart and Wellner (1996) yields that

$$\sup_{P \in \mathbf{P}} E_P[\|\mathbb{G}_{n,P} - \mathbb{G}_{n,P} \circ \Gamma_{n,P}\|_{\mathcal{F}_n}]$$
$$\lesssim \sup_{P \in \mathbf{P}} J_{[]}(\delta_n, \mathcal{F}_n, \|\cdot\|_{L_P^2})(1 + \sup_{P \in \mathbf{P}} \frac{J_{[]}(\delta_n, \mathcal{F}_n, \|\cdot\|_{L_P^2})K_n}{\delta_n^2 \sqrt{n}}) \equiv J_n(1 + \frac{J_n K_n}{\delta_n^2 \sqrt{n}}) \tag{G.9}$$

Therefore, combining results (G.7), (G.8), and (G.9) together with the decomposition

$$\|\mathbb{G}_{n,P} - \mathbb{W}_{n,P}\|_{\mathcal{F}_n}$$
$$\leq \|\mathbb{G}_{n,P} - \mathbb{W}_{n,P}\|_{\mathcal{F}_{P,n,\delta_n}} + \|\mathbb{G}_{n,P} - \mathbb{G}_{n,P} \circ \Gamma_{n,P}\|_{\mathcal{F}_n} + \|\mathbb{W}_{n,P} - \mathbb{W}_{n,P} \circ \Gamma_{n,P}\|_{\mathcal{F}_n} , \tag{G.10}$$

establishes the claim of the Theorem by Markov's inequality. ∎

PROOF OF COROLLARY G.1: Define the class $\mathcal{G}_n \equiv \{fq_{k,n,\jmath} :$ for some $f \in \mathcal{F}_n$, $1 \le \jmath \le \mathcal{J}$ and $1 \le k \le k_{n,\jmath}\}$, and note that Lemma B.1 implies that for any $\delta_n \downarrow 0$

$$\sup_{P \in \mathbf{P}} N_{[\,]}(\delta_n, \mathcal{G}_n, \| \cdot \|_{L_P^2}) \le k_n \times \sup_{P \in \mathbf{P}} N_{[\,]}(\delta_n/B_n, \mathcal{F}_n, \| \cdot \|_{L_P^2}) \le k_n \times (\frac{DB_n}{\delta_n})^{j_n} . \quad \text{(G.11)}$$

Similarly, exploiting (G.11), $\delta_n \downarrow 0$ and $\int_0^a \log(M/u)du = a\log(M/a) + a$ we conclude

$$\sup_{P \in \mathbf{P}} J_{[\,]}(\delta_n, \mathcal{G}_n, \| \cdot \|_{L_P^2}) \le \int_0^{\delta_n} \{\log(k_n) + j_n \log(\frac{DB_n}{\epsilon})\}^{1/2} d\epsilon$$

$$\lesssim (\sqrt{\log(k_n)} + \sqrt{j_n}) \int_0^{\delta_n} \log(\frac{DB_n}{\epsilon}) d\epsilon \lesssim (\sqrt{\log(k_n)} + \sqrt{j_n}) \times \delta_n \log(\frac{B_n}{\delta_n}) . \quad \text{(G.12)}$$

In turn, note that for S_n as defined in (G.2), we obtain from $\varphi_n(h) = A_n h^\gamma$ that

$$\{\sum_{i=0}^{\lceil \log_2 n \rceil} 2^i \times \frac{A_n^2}{2^{2i\gamma/d_v}}\}^{1/2} \lesssim A_n \times n^{\frac{1}{2} - \frac{\gamma}{d_v}} . \quad \text{(G.13)}$$

Finally, we also note that $\sup_{f \in \mathcal{G}_n} \|f\|_{L_P^\infty} \lesssim B_n$ due to Assumption 3.2(i) and $f \in \mathcal{F}_n$ being uniformly bounded by hypothesis. Therefore, setting $\delta_n = \sqrt{j_n}/\sqrt{n}$ in (G.11) and (G.12), and exploiting $\|a\|_r \le d^{1/r}\|a\|_\infty$ for any $a \in \mathbf{R}^d$ we obtain that

$$\sup_{f \in \mathcal{F}_n} \|\mathbb{G}_{n,P} f q_n^{k_n} - \mathbb{W}_{n,P} f q_n^{k_n}\|_r$$

$$\le k_n^{1/r} \|\mathbb{G}_{n,P} - \mathbb{W}_{n,P}\|_{\mathcal{G}_n} = O_p(k_n^{1/r} B_n \log(B_n k_n n)\{\frac{A_n \sqrt{j_n}}{n^{\gamma/d_v}} + \frac{j_n \log(B_n n)}{\sqrt{n}}\}) \quad \text{(G.14)}$$

uniformly in $P \in \mathbf{P}$ by Theorem G.1. ∎

Lemma G.1. *Let $\mathcal{B}_P$ denote the completion of the Borel $\sigma-$algebra on $\Omega(P)$ with respect to P. If Assumptions G.1(i)-(ii) and G.2(i)-(ii) hold, then for each $P \in \mathbf{P}$ there exists a sequence $\{\Delta_i(P)\}$ of partitions of the probability space $(\Omega(P), \mathcal{B}_P, P)$ such that:*
(i) $\Delta_i(P) = \{\Delta_{i,k}(P) : k = 0, \ldots, 2^i - 1\}$, $\Delta_{i,k}(P) \in \mathcal{B}_P$ and $\Delta_{0,0}(P) = \Omega(P)$.
(ii) $\Delta_{i,k}(P) = \Delta_{i+1,2k}(P) \cup \Delta_{i+1,2k+1}(P)$ and $\Delta_{i+1,2k}(P) \cap \Delta_{i+1,2k+1}(P) = \emptyset$ for any integers $k = 0, \ldots 2^i - 1$ and $i \ge 0$.
(iii) $P(\Delta_{i+1,2k}(P)) = P(\Delta_{i+1,2k+1}(P)) = 2^{-i-1}$ for $k = 0, \ldots 2^i - 1$, $i \ge 0$.
(iv) $\sup_{P \in \mathbf{P}} \max_{0 \le k \le 2^i - 1} \sup_{v,v' \in \Delta_{i,k}(P)} \|v - v'\|_2 = O(2^{-\frac{i}{d_v}})$.
(v) $\mathcal{B}_P$ equals the completion with respect to P of the σ-algebra generated by $\bigcup_{i \ge 0} \Delta_i(P)$.

PROOF: Let $\mathcal{A}$ denote the Borel σ-algebra on $[0,1]^{d_v}$, and for any $A \in \mathcal{A}$ define

$$Q_P(A) \equiv P(T_P(A)) , \quad \text{(G.15)}$$

where $T_P(A) \in \mathcal{B}_P$ due to T_P^{-1} being measurable. Moreover, $Q_P([0,1]^{d_v}) = 1$ due to T_P being surjective, and Q_P is σ-additive due to T_P being injective. Hence, we conclude

Q_P defined by (G.15) is a probability measure on $([0,1]^{d_v}, \mathcal{A})$. In addition, for λ the Lebesgue measure, we obtain from Theorem 3.7.1 in Bogachev (2007) that

$$Q_P(A) = P(T_P(A)) = \int_{T_P(A)} \frac{dP}{d\lambda}(v) d\lambda(v) = \int_A \frac{dP}{d\lambda}(T_P(a))|JT_P(a)| d\lambda(a) , \quad \text{(G.16)}$$

where $|JT_P(a)|$ denotes the Jacobian of T_P at any point $a \in [0,1]^{d_v}$. Hence, Q_P has density with respect to Lebesgue measure given by $g_P(a) \equiv \frac{dP}{d\lambda}(T_P(a))|JT_P(a)|$ for any $a \in [0,1]^{d_v}$. Next, let $a = (a_1, \dots, a_{d_v})' \in [0,1]^{d_v}$ and define for any $t \in [0,1]$

$$G_{l,P}(t|A) \equiv \frac{Q_P(a \in A : a_l \leq t)}{Q_P(A)} , \quad \text{(G.17)}$$

for any set $A \in \mathcal{A}$ and $1 \leq l \leq d_v$. Further let $\mathrm{m}(i) \equiv i - \lfloor \frac{i-1}{d_v} \rfloor \times d_v$ – i.e. $\mathrm{m}(i)$ equals i modulo d_v – and setting $\tilde{\Delta}_{0,0}(P) = [0,1]^{d_v}$ inductively define the partitions (of $[0,1]^{d_v}$)

$$\tilde{\Delta}_{i+1,2k}(P) \equiv \{a \in \tilde{\Delta}_{i,k}(P) : G_{\mathrm{m}(i+1),P}(a_{\mathrm{m}(i+1)}|\tilde{\Delta}_{i,k}(P)) \leq \frac{1}{2}\}$$
$$\tilde{\Delta}_{i+1,2k+1}(P) \equiv \tilde{\Delta}_{i,k}(P) \setminus \tilde{\Delta}_{i+1,2k}(P) \quad \text{(G.18)}$$

for $0 \leq k \leq 2^i - 1$. For $\mathrm{cl}(\tilde{\Delta}_{i,k}(P))$ the closure of $\tilde{\Delta}_{i,k}(P)$, we then note that by construction each $\tilde{\Delta}_{i,k}(P)$ is a hyper-rectangle in $[0,1]^{d_v}$ – i.e. it is of the general form

$$\mathrm{cl}(\tilde{\Delta}_{i,k}(P)) = \prod_{j=1}^{d_v} [l_{i,k,j}(P), u_{i,k,j}(P)] . \quad \text{(G.19)}$$

Moreover, since g_P is positive everywhere on $[0,1]^{d_v}$ by Assumptions G.1(ii) and G.2(ii), it follows that for any $i \geq 0$, $0 \leq k \leq 2^i - 1$ and $1 \leq j \leq d_v$, we have

$$l_{i+1,2k,j}(P) = l_{i,k,j}(P)$$
$$u_{i+1,2k,j}(P) = \begin{cases} u_{i,k,j}(P) \text{ if } j \neq \mathrm{m}(i+1) \\ \text{solves } G_{\mathrm{m}(i+1),P}(u_{i+1,2k,j}(P)|\tilde{\Delta}_{i,k}(P)) = \frac{1}{2} \text{ if } j = \mathrm{m}(i+1) \end{cases} \quad \text{(G.20)}$$

