cemmap

centre for microdata methods and practice

Constrained conditional moment restriction models

Victor Chernozhukov
Whitney K. Newey
Andres Santos

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP14/22

Economic and Social Research Council

Constrained Conditional Moment Restriction Models[*]

Victor Chernozhukov
M.I.T.
vchern@mit.edu

Whitney K. Newey[†]
M.I.T.
wnewey@mit.edu

Andres Santos[‡]
U.C.L.A.
andres@econ.ucla.edu

First Draft: September, 2015
This Draft: April 2022

Abstract

Shape restrictions have played a central role in economics as both testable implications of theory and sufficient conditions for obtaining informative counterfactual predictions. In this paper we provide a general procedure for inference under shape restrictions in identified and partially identified models defined by conditional moment restrictions. Our test statistics and proposed inference methods are based on the minimum of the generalized method of moments (GMM) objective function with and without shape restrictions. Uniformly valid critical values are obtained through a bootstrap procedure that approximates a subset of the true local parameter space. In an empirical analysis of the effect of childbearing on female labor supply, we show that employing shape restrictions in linear instrumental variables (IV) models can lead to shorter confidence regions for both local and average treatment effects. Other applications we discuss include inference for the variability of quantile IV treatment effects and for bounds on average equivalent variation in a demand model with general heterogeneity. We find in Monte Carlo examples that the critical values are conservatively accurate and that tests about objects of interest have good power relative to unrestricted GMM.

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

[*]We thank Riccardo D'amato for excellent research assistance. We are also indebted to three anonymous referees and numerous seminar participants for their valuable comments.

[†]Research supported by NSF Grant 1757140.

[‡]Research supported by NSF Grant SES-1426882.

1 Introduction

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; Chetverikov et al., 2018). A canonical example of this interplay between theory and practice is consumer demand analysis, where theoretical predictions such as Slutsky conditions have been extensively tested for and employed in estimation (Hausman and Newey, 1995, 2016; Blundell et al., 2012; Dette et al., 2016). 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).

Shape restrictions are often equivalent to inequality restrictions on parameters of interest and on certain unknown functions. For example, Slutsky negative semi-definiteness and monotonicity require that certain functions satisfy inequality restrictions. Inference with inequality restrictions is difficult. Such restrictions lead to discontinuities in (pointwise) limiting distributions where the inequality restrictions are "close" to binding, which makes inference challenging due to non-pivotal and potentially unreliable pointwise asymptotic approximations (Andrews, 2000, 2001). Limit discontinuities further make it difficult to construct confidence intervals with uniform coverage.

We address these challenges by obtaining critical values through a bootstrap procedure that uniformly approximates a subset of the local parameter space. The proposed critical values simultaneously deliver uniformly valid inference and pointwise limiting rejection probabilities that equal the nominal level of the test in many applications. Our results apply to a class of conditional moment restriction models (Ai and Chen, 2007, 2012) that encompasses parametric (Hansen, 1982), semiparametric (Ai and Chen, 2003), and nonparametric (Newey and Powell, 2003) instrumental variable (IV) models, as well as panel data applications (Chamberlain, 1992), and the study of plug-in functionals. For parametric IV our results deliver novel uniformly valid tests of inequality and equality restrictions as well as confidence intervals for parameters of interest in the presence of inequality restrictions in both identified and partially identified models.

Our test statistics and proposed inference methods are based on the difference of the minimum of a generalized method of moments (GMM) objective function with and without inequality restrictions. The value of the test statistic increases when more binding constraints are imposed. To ensure uniform validity, critical values are obtained through a bootstrap procedure that acknowledges that some inequalities that do not bind in the sample could have bound under a different draw of the sample. Intuitively, in the bootstrap, we impose the inequalities that are within a region of the boundary that shrinks slightly slower than the convergence rate of the shape restricted estimator. The bootstrap procedure can further be set to ignore inequalities that are outside this shrinking region, leading to pointwise rejection probabilities that equal the nominal level in many applications. As always, uniformity is essential for confidence intervals to be asymptotically valid over a set of unknown parameter values. The resulting inference is powerful in exploiting the large amount of information that inequality restrictions can provide in many cases relevant for applications.

Our tests and confidence intervals remain valid under partial identification. In this setting, the tests and confidence intervals give an accurate and computationally feasible method of doing inference for a subvector of parameters under partial identification. Indeed, these methods have already been used by Torgovitsky (2019) to construct informative confidence intervals for various partially identified state dependence parameters in the presence of unobserved heterogeneity. Also, Kline and Walters (2021) used these methods to test shape constraints implied by a model of callback probabilities for employment applications. By incorporating nuisance parameters into the definition of the parameter space, our results can further be applied to partially identified semi(non)-parametric models defined by conditional moment inequalities.

We demonstrate the usefulness of this approach in an empirical application. Specifically, we conduct inference on the causal effect of childbearing on female labor force participation by relying on the instrumental variables approach of Angrist and Evans (1998). We find that monotonicity of the local average treatment effect (LATE) in education is not rejected by the data and neither is monotonicity and negativity – these restrictions were discussed, but not formally tested, by Angrist and Evans (1998). We further find that imposing these shape restrictions yields narrower confidence intervals for the LATE at different schooling levels. Finally, we obtain similar results for the partially identified average treatment effect (ATE), though the data is less informative about the ATE because of the low proportion of compliers.

The inequalities associated with nonparametric shape restrictions necessitate consideration of parameter spaces that are sufficiently general yet endowed with enough structure to ensure a fruitful asymptotic analysis. An important theoretical insight of this paper is that this simultaneous flexibility and structure is possessed by sets defined by inequality restrictions on Abstract M (AM) spaces; i.e. Banach lattices whose norm obeys a condition discussed in Section 3. We also introduce potentially regularized approximations to the local parameter spaces in order to account for the curvature present in nonlinear constraints. While aspects of our analysis are specific to models defined by conditional moment restrictions, the role of the local parameter space is solely dictated by the shape restrictions. As such, we expect the insights of the set up here to be applicable to the study of shape restrictions in alternative models as well. The critical values are shown to be uniformly asymptotically valid by developing strong approximations to both the test and bootstrap statistics. Sufficient conditions are provided by adapting the coupling of Koltchinskii (1994). Our coupling arguments and the use of AM spaces are key features of the theory that enable us to show that inference is uniformly valid and that partial identification is permitted.

We illustrate the general applicability of our analysis by obtaining novel uniformly valid inference results in a variety of problems. Specifically, we: (i) Conduct inference about partially identified sets of average equivalent variation and other objects of interest in demand estimation with general heterogeneity and smooth demand functions; (ii) Test and impose shape restrictions on structural functions identified through quantile conditional moment restrictions; and (iii) Impose the Slutsky restrictions to conduct inference in a linear conditional moment restriction model. Additionally, while we do not pursue further examples in detail for conciseness, we note our results may be applied to conduct tests of homogeneity, supermodularity, and economies of scale or scope.

In a small Monte Carlo study, we examine instrumental variables estimation of a nonlinear structural function and consider the power of imposing monotonicity and/or convexity on the structural function. We find rejection frequencies for our test that are conservatively accurate when testing a point null hypothesis about the value or derivative of the structural function. In addition, we find that imposing shape restrictions leads to large increases in power relative to employing an unrestricted estimator, in moderately large samples. Our Monte Carlo analysis further examines the performance of our test in a partially identified parametric IV model with discrete data. In that context, we find that shape restrictions have substantial identifying power and that our test provides valid inference on the value of a function at a point. A similar partially identified IV setting was previously studied by Freyberger and Horowitz (2015), who also provide an inference procedure. However, their procedure is based on limiting distributions that are discontinuous in true parameters leading to nonuniform inference.

Our paper contributes to an extensive literature studying semiparametric and nonparametric models under partial identification (Manski, 2003; Molinari, 2020). When specialized to finite dimensional models, our results enable us to conduct inference on functionals of the identified set in models defined by moment (in)equalities (Canay and Shaikh, 2017; Ho and Rosen, 2017). In that context, our results are complementary to those of Bugni et al. (2017) and Kaido et al. (2019), who provide uniformly valid procedures for subvector inference. Their analysis is focused on convex models and can thus be invalid or conservative when conducting inference on nonlinear functionals or imposing non-convex restrictions – we emphasize, however, that their analysis is also motivated by a different set of models than the ones we consider. Our analysis is further related to Hong (2017), Santos (2012), Tao (2014), and Chen et al. (2011) who study inference on functionals of potentially partially identified structural functions, but do not allow for shape constraints as we do.

Following the original version of this paper, Zhu (2019) and Fang and Seo (2019) have proposed inference methods for convex restrictions which, while applicable to an important class of problems, rule out inference on nonlinear functionals or tests of certain shape restrictions. Also related is Freyberger and Reeves (2018) who have more recently developed uniform inference for functionals under shape restrictions while imposing point identification. Our paper is of course related to a large literature on shape restrictions; see Samworth and Sen (2018) and Chetverikov et al. (2018) for recent reviews. We highlight here an important literature on linear Gaussian models focused on adaptivity (which we do not establish), but not applicable to many of the models that motivate us (Dumbgen and Spokoiny, 2001; Cai et al., 2013; Armstrong, 2015).

The results here are also highly complementary to Chetverikov and Wilhelm (2017) in providing inference for nonparametric IV under shape restrictions while they showed that imposing monotonicity can greatly improve the convergence rate of the estimator – an observation that additionally motivates our use of test statistics based on shape constrained (instead of unconstrained) estimators. Finally, we note that our results do not lend themselves computationally for the construction of uniform confidence bands for shape restricted functions – a problem that has been addressed in different contexts by Chernozhukov et al. (2009) and Horowitz and Lee (2017).

The remainder of the paper is organized as follows. In Section 2 we show how to implement our tests in a linear instrumental variables model with inequality restrictions under both point and partial identification. Section 2 further illustrates our results by revisiting the analysis of Angrist and Evans (1998). Section 3 contains our main theoretical results, while Section 4 applies them to conduct inference in the heterogenous demand model of Hausman and Newey (2016). Finally, Section 5 contains a brief simulation study. All mathematical derivations are included in a series of appendices; see in particular Appendix A.2 for applications of our general results and Appendix S.6 for coupling results based on Koltchinskii (1994).

2 Application for Linear Instrumental Variables

To fix ideas, we first describe our test in a linear instrumental variables model and illustrate its implementation by revisiting the analysis of Angrist and Evans (1998). We reserve until later the full mathematical framework and focus here on implementation.

2.1 Linear Instrumental Variables

As perhaps the simplest possible example, we first consider a linear instrumental variable model in which $\theta_0 \in \Theta \subseteq \mathbf{R}^{d_\theta}$ is identified through the moment conditions

$$E_P[(Y - W'\theta_0)Z] = 0,$$

where Y is a scalar, W and Z are vectors, and P denotes the distribution of $V \equiv (Y, W, Z)$. We are interested in testing whether θ_0 belongs to a set R characterized by

$$R = \{\theta \in \mathbf{R}^{d_\theta} : F\theta = f,\ G\theta \leq g\}, \tag{1}$$

for known matrices F and G and known vectors f and g.

We consider tests based on minimizing the norm of the weighted sample moments as in Sargan (1958) and Hansen (1982). To this end, we define the criterion

$$Q_n(\theta) \equiv \|\hat{\Sigma}_n\{\frac{1}{n}\sum_{i=1}^{n}(Y_i - W_i'\theta)Z_i\}\|_2, \tag{2}$$

where $\|\cdot\|_2$ is the standard Euclidean norm and $\hat{\Sigma}_n$ is consistent for $(E[ZZ'U^2])^{-1/2}$ for $U \equiv Y - W'\theta_0$. Our analysis then enables us to employ tests based on the statistics

$$I_n(R) \equiv \min_{\theta \in \Theta \cap R} \sqrt{n} Q_n(\theta) \qquad\qquad I_n(\Theta) \equiv \min_{\theta \in \Theta} \sqrt{n} Q_n(\theta); \tag{3}$$

e.g., we may consider a test that rejects for large values of $I_n(R) - I_n(\Theta)$. In what follows it will also be helpful to let $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$ denote the minimizers of Q_n over $\Theta \cap R$ and Θ respectively – i.e. $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$ are the constrained and unconstrained estimators.

We construct critical values by relying on the multiplier bootstrap (Ledoux and Talagrand, 1988). Specifically, let $b \in \{1, \ldots, B\}$ index a bootstrap draw, $\{\omega_i^b\}_{i=1}^n$ be i.i.d. independent of the data with $\omega_i^b \sim N(0,1)$, and for any $\theta \in \mathbf{R}^{d_\theta}$ define

$$\hat{\mathbb{W}}_n^b(\theta) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_i^b\{(Y_i - W_i'\theta)Z_i - \frac{1}{n}\sum_{j=1}^{n}(Y_j - W_j'\theta)Z_j\},$$

which is a simulated draw of the true (centered) moment functions.[1] We also require an estimator of the derivative of the moment conditions, and to this end we set

$$\hat{\mathbb{D}}_n[h] \equiv -\frac{1}{n}\sum_{i=1}^{n} Z_i W_i' h.$$

[1]We follow previous work (Lewbel, 1995; Hansen, 1996) in considering Gaussian weights $\{\omega_i\}_{i=1}^n$ because it simplifies the proofs of our main results in Section 3. We expect our analysis extends to alternative specifications for the distribution of $\{\omega_i\}_{i=1}^n$ – e.g., for ω_i following an exponential distribution, which results in a version of the Bayesian bootstrap advocated by Chamberlain and Imbens (2003).

Here, we can think of h as a local parameter, representing the possible values that the random variable $\sqrt{n}\{\hat{\theta}_n - \theta_0\}$ may take (recall $\hat{\theta}_n$ is the minimizer of Q_n over $\Theta \cap R$).

Finally, we need to enforce the inequality constraints in the bootstrap in a way that delivers a uniformly valid critical value. To this end, we account for the variation in $G_j\hat{\theta}_n - g_j$ for each j, where G_j is the j^{th} row of G and g_j the j^{th} coordinate of g. That is, we account for the likelihood that a constraint will bind at the restricted estimator $\hat{\theta}_n$ when computing $I_n(R) = \sqrt{n}Q_n(\hat{\theta}_n)$. For this purpose we introduce the set

$$\hat{V}_n(\hat{\theta}_n, R) \equiv \{h \in \mathbf{R}^{d_\theta} : Fh = 0,\ G_j h \leq \sqrt{n}\max\{0, -(r_n + G_j\hat{\theta}_n - g_j)\} \text{ for all } j\}, \quad (4)$$

where $r_n > 0$ is a slackness parameter whose choice we discuss shortly. The set $\hat{V}_n(\hat{\theta}_n, R)$ can be thought of as a local version of R, approximating the set of values h that could equal $\sqrt{n}\{\hat{\theta}_n - \theta_0\}$. Our bootstrap approximations to $I_n(R)$ and $I_n(\Theta)$ are then

$$\hat{U}_n^b(R) \equiv \min_{h \in \hat{V}_n(\hat{\theta}_n, R)} \|\hat{\Sigma}_n\{\hat{\mathbb{W}}_n^b(\hat{\theta}_n) + \hat{\mathbb{D}}_n[h]\}\|_2 \quad (5)$$

$$\hat{U}_n^b(\Theta) \equiv \min_{h \in \mathbf{R}^{d_\theta}} \|\hat{\Sigma}_n\{\hat{\mathbb{W}}_n^b(\hat{\theta}_n^{\mathrm{u}}) + \hat{\mathbb{D}}_n[h]\}\|_2. \quad (6)$$

Thus, we may obtain a level α test by rejecting whenever the test statistic $I_n(R) - I_n(\Theta)$ exceeds the $1-\alpha$ quantile of $\hat{U}_n^b(R) - \hat{U}_n^b(\Theta)$ across the B bootstrap draws. The main assumption required for the test to be asymptotically valid is that θ_0 be strongly identified – i.e. θ_0 can be consistently estimated uniformly in P.

The critical value depends on the choice of r_n. When applied to linear instrumental variables, our asymptotic theory requires that r_n tend to zero slower than the convergence rate of the restricted estimator, which is $1/\sqrt{n}$. Heuristically, when r_n tends to zero any constraint that is not binding at θ_0 will also not be binding in the bootstrap with probability approaching one (under pointwise in P asymptotics). Consequently inference is not asymptotically conservative for a fixed data generating process. Setting $r_n \to 0$ while satisfying $r_n\sqrt{n} \to \infty$ leads to uniformly valid inference with constraints only being conservatively enforced when they are within order $1/\sqrt{n}$ of binding at θ_0. Setting $r_n = +\infty$ is always theoretically valid, but it may be conservative and result in a loss of power. Other, smaller choices of r_n can lead to smaller, valid critical values and so may result in more powerful tests and tighter confidence intervals than $r_n = +\infty$.

Intuitively, r_n is meant to quantify the sampling uncertainty in $G\{\hat{\theta}_n - \theta_0\}$. Since the distribution of $\hat{\theta}_n$ cannot be uniformly consistently estimated, we suggest linking r_n to the degree of sampling uncertainty in $G\{\hat{\theta}_n^{\mathrm{u}} - \theta_0\}$ instead. Specifically, for $\hat{\theta}_n^{\mathrm{u}\star}$ a "bootstrap" analogue of $\hat{\theta}_n^{\mathrm{u}}$ and some $\gamma_n \to 0$, we recommend setting r_n to satisfy

$$P(\max_j G_j\{\hat{\theta}_n^{\mathrm{u}} - \hat{\theta}_n^{\mathrm{u}\star}\} \leq r_n|\text{Data}) = 1 - \gamma_n. \quad (7)$$

This approach changes the problem of selecting r_n into the problem of selecting γ_n. However, γ_n is more interpretable: If we employed $\hat{V}_n(\hat{\theta}_n^{\mathrm{u}}, R)$ in place of $\hat{V}_n(\hat{\theta}_n, R)$ in (5), then a Bonferroni bound implies that the test that rejects whenever $I_n(R) - I_n(\Theta)$ exceeds the $1-\alpha$ quantile of $\hat{U}_n^b(R) - \hat{U}_n^b(\Theta)$ has asymptotic size at most $\alpha + \gamma_n$ even if γ_n is fixed with n.[2] In particular, if we employed the $1-\alpha+\gamma_n$ quantile of $\hat{U}_n^b(R) - \hat{U}_n^b(\Theta)$ as a critical value instead, then the resulting test would have asymptotic size at most α (even if γ_n is fixed). In simulations, however, we find the described bound to be pessimistic in that, when setting r_n according to (7), our test has a rejection probability under the null hypothesis of at most α for a wide range of choices of γ_n.

Remark 2.1. Our results may be employed to obtain confidence regions for a coordinate of θ_0 while imposing restrictions of the form $G\theta_0 \leq g$ on θ_0 (e.g., sign or monotonicity restrictions on $w \mapsto w'\theta_0$). For example, for θ_k the k^{th} coordinate of $\theta \in \mathbf{R}^{d_\theta}$ let

$$R_\lambda = \{\theta \in \mathbf{R}^{d_\theta} : \theta_k = \lambda,\ G\theta \leq g\},$$

which is a special case of (1). We may then obtain a confidence region for the k^{th} coordinate of θ_0 by conducting test inversion in λ employing the test based on $I_n(R_\lambda) - I_n(\Theta)$; see also Remark 3.1 for alternative constructions based on our analysis. ■

Remark 2.2. In certain applications it may be desirable to studentize the constraints in our bootstrap approximation – i.e. replace G_j and g_j by $G_j/\hat{\sigma}_j$ and $g_j/\hat{\sigma}_j$ everywhere in (4) (and in (7) if employed). In the empirical analysis below we proceed in this manner by setting $\hat{\sigma}_j^2$ to be an estimate of the asymptotic variance of $\sqrt{n}G_j\{\hat{\theta}_n^{\mathrm{u}} - \theta_0\}$. ■

2.1.1 Fertility and Labor Supply: LATE

We illustrate the preceding discussion by revisiting the study by Angrist and Evans (1998) on the causal effect of childbearing on female labor force participation. Like Angrist and Evans (1998), we employ the 1980 Census Public Use Micro Sample restricted to mothers aged 21-35 with at least two children, and set: (i) $D \in \{0,1\}$ to indicate whether a mother has more than two children (the treatment); (ii) $Y \in \{0,1\}$ to indicate whether a mother is employed (the outcome of interest); and (iii) $Z \in \{0,1\}$ to indicate whether the first two children are of the same sex (the instrument). We further adopt the heterogeneous treatment effects model of Imbens and Angrist (1994) and let Y_d denote the potential outcome under treatment status $d \in \{0,1\}$ and employ "`C`," "`NT`," and "`AT`" to denote compliers, never takers, and always takers.

Angrist and Evans (1998) document that the impact of childbearing on labor force participation depends on observable characteristics. In particular, their two stage least

[2] While we may replace $\hat{V}_n(\hat{\theta}_n, R)$ with $\hat{V}_n(\hat{\theta}_n^{\mathrm{u}}, R)$ in identified models, in partially identified models we employ $\hat{V}_n(\hat{\theta}_n, R)$ due to the identified set potentially not being a subset of R under the null hypothesis.

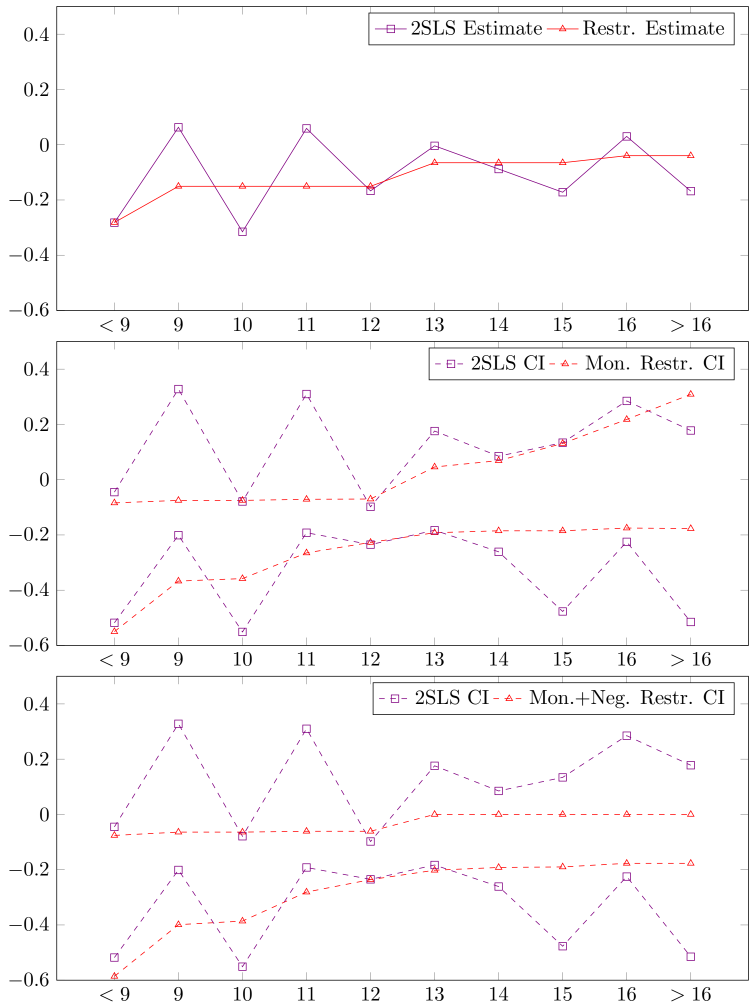


Figure 1: First Panel: Unconstrained and shape restricted LATE estimates (imposing monotonicty or monotonicity and negativity yield the same estimates). Second and Third Panels: 95% Confidence intervals for LATE at different education levels.

squares (2SLS) estimates suggest a negative impact of childbearing on labor force participation across different levels of schooling, but that the magnitude of the impact decreases with schooling – a phenomenon that may reflect that more educated mothers have a stronger attachment to the labor force. To formally examine this claim, we introduce dummy variables S for each year of schooling between 9 and 16 and for the categories "less than 9" and "more than 16." Defining the local average treatment effects

$$\mathrm{LATE}(S) \equiv E[Y_1 - Y_0 | S, \mathtt{C}]$$

we then test whether: (i) LATE$(\cdot)$ is increasing in schooling, and (ii) LATE$(\cdot)$ is increasing in schooling and nonpositive. Both hypotheses fall within the framework of the preceding section because LATE$(\cdot)$ is identified through linear moment restrictions and the hypothesized restrictions are linear in LATE$(\cdot)$. Employing five thousand bootstrap replications and setting $r_n = +\infty$ or r_n as suggested in (7) with $\gamma_n = 0.05$ yields in this case equal p-values that fail to reject either null hypothesis. The p-values for LATE$(\cdot)$ being nondecreasing is 0.21 and for it being nondecreasing and nonpositive is 0.394.

In Figure 1 we study the values of LATE(S) at different schooling levels S. The first panel displays the unconstrained 2SLS estimates and their monotonicity restricted counterparts – the latter are negative and hence additionally demanding nonpositivity does not change the estimates. Unfortunately, two sided confidence regions based on the (pointwise in P) asymptotic distribution of the shape-restricted 2SLS estimator can asymptotically undercover the true parameter. In the second panel of Figure 1 we instead proceed as in Remark 2.1 to obtain 95% confidence intervals while imposing monotonicity and again selecting r_n by setting $\gamma_n = 0.05$ in (7). Employing the monotonicity restriction in this manner yields confidence intervals that are sometimes substantially shorter than their 2SLS counterparts. Notably, we observe lower upper ends for the restricted confidence intervals at the lower education levels and higher lower ends at higher education levels. As shown in the third panel of Figure 1, additionally imposing that LATE$(\cdot)$ be nonpositive mostly reduces the upper bound of our confidence intervals at higher education levels.

2.2 Partial Identification

We next illustrate the implementation of our results in a partially identified setting. With an eye towards extending the preceding empirical analysis to study average treatment effects (ATEs), we maintain that the parameter of interest $\theta_0 \in \Theta \subseteq \mathbf{R}^{d_\theta}$ satisfies

$$E_P[(Y - W'\theta_0)Z] = 0, \tag{8}$$

but no longer assume θ_0 is identified by (8). Instead, we define the identified set

$$\Theta_0 \equiv \{\theta \in \Theta : E_P[(Y - W'\theta)Z] = 0\} \tag{9}$$

and consider the problem of testing whether the intersection of Θ_0 and R is nonempty (i.e. $\Theta_0 \cap R \neq \emptyset$). Such hypotheses can be employed, for instance, to build confidence regions for functionals of the identified set; see Remark 2.3 below. We also now set

$$R = \{\theta \in \mathbf{R}^{d_\theta} : \Upsilon_F(\theta) = 0,\ G\theta \leq g\}, \tag{10}$$

for Υ_F a known possibly nonlinear function – e.g., $\Upsilon_F(\theta) = F\theta - f$ recovers (1).