Similarly, since $\tilde{\Delta}_{i+1,2k+1}(P) = \tilde{\Delta}_{i,k}(P) \setminus \tilde{\Delta}_{i+1,2k}(P)$, it additionally follows that

$$u_{i+1,2k+1,j}(P) = u_{i,k,j}(P) \qquad l_{i+1,2k+1,j}(P) = \begin{cases} l_{i,k,j}(P) \text{ if } j \neq \mathrm{m}(i+1) \\ u_{i+1,2k,j}(P) \text{ if } j = \mathrm{m}(i+1) \end{cases} \quad \text{(G.21)}$$

Since $Q_P(\mathrm{cl}(\tilde{\Delta}_{i+1,2k}(P))) = Q_P(\tilde{\Delta}_{i+1,2k}(P))$ by $Q_P \ll \lambda$, (G.17) and (G.20) yield

$$\begin{aligned} Q_P(\tilde{\Delta}_{i+1,2k}(P)) &= Q_P(a \in \tilde{\Delta}_{i,k}(P) : a_{\mathrm{m}(i+1)} \leq u_{i+1,2k,\mathrm{m}(i+1)}(P)) \\ &= G_{\mathrm{m}(i+1),P}(u_{i+1,2k,\mathrm{m}(i+1)}(P)|\tilde{\Delta}_{i,k}(P))Q_P(\tilde{\Delta}_{i,k}(P)) \\ &= \frac{1}{2}Q_P(\tilde{\Delta}_{i,k}(P)) . \end{aligned} \quad \text{(G.22)}$$

Therefore, since $\tilde{\Delta}_{i,k}(P) = \tilde{\Delta}_{i+1,2k}(P) \cup \tilde{\Delta}_{i+1,2k+1}(P)$, it follows that $Q_P(\tilde{\Delta}_{i+1,2k+1}(P)) = \frac{1}{2}Q_P(\tilde{\Delta}_{i,k}(P))$ for $0 \leq k \leq 2^i - 1$ as well. In particular, $Q_P(\tilde{\Delta}_{0,0}(P)) = 1$ implies that

$$Q_P(\tilde{\Delta}_{i,k}(P)) = \frac{1}{2^i} \tag{G.23}$$

for any integers $i \geq 1$ and $0 \leq k \leq 2^i - 1$. Moreover, we note that result (G.16) and Assumptions G.1(ii) and G.2(ii) together imply that the densities g_P of Q_P satisfy

$$0 < \inf_{P \in \mathbf{P}} \inf_{a \in [0,1]^{d_v}} g_P(a) < \sup_{P \in \mathbf{P}} \sup_{a \in [0,1]^{d_v}} g_P(a) < \infty , \tag{G.24}$$

and therefore $Q_P(A) \asymp \lambda(A)$ uniformly in $A \in \mathcal{A}$ and $P \in \mathbf{P}$. Hence, since by (G.20) $u_{i+1,2k,j}(P) = u_{i,k,j}(P)$ and $l_{i+1,2k,j}(P) = l_{i,k,j}(P)$ for all $j \neq \mathrm{m}(i+1)$, we obtain

$$
\begin{aligned}
\frac{(u_{i+1,2k,\mathrm{m}(i+1)}(P) - l_{i+1,2k,\mathrm{m}(i+1)}(P))}{(u_{i,k,\mathrm{m}(i+1)}(P) - l_{i,k,\mathrm{m}(i+1)}(P))} &= \frac{\prod_{j=1}^{d_v}(u_{i+1,2k,j}(P) - l_{i+1,2k,j}(P))}{\prod_{j=1}^{d_v}(u_{i,k,j}(P) - l_{i,k,j}(P))} \\
&= \frac{\lambda(\tilde{\Delta}_{i+1,2k}(P))}{\lambda(\tilde{\Delta}_{i,k}(P))} \asymp \frac{Q_P(\tilde{\Delta}_{i+1,2k}(P))}{Q_P(\tilde{\Delta}_{i,k}(P))} = \frac{1}{2} \quad \text{(G.25)}
\end{aligned}
$$

uniformly in $P \in \mathbf{P}$, $i \geq 0$, and $0 \leq k \leq 2^i - 1$ by results (G.23) and (G.24). Moreover, by identical arguments but using (G.21) instead of (G.20) we conclude

$$\frac{(u_{i+1,2k+1,\mathrm{m}(i+1)}(P) - l_{i+1,2k+1,\mathrm{m}(i+1)}(P))}{(u_{i,k,\mathrm{m}(i+1)}(P) - l_{i,k,\mathrm{m}(i+1)}(P))} \asymp \frac{1}{2} \tag{G.26}$$

also uniformly in $P \in \mathbf{P}$, $i \geq 0$ and $0 \leq k \leq 2^i - 1$. Thus, since $(u_{i+1,2k,j}(P) - l_{i+1,2k,j}(P)) = (u_{i+1,2k+1,j}(P) - l_{i+1,2k+1,j}(P)) = (u_{i,k,j}(P) - l_{i,k,j}(P))$ for all $j \neq \mathrm{m}(i+1)$, and $u_{0,0,j}(P) - l_{0,0,j}(P) = 1$ for all $1 \leq j \leq d_v$ we obtain from $\mathrm{m}(i) = i - \lfloor \frac{i-1}{d_v} \rfloor \times d_v$, results (G.25) and (G.26), and proceeding inductively that

$$(u_{i,k,j}(P) - l_{i,k,j}(P)) \asymp 2^{-\frac{i}{d_v}} , \tag{G.27}$$

uniformly in $P \in \mathbf{P}$, $i \geq 0$, $0 \leq k \leq 2^i - 1$, and $1 \leq j \leq d_v$. Thus, result (G.27) yields

$$
\begin{aligned}
\sup_{P \in \mathbf{P}} &\max_{0 \leq k \leq 2^i - 1} \sup_{a,a' \in \tilde{\Delta}_{i,k}(P)} \|a - a'\|_2 \\
&\leq \sup_{P \in \mathbf{P}} \max_{0 \leq k \leq 2^i - 1} \max_{1 \leq j \leq d_v} \sqrt{d_v} \times (u_{i,j,k}(P) - l_{i,j,k}(P)) = O(2^{-\frac{i}{d_v}}) . \quad \text{(G.28)}
\end{aligned}
$$

We next obtain the desired sequence of partitions $\{\Delta_i(P)\}$ of $(\Omega(P), \mathcal{B}_P, P)$ by constructing them from the partition $\{\tilde{\Delta}_{i,k}(P)\}$ of $[0,1]^{d_v}$. To this end, set

$$\Delta_{i,k}(P) \equiv T_P(\tilde{\Delta}_{i,k}(P)) \tag{G.29}$$

111

for all $i \geq 0$ and $0 \leq k \leq 2^i - 1$. Note that $\{\Delta_i(P)\}$ satisfies conditions (i) and (ii) due to T_P^{-1} being a measurable map, T_P being bijective, and result (G.18). In addition, $\{\Delta_i(P)\}$ satisfies condition (iii) since by definition (G.15) and result (G.23):

$$P(\Delta_{i,k}(P)) = P(T_P(\tilde{\Delta}_{i,k}(P))) = Q_P(\tilde{\Delta}_{i,k}(P)) = 2^{-i} , \tag{G.30}$$

for all $0 \leq k \leq 2^i - 1$. Moreover, by Assumption G.2(ii), $\sup_{P \in \mathbf{P}} \sup_{a \in [0,1]^{d_v}} \|T_P'(a)\|_{o,2} < \infty$, and hence by the mean value theorem we can conclude that

$$\sup_{P \in \mathbf{P}} \max_{0 \leq k \leq 2^i - 1} \sup_{v,v' \in \Delta_{i,k}(P)} \|v - v'\|_2 = \sup_{P \in \mathbf{P}} \max_{0 \leq k \leq 2^i - 1} \sup_{a,a' \in \tilde{\Delta}_{i,k}(P)} \|T_P(a) - T_P(a')\|_2$$

$$\lesssim \sup_{P \in \mathbf{P}} \max_{0 \leq k \leq 2^i - 1} \sup_{a,a' \in \tilde{\Delta}_{i,k}(P)} \|a - a'\|_2 = O(2^{-\frac{i}{d_v}}) , \quad \text{(G.31)}$$

by result (G.28), which verifies that $\{\Delta_i(P)\}$ satisfies condition (iv). Also note that to verify $\{\Delta_i(P)\}$ satisfies condition (v) it suffices to show that $\bigcup_{i \geq 0} \Delta_i(P)$ generates the Borel σ-algebra on $\Omega(P)$. To this end, we first aim to show that

$$\mathcal{A} = \sigma(\bigcup_{i \geq 0} \tilde{\Delta}_i(P)) , \tag{G.32}$$

where for a collection of sets $\mathcal{C}$, $\sigma(\mathcal{C})$ denotes the σ-algebra generated by $\mathcal{C}$. For any closed set $A \in \mathcal{A}$, then define $D_i(P)$ to be given by

$$D_i(P) \equiv \bigcup_{k : \tilde{\Delta}_{i,k}(P) \cap A \neq \emptyset} \tilde{\Delta}_{i,k}(P) . \tag{G.33}$$