We continue to rely on the statistics $I_n(R)$ and $I_n(\Theta)$ (as in (3)) for inference. However, since in many settings in which θ_0 fails to be identified by (8) we will have that the dimension of Z is smaller than that of W, in what follows we assume for ease of exposition that $I_n(\Theta) = 0$ (almost surely); see Section 3.2.2 for a general discussion. Another distinction relative to Section 2.1 is that the choice of $\hat{\Sigma}_n$ (as in (2)) may need to be modified in settings in which $U \equiv Y - W'\theta_0$ cannot be consistently estimated due to θ_0 being partially identified. In such instances we may, for example, set

$$\hat{\Sigma}_n \equiv (\frac{1}{n}\sum_{i=1}^n Z_i Z_i'(Y_i - W_i'\hat{\theta}_n^{\mathrm{u}})^2)^{-1/2},$$

where we now interpret $\hat{\theta}_n^{\mathrm{u}}$ as the minimum norm minimizer of Q_n over Θ. While the choice of $\hat{\Sigma}_n$ has an impact on how local power is directed, we note that the test has correct asymptotic size provided $\hat{\Sigma}_n$ converges in probability to a non-stochastic limit.

Our bootstrap procedure requires two modifications relative to our preceding discussion. First, because in (10) we consider nonlinear equality constraints, we now set

$$\hat{V}_n(\theta, R) \equiv \{h \in \mathbf{R}^{d_\theta} : \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0,\ G_j h \leq \sqrt{n}\max\{0, -(r_n + G_j\theta - g_j)\} \text{ for all } j\}$$

(notice that if $\Upsilon_F(\theta) = F\theta - f$, then we recover (4)). A distinction with Section 2.1 is that if one aims to employ (7) to select r_n, then an alternative to an unrestricted estimator $\hat{\theta}_n^{\mathrm{u}}$ may be necessary; see Section 2.2.1 for an example. Second, our bootstrap approximation employs an estimator $\hat{\Theta}_n^{\mathrm{r}}$ for $\Theta_0 \cap R$. To this end, we set

$$\hat{\Theta}_n^{\mathrm{r}} \equiv \{\theta \in \Theta \cap R : Q_n(\theta) \leq \inf_{\theta \in \Theta \cap R} Q_n(\theta) + \tau_n\}$$

where $\tau_n \geq 0$ is a bandwidth whose choice we discuss shortly – i.e. $\hat{\Theta}_n^{\mathrm{r}}$ is the set of "near" minimizers of Q_n over $\Theta \cap R$. Our bootstrap approximation to $I_n(R)$ then equals

$$\hat{U}_n^b(R) \equiv \min_{\theta \in \hat{\Theta}_n^{\mathrm{r}}} \min_{h \in \hat{V}_n(\theta, R)} \|\hat{\Sigma}_n\{\hat{\mathbb{W}}_n^b(\theta) + \hat{\mathbb{D}}_n[h]\}\|_2.$$

Thus, to obtain a level α test we reject the null hypothesis whenever $I_n(R)$ exceeds the $1 - \alpha$ quantile of $\hat{U}_n^b(R)$ across bootstrap draws. Paralleling Section 2.1, a principal assumption for the test to be asymptotically valid is that Θ_0 be strongly identified.

When specialized to the current setting, our asymptotic theory requires that τ_n tend to zero. It is theoretically valid to set $\tau_n = 0$, which simplifies the computation of our bootstrap statistic – e.g., let $\hat{\Theta}_n^{\mathrm{r}} = \{\hat{\theta}_n\}$ for any $\hat{\theta}_n$ minimizing Q_n over $\Theta \cap R$ to recover (5). However, setting $\tau_n = 0$ can result in lower power in applications for which the corresponding $\hat{\Theta}_n^{\mathrm{r}}$ is not consistent for $\Theta_0 \cap R$ (in the Hausdorff metric) – to ensure consistency, τ_n must in addition satisfy $\tau_n\sqrt{n} \to \infty$. For applications in which it is desirable to set $\tau_n > 0$, we propose a procedure inspired by Romano and Shaikh (2010).

Specifically, for any set $K \subseteq \Theta \cap R$ we define the quantile $\hat{q}_n(K)$ according to

$$P(\sup_{\theta \in K} \|\hat{\Sigma}_n \hat{\mathbb{W}}_n(\theta)\|_2 \leq \hat{q}_n(K)|\text{Data}) = 1 - \gamma_n$$

where $\gamma_n \in (0,1)$. Letting $S_1 \equiv \Theta \cap R$, we then inductively define $S_{j+1} \equiv \{\theta \in \Theta \cap R : \sqrt{n}Q_n(\theta) \leq \hat{q}_n(S_j)\}$ noting that by construction $S_{j+1} \subseteq S_j$. To select τ_n, we proceed inductively until we find $S_j = \emptyset$, in which case we set $\tau_n = 0$, or $S_{j+1} = S_j \neq \emptyset$, in which case we set $\tau_n = \hat{q}_n(S_j)$. Heuristically, under such a choice of τ_n, the set $\hat{\Theta}_n^{\mathrm{r}}$ may be interpreted as a $1-\gamma_n$ confidence region for $\Theta_0 \cap R$. While power considerations suggest setting γ_n to tend to zero, for practical considerations we suggest simply setting $1-\gamma_n$ to be a high quantile fixed with n (e.g., $1-\gamma_n = 0.8$).

Remark 2.3. The introduced test can be employed to obtain confidence regions for functionals of the identified set satisfying the coverage requirement advocated by Imbens and Manski (2004). Specifically, given a functional $\Upsilon_F : \Theta \to \mathbf{R}$ we may set

$$R_\lambda = \{\theta \in \mathbf{R}^{d_\theta} : \Upsilon_F(\theta) = \lambda, G\theta \leq g\}$$

and obtain the desired confidence region by conducting test inversion in λ of the null hypothesis that the set $\Theta_0 \cap R_\lambda$ is not empty. ■

2.2.1 Fertility and Labor Supply: ATE

Returning to our analysis of the causal impact of fertility on female labor force participation, we next turn to estimating the average treatment effect at different education levels S (denoted ATE(S)). Following the literature, we decompose ATE(S) into

$$\text{LATE}(S)P(\texttt{C}|S) + E[Y_1 - Y_0|S, \texttt{AT}]P(\texttt{AT}|S) + E[Y_1 - Y_0|\texttt{NT}, S]P(\texttt{NT}|S), \qquad (11)$$

where recall `C`, `AT`, and `NT` denote "compliers," "always takers," and "never takers." With the exception of $E[Y_0|\texttt{AT}, S]$ and $E[Y_1|\texttt{NT}, S]$, all terms in (11) can be identified through linear moment restrictions.[3] Because S has ten support points, we obtain sixty moments and eighty parameters so that $I_n(\Theta) = 0$ almost surely.

Following our analysis of LATE(S) we conduct inference on ATE(S) under three increasingly stringent set of (linear) restrictions: (i) The logical bounds implied by $Y_d \in \{0,1\}$; (ii) Adding to (i) that the average treatment effect be increasing in schooling among all types (i.e. `C`, `NT`, and `AT`); (iii) Adding to (ii) that average treatment effects be nonpositive for all levels of education and types. Figure 2 reports the resulting

[3]Technically, the moment equations have the structure $E_P[(Y_\jmath - W_\jmath'\theta_0)Z_\jmath] = 0$ with the instruments $Z_\jmath$ not being common across all $1 \leq \jmath \leq \mathcal{J}$ equations. The bootstrap implementation in this case, formally studied in Section 3, is identical with only $\hat{\mathbb{W}}_n$ and $\hat{\mathbb{D}}_n$ being modified in the natural way.

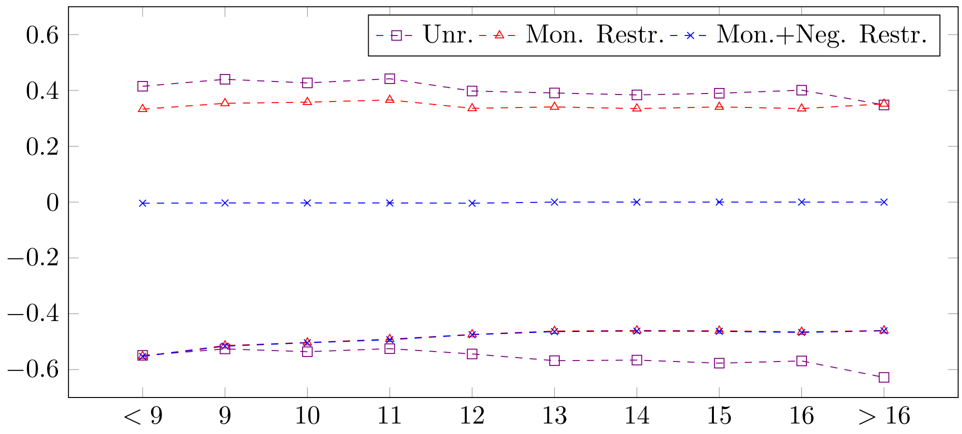


Figure 2: 95% Confidence intervals for ATE at different education levels. "Unr." uses bounds implied by $Y_d \in \{0, 1\}$; "Mon. Restr." adds that average treatment effects be increasing in education for all types; "Mon.+Neg. Restr." also requires they be negative.

95% confidence regions obtained through the approach described in Remark 2.3 – here, the restriction $G\theta \leq g$ imposes the described shape constraints while the nonlinear restriction $\Upsilon_F(\theta) = 0$ corresponds to imposing a hypothesized value for ATE(S) through (11). In our bootstrap approximation, we set $\tau_n = 0$ and selected r_n according to (7) with $\gamma_n = 0.05$ and where, when necessary, we used the distribution of estimators of identified parameters for their partially identified counterparts.[4] We do not report estimates of the identified sets for ATE(S) as they are very close to the obtained confidence intervals: On average the bounds of the confidence intervals exceed the bounds of estimates of the identified set by 0.011. Nonetheless, the unrestricted confidence intervals are large as the estimates for the identified set are themselves large – a result driven by the low proportion of compliers (5% on average across schooling levels). Imposing monotonicity across types carries identifying information on the upper end of the identified set at low levels of education and on the lower end of the identified set at high levels of education. Additionally imposing nonpositivity sharpens the upper bound of the identified set at all schooling levels. The resulting confidence regions sign ATE(S) at all education levels (weakly) smaller than 12 as strictly negative, though very close to zero.

Finally, as a preview of our general analysis in Section 3, in Table 1 we employ the same shape restrictions to report estimates and 95% confidence intervals for the identified sets of the average treatment effects for: High School Dropouts (edu $\in [9, 12)$), College Dropouts (edu $\in [13, 15)$), College Graduates (edu ≥ 16) and the overall average treatment effect. These confidence regions are obtained through test inversion after noting that a hypothesized value for the average treatment effect of a subgroup can be written as a nonlinear moment restriction in θ_0 through (11) – nonlinear moment restrictions fall within our general framework but outside the scope of Section 2.2. Overall

[4] E.g., for the constraint $E[Y_1|\texttt{NT}, S] \leq 1$ we substituted the corresponding $G_j\{\hat{\theta}_n^{\mathrm{u}} - \hat{\theta}_n^{\mathrm{u}\star}\}$ term in (7) with a mean zero normal distribution with the variance of the estimator for $E[Y_0|\texttt{NT}, S]$.

Subgroup	Unrestricted Estimate	Unrestricted 95% CI	Mon. Restr. Estimate	Mon. Restr. 95% CI	Mon.+Neg Restr. Estimate	Mon.+Neg Restr. 95% CI
HS Drop	[-0.520,0.426]	[-0.526,0.432]	[-0.489,0.346]	[-0.500,0.356]	[-0.489,-0.008]	[-0.501,-0.003]
Coll. Drop	[-0.561,0.380]	[-0.566,0.385]	[-0.447,0.325]	[-0.460,0.337]	[-0.447,-0.004]	[-0.462,0.000]
Coll. Grad	[-0.579,0.375]	[-0.586,0.382]	[-0.446,0.328]	[-0.462,0.339]	[-0.446,-0.002]	[-0.464,0.000]
All	[-0.545,0.395]	[-0.547,0.398]	[-0.467,0.328]	[-0.477,0.338]	[-0.467,-0.008]	[-0.478,-0.003]

Table 1: Point Estimates and 95% confidence intervals for the average treatment effect at different groups defined by schooling levels under different shape restrictions.

the impact of imposing shape restrictions parallels the results in Figure 2.

3 General Analysis

We next develop a general inferential framework that encompasses the tests discussed in Section 2. The class of models we consider are those in which the parameter of interest $\theta_0 \in \Theta$ satisfies a finite number $\mathcal{J}$ of conditional moment restrictions

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

with $\rho_\jmath : \mathbf{X} \times \Theta \to \mathbf{R}$, $X \in \mathbf{X}$, and $Z_\jmath \in \mathbf{Z}_\jmath$. For notational simplicity, we also let $Z \equiv (Z_1, \ldots, Z_{\mathcal{J}})$ and $V \equiv (X, Z)$ with $V \sim P \in \mathbf{P}$. In some of the applications that motivate us, the parameter θ_0 is not identified. We therefore define the identified set

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

and employ it as the basis of our statistical analysis – we emphasize that Θ_0 depends on P, but leave such dependence implicit to simplify notation. For a set R of parameters satisfying a conjectured restriction, we develop a test for the hypothesis

$$H_0 : \Theta_0 \cap R \neq \emptyset \qquad H_1 : \Theta_0 \cap R = \emptyset; \tag{12}$$

i.e. we devise a test of whether at least one element of the identified set satisfies the posited constraint. In what follows, we denote the set of distributions $P \in \mathbf{P}$ satisfying the null hypothesis in (12) by $\mathbf{P}_0$. We also note that in an identified model, a test of (12) is equivalent to a test of whether θ_0 itself satisfies the hypothesized constraint.

The defining elements determining the type of applications encompassed by (12) are the choices of Θ and R. In imposing restrictions on Θ and R we therefore aim to allow for a general framework while simultaneously ensuring enough structure for a fruitful asymptotic analysis. To this end, we require Θ to be a subset of a complete vector space

$\mathbf{B}$ with norm $\|\cdot\|_{\mathbf{B}}$ (i.e. $(\mathbf{B}, \|\cdot\|_{\mathbf{B}})$ is a Banach space) and consider sets R satisfying

$$R = \{\theta \in \mathbf{B} : \Upsilon_F(\theta) = 0 \text{ and } \Upsilon_G(\theta) \leq 0\}, \tag{13}$$

where $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ and $\Upsilon_G : \mathbf{B} \to \mathbf{G}$ are known maps. Our first assumption formalizes the basic structure of the hypothesis testing problem we study.

Assumption 3.1. *(i) $\{V_i\}_{i=1}^n$ is i.i.d. with $V \sim P \in \mathbf{P}$; (ii) $\Theta \subseteq \mathbf{B}$, where $(\mathbf{B}, \|\cdot\|_{\mathbf{B}})$ is a Banach space; (iii) $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ and $\Upsilon_G : \mathbf{B} \to \mathbf{G}$, where $(\mathbf{F}, \|\cdot\|_{\mathbf{F}})$ is a Banach space and $(\mathbf{G}, \|\cdot\|_{\mathbf{G}})$ is an AM space with order unit $\mathbf{1_G}$.*

Through Assumption 3.1(i) we focus on the i.i.d. setting, though extensions to other sampling frameworks are feasible. Assumption 3.1(ii) allows us to address parametric, semiparametric, and nonparametric models, while Assumption 3.1(iii) allows Υ_F to impose both finite dimensional or infinite dimensional equality restrictions. Assumption 3.1(iii) further requires that Υ_G take values in an AM space $\mathbf{G}$ – we provide an overview of AM spaces in the supplemental appendix. Heuristically, the key properties of $\mathbf{G}$ are: (i) $\mathbf{G}$ is a vector space equipped with a partial order "$\leq$"; (ii) The partial order and the vector space operations interact in the same manner they do on $\mathbf{R}$ (e.g. if $\theta_1 \leq \theta_2$, then $\theta_1 + \theta_3 \leq \theta_2 + \theta_3$); 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 $|\theta| \leq \lambda \mathbf{1_G}$ (e.g. when $\mathbf{G} = \mathbf{R}^d$ we may set $\mathbf{1_G} \equiv (1, \ldots, 1)' \in \mathbf{R}^d$). These properties of an AM space will prove instrumental in our analysis. In particular, the order unit $\mathbf{1_G}$ will provide a crucial link between the partial order "$\leq$", the norm $\|\cdot\|_{\mathbf{G}}$, and (through smoothness of Υ_G) allow us to leverage a rate of convergence in $\mathbf{B}$ to build a suitable sample analogue to the local parameter space.

3.1 Main Results

Our analysis centers around a statistic $I_n(R)$ that constitutes a "building block" for different tests of (12) – e.g., it may be employed to implement a generalization of the J-test of Sargan (1958) and Hansen (1982) or the incremental J-test of Eichenbaum et al. (1988). In this section we first introduce $I_n(R)$, obtain an approximation to its finite sample distribution, and devise a bootstrap procedure for estimating its quantiles. Together, these results allow us to establish the asymptotic validity of different tests.

3.1.1 The Building Block

We first introduce the statistic $I_n(R)$ that we employ to build different tests. To this end, for each instrument $Z_\jmath$ we consider transformations $\{q_{k,\jmath}\}_{k=1}^{k_{n,\jmath}}$ and let $q_\jmath^{k_{n,\jmath}}(z_\jmath) \equiv (q_{1,\jmath}(z_\jmath), \ldots, q_{k_{n,\jmath},\jmath}(z_\jmath))'$. Recalling that $Z \equiv (Z_1, \ldots, Z_{\mathcal{J}})$, we further set $k_n \equiv \sum_{\jmath=1}^{\mathcal{J}} k_{n,\jmath}$,

$q^{k_n}(z) \equiv (q_1^{k_{n,1}}(z_1)', \ldots, q_{\mathcal{J}}^{k_{n,\mathcal{J}}}(z_{\mathcal{J}})')'$, $\rho(x,\theta) \equiv (\rho_1(x,\theta), \ldots, \rho_{\mathcal{J}}(x,\theta))'$, and let

$$\rho(X_i,\theta) * q^{k_n}(Z_i) \equiv \begin{pmatrix} \rho_1(X_i,\theta) q_1^{k_{n,1}}(Z_{i,1}) \\ \vdots \\ \rho_{\mathcal{J}}(X_i,\theta) q_{\mathcal{J}}^{k_{n,\mathcal{J}}}(Z_{i,\mathcal{J}}) \end{pmatrix};$$

i.e. for each θ we take the product of each "residual" $\rho_\jmath(X,\theta)$ with the transformations of its respective instrument $Z_\jmath$. For a $k_n \times k_n$ matrix $\hat{\Sigma}_n$, we then define

$$Q_n(\theta) \equiv \|\frac{1}{n}\sum_{i=1}^n \rho(X_i,\theta) * q^{k_n}(Z_i)\|_{\hat{\Sigma}_n,p},$$

where $\|a\|_{\hat{\Sigma}_n,p} \equiv \|\hat{\Sigma}_n a\|_p$ and $\|\cdot\|_p$ is the p-norm on $\mathbf{R}^{k_n}$ for any $p \geq 2$ – i.e. $\|a\|_p \equiv (\sum_{i=1}^d |a^{(i)}|^p)^{1/p}$ for any $a \equiv (a^{(1)}, \ldots, a^{(d)})' \in \mathbf{R}^d$. Letting $\Theta_n \cap R$ be a finite dimensional subset of $\Theta \cap R$ that grows dense in $\Theta \cap R$ (Chen, 2007), we then define $I_n(R)$ to equal

$$I_n(R) \equiv \inf_{\theta \in \Theta_n \cap R} \sqrt{n} Q_n(\theta).$$

We note that setting $p = 2$ is often computationally attractive. However, we allow for $p > 2$ because higher values of p enable us to establish distributional approximations under weaker conditions on the number of unconditional moments k_n.

Heuristically, $\sqrt{n}Q_n$ should diverge to infinity when evaluated at any $\theta \notin \Theta_0$ and remain "stable" when evaluated at a $\theta \in \Theta_0$. Thus, examining the minimum of $\sqrt{n}Q_n$ over R should reveal whether there is a θ that simultaneously makes $\sqrt{n}Q_n(\theta)$ "stable" ($\theta \in \Theta_0$) and satisfies the conjectured restriction ($\theta \in R$). This intuition suggests $I_n(R)$ may be employed as a test statistic that is similar in spirit to the J-statistic of Hansen (1982). Alternatively, we may build a test by considering the recentered test statistic

$$I_n(R) - I_n(\Theta),$$

which aims power in a different direction than $I_n(R)$ (Chen and Santos, 2018). Conceptually, it is important to note that $I_n(\Theta)$ is a special case of $I_n(R)$ (i.e. set $R = \Theta$). We refer to $I_n(R)$ as a "building block" in the sense that, together with closely related variants like $I_n(\Theta)$, it may be employed to obtain a variety of different tests.

3.1.2 Strong Approximation

We first obtain a strong approximation to statistics of the form $I_n(R)$. Before proceeding, we introduce some additional notation. First, we define the class

$$\mathcal{F}_n \equiv \{\rho_\jmath(\cdot,\theta) : \theta \in \Theta_n \cap R \text{ and } 1 \leq \jmath \leq \mathcal{J}\}. \tag{14}$$

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

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

where $\|f\|_{P,2}^2 \equiv E_P[f^2(V)]$ and $N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2})$ is the smallest number of ϵ-brackets (under $\|\cdot\|_{P,2}$) required to cover $\mathcal{F}_n$. Finally, we denote the empirical process by

$$\mathbb{G}_n(\theta) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\rho(X_i, \theta) * q^{k_n}(Z_i) - E_P[\rho(X, \theta) * q^{k_n}(Z)]\}.$$

Our next assumptions imposes requirements on $\Theta_n \cap R$ and the transformation $q^{k_n}(Z)$.

Assumption 3.2. *(i)* $\max_{1 \leq \jmath \leq \mathcal{J}} \max_{1 \leq k \leq k_{n,\jmath}} \|q_{k,\jmath}\|_\infty \leq B_n$ *with* $B_n \geq 1$*; (ii) The eigenvalues of* $E_P[q_\jmath^{k_{n,\jmath}}(Z_\jmath) q_\jmath^{k_{n,\jmath}}(Z_\jmath)']$ *are bounded uniformly in* $k_{n,\jmath}$ *and* $P \in \mathbf{P}$*; (iii)* $\mathcal{F}_n$ *has envelope* F_n*,* $\sup_{P \in \mathbf{P}} \|F_n\|_{P,2} < \infty$*, and* $\sup_{P \in \mathbf{P}} J_{[]}(\|F_n\|_{P,2}, \mathcal{F}_n, \|\cdot\|_{P,2}) \leq J_n$ *with* $J_n < \infty$*.*

Assumption 3.3. *(i)* $\sup_{\theta \in \Theta_n \cap R} \|\mathbb{G}_n(\theta) - \mathbb{W}_P(\theta)\|_p = o_P(a_n)$ *uniformly in* $P \in \mathbf{P}$ *for some* $a_n = o(1)$ *and Gaussian* $\mathbb{W}_P$ *satisfying* $E[\mathbb{W}_P(\theta)] = 0$ *and* $\mathrm{Cov}\{\mathbb{W}_P(\theta), \mathbb{W}_P(\theta')\} = \mathrm{Cov}_P\{\mathbb{G}_n(\theta), \mathbb{G}_n(\theta')\}$*; (ii) There is a norm* $\|\cdot\|_{\mathbf{E}}$*,* $\kappa_\rho > 0$*, and* $K_\rho < \infty$ *such that* $E_P[\|\rho(X, \theta_1) - \rho(X, \theta_2)\|_2^2] \leq K_\rho^2 \|\theta_1 - \theta_2\|_{\mathbf{E}}^{2\kappa_\rho}$ *for all* $\theta_1, \theta_2 \in \Theta_n \cap R$ *and* $P \in \mathbf{P}$*.*

Assumptions 3.2(i)(ii) impose standard requirements on the transformations q^{k_n} – e.g., Assumption 3.2(i) holds with $B_n = 1$ for trigonometric series and $B_n \asymp \sqrt{k_n}$ for normalized B-splines. Assumption 3.2(iii) controls the "size" of $\mathcal{F}_n$. We allow J_n to depend on n to accommodate non-compact parameter spaces (Chen and Pouzo, 2012, 2015). Assumption 3.3(i) requires that the empirical process be approximately Gaussian. The sequence $\{a_n\}_{n=1}^\infty$ denotes a bound on the rate of coupling, which in turn characterizes the rate of convergence of our strong approximation. In the appendix, we verify Assumption 3.3(i) by relying on existing results (Yurinskii, 1977; Zhai, 2018) or a novel extension of Koltchinskii (1994). Assumption 3.3(ii) imposes a mild restriction on the moment functions that ensures $\mathbb{W}_P$ is equicontinuous with respect to $\|\cdot\|_{\mathbf{E}}$.

In establishing our strong approximation to $I_n(R)$, it is helpful to derive the rate of convergence of the minimizer of Q_n over $\Theta_n \cap R$. To this end, we follow the literature on set estimation (Chernozhukov et al., 2007; Beresteanu and Molinari, 2008; Santos, 2011; Kaido and Santos, 2014) and for any sets A and B we define

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

which is known as the directed Hausdorff distance. For each $\theta \in \Theta \cap R$, we further let

$\Pi_n\theta$ denote its approximation on $\Theta_n \cap R$ and denote the approximation to $\Theta_0 \cap R$ by

$$\Theta^{\mathrm{r}}_{0n} \equiv \{\Pi_n\theta : \theta \in \Theta_0 \cap R\}. \tag{15}$$

Our next assumption enables us to obtain a rate of convergence (under $\|\cdot\|_{\mathbf{E}}$) to Θ^{r}_{0n}.

Assumption 3.4. *There are $\mathcal{V}_n(P) \subseteq \Theta_n \cap R$ and a sequence constants $\{\nu_n\}$ with $0 < \nu_n^{-1} = O(1)$ such that (i) For any $\theta \in \mathcal{V}_n(P)$ it holds that*

$$\nu_n^{-1}\overrightarrow{d}_H(\theta, \Theta^{\mathrm{r}}_{0n}, \|\cdot\|_{\mathbf{E}}) \le \sup_{\tilde{\theta}\in\Theta^{\mathrm{r}}_{0n}} \|E_P[(\rho(X,\theta) - \rho(X,\tilde{\theta})) * q^{k_n}(Z)]\|_{\Sigma_P,p};$$

(ii) There is a $\hat{\theta}_n \in \mathcal{V}_n(P)$ satisfying $Q_n(\hat{\theta}_n) \le \inf_{\theta\in\Theta_n\cap R} Q_n(\theta) + o(a_n/\sqrt{n})$ with probability tending to one uniformly in $P \in \mathbf{P}_0$.