Notice that since $\{\tilde{\Delta}_i(P)\}$ is a partition of $[0,1]^{d_v}$, $A \subseteq D_i(P)$ for all $i \geq 0$ and hence $A \subseteq \bigcap_{i \geq 0} D_i(P)$. Moreover, if $a_0 \in A^c$, then A^c being open and (G.28) imply $a_0 \notin D_i(P)$ for i sufficiently large. Hence, $A^c \cap (\bigcap_{i \geq 0} D_i(P)) = \emptyset$ and therefore $A = \bigcap_{i \geq 0} D_i(P)$. It follows that if A is closed, then $A \in \sigma(\bigcup_{i \geq 0} \tilde{\Delta}_i(P))$, which implies $\mathcal{A} \subseteq \sigma(\bigcup_{i \geq 0} \tilde{\Delta}_i(P))$. On the other hand, since $\tilde{\Delta}_{i,k}(P)$ is Borel for all $i \geq 0$ and $0 \leq k \leq 2^i - 1$, we also have $\sigma(\bigcup_{i \geq 0} \tilde{\Delta}_i(P)) \subseteq \mathcal{A}$, and hence (G.33) follows. To conclude, we then note that

$$\sigma(\bigcup_{i \geq 0} \Delta_i(P)) = \sigma(\bigcup_{i \geq 0} T_P(\tilde{\Delta}_i(P))) = T_P(\sigma(\bigcup_{i \geq 0} \tilde{\Delta}_i(P))) = T_P(\mathcal{A}) , \tag{G.34}$$

by Corollary 1.2.9 in Bogachev (2007). However, T_P and T_P^{-1} being continuous implies $T_P(\mathcal{A})$ equals the Borel σ-algebra in $\Omega(P)$, and therefore (G.34) implies $\{\Delta_i(P)\}$ satisfies condition (v) establishing the Lemma. ∎

Lemma G.2. *Let $\{\Delta_i(P)\}$ be as in Lemma G.1, and $\mathcal{B}_{P,i}$ denote the σ-algebra generated by $\Delta_i(P)$. If Assumptions G.1(i)-(ii) and G.2(i)-(ii) hold, then there are constants*

$K_0 > 0$, $K_1 > 0$ *such that for all $P \in \mathbf{P}$ and any f satisfying $f \in L_P^2$ for all $P \in \mathbf{P}$:*

$$E_P[(f(V) - E_P[f(V)|\mathcal{B}_{P,i}])^2] \leq K_0 \times \varpi^2(f, K_1 \times 2^{-\frac{i}{d_v}}, P) .$$

PROOF: Since $\Delta_i(P)$ is a partition of $\Omega(P)$ and $P(\Delta_{i,k}(P)) = 2^{-i}$ for all $i \geq 0$ and $0 \leq k \leq 2^i - 1$, we may express $E_P[f(V)|\mathcal{B}_{P,i}]$ as an element of L_P^2 by

$$E_P[f(V)|\mathcal{B}_{P,i}] = 2^i \sum_{k=0}^{2^i-1} 1\{v \in \Delta_{i,k}(P)\} \int_{\Delta_{i,k}(P)} f(v)dP(v) . \tag{G.35}$$

Hence, result (G.35), $P(\Delta_{i,k}(P)) = 2^{-i}$ for all $i \geq 0$ and $0 \leq k \leq 2^i - 1$ together with $\Delta_i(P)$ being a partition of $\Omega(P)$, and applying Holder's inequality to the term $(f(v) - f(\tilde{v}))1\{v \in \Omega(P)\} \times 1\{\tilde{v} \in \Delta_{i,k}(P)\}$ we obtain that

$$E_P[(f(V) - E_P[f(V)|\mathcal{B}_{P,i}])^2]$$
$$= \sum_{k=0}^{2^i-1} \int_{\Delta_{i,k}(P)} (f(v) - 2^i \int_{\Delta_{i,k}(P)} f(\tilde{v})dP(\tilde{v}))^2 dP(v)$$
$$= \sum_{k=0}^{2^i-1} 2^{2i} \int_{\Delta_{i,k}(P)} (\int_{\Delta_{i,k}(P)} (f(v) - f(\tilde{v}))1\{v \in \Omega(P)\}dP(\tilde{v}))^2 dP(v)$$
$$\leq \sum_{k=0}^{2^i-1} 2^{2i} P(\Delta_{i,k}(P)) \int_{\Delta_{i,k}(P)} \int_{\Delta_{i,k}(P)} (f(v) - f(\tilde{v}))^2 1\{v \in \Omega(P)\}dP(\tilde{v})dP(v)$$
$$= \sum_{k=0}^{2^i-1} 2^i \int_{\Delta_{i,k}(P)} \int_{\Delta_{i,k}(P)} (f(v) - f(\tilde{v}))^2 1\{v \in \Omega(P)\}dP(\tilde{v})dP(v) . \tag{G.36}$$

Let $D_i \equiv \sup_{P \in \mathbf{P}} \max_{0 \leq k \leq 2^i-1} \text{diam}\{\Delta_{i,k}(P)\}$, where $\text{diam}\{\Delta_{i,k}(P)\}$ is the diameter of $\Delta_{i,k}(P)$. Further note that by Lemma G.1(iv), $D_i = O(2^{-\frac{i}{d_v}})$ and hence we have $\lambda(\{s \in \mathbf{R}^{d_v} : \|s\| \leq D_i\}) \leq M_1 2^{-i}$ for some $M_1 > 0$ and λ the Lebesgue measure. Noting that $\sup_{P \in \mathbf{P}} \sup_{v \in \Omega(P)} \frac{dP}{d\lambda}(v) < \infty$ by Assumption G.1(ii), and doing the change of variables $s = v - \tilde{v}$ we then obtain that for some constant $M_0 > 0$

$$E_P[(f(V) - E_P[f(V)|\mathcal{B}_{P,i}])^2]$$
$$\leq M_0 \sum_{k=0}^{2^i-1} 2^i \int_{\Delta_{i,k}(P)} \int_{\Delta_{i,k}(P)} (f(v) - f(\tilde{v}))^2 1\{v \in \Omega(P)\}d\lambda(\tilde{v})d\lambda(v)$$
$$\leq M_0 M_1 \sup_{\|s\| \leq D_i} \sum_{k=0}^{2^i-1} \int_{\Delta_{i,k}(P)} (f(\tilde{v} + s) - f(\tilde{v}))^2 1\{\tilde{v} + s \in \Omega(P)\}d\lambda(\tilde{v}) . \tag{G.37}$$

Next observe that $\varpi(f, h, P)$ is decreasing in h. Hence, since $\{\Delta_{i,k}(P) : k = 0 \ldots 2^i - 1\}$

is a partition of $\Omega(P)$, and $D_i \leq K_1 2^{-\frac{i}{d_v}}$ for some $K_1 > 0$ by Lemma G.1(iv), we obtain

$$E_P[(f(V) - E_P[f(V)|\mathcal{B}_{P,i}])^2] \leq M_0 M_1 \times \varpi^2(f, K_1 \times 2^{-\frac{i}{d_v}}, P) \tag{G.38}$$

by (G.37). Setting $K_0 \equiv M_0 \times M_1$ in (G.38) establishes the claim of the Lemma. ∎

APPENDIX H - Multiplier Bootstrap Results

In this Appendix, we develop results that enable us to provide sufficient conditions for verifying that Assumption 6.5 is satisfied. The results in this Appendix may be of independent interest, as they extend the validity of the multiplier bootstrap to suitable non-Donsker classes. In particular, applying Theorem H.1 below to the case $q_{k,n}(z) = 1$ and $k_n = 1$ for all n yields the consistency of the multiplier bootstrap for the law of the standard empirical process indexed by a expanding classes $\mathcal{F}_n$ of functions.

Our analysis requires the classes $\mathcal{F}_n$ be sufficiently "smooth" in that they satisfy:

Assumption H.1. *For each $P \in \mathbf{P}$ and n there exists a $\{p_{j,n,P}\}_{j=1}^\infty \subset L_P^2$ such that: (i) $\{p_{j,n,P}\}_{j=1}^\infty$ is orthonormal in L_P^2 and $\|p_{j,n,P}\|_{L_P^\infty}$ is uniformly bounded in $j, n \in \mathbb{N}$ and $P \in \mathbf{P}$; (ii) For any $j_n \uparrow \infty$ and, $p_{n,P}^{j_n}(v) \equiv (p_{1,n,P}(v), \ldots, p_{j_n,n,P}(v))'$ the eigenvalues of $E_P[(q_n^{k_n}(Z_i) \otimes p_{n,P}^{j_n}(V_i))(q_n^{k_n}(Z_i) \otimes p_{n,P}^{j_n}(V_i))']$ are bounded uniformly in n and $P \in \mathbf{P}$; (iii) For some $M_n \uparrow \infty$ and $\gamma_\rho > 3/2$ we have for all $P \in \mathbf{P}$ the inclusion*

$$\mathcal{F}_n \subseteq \{f = \sum_{j=1}^\infty p_{j,n,P}(v)\beta_j : \{\beta_j\}_{j=1}^\infty \text{ satisfies } |\beta_j| \leq \frac{M_n}{j^{\gamma_\rho}}\} . \tag{H.1}$$

Assumption H.1(i) demands the existence of orthonormal and bounded functions $\{p_{j,n,P}\}_{j=1}^\infty$ in L_P^2 that provide suitable approximations to the class $\mathcal{F}_n$ in the sense imposed in Assumption H.1(iii). Crucially, we emphasize that the array $\{p_{j,n,P}\}_{j=1}^\infty$ need not be known as it is merely employed in the theoretical construction of the bootstrap coupling, and not in the computation of the multiplier bootstrap process $\hat{\mathbb{W}}_n$. In certain applications, however, such as when ρ is linear in θ and linear sieves are employed as Θ_n, the functions $\{p_{j,n,P}\}_{j=1}^\infty$ may be set to equal a rotation of the sieve.[20] It is also worth pointing out that, as in Appendix G, the concept of "smoothness" employed does not necessitate that ρ be differentiable in its arguments. Finally, Assumption H.1(ii) constrains the eigenvalues of $E_P[(q_n^{k_n}(Z_i) \otimes p_{n,P}^{j_n}(V_i))(q_n^{k_n}(Z_i) \otimes p_{n,P}^{j_n}(V_i))']$ to be bounded from above. This requirement may be dispensed with, allowing the largest eigenvalues to diverge with n, at the cost of slowing the rate of convergence of the Gaussian multiplier bootstrap $\hat{\mathbb{W}}_n$ to the corresponding isonormal process $\mathbb{W}_{n,P}^\star$.