Assumption 3.4(ii) requires that an approximate minimum of Q_n over $\Theta_n \cap R$ be attained at a point $\hat{\theta}_n$ in a set $\mathcal{V}_n(P)$ with high probability. Typically, $\mathcal{V}_n(P)$ may be taken to equal the entire sieve in convex models, or it may be taken to equal a local neighborhood of Θ^{r}_{0n} after establishing the consistency of $\hat{\theta}_n$ through standard arguments; see, e.g., Lemma S.1.1 in the appendix. Assumption 3.4(i) introduces a local identification condition on $\mathcal{V}_n(P)$ by requiring that the moments "change" at a rate ν_n^{-1} as θ moves away from Θ^{r}_{0n}. The parameter ν_n^{-1}, which implicitly depends on k_n and the choice of sieve $\Theta_n \cap R$, is conceptually related to sieve measure of ill-posedness (Blundell et al., 2007).

By employing Assumption 3.4, we are able to show that with arbitrarily high probability, $\hat{\theta}_n$ is contained in a $\|\cdot\|_{\mathbf{E}}$-neighborhood of Θ^{r}_{0n} that shrinks at a rate

$$\mathcal{R}_n \equiv \nu_n\{\frac{k_n^{1/p}\sqrt{\log(1+k_n)}J_nB_n}{\sqrt{n}}\}, \tag{16}$$

where recall B_n and J_n where introduced in Assumption 3.2. Under assumptions on the (Hausdorff) distance between Θ^{r}_{0n} and $\Theta_0 \cap R$, the triangle inequality can yield a rate of convergence of $\hat{\theta}_n$ to $\Theta_0 \cap R$. Heuristically, we focus on convergence to Θ^{r}_{0n} (instead of $\Theta_0 \cap R$) because our strong approximation will rely on undersmoothing.

In our final assumptions, we follow the literature and accommodate non-differentiable moment functions by requiring that their conditional expectations be differentiable (Chen and Pouzo, 2009, 2012). Specifically, for each $1 \le \jmath \le \mathcal{J}$ and $\theta \in \Theta$ we set

$$m_{P,\jmath}(\theta)(Z_\jmath) \equiv E_P[\rho_\jmath(X,\theta)|Z_\jmath];$$

i.e. $m_{P,\jmath}$ maps each $\theta \in \Theta$ to a square integrable function of $Z_\jmath$. Letting $\mathbf{B}_n$ denote the vector subspace generated by $\Theta_n \cap R$, we then impose the following:

Assumption 3.5. *There is a norm* $\|\cdot\|_{\mathbf{L}}$ *on* $\mathbf{B}_n$*, linear maps* $\nabla m_{P,\jmath}(\theta) : \mathbf{B} \to L^2_P$*, and constants* $\epsilon > 0$ *and* $K_m, M < \infty$ *such that for all* $P \in \mathbf{P}$*,* $h \in \mathbf{B}_n$*, and elements* $\theta_1, \theta_2 \in \{\theta \in \Theta_n \cap R : \overrightarrow{d}_H(\theta, \Theta^{\mathrm{r}}_{0n}, \|\cdot\|_{\mathbf{E}}) \leq \epsilon\}$ *we have: (i)* $\|m_{P,\jmath}(\theta_1) - m_{P,\jmath}(\theta_2) - \nabla m_{P,\jmath}(\theta_2)[\theta_1 - \theta_2]\|_{P,2} \leq K_m \|\theta_1 - \theta_2\|_{\mathbf{L}} \|\theta_1 - \theta_2\|_{\mathbf{E}}$*; (ii)* $\|\nabla m_{P,\jmath}(\theta_1)[h] - \nabla m_{P,\jmath}(\theta_2)[h]\|_{P,2} \leq K_m \|\theta_1 - \theta_2\|_{\mathbf{L}} \|h\|_{\mathbf{E}}$*; (iii)* $\|\nabla m_{P,\jmath}(\theta_2)[h]\|_{P,2} \leq M \|h\|_{\mathbf{E}}$.

Assumption 3.6. *(i)* $k_n^{1/p} \sqrt{\log(1 + k_n)} B_n \sup_{P \in \mathbf{P}} J_{[\,]}(\mathcal{R}_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$*; (ii)* $\sup_{P \in \mathbf{P}_0} \sup_{\theta \in \Theta^{\mathrm{r}}_{0n}} \sqrt{n} \|E_P[\rho(X, \theta) * q^{k_n}(Z)]\|_{\Sigma_P, p} = o(a_n)$.

Assumption 3.7. *(i) For each* $P \in \mathbf{P}$ *there is a* $k_n \times k_n$ *matrix* $\Sigma_P > 0$ *such that* $\|\hat{\Sigma}_n - \Sigma_P\|_{o,p} = o_P(1 \wedge a_n \{k_n^{1/p} \sqrt{\log(1 + k_n)} B_n J_n\}^{-1})$ *uniformly in* $P \in \mathbf{P}$*; (ii)* $\|\Sigma_P\|_{o,p}$ *and* $\|\Sigma_P^{-1}\|_{o,p}$ *are uniformly bounded in* k_n *and* $P \in \mathbf{P}$.

Assumption 3.5(i) ensures $m_{P,\jmath}$ is approximated by linear maps $\nabla m_{P,\jmath}$ with an approximation error that is controlled by $\|\cdot\|_{\mathbf{E}}$ and a potentially stronger norm $\|\cdot\|_{\mathbf{L}}$. In turn, Assumptions 3.5(ii)(iii) impose continuity conditions on $\nabla m_{P,\jmath}$ – these assumptions are not used in this section, but will be needed for our bootstrap results. Assumption 3.6 contains our key rate restrictions. Assumption 3.6(i) ensures the rate of convergence $\mathcal{R}_n$ (as in (16)) is sufficiently fast to overcome an asymptotic loss of equicontinuity – a requirement that can hold even when $\mathcal{R}_n$ is slower than the traditional $o(n^{-1/4})$ rate employed to linearize nonlinear models.Assumption 3.6(ii) is an undersmoothing assumption, which ensures that $I_n(R)$ is properly centered under the null hypothesis. Finally, Assumption 3.7 requires $\hat{\Sigma}_n$ to converge to an invertible matrix Σ_P at a suitable rate – here, $\|\cdot\|_{o,p}$ denotes the operator norm when $\mathbf{R}^{k_n}$ is endowed with $\|\cdot\|_p$.

The introduced assumptions suffice for obtaining a strong approximation through a local reparametrization. Formally, we denote the local deviations from $\theta \in \Theta_n \cap R$ by

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

Recall $\mathbf{B}_n$ denotes the vector subspace generated by $\Theta_n \cap R$ and for any $h \in \mathbf{B}_n$ set

$$\mathbb{D}_P(\theta)[h] \equiv E_P[\nabla m_P(\theta)[h](Z) * q^{k_n}(Z)],$$

where $\nabla m_P(\theta)[h](Z) \equiv (\nabla m_{P,1}(\theta)[h](Z_1), \ldots, \nabla m_{P,\mathcal{J}}(\theta)[h](Z_{\mathcal{J}}))'$. For any given sequence ℓ_n, we then define a sequence of random variables $U_P(R|\ell_n)$ to be given by

$$U_P(R|\ell_n) \equiv \inf_{\theta \in \Theta^{\mathrm{r}}_{0n}} \inf_{h \in V_n(\theta, R|\ell_n)} \|\mathbb{W}_P(\theta) + \mathbb{D}_P(\theta)[h]\|_{\Sigma_P, p}. \tag{17}$$

As a final piece of notation, for any two norms $\|\cdot\|_{\mathbf{A}_1}$ and $\|\cdot\|_{\mathbf{A}_2}$ defined on $\mathbf{B}_n$, we set

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

which we note depends on the sample size n only through the choice of sieve $\Theta_n \cap R$.

The next result establishes the relation between $U_P(R|\ell_n)$ and $I_n(R)$. It is helpful to recall here that the norm $\|\cdot\|_{\mathbf{L}}$ and constants K_m, introduced in Assumption 3.5, control the linearization of the moments and that $K_m = 0$ for linear models.

Theorem 3.1. *Let Assumptions 3.1(i), 3.2, 3.3, 3.4, 3.5(i), 3.6, and 3.7 hold. Then: (i) For any $\ell_n \downarrow 0$ satisfying $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$ and $K_m\ell_n^2 \times \mathcal{S}_n(\mathbf{L},\mathbf{E}) = o(a_n n^{-1/2})$ it follows uniformly in $P \in \mathbf{P}_0$ that:*

$$I_n(R) \leq U_P(R|\ell_n) + o_P(a_n).$$

(ii) If in addition $K_m\mathcal{R}_n^2 \times \mathcal{S}_n(\mathbf{L},\mathbf{E}) = o(a_n n^{-1/2})$, then ℓ_n may be additionally chosen to satisfy $\mathcal{R}_n = o(\ell_n)$, in which case it follows uniformly in $P \in \mathbf{P}_0$ that:

$$I_n(R) = U_P(R|\ell_n) + o_P(a_n).$$

Theorem 3.1 is perhaps best understood as establishing the validity of a family (indexed by $\{\ell_n\}$) of strong approximations that differ on the size of the local neighborhoods of Θ_{0n}^{r} that they employ. Its proof crucially relies on the linearization

$$\mathbb{D}_P(\theta)[h] \approx \sqrt{n}\{E_P[\rho(X,\theta+\frac{h}{\sqrt{n}}) * q^{k_n}(Z)] - E_P[\rho(X,\theta) * q^{k_n}(Z)]\}, \qquad (18)$$

which holds for nonlinear moments ($K_m \neq 0$) when $h/\sqrt{n}$ is sufficiently small. In particular, if the infimum defining $I_n(R)$ is attained at a point $\hat{\theta}_n$ that converges to Θ_{0n}^{r} sufficiently fast, then we may apply (18) to establish Theorem 3.1(ii). Regrettably, in certain models the rate of convergence of $\hat{\theta}_n$ may be too slow to apply the approximation in (18) to $\hat{\theta}_n$. In such instances, we may instead rely on the inequality

$$I_n(R) = \inf_{\theta\in\Theta_n\cap R} \sqrt{n}Q_n(\theta) \leq \inf_{(\theta,h)\in(\Theta_{0n}^{\mathrm{r}}, V_n(\theta,R|\ell_n))} \sqrt{n}Q_n(\theta+\frac{h}{\sqrt{n}}) \qquad (19)$$

and successfully couple the right hand side of (19) by restricting attention to sequences ℓ_n for which (18) is accurate. Thus, by regularizing the local parameter space through a norm bound, we obtain in Theorem 3.1(i) a distributional approximation that, while potentially conservative, holds under weaker requirements on the rate of convergence.

3.1.3 Bootstrap Approximation

Theorem 3.1 shows that the distribution of $U_P(R|\ell_n)$ is a suitable approximation for the distribution of $I_n(R)$. We next develop a bootstrap procedure for estimating the distribution of $U_P(R|\ell_n)$ with the goal of obtaining valid critical values.

We estimate the distribution of $U_P(R|\ell_n)$ by replacing population parameters with suitable sample analogues. The key ingredients are: (i) A random variable $\hat{\mathbb{W}}_n$ whose distribution conditional on the data is consistent for the distribution of $\mathbb{W}_P$; (ii) An estimator $\hat{\mathbb{D}}_n(\theta)$ for $\mathbb{D}_P(\theta)$; (iii) An estimator $\hat{\Theta}_n^{\mathrm{r}}$ for Θ_{0n}^{r} (as in (15)); and (iv) A sample analogue $\hat{V}_n(\theta, R|\ell_n)$ for the local parameter space $V_n(\theta, R|\ell_n)$. We then approximate the distribution of $U_P(R|\ell_n)$ by the distribution (conditional on the data) of

$$\hat{U}_n(R|\ell_n) \equiv \inf_{\theta \in \hat{\Theta}_n^{\mathrm{r}}} \inf_{h \in \hat{V}_n(\theta, R|\ell_n)} \|\hat{\mathbb{W}}_n(\theta) + \hat{\mathbb{D}}_n(\theta)[h]\|_{\hat{\Sigma}_n, p}.$$

For concreteness, we employ the following sample analogues in our construction.

Gaussian Distribution: We estimate the distribution of $\mathbb{W}_P$ with the multiplier bootstrap. Specifically, for i.i.d. $\{\omega_i\}_{i=1}^n$ with $\omega_i \sim N(0,1)$ independent of $\{V_i\}_{i=1}^n$ we let

$$\hat{\mathbb{W}}_n(\theta) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \omega_i \{\rho(X_i, \theta) * q^{k_n}(Z_i) - \frac{1}{n} \sum_{j=1}^n \rho(X_j, \theta) * q^{k_n}(Z_j)\}.$$

We focus on the multiplier bootstrap due to its theoretical tractability, though we note that alternative bootstrap approaches can also be valid. ■

The Derivative: We estimate $\mathbb{D}_P(\theta)$ by employing a construction that is applicable to non-differentiable moments. Specifically, for any $\theta \in \Theta_n \cap R$ and $h \in \mathbf{B}_n$ we set

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

We employ $\hat{\mathbb{D}}_n(\theta)$ due to its general applicability, though alternative approaches may be preferable in some applications. In particular, if moments are differentiabile, then using

$$\frac{1}{n} \sum_{i=1}^n \nabla_\theta \rho(X_i, \theta)[h] * q^{k_n}(Z_i)$$

as an estimator for $\mathbb{D}_P(\theta)[h]$ leads to a computationally simpler bootstrap statistic. ■

The Identified Set: We estimate the identified set by employing the set of (approximate) minimizers of Q_n on $\Theta_n \cap R$. Formally, for a sequence $\tau_n \downarrow 0$, we let

$$\hat{\Theta}_n^{\mathrm{r}} \equiv \{\theta \in \Theta_n \cap R : Q_n(\theta) \leq \inf_{\theta \in \Theta_n \cap R} Q_n(\theta) + \tau_n\}. \tag{20}$$

We may set $\tau_n = 0$ in identified models, in which case $\hat{\Theta}_n^{\mathrm{r}}$ reduces to the minimizer of Q_n. In partially identified models, $\hat{\Theta}_n^{\mathrm{r}}$ can be shown to asymptotically lie in a shrinking neighborhood of Θ_{0n}^{r} provided $\tau_n \to 0$. In order for $\hat{\Theta}_n^{\mathrm{r}}$ to additionally be Hausdorff consistent for Θ_{0n}^{r}, however, τ_n must not tend to zero too fast; see Lemma S.1.1. ■

Local Parameter Space: We account for the role inequality constraints play in determining the local parameter space by estimating "binding" sets in analogy to approaches pursued in the moment inequalities literature (Chernozhukov et al., 2007; Andrews and Soares, 2010). Specifically, for a sequence r_n and any $\theta \in \Theta_n \cap R$ we define

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

where recall $\mathbf{1_G}$ is the order unit in $\mathbf{G}$ and $g_1 \vee g_2$ represents the supremum of any $g_1, g_2 \in \mathbf{G}$. The constant K_g, formally introduced in Assumption 3.8 below, is related to the curvature of Υ_G and equals zero for linear Υ_G. For any ℓ_n we then define

$$\hat{V}_n(\theta, R|\ell_n) \equiv \{h \in \mathbf{B}_n : h \in G_n(\theta),\ \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\}, \tag{21}$$

i.e. in comparison to $V_n(\theta, R|\ell_n)$ we: (i) Replace $\Upsilon_G(\theta + h/\sqrt{n}) \leq 0$ by $h \in G_n(\theta)$; (ii) Retain $\Upsilon_F(\theta + h/\sqrt{n}) = 0$; and (iii) Substitute $\|h/\sqrt{n}\|_{\mathbf{E}} \leq \ell_n$ with $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$. $\blacksquare$

Before establishing the asymptotic validity of the proposed bootstrap procedure, we require some additional notation. For any set $A \subseteq \mathbf{B}_n$, we denote its ϵ-neighborhood by

$$(A)^\epsilon \equiv \{\theta \in \mathbf{B}_n : \inf_{a \in A} \|a - \theta\|_{\mathbf{B}} \leq \epsilon\}.$$

We further denote the closure of the linear span of $\Upsilon_F(\mathbf{B}_n)$ by $\mathbf{F}_n$, and for any linear map Γ on $\mathbf{B}$ we let $\mathcal{N}(\Gamma) \equiv \{h \in \mathbf{B} : \Gamma(h) = 0\}$ denote its null space. In the assumptions that follow, it is helpful to recall that Θ_{0n}^{r} is implicitly a function of P.

Assumption 3.8. *For some $K_g, M < \infty$, $\epsilon > 0$ and all n, $P \in \mathbf{P}_0$, $\theta_1, \theta_2 \in (\Theta_{0n}^{\mathrm{r}})^\epsilon$ (i) Υ_G is Fréchet differentiable with $\|\Upsilon_G(\theta_1) - \Upsilon_G(\theta_2) - \nabla\Upsilon_G(\theta_1)[\theta_1 - \theta_2]\|_{\mathbf{G}} \leq K_g \|\theta_1 - \theta_2\|_{\mathbf{B}}^2$; (ii) $\|\nabla\Upsilon_G(\theta_1) - \nabla\Upsilon_G(\theta_2)\|_o \leq K_g\|\theta_1 - \theta_2\|_{\mathbf{B}}$; (iii) $\|\nabla\Upsilon_G(\theta_1)\|_o \leq M$.*

Assumption 3.9. *For some $K_f, M < \infty$, $\epsilon > 0$ and all n, $P \in \mathbf{P}_0$, $\theta_1, \theta_2 \in (\Theta_{0n}^{\mathrm{r}})^\epsilon$ (i) Υ_F is Fréchet differentiable with $\|\Upsilon_F(\theta_1) - \Upsilon_F(\theta_2) - \nabla\Upsilon_F(\theta_1)[\theta_1 - \theta_2]\|_{\mathbf{F}} \leq K_f \|\theta_1 - \theta_2\|_{\mathbf{B}}^2$; (ii) $\|\nabla\Upsilon_F(\theta_1) - \nabla\Upsilon_F(\theta_2)\|_o \leq K_f\|\theta_1 - \theta_2\|_{\mathbf{B}}$; (iii) $\|\nabla\Upsilon_F(\theta_1)\|_o \leq M$; (iv) $\nabla\Upsilon_F(\theta_1) : \mathbf{B}_n \to \mathbf{F}_n$ admits a right inverse $\nabla\Upsilon_F(\theta_1)^-$ with $K_f\|\nabla\Upsilon_F(\theta_1)^-\|_o \leq M$.*

Assumption 3.10. *Either (i) $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ is affine, or (ii) There are constants $\epsilon > 0$, $M < \infty$ such that for every $P \in \mathbf{P}_0$, n, and $\theta \in \Theta_{0n}^{\mathrm{r}}$ there exists a $h \in \mathbf{B}_n \cap \mathcal{N}(\nabla\Upsilon_F(\theta))$ satisfying $\Upsilon_G(\theta) + \nabla\Upsilon_G(\theta)[h] \leq -\epsilon\mathbf{1_G}$ and $\|h\|_{\mathbf{B}} \leq M$.*

Assumption 3.8 imposes that Υ_G be Fréchet differentiable. The constant K_g, employed in the construction of $\hat{V}_n(\theta, R|\ell_n)$, may be interpreted as a bound on the second derivative of Υ_G and equals zero when Υ_G is linear. Assumptions 3.9 and 3.10 mark an important difference between hypotheses in which Υ_F is linear and those in which Υ_F is nonlinear – note linear Υ_F automatically satisfy Assumptions 3.9 and 3.10. This distinction reflects that when Υ_F is linear its impact on the local parameter space is known and need not be estimated.[5] Thus, while Assumptions 3.9(i)-(iii) impose conditions analogous to those required of Υ_G, Assumption 3.9(iv) additionally demands that $\nabla\Upsilon_F(\theta)$ posses a norm bounded right inverse on $(\Theta_{0n}^{\mathrm{r}})^\epsilon$ – the existence of a right inverse is equivalent to a classical rank condition.[6] Finally, for nonlinear Υ_F, Assumption 3.10(ii) requires the existence of a local perturbation to any $\theta \in \Theta_{0n}^{\mathrm{r}}$ that relaxes "active" inequality constraints without a first order effect on the equality restrictions.

We impose a final set of assumptions in order to couple our bootstrap statistic.

Assumption 3.11. *$\sup_{\theta\in\Theta_n\cap R}\|\hat{\mathbb{W}}_n(\theta) - \mathbb{W}_P^\star(\theta)\|_p = o_P(a_n)$ uniformly in $\Phi \times P$ with $P \in \mathbf{P}$ for Φ the standard normal distribution, $a_n = o(1)$, and $\mathbb{W}_P^\star$ independent of $\{V_i\}_{i=1}^n$ and having the same distribution as $\mathbb{W}_P$.*

Assumption 3.12. *(i) For some $M < \infty$, $\|h\|_{\mathbf{E}} \leq M\|h\|_{\mathbf{B}}$ for all $h \in \mathbf{B}_n$; (ii) There is an $\epsilon > 0$ such that $P((\hat{\Theta}_n^{\mathrm{r}})^\epsilon \subseteq \Theta_n)$ tends to one uniformly in $P \in \mathbf{P}_0$; (iii) For $\mathcal{V}_n(P)$ as in Assumption 3.4, $P(\hat{\Theta}_n^{\mathrm{r}} \subseteq \mathcal{V}_n(P))$ tends to one uniformly in $P \in \mathbf{P}_0$.*

Assumption 3.13. *(i) Either Υ_F and Υ_G are affine or $(\mathcal{R}_n+\nu_n\tau_n)\times\mathcal{S}_n(\mathbf{B},\mathbf{E}) = o(1)$; (ii) The sequences ℓ_n, τ_n satisfy $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho} \vee (\nu_n\tau_n)^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$, $K_m\ell_n(\ell_n+\mathcal{R}_n+\nu_n\tau_n)\times\mathcal{S}_n(\mathbf{L},\mathbf{E}) = o(a_n n^{-1/2})$, and $\ell_n(\ell_n+\{\mathcal{R}_n+\nu_n\tau_n\}\times\mathcal{S}_n(\mathbf{B},\mathbf{E}))1\{K_f > 0\} = o(a_n n^{-1/2})$; (iii) The sequence r_n satisfies $\limsup_{n\to\infty} 1\{K_g > 0\}\ell_n/r_n < 1/2$ and $(\mathcal{R}_n + \nu_n\tau_n) \times \mathcal{S}_n(\mathbf{B},\mathbf{E}) = o(r_n)$.*

Assumption 3.11 demands that $\hat{\mathbb{W}}_n$ be coupled with a Gaussian $\mathbb{W}_P^\star$ independent of $\{V_i\}_{i=1}^n$. This condition implies the multiplier bootstrap is valid in our potentially non-Donsker setting; see Appendix S.7 for sufficient conditions. More generally, we note that our analysis remains valid if the multiplier bootstrap is replaced with any other resampling scheme (e.g., nonparametric bootstrap) satisfying a coupling requirement like Assumption 3.11. Assumption 3.12(i) ensures that $\|\cdot\|_{\mathbf{B}}$ is (weakly) stronger than $\|\cdot\|_{\mathbf{E}}$. Assumption 3.12(ii) demands that $\hat{\Theta}_n^{\mathrm{r}}$ be asymptotically contained in the interior of Θ_n. This requirement does not rule out that parameter space restrictions be binding at Θ_{0n}^{r} – instead, Assumption 3.12(ii) requires that all such restrictions be stated through R. Together with Assumption 3.4(i), Assumption 3.12(iii) enables us to obtain a rate of convergence for $\hat{\Theta}_n^{\mathrm{r}}$ and may be verified in the same manner as Assumption 3.4(ii).

Assumption 3.13 contains our main rate requirements. In particular, Assumption 3.13(i) ensures the one sided Hausdorff convergence of $\hat{\Theta}_n^{\mathrm{r}}$ to Θ_{0n}^{r} under $\|\cdot\|_{\mathbf{B}}$ whenever Υ_F or Υ_G are nonlinear. The main conditions on ℓ_n, employed in constructing

[5] For linear Υ_F, the requirement $\Upsilon_F(\theta + h/\sqrt{n}) = 0$ is equivalent to $\Upsilon_F(h) = 0$ for any $\theta \in R$.

[6] 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 3.9(iv) is satisfied as long as it holds for some right inverse of $\nabla\Upsilon_F(\theta) : \mathbf{B}_n \to \mathbf{F}_n$.

$\hat{V}_n(\theta, R|\ell_n)$, are contained in Assumption 3.13(ii). These conditions ensure the consistency of $\hat{\mathbb{D}}_n(\theta)[h]$, the applicability of Theorem 3.1, and that $\hat{V}_n(\theta, R|\ell_n)$ be well approximated by the true local parameter space. Heuristically, whenever the rate of convergence $\mathcal{R}_n$ is too slow, regularizing the local parameter space by selecting a small ℓ_n can ensure the asymptotic validity of the test. As in Section 2, however, we note that whenever the rate of convergence $\mathcal{R}_n$ is sufficiently fast such regularization is unnecessary and it is possible to set $\ell_n = +\infty$ – in such applications, setting ℓ_n to be too small can lead to a loss of power. In turn, Assumption 3.13(iii) requires that r_n not decrease to zero faster than the $\|\cdot\|_{\mathbf{B}}$-rate of convergence of $\hat{\Theta}^{\mathrm{r}}_n$. Assumption 3.13(iii) is always satisfied if $r_n = +\infty$, though setting $r_n \to 0$ can improve power against certain alternatives. Similarly, we note that the requirements on τ_n imposed by Assumption 3.13 can always be satisfied by setting $\tau_n = 0$, but such a choice can lead to a loss of power in certain partially identified models (recall the discussion in Section 2.2).

Our next result provides a coupling result for our bootstrap statistic. In its statement, $U^\star_P(R|\ell_n)$ is defined identically to $U_P(R|\ell_n)$ but with $\mathbb{W}^\star_P$ in place of $\mathbb{W}_P$.

Theorem 3.2. *If Assumptions 3.1, 3.2, 3.3, 3.4(i), 3.5, 3.6(ii), 3.7, 3.8, 3.9, 3.10, 3.11, 3.12, 3.13 hold, then there is $\tilde{\ell}_n \asymp \ell_n$ so that uniformly in $P \in \mathbf{P}_0$*

$$\hat{U}_n(R|\ell_n) \geq U^\star_P(R|\tilde{\ell}_n) + o_P(a_n).$$

Theorem 3.2 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 coupling to $I_n(R)$ obtained in Theorem 3.1. Thus, Theorems 3.1 and 3.2 provide the basis for constructing tests that employ increasing functions of $I_n(R)$ as a test statistic and the analogous bootstrap quantiles of $\hat{U}_n(R|\ell_n)$ as critical values. The resulting tests may be conservative, however, whenever the inequalities in Theorems 3.1 and 3.2 are not "sharp." In particular, in order for the pointwise (in P) rejection probability to equal the nominal level of the test under the null hypothesis we require: (i) The rate of convergence $\mathcal{R}_n$ must be sufficiently fast for Theorem 3.1(ii) to apply (in which case setting $\ell_n = +\infty$ is often valid); (ii) We should select r_n to tend to zero with the sample size; and (iii) In partially identified settings, τ_n must tend to zero sufficiently slowly so that $\hat{\Theta}^{\mathrm{r}}_n$ is Hausdorff consistent for Θ^{r}_{0n}.

3.2 The Tests

We next employ the theoretical results of Section 3.1 to establish the asymptotic validity of different tests of the null hypothesis defined in (12). In what follows, for any statistic $\hat{T}_n$ that is a function of $\{V_i\}_{i=1}^n$ and the bootstrap weights $\{\omega_i\}_{i=1}^n$, we let $\hat{q}_\tau(\hat{T}_n)$ denote its conditional τ quantile given $\{V_i\}_{i=1}^n$. For example, we have that

$$\hat{q}_{1-\alpha}(\hat{U}_n(R|\ell_n)) = \inf\{u : P(\hat{U}_n(R|\ell_n) \leq u|\{V_i\}_{i=1}^n) \geq 1-\alpha\}.$$

3.2.1 Tests Based on $I_n(R)$

We first examine a test that employs $I_n(R)$ as a test statistic and a bootstrap quantile of $\hat{U}_n(R|\ell_n)$ as a critical value. As has been shown in the literature, uniform consistent estimation of approximating distributions is not sufficient for characterizing the asymptotic size of a test (Romano and Shaikh, 2012). Heuristically, to establish the asymptotic validity of a test the approximating distributions must additionally be suitably uniformly continuous. Our next assumption suffices for verifying this final requirement.

Assumption 3.14. *There is $\eta \geq 0$ and $\varrho_n = o(a_n^{-1})$ such that for $\hat{c}_n = \hat{q}_{1-\alpha}(\hat{U}_n(R|\ell_n))$ and any $\tilde{\ell}_n \asymp \ell_n$: (i) $P(I_n(R) > \hat{c}_n) = P(I_n(R) > \hat{c}_n \vee \eta) + o(1)$ uniformly in $P \in \mathbf{P}_0$, and (ii) $\sup_{P\in\mathbf{P}_0} \sup_{t\in(\eta-a_n,+\infty)} P(|U_P(R|\tilde{\ell}_n) - t| \leq \epsilon) \leq \varrho_n(\epsilon \wedge 1) + o(1)$.*

Assumption 3.14(i) trivially holds with $\eta = 0$ since both $I_n(R)$ and $\hat{U}_n(R|\ell_n)$ are (weakly) positive almost surely. However, in some applications it is possible to verify Assumption 3.14(i) in fact holds with $\eta > 0$ by arguing that the bootstrap quantiles of $\hat{U}_n(R|\ell_n)$ are suitably bounded away from zero when $I_n(R)$ is strictly positive. Establishing Assumption 3.14(i) holds with $\eta > 0$ eases the verification of Assumption 3.14(ii), which intuitively requires that $U_P(R|\tilde{\ell}_n)$ be continuously distributed on $(\eta - a_n, +\infty)$ with a density bounded by a, possibly diverging, ϱ_n. Because $U_P(R|\tilde{\ell}_n)$ is a functional of the Gaussian measure $\mathbb{W}_P$, Assumption 3.14(ii) can in some applications be verified using available results in the literature (Davydov et al., 1998). For instance, when $U_P(R|\tilde{\ell}_n)$ is a convex function of $\mathbb{W}_P$, as in the application of Section 2.1.1, the distribution of $U_P(R|\tilde{\ell}_n)$ can readily be shown to be continuous in $(0, +\infty)$. We refer the reader to Chernozhukov et al. (2014) for further discussion and motivation of conditions such as Assumption 3.14(ii), called *anti-concentration* conditions.

The next result establishes the asymptotic validity of a test based on $I_n(R)$.

Corollary 3.1. *Let Assumption 3.14 hold and the conditions of Theorem 3.1(i) and Theorem 3.2 be satisfied. If $\hat{c}_n = \hat{q}_{1-\alpha}(\hat{U}_n(R|\ell_n))$, then it follows that:*

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(I_n(R) > \hat{c}_n) \leq \alpha.$$

In Algorithm 1 below we describe how to obtain p-values for the test described in Corollary 3.1 when the moments are differentiable. We note that if there are no inequality constraints, then it is possible to show that the test in Corollary 3.1 is similar and its asymptotic size equals the nominal level α whenever the conditions of Theorem

3.1(ii) are satisfied. The consistency of the test against any $P \in \mathbf{P} \setminus \mathbf{P}_0$ for which $\max_j \|E_P[\rho_j(X,\theta)|Z_j]\|_{P,2}$ is bounded away from zero (in $\theta \in \Theta \cap R$) is also straightforward to establish under suitable conditions. Finally, we also note that if we instead employ the critical value $\hat{c}_n = \hat{q}_{1-\alpha+\delta}(\hat{U}_n(R|\ell_n)) + \delta$ for any $\delta > 0$, then the conclusion of Corollary 3.1 holds without needing to impose Assumption 3.14; see Corollary S.3.1. This modification to the critical value was originally proposed in a different context by Andrews and Shi (2013), who suggest setting $\delta = 10^{-6}$.

Algorithm 1 Computing p-values for test based on $I_n(R)$

Require: Θ_n, Υ_F, Υ_G, $\{\rho(X_i,\theta) * q^{k_n}(Z_i)\}_{i=1}^n$, $\hat{\Sigma}_n$, r_n, τ_n, ℓ_n

$\triangleright$ Compute the Test Statistic

$Q_n(\theta) \leftarrow \|\hat{\Sigma}_n\{\frac{1}{n}\sum_{i=1}^n \rho(X_i,\theta) * q^{k_n}(Z_i)\}\|_p$ $\triangleright$ Criterion function

$R \leftarrow \{\theta : \Upsilon_F(\theta) = 0, \Upsilon_G(\theta) \leq 0\}$ $\triangleright$ Constraint Set

$I_n(R) \leftarrow \min_{\theta\in\Theta_n} \sqrt{n}Q_n(\theta)$ s.t. $\theta \in R$ $\triangleright$ Test Statistic

$\triangleright$ Prepare variables for bootstrap problem

$\hat{\mathbb{D}}_n(\theta)[h] \leftarrow \frac{1}{n}\sum_{i=1}^n \nabla_\theta \rho(X_i,\theta)[h] * q^{k_n}(Z_i)$ $\triangleright$ Moments Derivative

$\hat{\Theta}_n^{\mathrm{r}} \leftarrow \{\theta \in \Theta_n \cap R : Q_n(\theta) \leq I_n(R)/\sqrt{n} + \tau_n\}$ $\triangleright$ Boot Constraint θ

$G_n(\theta) \leftarrow \{h : \Upsilon_G(\theta + h/\sqrt{n}) \leq (\Upsilon_G(\theta) - K_g r_n \|h/\sqrt{n}\|_{\mathbf{B}} \mathbf{1}_{\mathbf{G}}) \vee (-r_n \mathbf{1}_{\mathbf{G}})\}$

$\hat{V}_n(\theta, R|\ell_n) \leftarrow \{h \in G_n(\theta) : \Upsilon_F(\theta + h/\sqrt{n}) = 0, \|h\|_{\mathbf{B}} \leq \ell_n\sqrt{n}\}$ $\triangleright$ Boot Constraint h

$\triangleright$ Compute B bootstrap statistics and obtain p-value

for $b = 1$ to B **do**

  $\{\omega_i^b\}_{i=1}^n \leftarrow$ Generate i.i.d. sample of $N(0,1)$ variables

  $\hat{\mathbb{W}}_n^b(\theta) \leftarrow \frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_i^b\{\rho(X_i,\theta) * q^{k_n}(Z_i) - \frac{1}{n}\sum_{j=1}^n \rho(X_j,\theta) * q^{k_n}(Z_j)\}$

  $F_n^b(\theta, h) \leftarrow \|\hat{\Sigma}_n\{\hat{\mathbb{W}}_n^b(\theta) + \hat{\mathbb{D}}_n(\theta)[h]\}\|_p$ $\triangleright$ Boot Criterion

  Boot[b] $\leftarrow \min_{\theta,h} F^b(\theta,h)$ s.t. $\theta \in \hat{\Theta}_n^{\mathrm{r}}, h \in \hat{V}_n(\theta, R|\ell_n)$ $\triangleright$ Boot Statistic

end for

pval $\leftarrow \frac{1}{B}\sum_{b=1}^B 1\{I_n(R) \leq \text{Boot}[b]\}$ $\triangleright$ Compute p-value

Remark 3.1. Suppose θ_0 is identified, we aim to test whether $\Upsilon_F(\theta_0) = 0$, and we are confident θ_0 satisfies $\Upsilon_G(\theta_0) \leq 0$. We could then set R to equal R_1 or R_2, where

$$R_1 = \{\theta \in \mathbf{B} : \Upsilon_G(\theta) \leq 0 \text{ and } \Upsilon_F(\theta) = 0\}$$
$$R_2 = \{\theta \in \mathbf{B} : \Upsilon_F(\theta) = 0\}.$$

The power functions of the corresponding tests are not necessarily ranked. As a result, it can be desirable to combine both tests by, for instance, using the test statistic $T_n \equiv \max\{F_1(I_n(R_1)), F_2(I_n(R_2))\}$ for F_1, F_2 increasing functions, and the quantiles of $\max\{F_1(\hat{U}_n(R_1|\ell_n)), F_2(\hat{U}_n(R_2|\ell_n))\}$ as critical values – e.g., F_j may be c.d.f. of $\hat{U}_n(R_j|\ell_n)$ conditional on the data. The asymptotic validity of such a test follows from Theorems 3.1 and 3.2 under a suitable modification of Assumption 3.14. ■

3.2.2 Tests Based on $I_n(R) - I_n(\Theta)$

We next establish the asymptotic validity of a test based on $I_n(R)-I_n(\Theta)$ by also relying on Theorems 3.1 and 3.2. In what follows, we signify parameters associated with setting $R = \Theta$ by a "u" superscript – e.g. $\mathcal{F}_n^{\mathrm{u}}$ is understood to be as in (14) but with $R = \Theta$.

In order to obtain a distributional approximation to the recentered statistic, we may simply apply Theorem 3.1(i) to $I_n(R)$ and Theorem 3.1(ii) to $I_n(\Theta)$ to conclude that

$$I_n(R) - I_n(\Theta) \leq U_P(R|\ell_n) - U_P(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n). \tag{22}$$

Moreover, by Theorem 3.2 we may approximate the distribution of $U_P(R|\ell_n)$ by using $\hat{U}_n(R|\ell_n)$. Similarly, to obtain a bootstrap approximation to $U_P(\Theta|+\infty)$, we define

$$\hat{\Theta}_n^{\mathrm{u}} \equiv \{\theta \in \Theta_n : Q_n(\theta) \leq \inf_{\theta \in \Theta_n} Q_n(\theta) + \tau_n^{\mathrm{u}}\};$$

i.e. $\hat{\Theta}_n^{\mathrm{u}}$ is simply the set estimator in (20) applied with $\Theta = R$. For $\mathbf{B}_n^{\mathrm{u}}$ the closed linear span of Θ_n, we then approximate the law of $U_P(\Theta|\ell_n^{\mathrm{u}})$ by employing

$$\hat{U}_n(\Theta|+\infty) \equiv \inf_{\theta \in \hat{\Theta}_n^{\mathrm{u}}} \inf_{h \in \mathbf{B}_n^{\mathrm{u}}} \|\hat{\mathbb{W}}_n(\theta) + \hat{\mathbb{D}}_n(\theta)[h]\|_{\hat{\Sigma}_n,p};$$

i.e. the bootstrap approximation equals that of Theorem 3.2, with the local parameter space being unconstrained due to the absence of equality or inequality restrictions.

The preceding discussion suggests that the quantiles of $\hat{U}_n(R|\ell_n) - \hat{U}_n(\Theta|+\infty)$ conditional on the data provide valid critical values for the recentered statistic. Our next result formally establishes that the resulting test is indeed asymptotically valid.

Corollary 3.2. *Let the conditions of Theorems 3.1(i) and 3.2 hold with R as in (13), the conditions of Theorems 3.1(ii) and 3.2 hold with $R = \Theta$, and Assumption 3.14 hold with $I_n(R)-I_n(\Theta)$, $\hat{U}_n(R|\ell_n)-\hat{U}_n(\Theta|+\infty)$, and $U_P(R|\tilde{\ell}_n)-U_P(\Theta|\tilde{\ell}_n^{\mathrm{u}})$ in place of $I_n(R)$, $\hat{U}_n(R|\ell_n)$, and $U_P(R|\tilde{\ell}_n)$ with $\tilde{\ell}_n^{\mathrm{u}}$ satisfying $\mathcal{R}_n^{\mathrm{u}} = o(\tilde{\ell}_n^{\mathrm{u}})$ and Assumption 3.13(ii) with $R = \Theta$. If $\tau_n^{\mathrm{u}} \downarrow 0$ satisfies $J_n^{\mathrm{u}} B_n k_n^{1/p}\sqrt{\log(1+k_n)/n} = o(\tau_n^{\mathrm{u}})$ and $\nu_n^{\mathrm{u}}\tau_n^{\mathrm{u}} \times \mathcal{S}_n^{\mathrm{u}}(\mathbf{B},\mathbf{E}) = o(1)$, then for $\hat{c}_n \equiv \hat{q}_{1-\alpha}(\hat{U}_n(R|\ell_n) - \hat{U}_n(\Theta|+\infty))$ it follows that*

$$\limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(I_n(R) - I_n(\Theta) > \hat{c}_n) \leq \alpha.$$

It is worth emphasizing that in coupling $I_n(\Theta)$ we must rely on Theorem 3.1(ii) instead of Theorem 3.1(i) in order to ensure that (22) holds. As a result, whenever moments are nonlinear, Corollary 3.2 requires the rate of convergence of the unconstrained estimator to be sufficiently fast for Theorem 3.1(ii) to apply. Similarly, in coupling $\hat{U}_n(\Theta|+\infty)$ it is important that $\hat{\Theta}_n^{\mathrm{u}}$ be consistent in the Hausdorff metric. Thus, while we may set $\tau_n^{\mathrm{u}} = 0$ in identified models, in partially identified models we require that τ_n^{u}

not tend to zero too fast; see Theorem S.1.1. Finally, we note that in identified models, it is possible to employ either $\hat{\mathbb{W}}_n(\hat{\theta}_n)$ or $\hat{\mathbb{W}}_n(\hat{\theta}_n^{\mathrm{u}})$ in constructing both $\hat{U}_n(R|\ell_n)$ and $\hat{U}_n(\Theta|+\infty)$ – a change that results in an asymptotically equivalent coupling but ensures that the bootstrap statistic $\hat{U}_n(R|\ell_n) - \hat{U}_n(\Theta|+\infty)$ is (weakly) positive.

4 Heterogeneity and Demand Analysis

For our final example, we illustrate how to conduct inference in the heterogeneous demand model of Hausman and Newey (2016) – alternative models of demand under conditional moment restrictions include the analysis in Hausman and Newey (1995), Blundell et al. (2012), and Chen and Christensen (2018). Specifically, for $Y \in [0,1]$ equal to the expenditure share on a commodity, $W \in \mathbf{W}$ a vector of prices, income, and covariates, and η representing unobserved individual heterogeneity we suppose

$$Y = g(W,\eta) \tag{23}$$

where g is a known function of (W,η). The unobserved heterogeneity η can potentially be infinite dimensional. For instance, Hausman and Newey (2016) set $\eta = \{\beta_j\}_{j=1}^\infty$ to be a random variable in the sequence space $\ell^2 \equiv \{\{a_j\}_{j=1}^\infty : \sum_j a_j^2 < \infty\}$, and let

$$g(W,\eta) = \sum_{j=1}^{\infty} \psi_j(W)\beta_j, \tag{24}$$

where $\{\psi_j\}_{j=1}^\infty$ is a known basis satisfying $\sum_{j=1}^\infty \psi_j^2(W) < \infty$ almost surely (in W).

If the covariates W are independent of η, then for any $c \in \mathbf{R}$ it follows that

$$P(Y \le c|W) = P(g(W,\eta) \le c|W) = \int 1\{g(W,\eta) \le c\}\mu_0(d\eta) \tag{25}$$

where μ_0 denotes the unknown distribution of η. Result (25) restricts the possible values of μ_0 and hence the identified set for functionals of μ_0, such as average exact consumer surplus or average share. Specifically, for $\Psi(g,\eta)$ an object of interest for preferences denoted by η, such as equivalent variation, Hausman and Newey (2016) study functionals

$$\int \Psi(g,\eta)\mu_0(d\eta), \tag{26}$$

which is the average across individuals. By evaluating the set of values of (26) which can be generated by a distribution μ_0 satisfying (25) at a grid $\{c_j\}_{j=1}^{\mathcal{J}}$, Hausman and Newey (2016) provide estimates of the identified set for the functional of interest. We further note bounds on the distribution of $\Psi(g,\eta)$ under μ_0 can be obtained by replacing $\Psi(g,\eta)$ in (26) with an indicator that $\Psi(g,\eta)$ be less than or equal to some number.

In what follows, we apply our results to conduct inference on functionals as in (26). To this end, we let $F_P(c_\jmath|W) \equiv P(Y \leq c_\jmath|W)$ for a given grid $\{c_\jmath\}_{\jmath=1}^{\mathcal{J}}$. To define $\mathbf{B}$, we suppose $\eta \in \Omega$ for some known Hausdorff space Ω, set $\mathcal{B}$ to be the Borel σ-algebra on Ω, let $\mathcal{M}$ be the space of regular signed Borel measures on Ω, and let $\|\cdot\|_{TV}$ denote the total variation norm. Assuming $F_P(c_\jmath|\cdot) \in C_B(\mathbf{W})$ for $C_B(\mathbf{W})$ the space of continuous and bounded functions on $\mathbf{W}$, we set $\mathbf{B} = (\bigotimes_{\jmath=1}^{\mathcal{J}} C_B(\mathbf{W})) \times \mathcal{M}$, for any $(\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu) = \theta \in \mathbf{B}$ let $\|\theta\|_{\mathbf{B}} = \sum_{\jmath=1}^{\mathcal{J}} \|F(c_\jmath|\cdot)\|_\infty + \|\mu\|_{TV}$, and set

$$\Theta = \{(\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu) = \theta \in \mathbf{B} : \max_{1\leq \jmath \leq \mathcal{J}} \|F(c_\jmath|\cdot)\|_\infty \leq 2\}, \tag{27}$$

where the "2" norm bound is simply selected to ensure Θ_0 is in the interior of Θ.

Letting $X = (Y, W)$ and setting $Z_\jmath = W$ for every $1 \leq \jmath \leq \mathcal{J}$ we then define

$$\rho_\jmath(X, \theta) = 1\{Y \leq c_\jmath\} - F(c_\jmath|W), \tag{28}$$

which yields conditional moment restrictions that identify $F_P(c_\jmath|W)$ – note, however, that μ_0 is potentially partially identified. For a grid $\{w_l\}_{l=1}^{\mathcal{L}} \subseteq \mathbf{W}$ we test whether a hypothesized value λ belongs to the identified set for the functional in (26) by setting

$$\begin{aligned} R = \Big\{ (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu) : \mu(\Omega) = 1,\ \mu(B) \geq 0 \text{ for all } B \in \mathcal{B},\ \int \Psi(g,\eta)\mu(d\eta) = \lambda, \\ \text{and } F(c_\jmath|w_l) = \int 1\{g(w_l,\eta) \leq c_\jmath\}\mu(d\eta) \text{ for all } 1 \leq \jmath \leq \mathcal{J}, 1 \leq l \leq \mathcal{L} \Big\}. \end{aligned} \tag{29}$$

Thus, the null hypothesis that $\Theta_0 \cap R$ be nonempty corresponds to requiring that there exist a distribution μ for η satisfying the restrictions in (25) at the points $(c_\jmath, w_l)$ and yielding a value for the functional in (26) of λ. By conducting test inversion in λ we can obtain a confidence region for the desired functional. To map R into the framework of Section 3, we set $\mathbf{G} = \ell^\infty(\mathcal{B})$ for $\ell^\infty(\mathcal{B})$ the set of bounded functions on $\mathcal{B}$ and for any $(\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu) = \theta \in \mathbf{B}$ let $\Upsilon_G : \mathbf{B} \to \ell^\infty(\mathcal{B})$ be given by

$$\Upsilon_G(\theta)(B) = -\mu(B). \tag{30}$$

Finally, we set $\Upsilon_F : \mathbf{B} \to \mathbf{R}^{\mathcal{J}\mathcal{L}+2}$ to equal $\Upsilon_F(\theta) = (\Upsilon_F^{(\mathrm{e})}(\theta), \Upsilon_F^{(\mu)}(\theta), \Upsilon_F^{(\mathrm{s})}(\theta))$, where

$$\begin{aligned} \Upsilon_F^{(\mathrm{e})}(\theta) &= \{F(c_\jmath|w_l) - \int 1\{g(w_l,\eta) \leq c_\jmath\}\mu(d\eta)\}_{1\leq \jmath\leq \mathcal{J}, 1\leq l\leq \mathcal{L}} \\ \Upsilon_F^{(\mu)}(\theta) &= \mu(\Omega) - 1 \\ \Upsilon_F^{(\mathrm{s})}(\theta) &= \int \Psi(g,\eta)\mu(d\eta) - \lambda. \end{aligned} \tag{31}$$

Given these definitions, we may then map R (as introduced in (29)) into the framework of Section 3 by noting that $R = \{\theta \in \mathbf{B} : \Upsilon_F(\theta) = 0 \text{ and } \Upsilon_G(\theta) \leq 0\}$.

As in Hausman and Newey (2016), we can impose utility maximization by requiring that the support Ω consist only of η such that $g(\cdot, \eta)$ satisfies the Slutsky conditions. One may sample from Ω by drawing randomly from sets of η that satisfy Slutsky symmetry and only keeping those where the compensated price effects matrix is negative semidefinite on a grid. This is the procedure followed in Hausman and Newey (2016) for two goods. Importantly, we emphasize that because the utility maximization restrictions are imposed through Ω, they do not affect the basic structure of Υ_F and Υ_G – i.e., Υ_F and Υ_G remain linear maps satisfying Assumptions 3.8-3.10. In this sense, as long as they are imposed through the support Ω of η, our procedure allows us to accommodate a wide array of shape restrictions on individual demand $g(\cdot, \eta)$.

Given a collection of orthogonal probability measures $\{\delta_s\}_{s=1}^{s_n} \subseteq \mathcal{M}$ we employ

$$\mathcal{M}_n = \{\mu \in \mathcal{M} : \mu = \sum_{s=1}^{s_n} \alpha_s \delta_s \text{ for some } \{\alpha_s\}_{s=1}^{s_n} \in \mathbf{R}^{s_n}\}$$

as a sieve for $\mathcal{M}$. Employing orthogonal measures, such as distinct Dirac measures, is computationally attractive as it simplifies imposing the nonnegativity constraint on any $\mu \in \mathcal{M}_n$. As a sieve for $\{F_P(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}$, we employ approximating functions $\{p_j\}_{j=1}^{j_n}$. In particular, setting $p^{j_n}(w) = (p_1(w), \ldots, p_{j_n}(w))'$, we set as our sieve

$$\Theta_n = \{(\{p^{j_n\prime}\beta_\jmath\}_{\jmath=1}^{\mathcal{J}}, \mu) : \mu \in \mathcal{M}_n \text{ and } \max_{1 \leq \jmath \leq \mathcal{J}} \|p^{j_n\prime}\beta_\jmath\|_\infty \leq 2\}.$$

Similarly, for a sequence $\{q_k\}_{k=1}^{k_n}$ and $k_n \times k_n$ positive definite matrices $\{\hat{\Sigma}_{\jmath,n}\}_{\jmath=1}^{\mathcal{J}}$, we set $q^{k_n}(w) = (q_1(w), \ldots, q_{k_n}(w))'$ and for any $(\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu) = \theta$ define

$$Q_n(\theta) = \{\sum_{\jmath=1}^{\mathcal{J}} \|\frac{1}{n}\sum_{i=1}^{n}(1\{Y_i \leq c_\jmath\} - F(c_\jmath|W_i))q^{k_n}(W_i)\|_{\hat{\Sigma}_{\jmath,n},2}^2\}^{1/2}. \tag{32}$$

The statistics $I_n(R)$ and $I_n(\Theta)$ then equal the minimums of $\sqrt{n}Q_n$ over $\Theta_n \cap R$ and Θ_n.

Our next set of assumptions enable us to couple $I_n(R)$ and $I_n(R) - I_n(\Theta)$.

Assumption 4.1. *(i) $\{Y_i, W_i\}_{i=1}^n$ is i.i.d. with $(Y, W) \sim P \in \mathbf{P}$; (ii) $\sup_w \|p^{j_n}(w)\|_2 \lesssim \sqrt{j_n}$; (iii) $E_P[p^{j_n}(W)p^{j_n}(W)']$ has eigenvalues bounded away from zero and infinity uniformly in $P \in \mathbf{P}$ and j_n; (iv) For each $P \in \mathbf{P}_0$ and $\theta \in \Theta_0 \cap R$, there exists a $\Pi_n\theta = (\{F_n(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu_n) \in \Theta_n \cap R$ such that $\sum_{\jmath=1}^{\mathcal{J}} \|E_P[(F_n(c_\jmath|W) - F_P(c_\jmath|W))q^{k_n}(W)]\|_2 = O((n\log(n))^{-1/2})$ uniformly in $P \in \mathbf{P}_0$ and $\theta \in \Theta_0 \cap R$.*

Assumption 4.2. *(i) $\max_{1 \leq k \leq k_n} \|q_k\|_\infty \lesssim \sqrt{k_n}$; (ii) $E_P[q^{k_n}(W)q^{k_n}(W)']$ has eigenvalues bounded uniformly in $P \in \mathbf{P}$ and k_n; (iii) $E_P[q^{k_n}(W)p^{j_n}(W)']$ has singular values bounded away from zero uniformly in $P \in \mathbf{P}$ and (k_n, j_n); (iv) $k_n^2 j_n \log^3(n) = o(n^{1/2})$.*

Assumption 4.3. *For all $1 \leq \jmath \leq \mathcal{J}$: (i) $\|\hat{\Sigma}_{\jmath,n} - \Sigma_{\jmath,P}\|_{o,2} = o_P(1/k_n\sqrt{j_n}\log^2(n))$*

uniformly in $P \in \mathbf{P}$; (ii) The $k_n \times k_n$ matrices $\Sigma_{\jmath,P}$ are invertible and $\|\Sigma_{\jmath,P}\|_{o,2}$ and $\|\Sigma_{\jmath,P}^{-1}\|_{o,2}$ are bounded uniformly in $P \in \mathbf{P}$ and k_n.