---

[20]Concretely, if $\Theta_n = \{f = \tilde{p}_n^{j_n}{}'\gamma : \text{ for some } \gamma \in \mathbf{R}^{j_n}\}$ and $X = (Y, W')'$ with $Y \in \mathbf{R}$ and $\rho(X, \theta) \equiv Y - \theta(W)$, then a candidate choice for $p_n^{j_n+1}(v)$ are the orthonormalized functions $(y, p_n^{j_n}(w)')'$.

As we next show, Assumption H.1 provides a sufficient condition for verifying Assumption 6.5. In the following, recall F_n is the envelope of $\mathcal{F}_n$ (as in Assumption 3.3(ii)).

Theorem H.1. *Let Assumptions 3.1, 3.2(i), 3.3(ii), H.1 hold, and $\{\omega_i\}_{i=1}^n$ be i.i.d. with $\omega_i \sim N(0,1)$ independent of $\{V_i\}_{i=1}^n$. Then, for any $j_n \uparrow \infty$ with $j_n k_n \log(j_n k_n) B_n = o(n)$ there is an isonormal $\mathbb{W}_{n,P}^\star$ independent of $\{V_i\}_{i=1}^n$ satisfying uniformly in $P \in \mathbf{P}$*

$$\sup_{f \in \mathcal{F}_n} \|\hat{\mathbb{W}}_n f q_n^{k_n} - \mathbb{W}_{n,P}^\star f q_n^{k_n}\|_r$$
$$= O_p\left(\frac{\sup_{P \in \mathbf{P}} \|F_n\|_{L_P^2} B_n^{\frac{1}{2}} k_n^{\frac{1}{4}+\frac{1}{r}} (j_n \log(k_n))^{\frac{3}{4}} (\log(j_n))^{\frac{1}{4}}}{n^{\frac{1}{4}}} + \frac{k_n^{1/r} B_n M_n \sqrt{\log(k_n)}}{j_n^{\gamma_\rho - 3/2}}\right) .$$

The rate of convergence derived in Theorem H.1 depends on the selected sequence $j_n \uparrow \infty$, which should be chosen optimally to deliver the best possible implied rate. Heuristically, the proof of Theorem H.1 proceeds in two steps. First, we construct a multivariate normal random variable $\mathbb{W}_{n,P}^\star(q_n^{k_n} \otimes p_{n,P}^{j_n}) \in \mathbf{R}^{j_n k_n}$ that is coupled with $\hat{\mathbb{W}}_n(q_n^{k_n} \otimes p_{n,P}^{j_n}) \in \mathbf{R}^{j_n k_n}$, and then exploit the linearity of $\hat{\mathbb{W}}_n$ to obtain a suitable coupling on the subspace $\mathbb{S}_{n,P} \equiv \overline{\mathrm{span}}\{q_n^{k_n} \otimes p_{n,P}^{j_n}\}$. Second, we employ Assumption H.1(iii) to show that a successful coupling on $\mathbb{S}_{n,P}$ leads to the desired construction since $\mathcal{F}_n$ is well approximated by $\{p_{j,n,P}\}_{j=1}^\infty$. We note that the rate obtained in Theorem H.1 may be improved upon whenever the smallest eigenvalues of the matrices

$$E_P[(q_n^{k_n}(Z_i) \otimes p_{n,P}^{j_n}(V_i))(q_n^{k_n}(Z_i) \otimes p_{n,P}^{j_n}(V_i))'] \tag{H.2}$$

are bounded away from zero uniformly in n and $P \in \mathbf{P}$; see Remark 8.1. Additionally, while we do not pursue it here for conciseness, it is also worth noting that the outlined heuristics can also be employed to verify Assumption 5.1 by coupling $\mathbb{G}_{n,P}(q_n^{k_n} \otimes p_{n,P}^{j_n})$ to $\mathbb{W}_{n,P}(q_n^{k_n} \otimes p_{n,P}^{j_n})$ through standard results (Yurinskii, 1977).

Remark 8.1. Under the additional requirement that the eigenvalues of (H.2) be bounded away from zero uniformly in n and $P \in \mathbf{P}$, Theorem H.1 can be modified to establish

$$\sup_{f \in \mathcal{F}_n} \|\hat{\mathbb{W}}_n f q_n^{k_n} - \mathbb{W}_{n,P}^\star f q_n^{k_n}\|_r$$
$$= O_p\left(\frac{\sup_{P \in \mathbf{P}} \|F_n\|_{L_P^2} B_n k_n^{\frac{1}{2}+\frac{1}{r}} j_n \log(k_n) \sqrt{\log(j_n)}}{\sqrt{n}} + \frac{k_n^{1/r} B_n M_n \sqrt{\log(k_n)}}{j_n^{\gamma_\rho - 3/2}}\right) . \tag{H.3}$$

Given the assumed orthonormality of the array $\{p_{j,n,P}\}_{j=1}^\infty$, the rate obtained in (H.3) is thus more appropriate when considering the multiplier bootstrap for the standard empirical process – i.e. $q_{k,n}(z) = 1$ and $k_n = 1$ for all n – since the smallest eigenvalue of the matrices in (H.2) then equals one. ∎

Below, we include the proof of Theorem H.1 and the necessary auxiliary results.

PROOF OF THEOREM H.1: We proceed by exploiting Lemma H.1 to couple $\hat{\mathbb{W}}_n$ on a finite dimensional subspace, and showing that such a result suffices for controlling both $\hat{\mathbb{W}}_n$ and $\mathbb{W}^\star_{n,P}$ on $\mathcal{F}_n$. To this end, let $\mathbb{S}_{n,P} \equiv \overline{\operatorname{span}}\{q_n^{k_n} \otimes p_{n,P}^{j_n}\}$ and note that Lemma H.1 and $j_n \uparrow \infty$ satisfying $j_n k_n \log(j_n k_n) B_n = o(n)$ by hypothesis imply that there exists a linear isonormal process $\mathbb{W}^{(1)}_{n,P}$ on $\mathbb{S}_{n,P}$ such that uniformly in $P \in \mathbf{P}$ we have

$$
\sup_{\|\beta\|_2 \le \sup_{P\in\mathbf{P}} \|F_n\|_{L^2_P}} \|\hat{\mathbb{W}}_n(\beta' p_{n,P}^{j_n})q_n^{k_n} - \mathbb{W}^{(1)}_{n,P}(\beta' p_{n,P}^{j_n})q_n^{k_n}\|_r
$$
$$
= O_p\Big(\frac{\sup_{P\in\mathbf{P}} \|F_n\|_{L^2_P} B_n^{\frac{1}{2}} k_n^{\frac{1}{4}+\frac{1}{r}} (j_n \log(k_n))^{\frac{3}{4}} (\log(j_n))^{\frac{1}{4}}}{n^{\frac{1}{4}}}\Big) . \quad \text{(H.4)}
$$

For any closed linear subspace $\mathbb{A}$ of L^2_P, let $\operatorname{Proj}\{f|\mathbb{A}\}$ denote the $\|\cdot\|_{L^2_P}$ projection of f onto $\mathbb{A}$ and set $\mathbb{A}^\perp \equiv \{f \in L^2_P : f = g - \operatorname{Proj}\{g|\mathbb{A}\}$ for some $g \in L^2_P\}$ – i.e. $\mathbb{A}^\perp$ denotes the orthocomplement of $\mathbb{A}$ in L^2_P. Assuming the underlying probability space is suitably enlarged to carry a linear isonormal process $\mathbb{W}^{(2)}_{n,P}$ on $\mathbb{S}_{n,P}^\perp$ independent of $\mathbb{W}^{(1)}_{n,P}$ and $\{V_i\}_{i=1}^n$, we then define the isonormal process $\mathbb{W}^\star_{n,P}$ on L^2_P pointwise by

$$
\mathbb{W}^\star_{n,P} f \equiv \mathbb{W}^{(1)}_{n,P}(\operatorname{Proj}\{f|\mathbb{S}_{n,P}\}) + \mathbb{W}^{(2)}_{n,P}(\operatorname{Proj}\{f|\mathbb{S}_{n,P}^\perp\}) . \quad \text{(H.5)}
$$

Next, set $\mathbb{P}_{n,P} \equiv \overline{\operatorname{span}}\{p_{n,P}^{j_n}\}$ and note that since $\operatorname{Proj}\{f|\mathbb{P}_{n,P}\} = \beta(f)' p_{n,P}^{j_n}$ for some $\beta(f) \in \mathbf{R}^{j_n}$, the orthonormality of $\{p_{j,n,P}\}_{j=1}^{j_n}$ imposed in Assumption H.1(i) implies $\|\beta(f)\|_2 \le \|f\|_{L^2_P} \le \|F_n\|_{L^2_P}$ by Assumption 3.3(ii). Since $(\operatorname{Proj}\{f|\mathbb{P}_{n,P}\})q_{k,n,\jmath} \in \mathbb{S}_{n,P}$ for any $f \in \mathcal{F}_n$, $1 \le \jmath \le \mathcal{J}$ and $1 \le k \le k_{n,\jmath}$, (H.4) and (H.5) imply uniformly in $P \in \mathbf{P}$

$$
\sup_{f\in\mathcal{F}_n} \|\hat{\mathbb{W}}_n(\operatorname{Proj}\{f|\mathbb{P}_{n,P}\})q_n^{k_n} - \mathbb{W}^\star_{n,P}(\operatorname{Proj}\{f|\mathbb{P}_{n,P}\})q_n^{k_n}\|_r
$$
$$
\le \sup_{\|\beta\|_2 \le \sup_{P\in\mathbf{P}} \|F_n\|_{L^2_P}} \|\hat{\mathbb{W}}_n(\beta' p_{n,P}^{j_n})q_n^{k_n} - \mathbb{W}^{(1)}_{n,P}(\beta' p_{n,P}^{j_n})q_n^{k_n}\|_r
$$
$$
= O_p\Big(\frac{\sup_{P\in\mathbf{P}} \|F_n\|_{L^2_P} B_n^{\frac{1}{2}} k_n^{\frac{1}{4}+\frac{1}{r}} (j_n \log(k_n))^{\frac{3}{4}} (\log(j_n))^{\frac{1}{4}}}{n^{\frac{1}{4}}}\Big) . \quad \text{(H.6)}
$$

Next, define the set of sequences $\mathcal{B}_n \equiv \{\{\beta_j\}_{j=j_n}^\infty : |\beta_j| \le M_n/j^{\gamma_\rho}\}$, and note that

$$
\sup_{f\in\mathcal{F}_n} \|\mathbb{W}^\star_{n,P}(\operatorname{Proj}\{f|\mathbb{P}_{n,P}^\perp\})q_n^{k_n}\|_r
$$
$$
\le k_n^{1/r} \sup_{\{\beta_j\}\in\mathcal{B}_n} \max_{1\le \jmath\le\mathcal{J}} \max_{1\le k\le k_{n,\jmath}} \Big|\mathbb{W}^\star_{n,P}\Big(q_{k,n,\jmath} \sum_{j=j_n}^\infty \beta_j p_{j,n,P}\Big)\Big| \quad \text{(H.7)}
$$

by Assumption H.1(iii). Moreover, also note that for any $\{\beta_j\}, \{\tilde{\beta}_j\} \in \mathcal{B}_n$, we have that

$$\{E_P[q_{k,n,\jmath}^2(Z_{i,\jmath})(\sum_{j=j_n}^{\infty}(\beta_j - \tilde{\beta}_j)p_{j,n,P}(V_i))^2]\}^{1/2} \lesssim B_n \sum_{j=j_n}^{\infty}|\beta_j - \tilde{\beta}_j| \qquad \text{(H.8)}$$

by Assumptions 3.2(i) and H.1(i). Hence, since $\mathbb{W}_{n,P}^{\star}$ is sub-Gaussian with respect to $\|\cdot\|_{L_P^2}$, defining $\mathcal{G}_n \equiv \{f \in L_P^2 : f = q_{k,n,\jmath}\sum_{j \geq j_n}\beta_j p_{j,n,P} \text{ for some } 1 \leq \jmath \leq \mathcal{J} \text{ and } 1 \leq k \leq k_{n,\jmath}, \{\beta_j\} \in \mathcal{B}_n\}$ we obtain from Corollary 2.2.8 in van der Vaart and Wellner (1996)

$$E_P[\sup_{g \in \mathcal{G}_n}|\mathbb{W}_{n,P}^{\star}g|] \lesssim \int_0^{\infty}\sqrt{\log N(\epsilon, \mathcal{G}_n, \|\cdot\|_{L_P^2})}d\epsilon$$

$$\lesssim \int_0^{B_n M_n j_n^{-(\gamma_\rho - 1)}}\sqrt{\log(k_n N(\epsilon/B_n, \mathcal{B}_n, \|\cdot\|_{\ell_1}))}d\epsilon , \quad \text{(H.9)}$$

where the final inequality holds for $\|\beta\|_{\ell_1} \equiv \sum_{j=j_n}^{\infty}|\beta_j|$ by (H.8) and noting that since $\{p_{j,n,P}\}$ are uniformly bounded by Assumption H.1(i), $\mathcal{G}_n$ has envelope G_n satisfying $\|G_n\|_{L_P^2} \lesssim B_n M_n \sum_{j \geq j_n} j^{-\gamma_\rho} \lesssim B_n M_n j_n^{-(\gamma_\rho - 1)}$. Furthermore, note that Lemma H.2, the change of variables $u = \epsilon j_n^{\gamma_\rho - 1}/B_n M_n$, and $\gamma_\rho > 3/2$ additionally yield

$$\int_0^{B_n M_n j_n^{-(\gamma_\rho - 1)}}\sqrt{\log(k_n N(\epsilon/B_n, \mathcal{B}_n, \|\cdot\|_{\ell_1}))}d\epsilon$$

$$\leq \frac{B_n M_n}{j_n^{\gamma_\rho - 1}}\int_0^1\{\log(k_n) + (\frac{2j_n^{\gamma_\rho - 1}}{u(\gamma_\rho - 1)})^{\frac{1}{\gamma_\rho - 1}}\log(\frac{4}{uj_n} + 1)\}^{\frac{1}{2}}du \lesssim \frac{B_n M_n \sqrt{\log(k_n)}}{j_n^{\gamma_\rho - 3/2}} \quad \text{(H.10)}$$

Therefore, we conclude by results (H.7), (H.9), (H.10), and Markov's inequality that

$$\sup_{f \in \mathcal{F}_n}\|\mathbb{W}_{n,P}^{\star}(\text{Proj}\{f|\mathbb{P}_{n,P}^{\perp}\})q_n^{k_n}\|_r = O_p(\frac{k_n^{1/r}B_n M_n \sqrt{\log(k_n)}}{j_n^{\gamma_\rho - 3/2}}) . \qquad \text{(H.11)}$$

In order to obtain an analogous result to (H.11) for $\hat{\mathbb{W}}_n$, we similarly note that

$$\sup_{f \in \mathcal{F}_n}\|\hat{\mathbb{W}}_n(\text{Proj}\{f|\mathbb{P}_{n,P}^{\perp}\})q_n^{k_n}\|_r$$

$$\leq k_n^{1/r}\sup_{\{\beta_j\} \in \mathcal{B}_n}\max_{1 \leq \jmath \leq \mathcal{J}}\max_{1 \leq k \leq k_{n,\jmath}}|\hat{\mathbb{W}}_n(q_{k,n,\jmath}\sum_{j=j_n}^{\infty}\beta_j p_{j,n,P})| . \quad \text{(H.12)}$$

Moreover, since $\{\omega_i\}_{i=1}^n$ is independent of $\{V_i\}_{i=1}^n$, we also obtain from Assumption 3.2(i)

and $\{p_{j,n,P}\}_{j \geq j_n}$ being uniformly bounded in j, n and $P \in \mathbf{P}$ by Assumption H.1(i) that

$$E[\{\hat{\mathbb{W}}_n(q_{k,n,\jmath} \sum_{j \geq j_n} \beta_j p_{j,n,P}) - \hat{\mathbb{W}}_n(q_{k,n,\jmath} \sum_{j \geq j_n} \beta_j p_{j,n,P})\}^2 | \{V_i\}_{i=1}^n]$$

$$\leq \frac{1}{n} \sum_{i=1}^n \{q_{k,n,\jmath}(Z_i, \jmath) \sum_{j \geq j_n} (\beta_j - \tilde{\beta}_j) p_{j,n,P}(V_i)\}^2 \lesssim B_n^2 \{\sum_{j \geq j_n} |\beta_j - \tilde{\beta}_j|\}^2 . \quad \text{(H.13)}$$

Hence, since $\hat{\mathbb{W}}_n$ is Gaussian conditional on $\{V_i\}_{i=1}^n$, applying Corollary 2.2.8 in van der Vaart and Wellner (1996) and arguing as in (H.9) and (H.10) implies

$$E[\sup_{\{\beta_j\} \in \mathcal{B}_n} \max_{1 \leq \jmath \leq \mathcal{J}} \max_{1 \leq k \leq k_{n,\jmath}} |\hat{\mathbb{W}}_n(q_{k,n,\jmath} \sum_{j=j_n}^\infty \beta_j p_{j,n,P})| | \{V_i\}_{i=1}^n] \lesssim \frac{B_n M_n \sqrt{\log(k_n)}}{j_n^{\gamma_\rho - 3/2}} . \quad \text{(H.14)}$$

Thus, results (H.12) and (H.14) together with Markov's inequality allow us to conclude

$$\sup_{f \in \mathcal{F}_n} \|\hat{\mathbb{W}}_n(\text{Proj}\{f | \mathbb{P}_{n,P}^\perp\}) q_n^{k_n}\|_r = O_p(\frac{k_n^{1/r} B_n M_n \sqrt{\log(k_n)}}{j_n^{\gamma_\rho - 3/2}})) \quad \text{(H.15)}$$

uniformly in $P \in \mathbf{P}$. Hence, the claim of the Theorem follows from noting that the linearity of $\hat{\mathbb{W}}_n$ and $\mathbb{W}_{n,P}^\star$ and $f = \text{Proj}\{f | \mathbb{P}_{n,P}\} + \text{Proj}\{f | \mathbb{P}_{n,P}^\perp\}$ together imply that

$$\sup_{f \in \mathcal{F}_n} \|\hat{\mathbb{W}}_n f q_n^{k_n} - \mathbb{W}_{n,P}^\star f q_n^{k_n}\|_r \leq \sup_{f \in \mathcal{F}_n} \|\hat{\mathbb{W}}_n(\text{Proj}\{f | \mathbb{P}_{n,P}\}) q_n^{k_n} - \mathbb{W}_{n,P}^\star(\text{Proj}\{f | \mathbb{P}_{n,P}\}) q_n^{k_n}\|_r$$

$$+ \sup_{f \in \mathcal{F}_n} \|\hat{\mathbb{W}}_n(\text{Proj}\{f | \mathbb{P}_{n,P}^\perp\}) q_n^{k_n}\|_r + \sup_{f \in \mathcal{F}_n} \|\mathbb{W}_{n,P}^\star(\text{Proj}\{f | \mathbb{P}_{n,P}^\perp\}) q_n^{k_n}\|_r , \quad \text{(H.16)}$$

which in conjunction with results (H.6), (H.11), and (H.15) conclude the proof. ∎

Lemma H.1. *Let Assumptions 3.1, 3.2(i), H.1(i)-(ii) hold, and $\{\omega_i\}_{i=1}^n$ be i.i.d. with $\omega_i \sim N(0,1)$ independent of $\{V_i\}_{i=1}^n$. If $j_n k_n \log(j_n k_n) B_n = o(n)$, then uniformly in $\mathbf{P}$*

$$\sup_{\|\beta\|_2 \leq D_n} \|\hat{\mathbb{W}}_n(p_{n,P}^{j_n\prime} \beta) q_n^{k_n} - \mathbb{W}_{n,P}^\star(p_{n,P}^{j_n\prime} \beta) q_n^{k_n}\|_r = O_p(\frac{D_n B_n^{\frac{1}{2}} k_n^{\frac{1}{4} + \frac{1}{r}} (j_n \log(k_n))^{\frac{3}{4}} (\log(j_n))^{\frac{1}{4}}}{n^{\frac{1}{4}}}) ,$$

for $\mathbb{W}_{n,P}^\star$ a linear isonormal process on $\mathbb{S}_{n,P} \equiv \overline{span}\{p_{n,P}^{j_n} \otimes q_n^{k_n}\}$ independent of $\{V_i\}_{i=1}^n$.