Assumptions 4.1(ii)-(iv) state the conditions on Θ_n, with Assumptions 4.1(ii)(iii) being satisfied by standard choices such as B-Splines or wavelets. Assumption 4.1(iv) is an asymptotic unbiasedness requirement – a condition that is eased by noting no requirements are imposed on the approximating space for μ_0. The requirements on $\{q_k\}_{k=1}^{k_n}$ are imposed in Assumption 4.2(i)(iii) and are again satisfied by standard choices. Assumption 4.2(iv) states a rate condition that suffices for verifying the coupling requirements of Theorem 3.1. Assumption 4.3 imposes the requirements on the weighting matrices.

Our next result employs Theorem 3.1(ii) to obtain strong approximations.

Theorem 4.1. *Let Assumptions 4.1, 4.2, 4.3 hold, $a_n = (\log(n))^{-1/2}$, and for any $\theta = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu) \in \mathbf{B}$ let $\|\theta\|_{\mathbf{E}} = \sum_{\jmath=1}^{\mathcal{J}} \sup_{P\in\mathbf{P}} \|F(c_\jmath|\cdot)\|_{P,2}$. If $\ell_n, \ell_n^{\mathrm{u}} \downarrow 0$ satisfy $k_n\sqrt{j_n}\log^2(n)(\ell_n \vee \ell_n^{\mathrm{u}}) = o(1)$, $k_n\sqrt{j_n}\log(n)/\sqrt{n} = o(\ell_n \wedge \ell_n^{\mathrm{u}})$, then uniformly in $P \in \mathbf{P}_0$:*

$$I_n(R) = U_P(R|\ell_n) + o_P(a_n)$$
$$I_n(R) - I_n(\Theta) = U_P(R|\ell_n) - U_P(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n).$$

In order to conduct inference, we next aim to estimate the distributions of $U_P(R|\ell_n)$ and $U_P(\Theta|\ell_n^{\mathrm{u}})$. To this end, we note that Θ_{0n}^{r} (as in (15)) is potentially non-singleton and we therefore employ a set estimator $\hat{\Theta}_n^{\mathrm{r}}$ (as in (20)) to estimate the distribution of $U_P(R|\ell_n)$. In contrast, since $U_P(\Theta|\ell_n^{\mathrm{u}})$ only depends on the identified component $\{F_P(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}$, for the unconstrained problem we employ any minimizer $\hat{\theta}_n^{\mathrm{u}}$ of Q_n over Θ_n. With regards to the local parameter space, we note that in this application

$$G_n(\theta) = \{(\{p^{j_n\prime}\beta_{\jmath,h}\}_{\jmath=1}^{\mathcal{J}}, \mu_h) : \mu_h(B) \geq \sqrt{n}\min\{r_n - \mu(B), 0\} \text{ for all } B \in \mathcal{B}\} \tag{33}$$

for any $\theta = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu)$. Computationally, since any $\mu, \mu_h \in \mathcal{M}_n$ has the structure $\mu = \sum_{s=1}^{s_n} \alpha_s\delta_s$ and $\mu_h = \sum_{s=1}^{s_n} \alpha_{sh}\delta_s$ it follows that the constraints in (33) reduce to $\alpha_{sh} \geq \min\{r_n - \alpha_s, 0\}$ for all $1 \leq s \leq s_n$ whenever $\{\delta_s\}_{s=1}^{s_n}$ are orthogonal. Furthermore, since moments and restrictions are linear, we may let $\ell_n = +\infty$ and set

$$\hat{V}_n(\theta, R| + \infty) = \{(\{p^{j_n\prime}\beta_{\jmath,h}\}_{\jmath=1}^{\mathcal{J}}, \mu_h) : h \in G_n(\theta),\ \Upsilon_F(h) = 0\}. \tag{34}$$

For each $\theta \in \Theta_n$, we denote the bootstrap process for the $\jmath^{th}$ conditional moment by

$$\hat{\mathbb{W}}_{\jmath,n}(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \omega_i\{\rho_\jmath(X_i, \theta)q^{k_n}(W_i) - \frac{1}{n}\sum_{j=1}^{n} \rho_\jmath(X_j, \theta)q^{k_n}(W_j)\}.$$

Similarly, we set $\hat{\mathbb{D}}_{\jmath,n}[h] = -\sum_{i=1}^{n} q^{k_n}(W_i)p^{j_n}(W_i)'\beta_{\jmath,h}/n$ for any $h = (\{p^{j_n\prime}\beta_{\jmath,h}\}_{\jmath=1}^{\mathcal{J}}, \mu_h)$.

Thus, the estimators of the strong approximations obtained in Theorem 4.1 equal

$$\hat{U}_n(R|+\infty) = \inf_{\theta\in\hat{\Theta}_n^{\mathrm{r}}} \inf_{h\in\hat{V}_n(\theta,R|+\infty)} \{\sum_{\jmath=1}^{\mathcal{J}} \|\hat{\mathbb{W}}_{\jmath,n}(\theta) + \hat{\mathbb{D}}_{\jmath,n}[h]\|_{\hat{\Sigma}_{\jmath,n},2}\}^{1/2}$$

$$\hat{U}_n(\Theta|+\infty) = \inf_h \{\sum_{\jmath=1}^{\mathcal{J}} \|\hat{\mathbb{W}}_{\jmath,n}(\hat{\theta}_n^{\mathrm{u}}) + \hat{\mathbb{D}}_{\jmath,n}[h]\|_{\hat{\Sigma}_{\jmath,n},2}\}^{1/2}.$$

Before stating our final assumption, we need an auxiliary result. To this end, define

$$\Gamma_n(\theta) \equiv \{\tilde{\mu} \in \mathcal{M}_n : \tilde{\theta} = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \tilde{\mu}) \text{ satisfies } \Upsilon_F(\tilde{\theta}) = 0,\ \Upsilon_G(\tilde{\theta}) \leq 0\} \tag{35}$$

for any $\theta = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu)$ – i.e. $\Gamma_n(\theta)$ is the set of distributions of η that agree with the restrictions implied by $\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}$. Our next result bounds the $\|\cdot\|_{TV}$-Hausdorff distance between $\Gamma_n(\theta_1)$ and $\Gamma_n(\theta_2)$, which we denote by $d_H(\Gamma_n(\theta_1), \Gamma_n(\theta_2), \|\cdot\|_{TV})$.

Lemma 4.1. *If the probability measures $\{\delta_s\}_{s=1}^{s_n}$ are orthogonal, then for every n there exists a constant $\zeta_n < \infty$ independent of $\mathbf{P}$ such that*

$$d_H(\Gamma_n(\theta_1), \Gamma_n(\theta_2), \|\cdot\|_{TV}) \leq \zeta_n \sum_{\jmath=1}^{\mathcal{J}} \|F_1(c_\jmath|\cdot) - F_2(c_\jmath|\cdot)\|_\infty$$

for any $(\{F_1(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu_1) = \theta_1 \in \Theta_n \cap R$ and $(\{F_2(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu_2) = \theta_2 \in \Theta_n \cap R$.

We introduce our final assumption to show the validity of our bootstrap procedure.

Assumption 4.4. *(i) $\Psi(g,\cdot)$ is bounded on Ω; (ii) The probability measures $\{\delta_s\}_{s=1}^{s_n}$ are orthogonal; (iii) $k_n^4 \jmath_n^5 \log^5(n)/n = o(1)$; (iv) $\Pi_n\theta = (\{F_n(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu_n)$ satisfies $\|F_n(c_\jmath|\cdot) - F_P(c_\jmath|\cdot)\|_\infty = o(1)$ uniformly in $\theta \in \Theta_0 \cap R$ and $P \in \mathbf{P}_0$; (v) $k_n\sqrt{\jmath_n}\log^2(n)\tau_n = o(1)$, and $\zeta_n(k_n \jmath_n \log(n)/\sqrt{n} + \sqrt{\jmath_n}\tau_n) = o(r_n)$.*

The boundedness of $\Psi(g,\cdot)$ on Ω ensures $\Upsilon_F^{(\mathrm{s})}$ (as in (31)) is continuous, while Assumption 4.4(ii) allows us to apply Lemma 4.1. Assumption 4.4(iii) is a low level sufficient condition for verifying the bootstrap coupling requirement of Assumption 3.11. These rate requirements could be improved under smoothness conditions on $F_P(c_\jmath|\cdot)$. Finally, Assumption 4.4(iv) imposes a mild requirement on the sieve, while Assumption 4.4(v) states conditions on τ_n and r_n – note $\tau_n = 0$ and $r_n = +\infty$ are always valid, though such choices can lead to lower local power against certain alternatives.

Our final result obtains a coupling for our bootstrap approximations.

Theorem 4.2. *Let the conditions of Theorem 4.1 hold and Assumption 4.4 be satisfied. Then: there are sequences $\ell_n, \ell_n^{\mathrm{u}} \downarrow 0$ satisfying $k_n\sqrt{\jmath_n}\log(n)/\sqrt{n} = o(\ell_n \wedge \ell_n^{\mathrm{u}})$ and*

$k_n\sqrt{j_n}\log^2(n)(\ell_n \vee \ell_n^{\mathrm{u}}) = o(1)$ *such that uniformly in* $P \in \mathbf{P}_0$

$$\hat{U}_n(R| + \infty) \geq U_P^\star(R|\ell_n) + o_P(a_n)$$
$$\hat{U}_n(R| + \infty) - \hat{U}_n(\Theta| + \infty) \geq U_P^\star(R|\ell_n) - U_P^\star(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n).$$

In particular, since the conditions on ℓ_n and ℓ_n^{u} imposed in Theorems 4.1 and 4.2 are the same, it follows that we may employ the quantiles of $\hat{U}_n(R|+\infty)$ and $\hat{U}_n(R|+\infty) - \hat{U}_n(\Theta| + \infty)$ conditional on the data as critical values for $I_n(R)$ and $I_n(R) - I_n(\Theta)$.

5 Simulation Evidence

To conclude, we study the finite sample performance of our inference procedure by revisiting the simulation design in Chetverikov and Wilhelm (2017).

5.1 Identified Model

We first consider a nonparametric instrumental variable model in which, for some unknown function θ_0, the distribution of $(Y, W, Z) \in \mathbf{R}^3$ satisfies the restriction

$$Y = \theta_0(W) + \varepsilon \qquad E[\varepsilon|Z] = 0; \tag{36}$$

see Appendix A.2 for a formal study of this model. Following Chetverikov and Wilhelm (2017), we set $\theta_0(w) \equiv 0.2w + w^2$ and for (ϵ, ζ, ν) independent standard normal random variables we let $Z = \Phi(\zeta)$, $W = \Phi(0.3\zeta + \sqrt{1-(0.3)^2}\epsilon)$, and $\varepsilon = (0.3\epsilon + \sqrt{1-(0.3)^2}\nu)/2$ for Φ the cumulative distribution function of a standard normal. All reported results are based on five thousand replications employing five hundred bootstrap draws each.

In what follows, we utilize the restriction $\Upsilon_F(\theta_0) = 0$ to impose a hypothesized value on the the level or the derivative of θ_0 at the point $w_0 = 0.5$ and use $\Upsilon_G(\theta_0) \leq 0$ to impose that θ_0 be either monotonically increasing or monotonically increasing and convex. We employ the test statistic $I_n(R) - I_n(\Theta)$ with $p = 2$ and $\hat{\Sigma}_n$ an estimate of the optimal weighting matrix based on a first stage unconstrained estimator. The implementation of the test is similar to that of the linear model of Section 2.1, with the difference that we must select the sieve $\Theta_n = \{p^{j_n\prime}\beta : \beta \in \mathbf{R}^{j_n}\}$ and q^{k_n}. We follow Chetverikov and Wilhelm (2017) in employing continuously differentiable piecewise quadratic splines with equally spaced knots for both p^{j_n} and q^{k_n}.

In computing critical values we set $\ell_n = +\infty$ since the model is linear and $\tau_n = 0$ since the model is identified. We select r_n by proceeding as in Section 2.1. Specifically, the choice of sieve implies that, for any $\theta = p^{j_n\prime}\beta$, the restriction $\Upsilon_G(\theta) \leq 0$ is equivalent

	$r_n/(j_n,k_n)$	Imposed: Mon.				Imposed: Mon.+ Conv.			
		Level		Derivative		Level		Derivative	
		(4,4)	(4,6)	(4,4)	(4,6)	(4,4)	(4,6)	(4,4)	(4,6)
	∞	1.90	1.72	1.88	2.02	1.44	1.52	2.74	2.84
	95%	1.74	1.68	1.90	2.08	1.46	1.54	2.68	2.84
$n = 500$	50%	1.74	1.70	1.90	2.10	1.46	1.54	2.68	2.84
	5%	2.18	2.90	2.20	2.96	1.52	1.82	2.74	2.98
	0	5.30	5.10	4.62	4.48	5.42	5.36	5.08	4.84
	∞	1.56	1.82	1.68	1.94	1.40	1.54	2.26	2.32
	95%	1.52	1.84	1.64	1.86	1.36	1.44	2.04	2.26
$n = 1000$	50%	1.52	1.86	1.64	1.86	1.36	1.44	2.04	2.26
	5%	2.02	2.84	2.06	3.06	1.44	1.86	2.14	2.38
	0	4.54	4.56	4.58	4.68	4.62	4.78	4.38	4.20
	∞	1.34	1.58	1.26	1.52	1.04	1.36	1.36	1.58
	95%	1.40	1.50	1.32	1.62	1.06	1.42	1.36	1.62
$n = 5000$	50%	1.42	1.52	1.32	1.62	1.06	1.42	1.36	1.62
	5%	2.20	3.62	2.36	3.36	1.42	2.38	1.46	1.86
	0	3.98	4.56	4.68	4.50	4.10	4.74	3.98	4.06

Table 2: Empirical rejection probabilities for 5%-level tests based on $I_n(R) - I_n(\Theta)$. Value of r_n set to a percentile corresponds to choice of $1 - \gamma_n$ in (37).

to $G\beta \leq 0$ for a known matrix G. For $p^{j_n\prime}\hat{\beta}_n^{\mathrm{u}}$ the minimizer of $I_n(\Theta)$ and $p^{j_n\prime}\hat{\beta}_n^{\mathrm{u}\star}$ its score bootstrap analogue (Kline and Santos, 2012), we therefore set r_n to satisfy

$$P(\max_j G_j\{\hat{\beta}_n^{\mathrm{u}\star\prime} - \hat{\beta}_n\} \leq r_n | \{V_i\}_{i=1}^n) = 1 - \gamma_n \tag{37}$$

where $\gamma_n \in (0,1)$ and the vectors $G_j \in \mathbf{R}^{j_n}$ depend on the shape restriction being imposed. We emphasize that the sequence γ_n must tend to zero in order for r_n to satisfy our assumptions. Finally, we employ the minimizer of $I_n(R)$ in obtaining bootstrap draws for both $\hat{U}_n(R|+\infty)$ and $\hat{U}_n(\Theta|+\infty)$; see discussion following Corollary 3.2.

Table 2 reports empirical rejection probabilities under the null hypothesis for 5%-level tests on the derivative and level of θ_0 at $w_0 = 0.5$ under different shape restrictions. With regards to r_n, we examine the extreme possible values (0 and ∞) and choices corresponding to (37) for different γ_n. In accord to theory, which requires $\gamma_n \downarrow 0$, we find that the rejection probability is no larger than the nominal level except for very small values of $1-\gamma_n$. Overall, we find the general lack of sensitivity to different choices of bandwidths to be reassuring for empirical practice.

In Figure 3 we report power curves for different 5%-level tests concerning the value of θ_0 and its derivative at $w_0 = 0.5$. For conciseness, we focus on the sample sizes $n \in \{1000, 5000\}$ and r_n chosen as in (37) with $1 - \gamma_n = 0.95$. The curves labeled "Mon" and "Mon+Conv" correspond to tests based on $I_n(R) - I_n(\Theta)$ with R imposing monotonicity and monotonicity and convexity while changing the conjectured value of

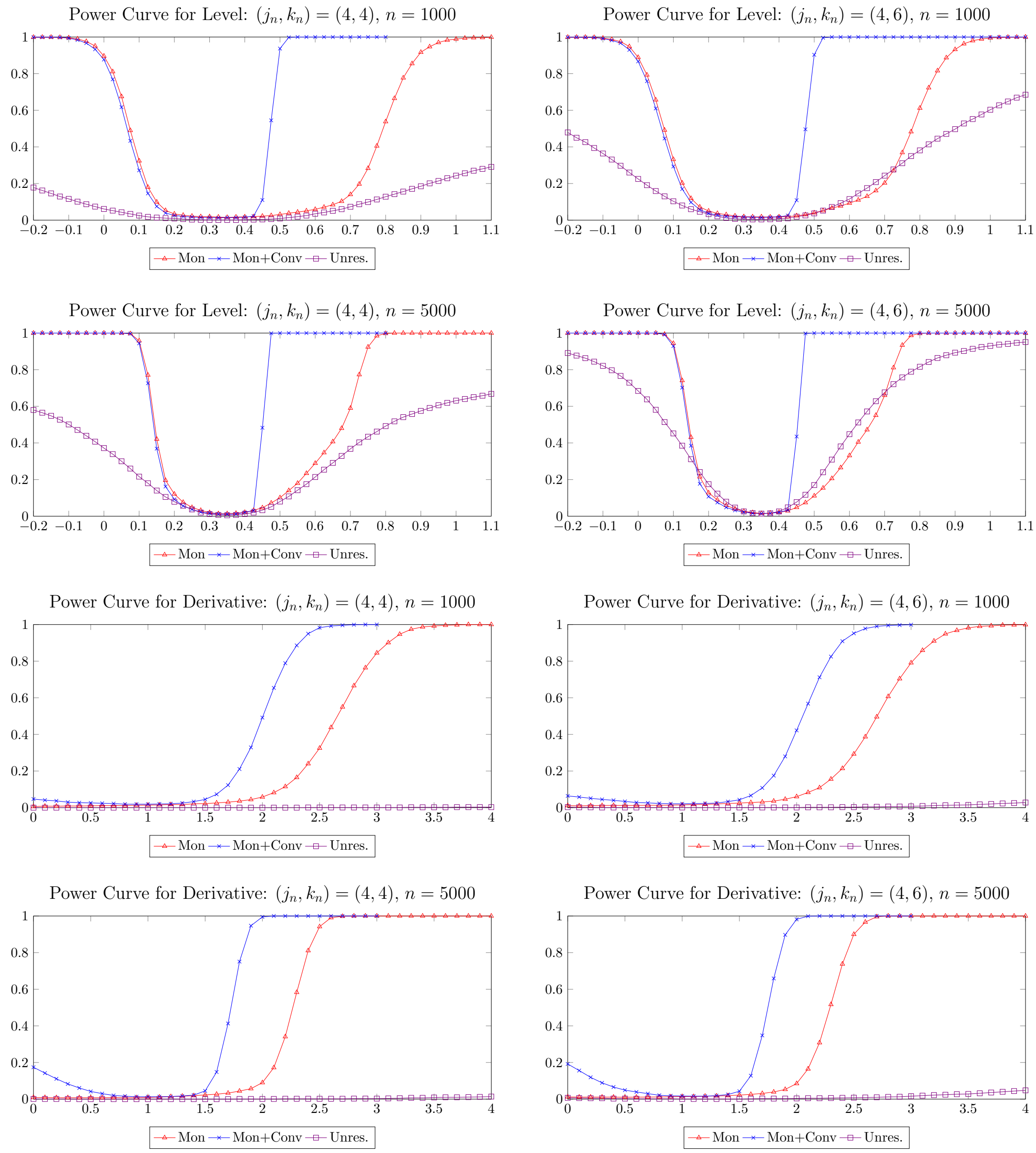


Figure 3: Rejection probabilities for 5%-level tests on conjectured value of $\theta_0(0.5)$ (true value 0.35) and $\theta_0'(0.5)$ (true value 1.2). Tests implemented with $1 - \gamma_n = 0.05$ in (37).

θ_0 and its derivative at $w_0 = 0.5$. The curve labeled "Unres." corresponds to a Wald test based on the unrestricted estimator. For all designs we find that imposing shape restrictions can improve power. The effect of imposing shape restrictions, however, depend on both the sampling uncertainty and how "close" the shape restrictions are to binding (Chetverikov et al., 2018). Since our design is fixed with n and θ_0 is strictly increasing and convex, in our simulations we see the advantages of imposing shape restrictions decrease with n as sample uncertainty decreases. Similarly, since estimating the derivative is a harder than estimating the level, we observe larger power gains when imposing shape restrictions in the former problem.

5.2 Partially Identified Model

We next examine the performance of our test in a partially identified setting by discretizing the simulation design in Chetverikov and Wilhelm (2017). Concretely, we generate $(W, Z, \epsilon) \in [0,1]^2 \times \mathbf{R}$ as in Section 5.1, divide $[0,1]$ into S_w and S_z equally spaced segments, and generate dummy variables D_w and D_z for the segment to which W and Z belong – e.g. if $(S_w, S_z) = (3,2)$, then $D_w(W) \equiv (1\{W \in [0,1/3]\}, 1\{W \in (1/3,2/3]\}, 1\{W \in (2/3,1]\})'$ and $D_z(Z) \equiv (1\{Z \in [0,1/2]\}, 1\{Z \in (1/2,1]\})'$. The outcome Y is generated according to (36) but employing D_w in place of W.

The discretized design is characterized by S_z linear unconditional moment restrictions in S_w unknowns. For conciseness, we focus on imposing that θ_0 be monotonically increasing and convex while conducting inference on the value of θ_0 at the point $d_0 \equiv D_w(0.5)$ – e.g, if $S_w = 3$, then $d_0 = (0,1,0)'$. The parameter $\theta_0(d_0)$ is generically not identified whenever $S_w > S_z$ but, as we report in Table 3, imposing a shape restriction on θ_0 partially identifies $\theta_0(d_0)$. A similar setting was previously studied by Freyberger and Horowitz (2015) who develop confidence regions for parameters such as $\theta_0(d_0)$. Their leading procedure is computationally simpler than ours, but can suffer from size distortions, for example, when the identified set for $\theta_0(d_0)$ is "small."

	(S_w, S_z)		
Restriction on θ_0	$(3,2)$	$(4,2)$	$(3,2)$
Mon.+Convex	[0.059, 0.252]	[0.100, 0.412]	[0.310, 0.388]
No Restriction	$(-\infty, +\infty)$	$(-\infty, +\infty)$	$(-\infty, +\infty)$

Table 3: Identified sets for $\theta_0(d_0)$ with and without shape restrictions.

We test whether a value λ belongs to the identified set for $\theta_0(d_0)$ by setting $\Upsilon_F(\theta) = \theta(d_0) - \lambda$ and employ the constraint $\Upsilon_G(\theta) \leq 0$ to impose that θ be monotonically increasing and convex. We base inference on $I_n(R)$ with $p = 2$, $\hat{\Sigma}_n$ the sample analogue to $E[D_z D_z']$, all moment restrictions ($k_n = S_z$), and a saturated model for θ_0 ($j_n = S_w$). To compute critical values we set $\ell_n = +\infty$ and $\tau_n = 0$ – though note $\hat{\Theta}_n^{\mathrm{r}}$ need not be a singleton when $\tau_n = 0$ because $j_n > k_n$. We select r_n by modifying the approach employed in Section 5.1. Specifically, we note that the constraint $\Upsilon_G(\theta) \leq 0$ may be written as $G\theta \leq 0$ for some matrix G, and for $\hat{\theta}_n^L$ and $\hat{\theta}_n^U$ the minimizer and maximizers of $\theta(d_0)$ over the set of θ that are monotonically increasing, convex, and minimize $\| \sum_{i=1}^n (Y_i - \theta(D_{w,i})) D_{z,i}/n \|_\infty$, we set r_n according to

$$P(\max_j \max\{G_j(\hat{\theta}_n^{L\star} - \hat{\theta}_n^L), G_j(\hat{\theta}_n^{U\star} - \hat{\theta}_n^U)\} \leq r_n | \{V_i\}_{i=1}^n) = 1 - \gamma_n, \tag{38}$$

where $\hat{\theta}_n^{L\star}$ and $\hat{\theta}_n^{U\star}$ are again computed employing the score bootstrap. As in our previous analysis, γ_n must tend to zero with n in order for r_n to satisfy our assumptions.

		Lower Endpoint			Midpoint			Upper Endpoint		
		(S_w, S_z)			(S_w, S_z)			(S_w, S_z)		
	r_n	(3,2)	(4,2)	(4,3)	(3,2)	(4,2)	(4,3)	(3,2)	(4,2)	(4,3)
$n = 500$	∞	1.96	3.34	1.48	0.10	0.02	1.48	1.88	3.10	2.00
	95%	3.64	4.70	1.46	0.10	0.02	1.46	2.26	3.12	1.98
	50%	5.34	5.24	1.46	0.50	0.06	1.50	5.22	5.02	2.04
	5%	5.36	5.24	3.56	0.50	0.06	3.44	5.24	5.02	3.54
	0	5.34	5.26	4.64	0.50	0.06	4.48	5.24	5.16	4.60
$n = 1000$	∞	1.84	3.06	1.12	0.00	0.00	1.10	1.96	2.90	1.34
	95%	4.98	4.84	1.12	0.02	0.00	1.08	2.98	2.90	1.34
	50%	5.10	4.88	1.20	0.12	0.00	1.14	5.00	4.86	1.44
	5%	5.10	4.88	3.48	0.12	0.00	3.12	5.00	4.86	2.78
	0	5.28	4.88	4.42	0.08	0.00	4.14	5.10	4.86	3.82
$n = 5000$	∞	1.98	4.40	1.34	0.00	0.00	1.22	1.98	2.80	1.36
	95%	5.08	6.76	1.34	0.00	0.00	1.26	4.56	4.86	1.34
	50%	5.08	8.30	1.48	0.00	0.00	1.44	4.58	4.84	1.52
	5%	5.08	9.00	4.28	0.00	0.00	4.14	4.58	4.84	3.58
	0	4.96	8.84	4.70	0.00	0.00	4.38	4.64	5.02	4.46

Table 4: Empirical rejection probabilities for 5%-level tests based on $I_n(R)$ for different points in the null hypothesis. Lower and upper endpoints correspond to Table 3.

Table 4 reports empirical rejection rates for testing whether λ belongs to the identified set, with the lower and upper endpoint columns corresponding to setting λ to equal the lower and upper endpoints in Table 3. All tests are conducted at a 5% nominal level. Across designs, we find that setting $r_n = +\infty$ always delivers tests with rejection probabilities below their nominal level. Setting r_n according to (38) with $1 - \gamma_n = 0.95$ also delivers adequate size control, with the exception of $n = 5000$ and $(S_w, S_z) = (4, 2)$ where we see a modest over-rejection at the lower endpoint of the identified set. Overall, the degree of sensitivity to the choice of r_n is similar to that found in Section 5.1.

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.

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.

ANGRIST, J. D. and EVANS, W. N. (1998). Children and their parents' labor supply: evidence from exogenous variation in family size. *The American Economic Review*, **88** 450–477.

ARMSTRONG, T. (2015). Adaptive testing on a regression function at a point. *The Annals of Statistics*, **43** 2086–2101.

ATHEY, S. and STERN, S. (1998). An empirical framework for testing theories about complimentarity in organizational design. Tech. rep., National Bureau of Economic Research.

BERESTEANU, A. and MOLINARI, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, **76** 763–814.

BLUNDELL, R., CHEN, X. and KRISTENSEN, D. (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, **75** 1613–1669.

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.

BUGNI, F. A., CANAY, I. A. and SHI, X. (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*, **8** 1–38.

CAI, T. T., LOW, M. G. and XIA, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *The Annals of Statistics*, **41** 722–750.

CANAY, I. A. and SHAIKH, A. M. (2017). Practical and theoretical advances in inference for partially identified models. In *Advances in Economics and Econometrics: Eleventh World Congress*, vol. 2. Cambridge University Press Cambridge, 271–306.

CHAMBERLAIN, G. (1992). Comment: Sequential moment restrictions in panel data. *Journal of Business & Economic Statistics*, **10** 20–26.

CHAMBERLAIN, G. and IMBENS, G. W. (2003). Nonparametric applications of bayesian inference. *Journal of Business & Economic Statistics*, **21** 12–18.

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. and CHRISTENSEN, T. M. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric iv regression. *Quantitative Economics*, **9** 39–84.

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. and SANTOS, A. (2018). Overidentification in regular models. *Econometrica*, **86** 1771–1817.

Chen, X., Tamer, E. and Torgovitsky, A. (2011). Sensitivity analysis in semiparametric likelihood models. Working paper, Yale University.

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., Hong, H. and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, **75** 1243–1284.

Chetverikov, D., Santos, A. and Shaikh, A. M. (2018). The econometrics of shape restrictions. *Annual Review of Economics*, **10** 31–63.

Chetverikov, D. and Wilhelm, D. (2017). Nonparametric instrumental variable estimation under monotonicity. *Econometrica*, **85** 1303–1320.

Davydov, Y. A., Lifshits, M. A. and Smorodina, N. V. (1998). *Local Properties of Distribuions of Stochastic Functionals*. American Mathematical Society, Providence.

Dette, H., Hoderlein, S. and Neumeyer, N. (2016). Testing multivariate economic restrictions using quantiles: the example of slutsky negative semidefiniteness. *Journal of Econometrics*, **191** 129–144.

Dumbgen, L. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. *Annals of Statistics* 124–152.

Eichenbaum, M. S., Hansen, L. P. and Singleton, K. J. (1988). A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *The Quarterly Journal of Economics*, **103** 51–78.

Fang, Z. and Seo, J. (2019). A general framework for inference on shape restrictions. *arXiv preprint arXiv:1910.07689*.

Freyberger, J. and Horowitz, J. L. (2015). Identification and shape restrictions in nonparametric instrumental variables estimation. *Journal of Econometrics*, **189** 41–53.

Freyberger, J. and Reeves, B. (2018). Inference under shape restrictions. *Available at SSRN 3011474*.

Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*, **97** 713–744.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the econometric society* 413–430.

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.

Hausman, J. A. and Newey, W. K. (2016). Individual heterogeneity and average welfare. *Econometrica*, **84** 1225–1248.

Ho, K. and Rosen, A. M. (2017). Partial identification in applied research: Benefits and challenges. In *Advances in Economics and Econometrics: Volume 2: Eleventh World Congress*, vol. 2. Cambridge University Press, 307.

Hong, S. (2017). Inference in semiparametric conditional moment models with partial identification. *Journal of econometrics*, **196** 156–179.

Horowitz, J. L. and Lee, S. (2017). Nonparametric estimation and inference under shape restrictions. *Journal of econometrics*, **201** 108–126.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** 467–475.

Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. **72** 1845–1857.

Jackwerth, J. C. (2000). Recovering risk aversion from option prices and realized returns. *Review of Financial Studies*, **13** 433–451.

Kaido, H., Molinari, F. and Stoye, J. (2019). Confidence intervals for projections of partially identified parameters. *Econometrica*, **87** 1397–1432.

Kaido, H. and Santos, A. (2014). Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities. *Econometrica*, **82** 387–413.

Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, **1** 23–41.

Kline, P. and Walters, C. (2021). Reasonable doubt: Experimental detection of job-level employment discrimination. *Econometrica*, **89** 765–792.

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.

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.

Lewbel, A. (1995). Consistent nonparametric hypothesis tests with an application to slutsky symmetry. *Journal of Econometrics*, **67** 379–401. URL https://www.sciencedirect.com/science/article/pii/030440769401637F.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer-Verlag, New York.

Matzkin, R. L. (1994). Restrictions of economic theory in nonparametric methods. In *Handbood of Econometrics* (R. Engle and D. McFadden, eds.), vol. IV. Elsevier.

Molinari, F. (2020). Microeconometrics with partial identification. *Handbook of econometrics*, **7** 355–486.

Newey, W. K. and Powell, J. (2003). Instrumental variables estimation of nonparametric models. *Econometrica*, **71** 1565–1578.

Reguant, M. (2014). Complementary bidding mechanisms and startup costs in electricity markets. *The Review of Economic Studies*, **81** 1708–1742.

Romano, J. P. and Shaikh, A. M. (2010). Inference for the identified set in partially identified econometric models. *Econometrica*, **78** 169–211.

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.

SAMWORTH, R. and SEN, B. (2018). Special issue on "nonparametric inference under shape constraints". *Statistical Science*.

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.

TORGOVITSKY, A. (2019). Nonparametric inference on state dependence in unemployment. *Econometrica*, **87** 1475–1505.

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.

ZHAI, A. (2018). A high-dimensional clt in w_2 distance with near optimal convergence rate. *Probability Theory and Related Fields*, **170** 821–845.

ZHU, Y. (2019). Inference in non-parametric/semi-parametric moment equality models with shape restrictions. *Semi-Parametric Moment Equality Models with Shape Restrictions (October 23, 2019)*.

Supplemental Appendix I

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.L.A.
andres@econ.ucla.edu

April, 2022

This Supplemental Appendix to "Constrained Conditional Moment Restriction Models" is organized as follows. Sections A.1 provides a review of AM spaces. Section A.2 specializes the general results of Section 3 to three additional examples: (i) GMM, (ii) Quantile Treatment Effects, and (iii) The Slutsky restriction in a partially linear model. The proofs for all results are included in Supplmental Appendix II.

*Research supported by NSF Grant 1757140.
†Research supported by NSF Grant SES-1426882.

A.1 AM Spaces

We provide a brief introduction to AM spaces and refer the reader to Chapters 8 and 9 of Aliprantis and Border (2006) for a more detailed exposition. Before proceeding, we first recall the definitions of a partially ordered set and a lattice.

Definition A.1.1. 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.1.2. 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 $x \in \mathbf{G}$ we define the positive part $x^+ \equiv x \vee 0$, the negative part $x^- \equiv (-x) \vee 0$, and the absolute value $|x| \equiv x \vee (-x)$. It is also natural to demand that the order relation $\geq$ interact with the algebraic operations in a manner analogous to that of $\mathbf{R}$ – i.e. to have

$$x \geq y \text{ implies } x + z \geq y + z \text{ for each } z \in \mathbf{G} \tag{A.1}$$

$$x \geq y \text{ implies } \alpha x \geq \alpha y \text{ for each } 0 \leq \alpha \in \mathbf{R}\ . \tag{A.2}$$

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.1.3. 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.1) and (A.2) hold. ■

An AM space is a Banach lattice in which the maximum of the norms of any two positive elements is equal to the norm of the maximums of the two elements.

Definition A.1.4. 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

$$\|x\|_\infty \equiv \inf\{\lambda > 0 : |x| \leq \lambda \mathbf{1_G}\}, \tag{A.3}$$

which is a norm on $\mathbf{G}$. In principle, $\|\cdot\|_\infty$ need not be related to the original norm $\|\cdot\|_{\mathbf{G}}$. However, if $\mathbf{G}$ is an AM space, then $\|\cdot\|_{\mathbf{G}}$ and $\|\cdot\|_\infty$ are equivalent in that they generate the same topology. Hence, we refer to $\mathbf{G}$ as an *AM space with unit* $\mathbf{1_G}$ if: (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 $\|\cdot\|_\infty$.

A.2 Illustrative Examples

In this Section, we examine special cases of our general analysis and illustrate both how to implement our procedure and verify the assumptions in the main text.

A.2.1 Generalized Method of Moments

Our first example concerns the generalized method of moments (GMM) model of Hansen (1982). We assume the parameter of interest θ_0 is identified as the unique solution to

$$E_P[\rho(X, \theta_0)] = 0, \tag{A.4}$$

where $X \in \mathbf{X}$ is distributed according to $P \in \mathbf{P}$ and $\rho : \mathbf{X} \times \Theta \to \mathbf{R}^{\mathcal{J}}$. This model maps into our general framework by letting $Z_\jmath = 1$ for all $1 \leq \jmath \leq \mathcal{J}$. Moreover, since we have assumed θ_0 is identified, the hypothesis testing problem simplifies to

$$H_0 : \theta_0 \in R \qquad H_1 : \theta_0 \notin R.$$

The set R is, as in the main text, defined by equality and inequality restrictions. In particular, for known functions $\Upsilon_F : \mathbf{R}^{d_\theta} \to \mathbf{R}^{d_F}$ and $\Upsilon_G : \mathbf{R}^{d_\theta} \to \mathbf{R}^{d_G}$ we set

$$R \equiv \{\theta \in \mathbf{R}^{d_\theta} : \Upsilon_F(\theta) = 0 \text{ and } \Upsilon_G(\theta) \leq 0\}. \tag{A.5}$$

To verify Assumptions 3.1(ii)(iii), note $\mathbf{R}^d$ is a Banach space under any norm $\|\cdot\|_p$ with $1 \leq p \leq \infty$, so for concreteness we set $\mathbf{B} = \mathbf{R}^{d_\theta}$, $\mathbf{F} = \mathbf{R}^{d_F}$, and $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{\mathbf{F}} = \|\cdot\|_2$. The space $\mathbf{R}^d$ is in addition a lattice under the standard pointwise partial order

$$a \leq b \text{ if and only if } a_i \leq b_i \text{ for all } 1 \leq i \leq d \tag{A.6}$$

for any $(a_1, \ldots, a_d)' = a$ and $(b_1, \ldots, b_d)' = b$ in $\mathbf{R}^d$, while the least upper bound equals

$$a \vee b = (\max\{a_1, b_1\}, \ldots, \max\{a_d, b_d\})'.$$

The vector $(1, \ldots, 1)'$ is an order unit in $\mathbf{R}^d$ under the partial order in (A.6). As discussed in Section A.1 of this Supplemental Appendix, the order unit induces the norm

$$\{\inf \lambda > 0 : |a| \leq \lambda(1, \ldots, 1)'\} = \max_{1 \leq i \leq d} |a_i|,$$

which corresponds to the usual $\|\cdot\|_\infty$ norm on $\mathbf{R}^d$. Hence, by setting $\mathbf{G} = \mathbf{R}^{d_G}$, $\|\cdot\|_{\mathbf{G}} = \|\cdot\|_\infty$, and $\mathbf{1_G} = (1, \ldots, 1)'$ we verify the requirements of Assumption 3.1(ii)(iii).

Since the parameter space Θ is finite dimensional and all moment restrictions are unconditional, we may set $\Theta_n = \Theta$ and $k_n = \mathcal{J}$ for all n. We base our test statistic on quadratic forms in the moments ($p = 2$), which implies $Q_n(\theta)$ is given by

$$Q_n(\theta) \equiv \|\hat{\Sigma}_n\{\frac{1}{n}\sum_{i=1}^{n} \rho(X_i, \theta)\}\|_2.$$

In what follows, we consider tests based on both the un-centered statistic $I_n(R)$ and the re-centered statistic $I_n(R) - I_n(\Theta)$. To this end, we impose the following:

Assumption A.2.1. *(i) $\{X_i\}_{i=1}^n$ is i.i.d. with $X_i \sim P \in \mathbf{P}$; (ii) For each $P \in \mathbf{P}_0$ there exists a unique $\theta_0 \in \Theta$ solving (A.4); (iii) Θ is convex and compact.*

Assumption A.2.2. *(i) The function $\rho(x, \cdot) : \Theta \to \mathbf{R}^{\mathcal{J}}$ is twice differentiable for all x; (ii) $E_P[\sup_{\theta\in\Theta} \|\rho(X, \theta)\|_2^3]$, $E_P[\sup_{\theta\in\Theta} \|\nabla_\theta \rho(X, \theta)\|_{o,2}^2]$, $E_P[\sup_{\theta\in\Theta} \|\nabla_\theta^2 \rho_\jmath(X, \theta)\|_{o,2}^{1+\delta}]$ are finite and bounded uniformly in $P \in \mathbf{P}$ for some $\delta > 0$.*

Assumption A.2.3. *(i) $\inf_{P\in\mathbf{P}_0} \inf_{\theta\in\Theta:\|\theta-\theta_0\|_2\geq\epsilon} \|E_P[\rho(X, \theta)]\|_2 > 0$ for all $\epsilon > 0$; (ii) The singular values of $E_P[\nabla_\theta \rho(X, \theta_0)]$ are bounded away from zero in $P \in \mathbf{P}_0$.*

Assumption A.2.4. *(i) $\|\hat{\Sigma}_n - \Sigma_P\|_{o,2} = O_P(n^{-1/2})$ uniformly in $P \in \mathbf{P}$; (ii) Σ_P is invertible and $\|\Sigma_P\|_{o,2}$ and $\|\Sigma_P^{-1}\|_{o,2}$ are bounded uniformly in $P \in \mathbf{P}$.*

In Assumption A.2.2 we focus on differentiable moments for simplicity. Assumption A.2.3 essentially imposes strong identification of θ_0 and hence guarantees that θ_0 can be consistently estimated uniformly in $P \in \mathbf{P}_0$ – recall that θ_0 depends on P through (A.4), though the dependence is left implicit in the notation. Finally, Assumption A.2.4 states the requirements on the $\mathcal{J} \times \mathcal{J}$ weighting matrix $\hat{\Sigma}_n$.

In what follows, we set the local parameter spaces $V_n(\theta, R|\ell)$ and $V_n(\theta, \Theta|\ell)$ to equal

$$V_n(\theta, R|\ell) = \{h \in \mathbf{R}^{d_\theta} : \theta + h/\sqrt{n} \in \Theta \cap R \text{ and } \|h/\sqrt{n}\|_2 \leq \ell\}$$
$$V_n(\theta, \Theta|\ell) = \{h \in \mathbf{R}^{d_\theta} : \theta + h/\sqrt{n} \in \Theta \text{ and } \|h/\sqrt{n}\|_2 \leq \ell\}.$$

Setting $\mathbb{D}_P(\theta_0)[h] \equiv E_P[\nabla_\theta \rho(X, \theta_0)]h$ and letting $\mathbb{W}_P(\theta_0) \sim N(0, \text{Var}_P\{\rho(X, \theta_0)\})$ we then denote the variables to which $I_n(R)$ and $I_n(\Theta)$ will be coupled to by

$$U_P(R|\ell_n) \equiv \inf_{h\in V_n(\theta_0,R|\ell_n)} \|\mathbb{W}_P(\theta_0) + \mathbb{D}_P(\theta_0)[h]\|_{\Sigma_P,2}$$
$$U_P(\Theta|\ell_n) \equiv \inf_{h\in V_n(\theta_0,\Theta|\ell_n)} \|\mathbb{W}_P(\theta_0) + \mathbb{D}_P(\theta_0)[h]\|_{\Sigma_P,2}.$$

Our distributional approximations follow immediately from Theorem 3.1(ii).

Theorem A.2.1. *Let Assumptions A.2.1, A.2.2, A.2.3, and A.2.4 hold, Υ_F and Υ_G be continuous, and set $a_n = \sqrt{\log(n)}/n^{\frac{1}{10+5d_\theta}}$. Then: For any $\ell_n, \ell_n^{\mathrm{u}} \downarrow 0$ satisfying*

$(\ell_n \vee \ell_n^{\mathrm{u}})\sqrt{\log(1/\ell_n \vee \ell_n^{\mathrm{u}})} = o(a_n)$ *and* $n^{-1/2} = o(\ell_n \vee \ell_n^{\mathrm{u}})$ *we have uniformly in* $P \in \mathbf{P}_0$

$$I_n(R) = U_P(R|\ell_n) + o_P(a_n)$$
$$I_n(R) - I_n(\Theta) = U_P(R|\ell_n) - U_P(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n).$$

The rate of coupling $a_n = \sqrt{\log(n)}/n^{\frac{1}{10+5d_\theta}}$ obtained in Theorem A.2.1 suffices for both the empirical process and bootstrap coupling; see Lemmas S.4.12 and S.4.13 in Supplemental Appendix II. While the rate is adequate for our purposes, it can be improved under additional moment restrictions. Here, we rely in Yurinskii (1977) both to illustrate the diversity of coupling arguments that can be employed to verify Assumption 3.3(i) and to impose only the weak third moment restriction of Assumption A.2.2(ii).

Our next goal is to obtain bootstrap approximations to the distributions of $U_P(R|\ell_n)$ and $U_P(\Theta|\ell_n^{\mathrm{u}})$. To this end, we write $\Upsilon_F(\theta) = (\Upsilon_{F,1}(\theta), \ldots, \Upsilon_{F,d_F}(\theta))'$ and $\Upsilon_G(\theta) = (\Upsilon_{G,1}(\theta), \ldots, \Upsilon_{G,d_G}(\theta))'$, for any $\epsilon > 0$ we define $B^\epsilon \equiv \bigcup_{P\in\mathbf{P}_0}\{\theta : \|\theta - \theta_0\|_2 \leq \epsilon\}$ (where recall θ_0 implicitly depends on P through (A.4)), and impose:

Assumption A.2.5. *For some $\epsilon > 0$: (i) $B^\epsilon \subseteq \Theta$; (ii) Υ_F and Υ_G are twice differentiable on B^ϵ; (iii) $\|\nabla\Upsilon_F(\theta)\|_{o,2}$ and $\|\nabla\Upsilon_G(\theta)\|_{o,2}$ are bounded on B^ϵ; (iv) $\|\nabla^2\Upsilon_{F,j}(\theta)\|_{o,2}$ is bounded on B^ϵ for $1 \leq j \leq d_F$; (v) $\|\nabla^2\Upsilon_{G,j}(\theta)\|_{o,2}$ is bounded on B^ϵ for $1 \leq j \leq d_G$; (vi) $\nabla\Upsilon_F(\theta)$ has full row-rank on B^ϵ.*

Assumption A.2.6. *Either (i) $\Upsilon_F : \mathbf{R}^{d_\theta} \to \mathbf{R}^{d_F}$ is affine, or (ii) There is an $\epsilon > 0$ and $M < \infty$ such that the singular values of $\nabla\Upsilon_F(\theta)'$ are bounded away from zero uniformly in $\theta \in B^\epsilon$, and for every $P \in \mathbf{P}_0$ there is an $h \in \mathcal{N}(\nabla\Upsilon_F(\theta_0))$ with $\|h\|_2 \leq M$ satisfying $\Upsilon_{G,j}(\theta_0) + \nabla\Upsilon_{G,j}(\theta_0)[h] \leq -\epsilon$ for all $1 \leq j \leq d_G$.*

In order to describe our bootstrap procedure in this application, we let $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$ denote the minimizers of Q_n over $\Theta \cap R$ and Θ respectively. Employing $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$ we obtain estimators for the distribution of $\mathbb{W}_P(\theta_0)$ and for $\mathbb{D}_P(\theta_0)$ by evaluating

$$\hat{\mathbb{W}}_n(\theta) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_i\{\rho(X_i,\theta) - \frac{1}{n}\sum_{j=1}^n \rho(X_j,\theta)\} \tag{A.7}$$
$$\hat{\mathbb{D}}_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^n \nabla_\theta\rho(X_i,\theta), \tag{A.8}$$

at $\theta = \hat{\theta}_n$ and $\theta = \hat{\theta}_n^{\mathrm{u}}$, where recall $\{\omega_i\}_{i=1}^n$ is an i.i.d. sample independent of $\{X_i\}_{i=1}^n$ with $\omega_i \sim N(0,1)$. We note that because moments are differentiable, we employ an analytical derivative in (A.8) instead of the numerical derivative studied in Section 3.

With regards to the local parameter space, we note that the construction of $\hat{V}_n(\theta, R|\ell)$ requires the bound K_g on the second derivative of Υ_G (as specified in Assumption 3.8).

In particular, Assumption A.2.5(v) implies Assumption 3.8 is satisfied with

$$K_g \equiv \max_{1 \leq j \leq d_G} \sup_{\theta \in B^\epsilon} \|\nabla_\theta^2 \Upsilon_{G,j}(\theta)\|_{o,2}$$

(see Lemma S.4.14). If an a-priory bound on the second derivative is not available, then it is also possible to simply substitute K_g with the data driven choice

$$\hat{K}_g \equiv \max_{1 \leq j \leq d_G} \sup_{\theta \in \Theta : \|\theta - \hat{\theta}_n\|_2 \leq r_n} \|\nabla_\theta^2 \Upsilon_{G,j}(\theta)\|_{o,2},$$

where we discuss the choice of r_n below. Given K_g (or $\hat{K}_g$), we set $G_n(\theta)$ to equal

$$G_n(\theta) = \{h \in \mathbf{R}^{d_\theta} : \Upsilon_{j,G}(\theta + \frac{h}{\sqrt{n}}) \leq \max\{\Upsilon_{j,G}(\theta) - K_g r_n \|\frac{h}{\sqrt{n}}\|_2, -r_n\} \text{ for all } j\}$$

In this application we may additionally specify ℓ_n to be infinite, and hence we set

$$\hat{V}_n(\theta, R| + \infty) = \{h \in \mathbf{R}^{d_\theta} : h \in G_n(\theta) \text{ and } \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0\}.$$

The approximations to the distributions of $I_n(R)$ and $I_n(\Theta)$ are then given by the laws of $\hat{U}_n(R| + \infty)$ and $\hat{U}_n(\Theta| + \infty)$ conditional on the data, where

$$\hat{U}_n(R| + \infty) \equiv \inf_{h \in \hat{V}_n(\hat{\theta}_n, R| + \infty)} \|\hat{\mathbb{W}}_n(\hat{\theta}_n) + \hat{\mathbb{D}}_n(\hat{\theta}_n)[h]\|_{\hat{\Sigma}_n,2}$$
$$\hat{U}_n(\Theta| + \infty) \equiv \inf_{h \in \mathbf{R}^{d_\theta}} \|\hat{\mathbb{W}}_n(\hat{\theta}_n^{\mathrm{u}}) + \hat{\mathbb{D}}_n(\hat{\theta}_n^{\mathrm{u}})[h]\|_{\hat{\Sigma}_n,2}.$$

The validity of these distributional approximations follows from Theorem 3.2.

Theorem A.2.2. *Let Assumptions A.2.1, A.2.2, A.2.3, A.2.4, A.2.5, and A.2.6 hold, set* $a_n = \sqrt{\log(n)}/n^{\frac{1}{10+5d_\theta}}$, *and let* $n^{-1/2} = o(r_n)$. *Then: there are sequences* $\ell_n, \ell_n^{\mathrm{u}} \downarrow 0$ *satisfying* $(\ell_n \vee \ell_n^{\mathrm{u}})^2 \sqrt{\log(1/(\ell_n \vee \ell_n^{\mathrm{u}}))} = o(a_n n^{-\frac{1}{2}})$, $\ell_n = o(r_n)$, *and* $n^{-\frac{1}{2}} = o(\ell_n \wedge \ell_n^{\mathrm{u}})$ *for which it follows uniformly in* $P \in \mathbf{P}_0$ *that*

$$\hat{U}_n(R| + \infty) \geq U_P^\star(R|\ell_n) + o_P(a_n)$$
$$\hat{U}_n(R| + \infty) - \hat{U}_n(\Theta| + \infty) \geq U_P^\star(R|\ell_n) - U_P^\star(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n).$$

Crucially, note that any sequences ℓ_n and ℓ_n^{u} satisfying the conditions of Theorem A.2.2 also satisfy the conditions of Theorem A.2.1. Therefore, Theorems A.2.2 and A.2.1 together establish the validity of employing the laws of $\hat{U}_n(R| + \infty)$ and $\hat{U}_n(\Theta| + \infty)$ conditional on the data to approximate the laws of $I_n(R)$ and $I_n(\Theta)$. In particular, for a level α test we may compare the test statistic $I_n(R)$ to the critical value

$$\hat{q}_{1-\alpha}(\hat{U}_n(R| + \infty)) \equiv \inf\{c : P(\hat{U}_n(R| + \infty) \leq c | \{X_i\}_{i=1}^n) \geq 1 - \alpha\}.$$

Similarly, for the re-centered statistic $I_n(R) - I_n(\Theta)$, valid critical values are given by:

$$\hat{q}_{1-\alpha}(\hat{U}_n(R|+\infty) - \hat{U}_n(\Theta|+\infty))$$
$$\equiv \inf\{c : P(\hat{U}_n(R|+\infty) - \hat{U}_n(\Theta|+\infty) \leq c|\{X_i\}_{i=1}^n) \geq 1-\alpha\}.$$

These approximations are valid under the requirement that r_n satisfy $r_n\sqrt{n} \to \infty$. Intuitively, the bandwidth r_n is meant to reflect a bound on the distance between $\hat{\theta}_n$ and θ_0. For a data driven choice of r_n we may therefore employ a bootstrap estimate of an upper quantile of the distribution of the *unconstrained* estimator. Specifically, for $\hat{\theta}_n^{\mathrm{u}\star}$ the bootstrapped version of $\hat{\theta}_n^{\mathrm{u}}$, we may set $\hat{r}_n$ to be given by

$$\hat{r}_n \equiv \inf\{c : P(\|\hat{\theta}_n^{\mathrm{u}\star} - \hat{\theta}_n^{\mathrm{u}}\|_2 \leq c|\{X_i\}_{i=1}^n)\} \geq 1-\gamma_n$$

for $\gamma_n \to 0$ as the sample size n tends to infinity, and employ $\hat{r}_n$ in place of r_n.

A.2.2 Consumer Demand

We base our next example on a long-standing literature aiming to replace parametric assumptions with shape restrictions implied by economic theory (Matzkin, 1994). Specifically, suppose that quantity demanded by individual i, denoted Q_i, satisfies

$$Q_i = g_0(S_i, Y_i) + W_i'\gamma_0 + U_i,$$

where $S_i \in \mathbf{R}_+$ denotes price, $Y_i \in \mathbf{R}_+$ denotes income, and $W_i \in \mathbf{R}^{d_w}$ is a set of covariates. In addition, we assume there is an instrument Z_i yielding the restriction

$$E_P[Q - g_0(S,Y) - W'\gamma_0|Z] = 0. \tag{A.9}$$

For instance, under exogeneity of prices we may let $Z = (S, Y, W')'$ as in Blundell et al. (2012). Alternatively, if there is a concern that prices are endogenous, then we may set $Z = (I, Y, W')'$ for I an instrument for S, as in Blundell et al. (2017).