PROOF: For notational simplicity, let $d_n \equiv j_n k_n$, set $r_{n,P}^{d_n}(v) \equiv q_n^{k_n}(z) \otimes p_{n,P}^{j_n}(v)$, and

$$\hat{\Sigma}_n(P) \equiv \frac{1}{n} \sum_{i=1}^n r_{n,P}^{d_n}(V_i) r_{n,P}^{d_n}(V_i)' \qquad \Sigma_n(P) \equiv E_P[r_{n,P}^{d_n}(V_i) r_{n,P}^{d_n}(V_i)'] . \quad \text{(H.17)}$$

Letting $r_{d,n,P}(v)$ denote the dth coordinate of $r_{n,P}^{d_n}(v)$ further note that $\|r_{d,n,P}\|_{L_P^\infty} \lesssim B_n$ since $\|p_{j,n,P}\|_{L_P^\infty}$ is uniformly bounded by Assumption H.1(ii) and $\|q_{k,n,\jmath}\|_{L_P^\infty} \leq B_n$ by

Assumption 3.2(i). Therefore, if for every $M > 0$ and $P \in \mathbf{P}$ we define the event

$$A_{n,P}(M) \equiv \{\|\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P)\|_{o,2} \le MR_n\} \ , \tag{H.18}$$

for $R_n \equiv \{j_n k_n \log(j_n k_n) B_n^2/n\}^{1/4}$, then Assumption H.1(ii) and Lemma H.3 yield

$$\liminf_{M \uparrow \infty} \liminf_{n \to \infty} \inf_{P \in \mathbf{P}} P(\{V_i\}_{i=1}^n \in A_{n,P}(M)) = 1 \ . \tag{H.19}$$

Next, let $\mathcal{N}_{d_n} \in \mathbf{R}^{d_n}$ follow a standard normal distribution and be independent of $\{(\omega_i, V_i)\}_{i=1}^n$ (defined on the same suitably enlarged probability space). Further let $\{\hat{\nu}_d\}_{d=1}^{d_n}$ denote eigenvectors of $\hat{\Sigma}_n(P)$, $\{\hat{\lambda}_d\}_{d=1}^{d_n}$ represent the corresponding (possibly zero) eigenvalues and define the random variable $\mathbb{Z}_{n,P} \in \mathbf{R}^{d_n}$ to be given by

$$\mathbb{Z}_{n,P} \equiv \sum_{d:\hat{\lambda}_d \ne 0} \hat{\nu}_d \times \frac{(\hat{\nu}_d' \hat{\mathbb{W}}_n(r_{n,P}^{d_n}))}{\sqrt{\hat{\lambda}_d}} + \sum_{d:\hat{\lambda}_d = 0} \hat{\nu}_d \times (\hat{\nu}_d' \mathcal{N}_{d_n}) \ . \tag{H.20}$$

Then note that since $\hat{\mathbb{W}}_n(r_{n,P}^{d_n}) \sim N(0, \hat{\Sigma}_n(P))$ conditional on $\{V_i\}_{i=1}^n$, and $\mathcal{N}_{d_n}$ is independent of $\{(\omega_i, V_i)\}_{i=1}^n$, $\mathbb{Z}_{n,P}$ is Gaussian conditional on $\{V_i\}_{i=1}^n$. Furthermore,

$$E[\mathbb{Z}_{n,P} \mathbb{Z}_{n,P}' | \{V_i\}_{i=1}^n] = \sum_{d=1}^{d_n} \hat{\nu}_d \hat{\nu}_d' = I_{d_n} \tag{H.21}$$

by direct calculation for I_{d_n} the $d_n \times d_n$ identity matrix, and hence $\mathbb{Z}_{n,P} \sim N(0, I_{d_n})$ conditional on $\{V_i\}_{i=1}^n$ almost surely in $\{V_i\}_{i=1}^n$ and is thus independent of $\{V_i\}_{i=1}^n$. Moreover, we also note that by Theorem 3.6.1 in Bogachev (1998) and $\hat{\mathbb{W}}_n(r_{n,P}^{d_n}) \sim N(0, \hat{\Sigma}_n(P))$ conditional on $\{V_i\}_{i=1}^n$, it follows that $\hat{\mathbb{W}}_n(r_{n,P}^{d_n})$ belongs to the range of $\hat{\Sigma}_n(P) : \mathbf{R}^{d_n} \to \mathbf{R}^{d_n}$ almost surely in $\{(\omega_i, V_i)\}_{i=1}^n$. Therefore, since $\{\hat{\nu}_d : \hat{\lambda}_d \ne 0\}_{d=1}^{d_n}$ spans the range of $\hat{\Sigma}_n(P)$, we conclude from (H.20) that for any $\gamma \in \mathbf{R}^{d_n}$ we must have

$$\gamma' \hat{\Sigma}_n^{1/2}(P) \mathbb{Z}_{n,P} = \gamma' \sum_{d:\hat{\lambda}_d \ne 0} \hat{\nu}_d (\hat{\nu}_d' \hat{\mathbb{W}}_n(r_{n,P}^{d_n})) = \hat{\mathbb{W}}_n(\gamma' r_{n,P}^{d_n}) \ . \tag{H.22}$$

Analogously, we also define for any $\gamma \in \mathbf{R}^{d_n}$ the isonormal process $\mathbb{W}_{n,P}^\star$ on $\mathbb{S}_{n,P}$ by

$$\mathbb{W}_{n,P}^\star(\gamma' r_{n,P}^{d_n}) \equiv \gamma' \Sigma_n^{1/2}(P) \mathbb{Z}_{n,P} \ , \tag{H.23}$$

which is trivially independent of $\{V_i\}_{i=1}^n$ due to the independence of $\mathbb{Z}_{n,P}$. Hence, letting $e_k \in \mathbf{R}^{k_n}$ denote the vector whose i^{th} coordinate equals $1\{i = k\}$, and $1\{A_{n,P}(M)\}$ be

an indicator for whether the event $\{V_i\}_{i=1}^n \in A_{n,P}(M)$ occurs, we obtain that

$$\sup_{\|\beta\|_2 \le D_n} \|\hat{\mathbb{W}}_n(p_{n,P}^{j_n\prime}\beta)q_n^{k_n} - \mathbb{W}_{n,P}^\star(p_{n,P}^{j_n\prime}\beta)q_n^{k_n}\|_r 1\{A_{n,P}(M)\}$$

$$\le k_n^{1/r} \times \sup_{\|\beta\|_2 \le D_n} \max_{1 \le k \le k_n} |(e_k \otimes \beta)'(\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P))\mathbb{Z}_{n,P}|1\{A_{n,P}(M)\} . \quad \text{(H.24)}$$

Defining $T_n \equiv \{1, \ldots, k_n\} \times \{\beta \in \mathbf{R}^{j_n} : \|\beta\|_2 \le D_n\}$, next set for any $(k, \beta) = t \in T_n$

$$\bar{\mathbb{W}}_{n,P}(t) \equiv |(e_k \otimes \beta)'(\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P))\mathbb{Z}_{n,P}|1\{A_{n,P}(M)\} , \quad \text{(H.25)}$$

and observe that conditional on $\{V_i\}_{i=1}^n$, $\bar{\mathbb{W}}_{n,P}(t)$ is sub-Gaussian under $d_n(\tilde{t}, t) \equiv \|(\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P))(e_{\tilde{k}} \otimes \tilde{\beta} - e_k \otimes \beta)\|_2$ for any $\tilde{t} = (\tilde{k}, \tilde{\beta})$ and $t = (k, \beta)$. Moreover, by standard arguments and definition (H.18), we obtain that under $A_{n,P}(M)$

$$N(\epsilon, T_n, d_n) \lesssim k_n \times (\frac{\|\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P)\|_{o,2} D_n}{\epsilon})^{j_n} \le k_n \times (\frac{MR_nD_n}{\epsilon})^{j_n} . \quad \text{(H.26)}$$

Therefore, noting that $\sup_{t,\tilde{t} \in T_n} d_n(t, \tilde{t}_n) \le 2MD_nR_n$ under the event $A_{n,P}(M)$, we obtain from Corollary 2.2.8 in van der Vaart and Wellner (1996) and (H.26) that

$$E[\sup_{t \in T_n} |\bar{\mathbb{W}}_{n,P}(t)| | \{V_i\}_{i=1}^n] \lesssim \int_0^\infty \sqrt{\log(N(\epsilon, T_n, d_n))} d\epsilon$$

$$\lesssim \int_0^{2MD_nR_n} \{\log(k_n) + j_n \log(\frac{MD_nR_n}{\epsilon})\}^{\frac{1}{2}} d\epsilon . \quad \text{(H.27)}$$

Hence, exploiting (H.27) and the change of variables $u = \epsilon/MD_nR_n$ we can conclude

$$E[\sup_{t \in T_n} |\bar{\mathbb{W}}_{n,P}(t)| | \{V_i\}_{i=1}^n]$$

$$\lesssim MD_nR_n \int_0^2 \{\log(k_n) + j_n \log(\frac{1}{u})\}^{\frac{1}{2}} du \lesssim MD_nR_n \times \sqrt{\log(k_n) + j_n} . \quad \text{(H.28)}$$