Our goal is to conduct inference on the level of demand at particular price income pair (s_0, y_0) while imposing that the function g_0 satisfies the Slutsky restriction

$$\frac{\partial}{\partial s}g_0(s,y) + g_0(s,y)\frac{\partial}{\partial y}g_0(s,y) \leq 0. \tag{A.10}$$

To map this problem into our framework, we assume that for some set Ω, $(S,Y) \in \Omega \subseteq \mathbf{R}_+^2$ with probability one for all $P \in \mathbf{P}$ and impose that $g_0 \in C_B^1(\Omega)$, where

$$C_B^m(\Omega) \equiv \{g : \Omega \to \mathbf{R} \text{ s.t. } \|g\|_{m,\infty} < \infty\} \qquad \|g\|_{m,\infty} \equiv \sup_{0\leq\alpha\leq m}\sup_{(s,y)\in\Omega}|\nabla^\alpha g(s,y)|.$$

Since $\theta_0 \equiv (g_0, \gamma_0)$ with $\gamma_0 \in \mathbf{R}^{d_w}$, we set $\mathbf{B} = C^1_B(\Omega) \times \mathbf{R}^{d_w}$ and for any $(g, \gamma) = \theta \in \mathbf{B}$ let $\|\theta\|_{\mathbf{B}} = \max\{\|g\|_{1,\infty}, \|\gamma\|_2\}$. We also note that $X = (Q, S, Y, W)$ and

$$\rho(X, \theta) = Q - g(S, Y) - W'\gamma. \tag{A.11}$$

We will assume $\theta_0 \equiv (g_0, \gamma_0)$ is identified by (A.9). Hence, we can think of θ_0 as a function of P through (A.9), though we leave such dependence implicit in the notation.

In order to impose the Slutsky restriction in (A.10) we let $\mathbf{G} = C^0_B(\Omega)$ and $\|\cdot\|_{\mathbf{G}} = \|\cdot\|_\infty$, where with some abuse of notation we write $\|\cdot\|_\infty$ in place of $\|\cdot\|_{0,\infty}$. The space $C^0_B(\Omega)$ is a Banach lattice under the standard pointwise ordering given by

$$a \leq b \text{ if and only if } a(s, y) \leq b(s, y) \text{ for all } (s, y) \in \Omega \tag{A.12}$$

for any $a, b \in C^0_B(\Omega)$. The constant function $\mathbf{c} \in C^0_B(\Omega)$ satisfying $\mathbf{c}(s, y) = 1$ for all $(s, y) \in \Omega$ is an order unit under the partial ordering in (A.12). Its induced norm is

$$\{\inf \lambda > 0 : |a| \leq \lambda \mathbf{c}\} = \sup_{(s,y)\in\Omega} |a(s, y)|,$$

which coincides with the norm $\|\cdot\|_\infty$ on $C^0_B(\Omega)$, and we therefore set $\mathbf{1}_{\mathbf{G}} = \mathbf{c}$. To encode the Slutsky restriction in (A.10) we then let the map $\Upsilon_G : \mathbf{B} \to \mathbf{G}$ equal

$$\Upsilon_G(\theta)(s, y) = \frac{\partial}{\partial s} g(s, y) + g(s, y) \frac{\partial}{\partial y} g(s, y) \tag{A.13}$$

for any $\theta = (g, \gamma) \in \mathbf{B}$. Finally, to test whether the level of demand at a prescribed price s_0 and income y_0 equals a hypothesized value c_0, we set $\mathbf{F} = \mathbf{R}$, $\|\cdot\|_{\mathbf{F}} = |\cdot|$, and

$$\Upsilon_F(\theta) = g(s_0, y_0) - c_0 \tag{A.14}$$

for any $\theta = (g, \gamma) \in \mathbf{B}$. By setting $R = \{\theta \in \mathbf{B} : \Upsilon_G(\theta) \leq 0 \text{ and } \Upsilon_F(\theta) = 0\}$ and conducting test inversion (over different values of c_0) of the null hypothesis

$$H_0 : \theta_0 \in R \qquad\qquad H_1 : \theta_0 \notin R$$

we may obtain a confidence region for the level of demand at price s_0 and income y_0.

We set the parameter space to be a ball in $\mathbf{B}$ under $\|\cdot\|_{\mathbf{B}}$ by letting Θ be equal to

$$\Theta \equiv \{(g, \gamma) \in C^1_B(\Omega) \times \mathbf{R}^{d_w} : \|g\|_{1,\infty} \leq C_0 \text{ and } \|\gamma\|_2 \leq C_0\} \tag{A.15}$$

for some $C_0 < \infty$. Given a sequence of approximating functions $\{p_j\}_{j=1}^{j_n}$, we then let $p^{j_n}(s, y) \equiv (p_1(s, y), \ldots, p_{j_n}(s, y))'$ and set the sieve Θ_n to equal

$$\Theta_n \equiv \{(p^{j_n\prime}\beta, \gamma) : \|p^{j_n\prime}\beta\|_{1,\infty} \leq C_0 \text{ and } \|\gamma\|_2 \leq C_0\}.$$

Similarly, for a sequence $\{q_k\}_{k=1}^{k_n}$ of transformations of the conditioning variable Z, we let $q^{k_n}(z) \equiv (q_1(z), \ldots, q_{k_n}(z))'$. We base our test statistic on the quadratic forms

$$Q_n(\theta) \equiv \|\hat{\Sigma}_n\{\frac{1}{n}\sum_{i=1}^{n}(Q_i - g(S_i, Y_i) - W_i'\gamma)q^{k_n}(Z_i)\}\|_2$$

for some $k_n \times k_n$ weighting matrix $\hat{\Sigma}_n$ and every $(g, \gamma) = \theta \in \Theta$. The statistics $I_n(R)$ and $I_n(\Theta)$ simply equal the minimums of $\sqrt{n}Q_n(\theta)$ over $\Theta_n \cap R$ and Θ_n respectively.

The next assumptions suffice for obtaining a strong approximation. In their statement, the notation $\underline{\text{sing}}\{A\}$ denotes the smallest singular value of a matrix A.

Assumption A.2.7. *(i) $\{X_i, Z_i\}_{i=1}^n$ is i.i.d. with (X, Z) distributed according to $P \in \mathbf{P}$; (ii) For Θ as in (A.15) and each $P \in \mathbf{P}_0$ there exists a unique $\theta_0 \in \Theta$ satisfying $E_P[\rho(X, \theta_0)|Z] = 0$; (iii) The support of (Q, W) is bounded uniformly in $P \in \mathbf{P}$.*

Assumption A.2.8. *(i) $\sup_{(s,y)} \|p^{j_n}(s, y)\|_2 \lesssim \sqrt{j_n}$; (ii) $\sup_{(s,y)} \|\partial_a p^{j_n}(s, y)\|_2 \lesssim j_n^{3/2}$ for $a \in \{s, y\}$; (iii) The eigenvalues of $E_P[p^{j_n}(S, Y)p^{j_n}(S, Y)']$ are bounded away from zero and infinity uniformly in $P \in \mathbf{P}$ and j_n; (iv) For each $P \in \mathbf{P}_0$ there is a $\Pi_n\theta_0 = (g_n, \gamma_0) \in \Theta_n \cap R$ with $\sup_{P\in\mathbf{P}_0} \|E_P[(g_0(S, Y) - g_n(S, Y))q^{k_n}(Z)\|_2 = o((n\log(n))^{-1/2})$.*

Assumption A.2.9. *(i) $\max_{1\le k\le k_n} \|q_k\|_\infty \lesssim \sqrt{k_n}$; (ii) $E_P[q^{k_n}(Z)q^{k_n}(Z)']$ has eigenvalues bounded uniformly in $P \in \mathbf{P}$, k_n; (iii) $\mathrm{s}_n \equiv \inf_{P\in\mathbf{P}} \underline{\text{sing}}\{E_P[q^{k_n}(Z)(p^{j_n}(S, Y)'\ W')]\}$ satisfies $0 < \mathrm{s}_n = O(1)$; (iv) $j_n^2 k_n^3 \log^3(n) = o(n)$ and $k_n^2 j_n \log^{3/2}(1 + k_n)/(\mathrm{s}_n\sqrt{n})(1 \vee \sqrt{\log(\mathrm{s}_n\sqrt{n}/k_n)}) = o((\log(n))^{-1/2})$.*

Assumption A.2.10. *(i) $\|\hat{\Sigma}_n - \Sigma_P\|_{o,2} = o_P((k_n\sqrt{j_n}\log^{3/2}(n))^{-1})$ uniformly in $P \in \mathbf{P}$; (ii) Σ_P is invertible and $\|\Sigma_P\|_{o,2}$ and $\|\Sigma_P^{-1}\|_{o,2}$ are bounded in $P \in \mathbf{P}$ and k_n.*

Assumption A.2.7(iii) requires (Q, W) to be bounded, which enables us to apply the recent coupling results by Zhai (2018). Alternatively, Assumption A.2.7(iii) can be relaxed under appropriate tail conditions. Assumptions A.2.8(i)-(iii) are standard requirements on Θ_n that can be satisfied by, e.g., tensor product wavelets or B-splines (Newey, 1997; Chen, 2007; Belloni et al., 2015; Chen and Christensen, 2018). Assumption A.2.8(iv) pertains the approximating requirements on the sieve; see Remarks A.2.1 and A.2.2 below. In turn, Assumption A.2.9(i)(ii) imposes standard requirements on $\{q_k\}_{k=1}^{k_n}$. Assumption A.2.9(iii)(iv) contains the required rate conditions, which are governed by s_n – a parameter that is proportional to ν_n^{-1} (as in Assumption 3.4) and is closely linked the degree of ill-posedness; see Remark A.2.2 below. Finally, Assumption A.2.10 states the conditions on the weighting matrix $\hat{\Sigma}_n$.

In this application, we may set $\|\theta\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}} \|g\|_{P,2} + \|\gamma\|_2$ for any $(g, \gamma) \in \Theta$.

Since in addition any $\theta = (g, \gamma) \in \Theta_n \cap R$ has the structure $g = p^{j_n \prime}\beta$, we have

$$V_n(\theta, R|\ell) = \Big\{(p^{j_n \prime}\beta_h, \gamma_h) : \ \|g + \frac{p^{j_n \prime}\beta_h}{\sqrt{n}}\|_{1,\infty} \leq C_0 \text{ and } \|\gamma + \frac{\gamma_h}{\sqrt{n}}\|_2 \leq C_0 \tag{A.16}$$

$$p^{j_n}(s_0, y_0)'\beta_h = 0 \tag{A.17}$$

$$\frac{\partial}{\partial s}(g + \frac{p^{j_n \prime}\beta_h}{\sqrt{n}}) + (g + \frac{p^{j_n \prime}\beta_h}{\sqrt{n}})\frac{\partial}{\partial y}(g + \frac{p^{j_n \prime}\beta_h}{\sqrt{n}}) \leq 0 \tag{A.18}$$

$$\sup_{P \in \mathbf{P}} \|p^{j_n \prime}\beta_h\|_{P,2} + \|\gamma_h\|_2 \leq \ell\sqrt{n}\Big\}, \tag{A.19}$$

where constraint (A.16) corresponds to $(\theta + h/\sqrt{n}) \in \Theta_n$, constraints (A.17) and (A.18) impose $\theta + h/\sqrt{n} \in R$, and constraint (A.19) imposes $\|h/\sqrt{n}\|_{\mathbf{E}} \leq \ell$. Similarly,

$$V_n(\theta, \Theta|\ell) = \Big\{(p^{j_n \prime}\beta_h, \gamma_h) : \ \|g + \frac{p^{j_n \prime}\beta_h}{\sqrt{n}}\|_{1,\infty} \leq C_0 \text{ and } \|\gamma + \frac{\gamma_h}{\sqrt{n}}\|_2 \leq C_0 \tag{A.20}$$

$$\sup_{P \in \mathbf{P}} \|p^{j_n \prime}\beta_h\|_{P,2} + \|\gamma_h\|_2 \leq \ell\sqrt{n}\Big\}. \tag{A.21}$$

Finally, recall that $\mathbb{W}_P(\theta) \sim N(0, \text{Var}_P\{\rho(X,\theta)q^{k_n}(Z)\})$ and define $\mathbb{D}_P$ to be given by

$$\mathbb{D}_P[h] \equiv -E_P[q^{k_n}(Z)(p^{j_n}(S,Y)'\beta_h + W'\gamma_h)]$$

for any $h = (p^{j_n \prime}\beta_h, \gamma_h)$. Given these definitions, note that for any ℓ_n we have that

$$U_P(R|\ell_n) \equiv \inf_{h \in V_n(\Pi_n\theta_0, R|\ell_n)} \|\mathbb{W}_P(\Pi_n\theta_0) + \mathbb{D}_P[h]\|_{\Sigma_P,2}$$
$$U_P(\Theta|\ell_n) \equiv \inf_{h \in V_n(\Pi_n\theta_0, \Theta|\ell_n)} \|\mathbb{W}_P(\Pi_n\theta_0) + \mathbb{D}_P[h]\|_{\Sigma_P,2}.$$

Theorem 3.1(ii) immediately yields the following distributional approximations.

Theorem A.2.3. *Let Assumptions A.2.7-A.2.10 hold, and* $a_n = (\log(n))^{-1/2}$*. Then: for any* $\ell_n, \ell_n^{\text{u}} \downarrow 0$ *satisfying* $k_n\sqrt{j_n \log(1+k_n)}(\ell_n \vee \ell_n^{\text{u}})\sqrt{\log(\sqrt{j_n}/(\ell_n \vee \ell_n^{\text{u}}))} = o(a_n)$ *and* $k_n\sqrt{j_n}\log(1+k_n)/\mathrm{s}_n\sqrt{n} = o(\ell_n \wedge \ell_n^{\text{u}})$ *it follows uniformly in* $P \in \mathbf{P}_0$ *that*

$$I_n(R) = U_P(R|\ell_n) + o_P(a_n)$$
$$I_n(R) - I_n(\Theta) = U_P(R|\ell_n) - U_P(\Theta|\ell_n^{\text{u}}) + o_P(a_n).$$

To obtain bootstrap estimates of the distributional approximations in Theorem A.2.3 we let $\hat{\theta}_n$ and $\hat{\theta}_n^{\text{u}}$ denote the minimizers of Q_n over $\Theta_n \cap R$ and Θ_n respectively. For $\rho(\cdot, \theta)$ as in (A.11), we approximate the law of $\mathbb{W}_P(\Pi_n\theta_0)$ by evaluating

$$\hat{\mathbb{W}}_n(\theta) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_i\{q^{k_n}(Z_i)\rho(X_i,\theta) - \frac{1}{n}\sum_{j=1}^n q^{k_n}(Z_j)\rho(X_j,\theta)\},$$

at $\theta = \hat{\theta}_n$ and $\theta = \hat{\theta}_n^{\mathrm{u}}$, where $\{\omega_i\}_{i=1}^n$ is an i.i.d. sample independent of the data satisfying $\omega_i \sim N(0,1)$. As our estimator for $\mathbb{D}_P[h]$, for any $h = (p^{j_n\prime}\beta_h, \gamma_h)$, we let

$$\hat{\mathbb{D}}_n[h] = -\frac{1}{n}\sum_{i=1}^n q^{k_n}(Z_i)(W_i'\gamma_h + p^{j_n}(S_i, Y_i)'\beta_h).$$

With regards to the local parameter space, we note that in this application Assumptions 3.8(i)(ii) are satisfied with $K_g = 2$ (see Lemma S.4.20). Therefore, we have

$$\begin{aligned} G_n(\hat{\theta}_n) = \Big\{h : \frac{\partial}{\partial s}p^{j_n}(s,y)'(\hat{\beta}_n + \frac{\beta_h}{\sqrt{n}}) + p^{j_n}(s,y)'(\hat{\beta}_n + \frac{\beta_h}{\sqrt{n}})\frac{\partial}{\partial y}p^{j_n}(s,y)'(\hat{\beta}_n + \frac{\beta_h}{\sqrt{n}}) \\ \leq \max\{\frac{\partial}{\partial s}p^{j_n}(s,y)'\hat{\beta}_n + p^{j_n}(s,y)'\hat{\beta}_n\frac{\partial}{\partial y}p^{j_n}(s,y)'\hat{\beta}_n - 2r_n\|\frac{p^{j_n\prime}\beta_h}{\sqrt{n}}\|_{1,\infty}, -r_n\}\Big\}. \quad \text{(A.22)} \end{aligned}$$

Moreover, because $\rho(X,\cdot)$ and Υ_F are linear, we may set $\ell_n = +\infty$ and obtain that

$$\hat{V}_n(\hat{\theta}_n, R| + \infty) = \{h = (p^{j_n\prime}\beta_h, \gamma_h) : h \in G_n(\hat{\theta}_n) \text{ and } p^{j_n}(s_0, y_0)'\beta_h = 0\}.$$

Given the introduced notation, we define the statistics $\hat{U}_n(R|+\infty)$ and $\hat{U}_n(\Theta|+\infty)$ by

$$\hat{U}_n(R|+\infty) \equiv \inf_{h\in\hat{V}_n(\hat{\theta}_n, R|+\infty)} \|\hat{\mathbb{W}}_n(\hat{\theta}_n) + \hat{\mathbb{D}}_n[h]\|_{\hat{\Sigma}_n,2}$$
$$\hat{U}_n(\Theta|+\infty) \equiv \inf_{h=(p^{j_n\prime}\beta_h,\gamma_h)} \|\hat{\mathbb{W}}_n(\hat{\theta}_n^{\mathrm{u}}) + \hat{\mathbb{D}}_n[h]\|_{\hat{\Sigma}_n,2}.$$

We impose one final assumption to establish the validity of the bootstrap. In the requirements below, it is helpful to recall θ_0 is implicitly a function of P through (A.9).

Assumption A.2.11. *(i) There is an $\epsilon > 0$ such that $\|g_0\|_{1,\infty} \vee \|\gamma_0\|_2 \leq C_0 - \epsilon$ for all $P \in \mathbf{P}_0$; (ii) $\Pi_n\theta_0 = (g_n, \gamma_0) \in \Theta_n \cap R$ satisfies $\|g_n - g_0\|_{1,\infty} = o(1)$ uniformly in $P \in \mathbf{P}_0$; (iii) The sequence $r_n \downarrow 0$ satisfies $k_n j_n^2\sqrt{\log(1+k_n)}/\mathrm{s}_n\sqrt{n} = o(r_n/\sqrt{\log(n)})$; (iv) $k_n j_n^{3/4}(\mathcal{E}_n \vee \sqrt{\log(k_n)})\log^{1/4}(1+k_n) = o(n^{1/4}/\sqrt{\log(n)})$, where $\mathcal{E}_n \equiv \int_0^\infty \sqrt{\log(\epsilon, \mathcal{C}_n, \|\cdot\|_2)}d\epsilon$ and $\mathcal{C}_n \equiv \{\beta : \|p^{j_n\prime}\beta\|_{1,\infty} \leq C_0\}$.*

Assumptions A.2.11(i)(ii) suffice for verifying Assumption 3.12(ii). These requirements may be dropped at the expense of modifying $\hat{V}_n(\hat{\theta}_n, R|+\infty)$ to reflect the possible impact of $\Pi_n\theta_0$ being "near" the boundary of Θ_n. Assumption A.2.11(iii) imposes the rate conditions on r_n. Finally, Assumption A.2.11(iv) controls the "size" of the set of coefficients β corresponding to elements $p^{j_n\prime}\beta \in \Theta_n$ and suffices for verifying the bootstrap coupling requirement of Assumption 3.11. For instance, $\mathcal{E}_n \asymp j_n^{1/4}$ for tensor product B-splines (see Lemma S.4.23), which implies a sufficient condition for Assumption A.2.11(iv) is that $k_n^4 j_n^4 \log^4(k_n) = o(n/\log^2(n))$. The rate requirements for a bootstrap coupling can be weakened if the test statistic is based on the $\|\cdot\|_\infty$-norm (see Lemma S.4.19) or under additional smoothness assumptions (see Theorem S.7.1(ii)).

Our next result characterizes the properties of the proposed bootstrap statistics.

Theorem A.2.4. *Let Assumptions A.2.7, A.2.8, A.2.9, A.2.10, A.2.11 hold, and* $a_n = (\log(n))^{-1/2}$*. Then: there are sequences* $\ell_n, \ell_n^{\mathrm{u}} \downarrow 0$ *satisfying* $k_n j_n^2 \log(1+k_n)/\mathrm{s}_n\sqrt{n} = o(\ell_n \wedge \ell_n^{\mathrm{u}})$*,* $\ell_n = o(r_n)$*, and* $k_n\sqrt{j_n \log(1+k_n)}(\ell_n \vee \ell_n^{\mathrm{u}})\sqrt{\log(\sqrt{j_n}/(\ell_n \vee \ell_n^{\mathrm{u}}))} = o(a_n)$ *for which it follows that uniformly in* $P \in \mathbf{P}_0$ *we have*

$$\hat{U}_n(R|+\infty) \geq U_P^\star(R|\ell_n) + o_P(a_n)$$
$$\hat{U}_n(R|+\infty) - \hat{U}_n(\Theta|+\infty) \geq U_P^\star(R|\ell_n) - U_P^\star(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n).$$

Importantly, any sequences ℓ_n and ℓ_n^{u} satisfying the requirements of Theorem A.2.4 also satisfy the requirements of Theorem A.2.3. Hence, we may employ

$$\hat{q}_{1-\alpha}(\hat{U}_n(R|+\infty)) \equiv \inf\{c : P(\hat{U}_n(R|+\infty) \leq c|\{V_i\}_{i=1}^n) \geq 1-\alpha\}$$

as a critical value for $I_n(R)$. Similarly, for the statistic $I_n(R) - I_n(\Theta)$ we may employ

$$\hat{q}_{1-\alpha}(\hat{U}_n(R|+\infty) - \hat{U}_n(\Theta|+\infty))$$
$$\equiv \inf\{c : P(\hat{U}_n(R|+\infty) - \hat{U}_n(\Theta|+\infty) \leq c|\{V_i\}_{i=1}^n) \geq 1-\alpha\}.$$

Remark A.2.1. Suppose for notational simplicity that there are no covariates W and let the marginal distribution of (S, Y, Z) be constant in $P \in \mathbf{P}$. If $Z = (S, Y)$ (i.e. (S, Y) is exogenous), we may set $q^{k_n}(Z) = p^{k_n}(S, Y)'$ for some $k_n \geq j_n$. The singular value s_n can then be assumed to be bounded away from zero, and a sufficient condition for Assumption A.2.9(iv) is that $k_n^4 j_n^2 \log^5(n) = o(n)$. In order to appreciate the content of Assumption A.2.8(iv), suppose $\{p_j\}_{j=1}^\infty$ is an orthonormal basis such that

$$g_0 = \sum_{j=1}^\infty \beta_j p_j \text{ with } |\beta_j| = O(j^{-\gamma_\beta}).$$

Setting $\Pi_n^{\mathrm{u}} g_0 = \sum_{j=1}^{j_n} p_j \beta_j$, we obtain from a standard integral bound for a sum that

$$\|E_P[(g_0(S,Y) - \Pi_n^{\mathrm{u}} g_0(S,Y))q^{k_n}(Z)]\|_2^2 \lesssim \sum_{j=j_n+1}^{k_n} \frac{1}{j^{2\gamma_\beta}} \lesssim \frac{1}{j_n^{2\gamma_\beta-1}} - \frac{1}{k_n^{2\gamma_\beta-1}}. \tag{A.23}$$

For instance, if $k_n - j_n = O(1)$, then the bound in (A.23) is of order $1/j_n^{2\gamma_\beta}$. Hence, provided the approximation error by $\Pi_n^{\mathrm{u}} g_0$ and g_n (as in Assumption A.2.8(iv)) are of the same order when $g_0 \in R$, we obtain that Assumption A.2.8(iv) is equivalent to $\sqrt{n\log(n)}/j_n^{\gamma_\beta} = o(1)$ when $k_n - j_n = O(1)$. This approximation requirement is compatible with the condition $k_n^4 j_n^2 \log^5(n) = o(n)$ provided $\gamma_\beta > 3$. ■

Remark A.2.2. Building on Remark A.2.1, suppose again there are no covariates W

and the marginal distribution of (S, Y, Z) is constant in $P \in \mathbf{P}$, but now let (S, Y) be endogenous. A standard benchmark for nonparametric models with endogeneity is to assume the operator $g \mapsto E_P[g(S, Y)|Z]$ is compact, in which case there are orthonormal sequences of functions $\{\phi_j\}_{j=1}^{\infty}$ of (S, Y) and $\{\psi_j\}_{j=1}^{\infty}$ of Z satisfying

$$E_P[\phi_j(S, Y)|Z] = \lambda_j \psi_j(Z) \qquad E_P[\psi_j(Z)|S, Y] = \lambda_j \phi_j(S, Y)$$

where $\lambda_j > 0$ tends to zero. In addition suppose g_0 admits for an expansion satisfying

$$g_0 = \sum_{j=1}^{\infty} \beta_j \phi_j \text{ with } |\beta_j| = O(j^{-\gamma_\beta}),$$

and let $p^{j_n} = (\phi_1, \ldots, \phi_{j_n})'$, $q^{k_n} = (\psi_1, \ldots, \psi_{k_n})'$ with $k_n \geq j_n$ and $k_n - j_n = O(1)$, and set $\Pi_n^{\mathrm{u}} g_0 = \sum_{j=1}^{j_n} \phi_j \beta_j$. Provided the approximation error of $\Pi_n^{\mathrm{u}} g_0$ and g_n (as in Assumption A.2.8(iv)) are of the same order when $g_0 \in R$, we then obtain

$$\|E_P[(g_0(S, Y) - g_n(S, Y))q^{k_n}(Z)]\|_2 \lesssim \frac{\lambda_{j_n}}{j_n^{\gamma_\beta}}.$$

Moreover, direct calculation shows s_n, which is proportional to ν_n^{-1} as in Assumption 3.4, satisfies $\mathrm{s}_n = \lambda_{j_n}$ and hence equals the reciprocal of the sieve measure of ill-posedness (Blundell et al., 2007). It follows that if $\lambda_j \asymp j^{-\gamma_\lambda}$, and $\gamma_\beta > 3$, then Assumptions A.2.8(iv) and A.2.9(iv) can be satisfied by setting $j_n \asymp n^{\kappa}$ with $(\gamma_\lambda + \gamma_\beta)^{-1} < 2\kappa < (3 + \gamma_\lambda)^{-1}$ and $k_n - j_n = O(1)$. Alternatively, if $\lambda_j = \exp\{-\gamma_\lambda j\}$, then Assumption A.2.8(iv) and A.2.9(iv) can be satisfied when $\gamma_\beta > 4$ by setting, for example, $j_n = (\log(n) - \kappa \log(\log(n)))/2\gamma_\lambda$ with $7 < \kappa < 2\gamma_\beta - 1$ and $k_n - j_n = O(1)$. ■

A.2.3 Quantile Treatment Effects

For our next example, we study a nonparametric quantile treatment effect (QTE) model. Specifically, for an outcome $Y \in \mathbf{R}$, treatment $D \in [0, 1]$, instrument $Z \in \mathbf{R}$, and quantile $\tau \in (0, 1)$, we assume the parameter of interest θ_0 satisfies

$$P(Y \leq \theta_0(D)|Z) = \tau. \tag{A.24}$$

If D is randomly assigned, then we may set $D = Z$ and interpret $\nabla\theta_0$ as the τ^{th} quantile treatment effect (QTE). Alternatively, if $D \neq Z$, then we obtain the QTE model of Chernozhukov and Hansen (2005). To map (A.24) into our framework, we set

$$\rho(X, \theta) = 1\{Y \leq \theta(D)\} - \tau, \tag{A.25}$$

where $X = (Y, D) \in \mathbf{X} \equiv \mathbf{R} \times [0, 1]$. In order to illustrate our conditions in a number of different settings, we focus on conducting inference on a nonlinear function of θ_0.