Next, for notational simplicity let $\delta_n \equiv k_n^{1/r} D_n R_n \sqrt{\log(k_n) + j_n}$, and then note that results (H.24), (H.25), and (H.28) together with Markov's inequality yield

$$P(\sup_{\|\beta\|_2 \le D_n} \|\hat{\mathbb{W}}_n(p_{n,P}^{j_n\prime}\beta)q_n^{k_n} - \mathbb{W}_{n,P}^\star(p_{n,P}^{j_n\prime}\beta)q_n^{k_n}\|_r > M^2\delta_n; \ A_{n,P}(M))$$

$$\le P(k_n^{1/r} \sup_{t \in T_n} |\bar{\mathbb{W}}_{n,P}(t)| > M^2\delta_n) \le E_P[\frac{k_n^{1/r}}{M^2\delta_n} E[\sup_{t \in T_n} |\bar{\mathbb{W}}_{n,P}(t)| | \{V_i\}_{i=1}^n]] \lesssim \frac{1}{M} .$$

$$\text{(H.29)}$$

Therefore, combining results (H.19) and (H.29), we can finally conclude that

$$\limsup_{M\uparrow\infty}\limsup_{n\to\infty}\sup_{P\in\mathbf{P}} P(\sup_{\|\beta\|_2\leq D_n}\|\hat{\mathbb{W}}_n(p_{n,P}^{j_n\prime}\beta)q_n^{k_n} - \mathbb{W}_n^\star(p_{n,P}^{j_n\prime}\beta)q_n^{k_n}\|_r > M^2\delta_n)$$

$$\lesssim \limsup_{M\uparrow\infty}\limsup_{n\to\infty}\sup_{P\in\mathbf{P}}\{\frac{1}{M} + P(\{V_i\}_{i=1}^n \notin A_{n,P}(M))\} = 0 \ , \quad \text{(H.30)}$$

which establishes the claim of the Lemma given the definitions of δ_n and R_n. ∎

Lemma H.2. *Let $\mathcal{B}_n \equiv \{\{\beta_j\}_{j=j_n}^\infty : \beta_j \leq M_n/j^{\gamma_\rho}\}$ for some $j_n \uparrow \infty$, $M_n > 0$, and $\gamma_\rho > 1$, and define the metric $\|\beta\|_{\ell_1} \equiv \sum_{j\geq j_n} |\beta_j|$. For any $\epsilon > 0$ it then follows that*

$$\log N(\epsilon,\mathcal{B}_n,\|\cdot\|_{\ell_1}) \leq \{(\frac{2M_n}{\epsilon(\gamma_\rho - 1)})^{\frac{1}{\gamma_\rho-1}} + 1 - j_n \vee 0\} \times \log(\frac{4M_n}{j_n^{\gamma_\rho}\epsilon} + 1) \ .$$

PROOF: For any $\{\beta_j\} \in \mathcal{B}_n$ and integer $k \geq (j_n-1)$ we first obtain the standard estimate

$$\sum_{j=k+1}^\infty |\beta_j| \leq \sum_{j=k+1}^\infty \frac{M_n}{j^{\gamma_\rho}} \leq M_n \int_k^\infty u^{-\gamma_\rho} du = M_n \frac{k^{-(\gamma_\rho-1)}}{(\gamma_\rho - 1)} \ . \quad \text{(H.31)}$$

For any $a \in \mathbf{R}$, let $\lceil a \rceil$ denote the smallest integer larger than a, and further define

$$j_n^\star(\epsilon) \equiv \lceil(\frac{2M_n}{\epsilon(\gamma_\rho - 1)})^{\frac{1}{\gamma_\rho-1}}\rceil \vee (j_n - 1) \ . \quad \text{(H.32)}$$

Then note that (H.31) implies that for any $\{\beta_j\} \in \mathcal{B}_n$ we have $\sum_{j>j^\star(\epsilon)} |\beta_j| \leq \epsilon/2$. Hence, letting $\mathcal{A}_n(\epsilon) \equiv \{\{\beta_j\} \in \mathcal{B}_n : \beta_j = 0$ for all $j > j_n^\star(\epsilon)\}$, we obtain

$$N(\epsilon,\mathcal{B}_n,\|\cdot\|_{\ell_1}) \leq N(\epsilon/2,\mathcal{A}_n(\epsilon),\|\cdot\|_{\ell_1})$$

$$\leq \prod_{j=j_n}^{j^\star(\epsilon)} N(\epsilon/2,[-\frac{M_n}{j^{\gamma_\rho}},\frac{M_n}{j^{\gamma_\rho}}],|\cdot|) \leq (\lceil\frac{4M_n}{j_n^{\gamma_\rho}\epsilon}\rceil)^{(j_n^\star(\epsilon)-j_n)\vee 0} \ , \quad \text{(H.33)}$$

where the product should be understood to equal one if $j^\star(\epsilon) = j_n - 1$. Thus, the claim of the Lemma follows from the bound $\lceil a \rceil \leq a + 1$ and results (H.32) and (H.33). ∎

Lemma H.3. *Let Assumption 3.1 hold, $\{f_{d,n,P}\}_{d=1}^{d_n}$ be a triangular array of functions $f_{d,n,P}: \mathbf{R}^{d_v} \to \mathbf{R}$, and define $f_{n,P}^{d_n}(v) \equiv (f_{1,n,P}(v),\ldots,f_{d_n,n,P}(v))'$ as well as*

$$\Sigma_n(P) \equiv E_P[f_{n,P}^{d_n}(V_i)f_{n,P}^{d_n}(V_i)'] \qquad \hat{\Sigma}_n(P) \equiv \frac{1}{n}\sum_{i=1}^n f_{n,P}^{d_n}(V_i)f_{n,P}^{d_n}(V_i)' \ .$$

If $\sup_{1\leq d\leq d_n} \|f_{d,n,P}\|_{L_P^\infty} \leq C_n$ for all $P \in \mathbf{P}$, the eigenvalues of $\Sigma_n(P)$ are bounded uniformly in n and $P \in \mathbf{P}$, and $d_n \log(d_n)C_n = o(n)$, then it follows that

$$\limsup_{M\uparrow\infty}\limsup_{n\to\infty}\sup_{P\in\mathbf{P}} P(\|\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P)\|_{o,2} > M\{\frac{d_n \log(d_n)C_n^2}{n}\}^{1/4}) = 0 \ .$$

PROOF: Set K_0 so that $\|\Sigma_n(P)\|_{o,2} \le K_0$ for all n and $P \in \mathbf{P}$, and then note that

$$\|\frac{1}{n}\{f_{n,P}^{d_n}(V_i)f_{n,P}^{d_n}(V_i)' - \Sigma_n(P)\}\|_{o,2} \le \frac{d_n C_n^2}{n} + \frac{K_0}{n} \tag{H.34}$$

almost surely for all $P \in \mathbf{P}$ since each entry of the matrix $f_{n,P}^{d_n}(V_i)f_{n,P}^{d_n}(V_i)'$ is bounded by C_n^2. Similarly, exploiting $\|f_{n,P}^{d_n}(V_i)f_{n,P}^{d_n}(V_i)'\|_{o,2} \le d_n C_n^2$ almost surely we obtain

$$\|\frac{1}{n}E_P[\{f_{n,P}^{d_n}(V_i)f_{n,P}^{d_n}(V_i)' - \Sigma_n(P)\}^2]\|_{o,2}$$
$$\le \frac{1}{n}\|E_P[\{f_{n,P}^{d_n}(V_i)f_{n,P}^{d_n}(V_i)'\}^2]\|_{o,2} + \frac{1}{n}\|\Sigma_n^2(P)\|_{o,2} \le \frac{d_n C_n^2 K_0}{n} + \frac{K_0^2}{n} \ . \tag{H.35}$$

Thus, employing results (H.34) and (H.35), together with $d_n \log(d_n)C_n = o(n)$, we obtain from Theorem 6.1 in Tropp (2012) (Bernstein's inequality for matrices) that

$$\limsup_{M\uparrow\infty}\limsup_{n\to\infty}\sup_{P\in\mathbf{P}} P(\|\hat{\Sigma}_n(P) - \Sigma_n(P)\|_{o,2} > M\frac{\sqrt{d_n \log(d_n)}C_n}{\sqrt{n}})$$
$$\le \limsup_{M\uparrow\infty}\limsup_{n\to\infty} d_n \exp\{-\frac{M^2 d_n \log(d_n)C_n^2}{2n}\frac{n}{(d_n C_n^2 + K_0)(K_0 + M)}\} = 0 \ . \tag{H.36}$$

Since $\hat{\Sigma}_n(P) \ge 0$ and $\Sigma_n(P) \ge 0$, Theorem X.1.1 in Bhatia (1997) in turn implies that

$$\|\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P)\|_{o,2} \le \|\hat{\Sigma}_n(P) - \Sigma_n(P)\|_{o,2}^{1/2} \tag{H.37}$$

almost surely, and hence the claim of the Lemma follows from (H.36) and (H.37). $\blacksquare$

References

AI, C. and CHEN, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, **71** 1795–1844.

AI, C. and CHEN, X. (2007). Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics*, **141** 5–43.

AI, C. and CHEN, X. (2012). The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *Journal of Econometrics*, **170** 442–457.

AIT-SAHALIA, Y. and DUARTE, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, **116** 9–47.

ALIPRANTIS, C. D. and BORDER, K. C. (2006). *Infinite Dimensional Analysis – A Hitchhiker's Guide*. Springer-Verlag, Berlin.

ANDREWS, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, **68** 399–405.

ANDREWS, D. W. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 683–734.

ANDREWS, D. W. K. and SHI, X. (2013). Inference based on conditional moment inequalities. *Econometrica*, **81** 609–666.

ANDREWS, D. W. K. and SOARES, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, **78** 119–157.

ARMSTRONG, T. B. (2015). Adaptive testing on a regression function at a point. Working paper, Yale University.

ATHEY, S. and STERN, S. (1998). An empirical framework for testing theories about complimentarity in organizational design. Tech. rep., National Bureau of Economic Research.

BEARE, B. K. and SCHMIDT, L. D. (2014). An empirical test of pricing kernel monotonicity. *Journal of Applied Econometrics* (in press).

BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). On the asymptotic theory for least squares series: pointwise and uniform results. *arXiv preprint arXiv:1212.0442v4*.

BERESTEANU, A. and MOLINARI, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, **76** 763–814.

BHATIA, R. (1997). *Matrix Analysis*. Springer, New York.

BLACK, D. A., GALDO, J. and SMITH, J. A. (2007). Evaluating the worker profiling and reemployment services system ysing a regression discontinuity approach. *The American Economic Review*, **97** 104–107.

BLUNDELL, R., HOROWITZ, J. L. and PAREY, M. (2012). Measuring the price responsiveness of gasoline demand: Economic shapre restrictions and nonparametric demand estimation. *Quantitative Economics*, **3** 29–51.

BLUNDELL, R., HOROWITZ, J. L. and PAREY, M. (2013). Nonparametric estimation of a heterogenous demand function udner the slutsky inequality restriction. Cemmap working paper cwp54/13.

BOGACHEV, V. I. (1998). *Gaussian Measures*. American Mathematical Society, Providence.

BOGACHEV, V. I. (2007). *Measure Theory: Volume I*. Springer-Verlag, Berlin.

BUGNI, F. A., CANAY, I. A. and SHI, X. (2014). Inference for functions of partially identified parameters in moment inequality models. Cemmap working papers, cwp05/14.

CALONICO, S., CATTANEO, M. D. and TITIUNIK, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, **82** 2295–2326.

CANAY, I. A., SANTOS, A. and SHAIKH, A. M. (2013). On the testability of identification in some nonparametric models with endogeneity. *Econometrica*, **81** 2535–2559.

CARD, D., LEE, D., PEI, Z. and WEBER, A. (2012). Nonlinear policy rules and the identification and estimation of causal effects in a generalized regression kink design. Working paper 18564, National Bureau of Economic Research.

CHAMBERLAIN, G. (1992). Comment: Sequential moment restrictions in panel data. *Journal of Business & Economic Statistics*, **10** 20–26.

CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2013). Improved risk bounds in isotonic regression. *arXiv preprint arXiv:1311.3765*.

CHEN, X. (2007). Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics 6B* (J. J. Heckman and E. E. Leamer, eds.). North Holland, Elsevier.

CHEN, X., CHERNOZHUKOV, V., LEE, S. and NEWEY, W. K. (2014). Local identification of nonparametric and semiparametric models. *Econometrica*, **82** 785–809.

Chen, X. and Christensen, T. (2013). Optimal uniform convergence rates for sieve nonparametric instrumental variables regression. Tech. rep., Working Paper, Yale University.

Chen, X. and Pouzo, D. (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics*, **152** 46–60.

Chen, X. and Pouzo, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, **80** 277–321.

Chen, X. and Pouzo, D. (2015). Sieve wald and qlr inferences on semi/nonparametric conditional moment models. *Econometrica*, **83** 1013–1079.

Chen, X., Pouzo, D. and Tamer, E. (2011a). Qlr inference on partially identified nonparametric conditional moment models. Working paper, Yale University.

Chen, X. and Reiss, M. (2011). On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, **27** 497–521.

Chen, X., Tamer, E. T. and Torgovitsky, A. (2011b). Sensitivity analysis in semiparametric likelihood models.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, **162** 47–70.

Chernozhukov, V., Fernandez-Val, I. and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, **96** 559–575.

Chernozhukov, V. and Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, **73** 245–261.

Chernozhukov, V., Hong, H. and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, **75** 1243–1284.

Chernozhukov, V., Lee, S. and Rosen, A. M. (2013). Intersection bounds: Estimation and inference. *Econometrica*, **81** 667–737.

Chetverikov, D. (2012). Testing regression monotonicity in econometric models. *arXiv preprint arXiv:1212.6757.*

Chetverikov, D. and Wilhelm, D. (2014). Nonparametric instrumental variable estimation under monotonicity. Tech. rep., Working paper, University College London.

CILIBERTO, F. and TAMER, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, **77** 1791–1828.

DE PAULA, A. and TANG, X. (2012). Inference of signs of interaction effects in simultaneous games with incomplete information. *Econometrica*, **80** 143–172.

DEUTSCH, F. (1982). Linear selections for the metric projection. *Journal of Functional Analysis*, **49** 269–292.

EVDOKIMOV, K. (2010). Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. *Department of Economics, Princeton University*.

FOX, J. T. and LAZZATI, N. (2014). Identification of discrete choice models for bundles and binary games. Working paper, University of Michigan.

FREYBERGER, J. and HOROWITZ, J. L. (2012). Identification and shape restrictions in nonparametric instrumental variables estimation. Cemmap working papers, cwp15/12.

GAGLIARDINI, P. and SCAILLET, O. (2012). Nonparametric instrumental variable estimation of structural quantile effects. *Econometrica*, **80** 1533–1562.

GALICHON, A. and HENRY, M. (2009). A test of non-identifying restrictions and confidence regions for partially identified parameters. *Journal of Econometrics*, **152** 186–196.

GENTZKOW, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*, **97** 713–744.

HAAG, B. R., HODERLEIN, S. and PENDAKUR, K. (2009). Testing and imposing slutsky symmetry in nonparametric demand systems. *Journal of Econometrics*, **153** 33–50.

HAHN, J., TODD, P. and VAN DER KLAAUW, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, **69** 201–209.

HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50** 891–916.

HAUSMAN, J. A. and NEWEY, W. K. (1995). Nonparametric estimation of exact consumers surplus and deadweight loss. *Econometrica* 1445–1476.

HO, K. and PAKES, A. (2014). Hospital choices, hospital prices, and financial incentives to physicians. *American Economic Review*, **104** 3841–3884.

HONG, H., MAHAJAN, A. and NEKIPELOV, D. (2010). Extremum estimation and numerical derivatives. Working paper, Stanford University.

HONG, S. (2011). Inference in semiparametric conditional moment models with partial identification. Tech. rep., Working Paper, Tinghua University.

HOROWITZ, J. L. (2007). Asymptotic normality of a nonparametric instrumental variables estimator. *International Economic Review*, **48** 1329–1349.

JACKWERTH, J. C. (2000). Recovering risk aversion from option prices and realized returns. *Review of Financial Studies*, **13** 433–451.

KAIDO, H. and SANTOS, A. (2014). Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities. *Econometrica*, **82** 387–413.

KOLTCHINSKII, V. I. (1994). Komlos-major-tusnady approximation for the general empirical process and haar expansions of classes of functions. *Journal of Theoretical Probability*, **7** 73–118.

KRESS, R. (1999). *Linear Integral Equations*. Springer, New York.

KRETSCHMER, T., MIRAVETE, E. J. and PERNÍAS, J. C. (2012). Competitive pressure and the adoption of complementary innovations. *The American Economic Review*, **102** 1540.

LEDOUX, M. and TALAGRAND, M. (1988). Un critère sur les petites boules dans le théorème limite central. *Probability Theory and Related Fields*, **77** 29–47.

LEE, D. S., MORETTI, E. and BUTLER, M. J. (2004). Do voters affect or elect policies? evidence from the u.s. house. *The Quarterly Journal of Economics*, **119** 808–859.

LEWBEL, A. (1995). Consistent nonparametric hypothesis tests with an application to slutsky symmetry. *Journal of Econometrics*, **67** 379–401.

LINTON, O., SONG, K. and WHANG, Y.-J. (2010). An improved bootstrap test of stochastic dominance. *Journal of Econometrics*, **154** 186–202.

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

MANSKI, C. F. (2003). *Partial Identification of Probability Distributions*. Springer-Verlag, New York.

MASSART, P. (1989). Strong approximation for multivariate empirical and related processes, via kmt constructions. *The Annals of Probability*, **17** 266–291.

MATZKIN, R. L. (1994). Restrictions of economic theory in nonparametric methods. In *Handbood of Econometrics* (R. Engle and D. McFadden, eds.), vol. IV. Elsevier.

NEWEY, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica* 1047–1070.

NEWEY, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, **79** 147–168.

NEWEY, W. K. and MCFADDEN, D. L. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics* (R. F. Engle and D. L. McFadden, eds.), vol. IV. Elsevier Science B.V., 2113–2245.

NEWEY, W. K. and POWELL, J. (2003). Instrumental variables estimation of nonparametric models. *Econometrica*, **71** 1565–1578.

PAKES, A. (2010). Alternative models for moment inequalities. *Econometrica*, **78** 1783–1822.

REGUANT, M. (2014). Complementary bidding mechanisms and startup costs in electricity markets. *The Review of Economic Studies*, **81** 1708–1742.

RIO, E. (1994). Local invariance principles and their application to density estimation. *Probability Theory and Related Fields*, **98** 21–45.

ROMANO, J. P. and SHAIKH, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference – Special Issue in Honor of Ted Anderson.*

ROMANO, J. P. and SHAIKH, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics*, **40** 2798–2822.

SANTOS, A. (2007). Inference in nonparametric instrumental variables with partial identification. Working paper, University of California at San Diego.

SANTOS, A. (2011). Instrumental variables methods for recovering continuous linear functionals. *Journal of Econometrics*, **161** 129–146.

SANTOS, A. (2012). Inference in nonparametric instrumental variables with partial identification. *Econometrica*, **80** 213–275.

SARGAN, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, **26** 393–415.

TAO, J. (2014). Inference for point and partially identified semi-nonparametric conditional moment models. Working paper, University of Washington.

TOPKIS, D. M. (1998). *Supermodularity and complementarity.* Princeton university press.

TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, **4** 389–434.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics.* Springer, New York.

Wolak, F. A. (2007). Quantifying the supply-side benefits from forward contracting in wholesale electricity markets. *Journal of Applied Econometrics*, **22** 1179–1209.

Yurinskii, V. V. (1977). On the error of the gaussian approximation for convolutions. *Theory of Probability and Its Applications*, **2** 236–247.

Zeidler, E. (1985). *Nonlinear Functional Analysis and its Applications I.* Springer-Verlag, Berlin.