Specifically, we conduct inference on the variance of the quantile treatment effects:

$$\int_0^1 (\nabla\theta_0(u))^2 du - (\int_0^1 \nabla\theta_0(u)du)^2$$

while imposing that the QTE be increasing in treatment intensity (i.e. $d \mapsto \nabla\theta_0(d)$ is increasing). To map this problem into our framework we define

$$C_B^m([0,1]) \equiv \{\theta : [0,1] \to \mathbf{R} \text{ s.t. } \|\theta\|_{m,\infty} < \infty\} \qquad \|\theta\|_{m,\infty} \equiv \sup_{0\le\alpha\le m} \sup_{d\in[0,1]} |\nabla^\alpha \theta(d)|,$$

and set $\mathbf{B} = C_B^2([0,1])$ and $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{2,\infty}$. We impose the restriction that the quantile treatment effect be increasing in the intensity of treatment by letting $\mathbf{G} = C_B^0([0,1])$, $\|\cdot\|_{\mathbf{G}} = \|\cdot\|_\infty$ (where we write $\|\cdot\|_\infty$ in place of $\|\cdot\|_{0,\infty}$), and defining

$$\Upsilon_G(\theta) \equiv -\nabla^2\theta. \tag{A.26}$$

As shown in Section A.2.2, $\mathbf{G}$ is a lattice with order unit $\mathbf{1_G} = \mathbf{c}$ for $\mathbf{c}$ the constant function $\mathbf{c}(d) = 1$ for all $d \in [0,1]$. Setting $\mathbf{F} = \mathbf{R}$ with $\|\cdot\|_{\mathbf{F}} = |\cdot|$, we test whether the variance of the quantile treatment effects equals a hypothesized value $\lambda \neq 0$ by setting

$$\Upsilon_F(\theta) = \int_0^1 (\nabla\theta(u))^2 du - (\int_0^1 \nabla\theta(u)du)^2 - \lambda. \tag{A.27}$$

For the parameter space for θ_0 we employ a ball in $\mathbf{B}$ and we thus set Θ to equal

$$\Theta \equiv \{\theta \in C_B^2([0,1]) \text{ s.t. } \|\theta\|_{2,\infty} \le C_0\} \tag{A.28}$$

for some $C_0 < \infty$. For a sequence of approximating functions $\{p_j\}_{j=1}^{j_n}$ defined on $[0,1]$ we then let $p^{j_n}(d) \equiv (p_1(d), \ldots, p_{j_n}(d))'$ and define Θ_n to equal

$$\Theta_n \equiv \{p^{j_n\prime}\beta \in C_B^2([0,1]) : \|p^{j_n\prime}\beta\|_{2,\infty} \le C_0\}. \tag{A.29}$$

Similarly for a sequence $\{q_k\}_{k=1}^{k_n}$, we set $q^{k_n}(z) \equiv (q_1(z), \ldots, q_{k_n}(z))'$ and define

$$Q_n(\theta) \equiv \|\hat{\Sigma}_n\{\frac{1}{n}\sum_{i=1}^n (1\{Y_i \le \theta(D_i)\} - \tau)q^{k_n}(Z_i)\}\|_p$$

for some $2 \le p \le \infty$ and weighting matrix $\hat{\Sigma}_n$. The statistics $I_n(R)$ and $I_n(\Theta)$ then equal the minimums of $\sqrt{n}Q_n$ over $\Theta_n \cap R$ and Θ_n respectively.

In what follows, we will assume for simplicity that θ_0 is identified. As a result, we may think of θ_0 as a function of P through (A.24), though we leave such dependence implicit in the notation. We next impose the following assumptions:

Assumption A.2.12. *(i)* $\{Y_i, D_i, Z_i\}_{i=1}^n$ *is i.i.d. with* $(Y, D, Z) \in \mathbf{R} \times [0,1] \times \mathbf{R}$ *dis-*

tributed according to $P \in \mathbf{P}$; (ii) For Θ as in (A.28) and each $P \in \mathbf{P}_0$ there exists a unique $\theta_0 \in \Theta$ satisfying (A.24); (iii) The distribution of Y conditional on (D, Z) is absolutely continuous with density $f_{Y|DZ,P}(\cdot|D, Z)$ that is bounded and Lipschitz uniformly in (D, Z) and $P \in \mathbf{P}$; (iv) Assumptions S.6.1 and S.6.2 hold.

Assumption A.2.13. *(i) $\sup_d \|p^{j_n}(d)\|_2 \lesssim \sqrt{j_n}$; (ii) $E_P[p^{j_n}(D)p^{j_n}(D)']$ has eigenvalues bounded away from zero and infinity uniformly in $P \in \mathbf{P}$ and j_n; (iii) For each $P \in \mathbf{P}_0$ there is a $\Pi_n\theta_0 \in \Theta_n \cap R$ satisfying $\sup_{P\in\mathbf{P}_0} \|E_P[(1\{Y \leq \Pi_n\theta_0(D)\} - 1\{Y \leq \theta_0(D)\})q^{k_n}(Z)]\|_p = O((n\log(n))^{-1/2})$ and $\sup_{P\in\mathbf{P}_0} \|\theta_0 - \Pi_n\theta_0\|_{1,\infty} = o(1)$.*

Assumption A.2.14. *(i) $\inf_{P\in\mathbf{P}_0} \inf_{\theta\in\Theta:\|\theta-\theta_0\|_{1,\infty}\geq\epsilon} E_P[(P(Y \leq \theta(D)|Z) - \tau)^2] > 0$ for every $\epsilon > 0$; (ii) There are ϵ and $\mathrm{s}_n > 0$ satisfying for all $P \in \mathbf{P}_0$ and $\|\theta - \Pi_n\theta_0\|_{1,\infty} \leq \epsilon$, $\mathrm{s}_n \leq \underline{\mathrm{sing}}\{E_P[f_{Y|D,Z}(\theta(D)|D, Z)q^{k_n}(Z)p^{j_n}(D)']\}$ and $\mathrm{s}_n = O(1)$.*

Assumption A.2.15. *(i) $\max_{1\leq k\leq k_n} \|q_k\|_\infty = O(1)$; (ii) $\max_{1\leq k\leq k_n} \|q_k\|_{1,\infty} = O(k_n)$; (iii) $E_P[q^{k_n}(Z)q^{k_n}(Z)']$ has eigenvalues bounded away from zero and infinity uniformly in $P \in \mathbf{P}$ and k_n; (iv) For each $\theta \in \Theta$ there is a $\pi_n(\theta) \in \mathbf{R}^{k_n}$ with $E_P[(E_P[\rho(X, \theta)|Z] - q^{k_n}(Z)'\pi_n(\theta))^2] = o(1)$ uniformly in $P \in \mathbf{P}$ and $\theta \in \Theta$; (v) $k_n^{1/p}\sqrt{j_n}\log^{3/2}(n)(n^{1/6} \vee k_n)/n^{1/3} = o(1)$ and $j_n \log^{3/2}(1 + k_n)k_n^{2/p+1/2}/\mathrm{s}_n\sqrt{n} = o((\log(n))^{-2})$.*

Assumption A.2.16. *(i) $\|\hat{\Sigma}_n - \Sigma_P\|_{o,p} = o_P((k_n^{1/p}\log(n))^{-1})$ uniformly in $P \in \mathbf{P}$; (ii) Σ_P is invertible and $\|\Sigma_P\|_{o,p}$ and $\|\Sigma_P^{-1}\|_{o,p}$ are bounded in $P \in \mathbf{P}$ and k_n.*

Assumption A.2.12 imposes regularity conditions on the distribution P that enable us to apply the empirical process coupling results of Appendix S.6. Assumption A.2.13 states the requirements on Θ_n, including demanding an asymptotically negligible bias in Assumption A.2.13(iii). Assumption A.2.14(i) holds pointwise in $P \in \mathbf{P}_0$ due to Θ being compact under $\|\cdot\|_{1,\infty}$, and hence the uniformity in $P \in \mathbf{P}_0$ demanded by Assumption A.2.14(i) corresponds to imposing strong identification. Assumption A.2.14(ii) enables us to obtain a uniform rate of convergence under $\|\cdot\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}} \|\cdot\|_{P,2}$. As in Section A.2.2, s_n can be shown to be related to the degree of ill-posedness. Assumptions A.2.15(i)-(iv) impose conditions on $\{q_k\}_{k=1}^{k_n}$ including that they be bounded – this requirement can be relaxed at the cost of more stringent rate restrictions to ensure a coupling of the empirical process (see Lemma S.4.28). Finally, Assumption A.2.15(v) states our rate restrictions, which we note are easier to satisfy for higher values of p.

For any $\theta = p^{j_n\prime}\beta \in \Theta_n \cap R$, in this application the local parameter space equals

$$\begin{aligned} V_n(\theta, R|\ell) = \Big\{h = p^{j_n\prime}\beta_h :& \|\theta + \frac{h}{\sqrt{n}}\|_{2,\infty} \leq C_0, \ \sup_{P\in\mathbf{P}} \|h\|_{P,2} \leq \ell\sqrt{n}, \\ & \int_0^1 (\nabla\theta(u) + \frac{\nabla h(u)}{\sqrt{n}})^2 du - (\int_0^1 \{\nabla\theta(u) + \frac{\nabla h(u)}{\sqrt{n}}\}du)^2 = \lambda, \\ & -\nabla^2\theta(d) - \frac{\nabla^2 h(d)}{\sqrt{n}} \leq 0 \text{ for all } d \in [0, 1]\Big\}, \end{aligned} \tag{A.30}$$

where the first two constraints impose that $\theta + h/\sqrt{n} \in \Theta_n$ and $\|h/\sqrt{n}\|_{\mathbf{E}} \leq \ell$, while the final two constraints require that $\theta + h/\sqrt{n} \in R$. Similarly, here

$$V_n(\theta, \Theta|\ell) = \Big\{h = p^{j_n\prime}\beta_h : \|\theta + \frac{h}{\sqrt{n}}\|_{2,\infty} \leq C_0 \text{ and } \sup_{P\in\mathbf{P}} \|h\|_{P,2} \leq \ell\sqrt{n}\Big\}.$$

Also recall that $\mathbb{W}_P(\theta) \sim N(0, \mathrm{Var}_P\{\rho(X,\theta)q^{k_n}(Z)\})$ and for any $h = p^{j_n\prime}\beta_h$ define

$$\mathbb{D}_P(\theta)[h] \equiv E_P[q^{k_n}(Z) f_{Y|DZ,P}(\theta(D)|D,Z) p^{j_n}(D)'\beta_h]. \tag{A.31}$$

The random variables to which $I_n(R)$ and $I_n(\Theta)$ will be coupled are then given by

$$U_P(R|\ell_n) \equiv \inf_{h\in V_n(\Pi_n\theta_0, R|\ell_n)} \|\mathbb{W}_P(\Pi_n\theta_0) + \mathbb{D}_P(\Pi_n\theta_0)[h]\|_{\Sigma_P,2}$$
$$U_P(\Theta|\ell_n) \equiv \inf_{h\in V_n(\Pi_n\theta_0, \Theta|\ell_n)} \|\mathbb{W}_P(\Pi_n\theta_0) + \mathbb{D}_P(\Pi_n\theta_0)[h]\|_{\Sigma_P,2}.$$

Our next result obtains distributional approximations by applying Theorem 3.1.

Theorem A.2.5. *Let Assumptions A.2.12, A.2.13, A.2.14, A.2.15, and A.2.16 hold, $a_n = (\log(n))^{-1/2}$, and $\ell_n \downarrow 0$ satisfy $k_n^{1/p}\sqrt{j_n\ell_n \log(1+k_n)\log(1/\ell_n)} = o((\log(n))^{-1/2})$ and $\ell_n^2\sqrt{nj_n\log(n)} = o(1)$. Then: (i) Uniformly in $P \in \mathbf{P}_0$ it follows that*

$$I_n(R) \leq U_P(R|\ell_n) + o_P(a_n).$$

(ii) If in addition $k_n\log(1+k_n)\sqrt{j_n\log(n)}/\mathrm{s}_n^2\sqrt{n} = o(1)$, then for any $\ell_n^{\mathrm{u}} \downarrow 0$ satisfying $k_n^{1/p}\sqrt{j_n\ell_n^{\mathrm{u}}\log(1+k_n)\log(1/\ell_n^{\mathrm{u}})} = o((\log(n))^{-1/2})$, $(\ell_n^{\mathrm{u}})^2\sqrt{nj_n\log(n)} = o(1)$, and $\sqrt{k_n\log(1+k_n)}/\mathrm{s}_n\sqrt{n} = o(\ell_n^{\mathrm{u}})$, it follows uniformly in $P \in \mathbf{P}_0$ that

$$I_n(R) - I_n(\Theta) \leq U_P(R|\ell_n) - U_P(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n).$$

Theorem A.2.5(i) obtains an upper bound for $I_n(R)$ by relying on Theorem 3.1(i). In order to approximate the re-centered statistic $I_n(R) - I_n(\Theta)$, we cannot rely on an upper bound for $I_n(\Theta)$ as the resulting approximation could fail to control size. Therefore, Theorem A.2.5(ii) instead relies on Theorem 3.1(ii). Applying Theorem 3.1(ii), however, requires an additional rate condition in order to establish the linearization of the moment conditions is asymptotically valid. We also note that the conclusion of Theorem A.2.5(ii) in fact holds with equality if ℓ_n satisfies the same rate restrictions as ℓ_n^{u}.

In order to provide bootstrap estimates for these distributional approximations, we let $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$ denote minimizers of Q_n over $\Theta_n \cap R$ and Θ_n respectively. Our bootstrap approximation estimates the law of $\mathbb{W}_P(\theta_0)$ and the derivative $\mathbb{D}_P(\theta_0)$ by evaluating

$$\hat{\mathbb{W}}_n(\theta) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_i\{q^{k_n}(Z_i)(1\{Y_i \leq \theta(D_i)\} - \tau) - \frac{1}{n}\sum_{j=1}^n q^{k_n}(Z_j)(1\{Y_j \leq \theta(D_j)\} - \tau)\}$$

$$\hat{\mathbb{D}}_n(\theta)[h] \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n q^{k_n}(Z_i)(1\{Y_i \leq \theta(D_i) + \frac{h(D_i)}{\sqrt{n}}\} - 1\{Y_i \leq \theta(D_i)\})$$

at $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$. An unappealing feature of $\hat{\mathbb{D}}_n(\theta)$ is that it is not linear in h, which complicates computation. Alternatively, a plug-in estimator based on (A.31) could be used, though at the expense of having to estimate the density $f_{Y|DZ,P}$.

With regards to the local parameter space, we note that in this application

$$G_n(\hat{\theta}_n) \equiv \{h = p^{j_n\prime}\beta_h : -\nabla^2\hat{\theta}_n(d) - \frac{\nabla^2 h(d)}{\sqrt{n}} \leq \max\{-\nabla^2\hat{\theta}_n(d) \vee -r_n\} \text{ for all } d \in [0,1]\}.$$

Employing that $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{2,\infty}$ and the expression for Υ_F in (A.27), we obtain that

$$\hat{V}_n(\hat{\theta}_n, R|\ell_n) = \Big\{h = p^{j_n\prime}\beta_h : h \in G_n(\hat{\theta}_n),\ \|\frac{h}{\sqrt{n}}\|_{2,\infty} \leq \ell_n$$
$$\int_0^1 (\nabla\hat{\theta}_n(u) + \frac{\nabla h(u)}{\sqrt{n}})^2 du - (\int_0^1 (\nabla\hat{\theta}_n(u) + \frac{\nabla h(u)}{\sqrt{n}}) du)^2 = \lambda\Big\},$$

where ℓ_n is chosen to satisfy conditions stated below. The bootstrap statistics $\hat{U}_n(R|\ell_n)$ and $\hat{U}_n(\Theta|+\infty)$ for approximating the distributions in Theorem A.2.5 are then

$$\hat{U}_n(R|\ell_n) \equiv \inf_{h\in\hat{V}_n(\hat{\theta}_n,R|\ell_n)} \|\hat{\mathbb{W}}_n(\hat{\theta}_n) + \hat{\mathbb{D}}_n(\hat{\theta}_n)[h]\|_{\hat{\Sigma}_n,p}$$
$$\hat{U}_n(\Theta|+\infty) \equiv \inf_{h=p^{j_n\prime}\beta_h} \|\hat{\mathbb{W}}_n(\hat{\theta}_n^{\mathrm{u}}) + \hat{\mathbb{D}}_n(\hat{\theta}_n^{\mathrm{u}})[h]\|_{\hat{\Sigma}_n,p}.$$

The following final assumption will enable us to establish bootstrap validity. In the requirements below, it is helpful to recall θ_0 is implicitly a function of P through (A.24).

Assumption A.2.17. *(i) The functions $\theta(d) = 1$, $\theta(d) = d^2$ are in $\mathbf{B}_n$; (ii) $\|\theta_0 - \Pi_n\theta_0\|_{2,\infty} = o(1)$ uniformly in $P \in \mathbf{P}_0$ and $\sup_{P\in\mathbf{P}_0}\|\theta_0\|_{2,\infty} < C_0$; (iii) k_n satisfies $k_n^{1/p+12/26} = o(n^{1/26}/\log(n))$; (iv) $\sup_d \|\nabla^2 p^{j_n}(d)\|_2 \vee \|\nabla p^{j_n}(d)\|_2 \lesssim j_n^{5/2}$; (v) r_n, ℓ_n satisfy $k_n^{1/p}\sqrt{j_n\ell_n\log(1+k_n)\log(1/\ell_n)} = o((\log(n))^{-1/2})$, $j_n^{5/2}\sqrt{k_n\log(1+k_n)}/\mathrm{s}_n\sqrt{n} = o(1 \wedge r_n)$, and $\ell_n(\sqrt{j_n n}\ell_n + j_n^{5/2}\sqrt{k_n\log(1+k_n)}/\mathrm{s}_n) = o((\log(n))^{-1/2})$.*

Assumption A.2.17(i) requires that the quadratic functions belong to $\mathbf{B}_n$ – a condition that holds if quadratic functions belong to the span of $\{p_j\}_{j=1}^{j_n}$. Assumption A.2.17(ii) implies that θ_0 and its approximation $\Pi_n\theta_0$ belong to the interior of Θ. Assumption A.2.17(iii) enables us to verify the bootstrap coupling requirement of Assumption 3.11 by applying the results in Appendix S.7 to a Haar basis expansion. While condition A.2.17(iii) suffices for verifying Assumption 3.11 in both the endogenous ($Z \neq D$) and exogenous ($Z = D$) settings, we note that in both cases better rate conditions can be obtained.[1] Finally, Assumption A.2.17(iv) ensures $\mathcal{S}_n(\mathbf{B},\mathbf{E}) \asymp j_n^{5/2}$, while Assumption

[1]For instance under endogeneity, a better rate could be obtained by conducting a basis expansion using the tensor product of a Haar Basis for (Y, D) and the functions $\{q_k\}_{k=1}^{k_n}$.

A.2.17(v) imposes the requirements on ℓ_n and r_n.

The next theorem establishes the validity of the bootstrap procedure.

Theorem A.2.6. *Let Assumptions A.2.12, A.2.13, A.2.14, A.2.15, A.2.16, and A.2.17 hold and* $a_n = (\log(n))^{-1/2}$. *Then, there is a sequence* $\tilde{\ell}_n \asymp \ell_n$ *satisfying*

$$\hat{U}_n(R|\ell_n) \geq U_P^\star(R|\tilde{\ell}_n) + o_P(a_n)$$

uniformly in $P \in \mathbf{P}_0$. *(ii) If in addition* $k_n \log(1+k_n)\sqrt{j_n \log(n)}/\mathrm{s}_n^2\sqrt{n} = o(1)$, *then for any* $\tilde{\ell}_n^{\mathrm{u}}$ *satisfying the conditions of Theorem A.2.5(ii) we have uniformly in* $P \in \mathbf{P}_0$

$$\hat{U}_n(R|\ell_n) - \hat{U}_n(\Theta|+\infty) \geq U_P^\star(R|\tilde{\ell}_n) - U_P^\star(\Theta|\tilde{\ell}_n^{\mathrm{u}}) + o_P(a_n).$$

Theorems A.2.5(i) and A.2.6(i) imply that as critical value for $I_n(R)$ we may employ

$$\hat{q}_{1-\alpha}(\hat{U}_n(R|\ell_n)) \equiv \inf\{c : P(\hat{U}_n(R|\ell_n) \leq c|\{V_i\}_{i=1}^n) \geq 1-\alpha\}.$$

If in addition $k_n \log(1+k_n)\sqrt{j_n \log(n)}/\mathrm{s}_n^2\sqrt{n} = o(1)$, then Theorems A.2.5(ii) and A.2.6(ii) imply a valid test can be obtained by rejecting whenever $I_n(R) - I_n(\Theta)$ exceeds

$$\hat{q}_{1-\alpha}(\hat{U}_n(R|\ell_n) - \hat{U}_n(\Theta|+\infty)) \equiv \inf\{c : P(\hat{U}_n(R|\ell_n) - \hat{U}_n(\Theta|+\infty) \leq c|\{V_i\}_{i=1}^n) \geq 1-\alpha\}.$$

Our critical values depend on the choices of r_n and ℓ_n. The slackness parameter r_n again measures sampling uncertainty in whether constraints "bind." Following the discussion in Section 2.1, for $\hat{\theta}_n^{\mathrm{u}\star}$ a "bootstrap" analogue to $\hat{\theta}_n^{\mathrm{u}}$, we may thus set

$$P(\max_{d\in[0,1]} \nabla^2\hat{\theta}_n^{\mathrm{u}}(d) - \nabla^2\hat{\theta}_n^{\mathrm{u}\star}(d) \leq r_n|\{V_i\}_{i=1}^n) = 1-\gamma_n \tag{A.32}$$

with $\gamma_n \to 0$. With regards to ℓ_n, we note that its main role in this application is to ensure that $\hat{V}_n(\hat{\theta}_n, R|\ell_n)$ is well approximated by the true local parameter space despite the nonlinearity of Υ_F. To this end, the requirements on ℓ_n imposed in Assumption A.2.6 ensure $\sqrt{n}\ell_n\|\hat{\theta}_n - \Pi_n\theta_0\|_{\mathbf{B}} = o_P(a_n)$ uniformly in $P \in \mathbf{P}_0$. Since $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{2,\infty}$ in this application, we may select ℓ_n in a data driven way by setting it to satisfy

$$P(\max_{d\in[0,1]} |\nabla^2\hat{\theta}_n^{\mathrm{u}}(d) - \nabla^2\hat{\theta}_n^{\mathrm{u}\star}(d)| \leq \frac{1}{\sqrt{n}\ell_n}|\{V_i\}_{i=1}^n) = 1-\gamma_n \tag{A.33}$$

for some $\gamma_n \to 0$. While we set γ_n in (A.32) and (A.33) to be the same, it is worth noting they could be different. In fact, r_n and ℓ_n do not "interact" in the requirements of Assumption A.2.17(v) and, in this sense, can be set independently. We also note that in settings in which the rate of convergence is sufficiently fast, (A.33) should deliver a "large" ℓ_n in the sense that $\hat{U}_n(R|\ell_n)$ and $\hat{U}_n(R|+\infty)$ are asymptotically equivalent. Moreover, in applications in which we expect the rate of convergence of $\hat{\theta}_n$ to be sufficiently fast, we may directly set $\ell_n = +\infty$; see Lemma S.3.1.

Remark A.2.3. To illustrate the role of ℓ_n, it is helpful to conduct a pointwise (in P) analysis, set $p = 2$, and connect our assumptions to the literature on estimation of conditional moment restriction models (Chen and Pouzo, 2012). We follow the literature in imposing a local curvature assumption, which in our application, corresponds to

$$\begin{aligned}&\|E_P[(P(Y \leq h(D)|Z) - \tau)q^{k_n}(Z)]\|_2 \\ &\qquad \asymp \|E_P[f_{Y|DZ,P}(\bar{\theta}(D)|D,Z)(\theta_0(D) - h(D))q^{k_n}(Z)]\|_2 \quad \text{(A.34)}\end{aligned}$$

for all $h \in \Theta_n$ and $\bar{\theta} \in \Theta$ that are in a neighborhood of θ_0. We further suppose the linear operator $h \mapsto E_P[f_{Y|DZ,P}(\theta_0(D)|D,Z)h(D)|Z]$ is compact, in which case there exist orthonormal bases $\{\psi_j\}$ and $\{\phi_k\}$ and a sequence $\lambda_j \downarrow 0$ satisfying

$$E_P[f_{Y|DZ,P}(\theta_0(D)|D,Z)\phi_j(D)|Z] = \lambda_j \psi_j(Z). \quad \text{(A.35)}$$

Setting $k_n \geq j_n$ with $k_n - j_n = O(1)$, $p^{j_n} = (\phi_1, \ldots, \phi_{j_n})'$, $q^{k_n} = (\psi_1, \ldots, \psi_{k_n})'$, and $\Pi_n^{\mathrm{u}}\theta_0 = \sum_{j=1}^{j_n} \phi_j \beta_j$, we also suppose θ_0 admits an expansion

$$\theta_0 = \sum_{j=1}^{\infty} \beta_j \phi_j \text{ with } |\beta_j| = O(j^{-\gamma_\beta}). \quad \text{(A.36)}$$

Provided that the approximation error of $\Pi_n\theta_0$ (as in Assumption A.2.13(iii)) and $\Pi_n^{\mathrm{u}}\theta_0$ are of the same order, it then follows from (A.34) and (A.35) that

$$\|E_P[(1\{Y \leq \Pi_n\theta_0(D)\} - 1\{Y \leq \theta_0(D)\})q^{k_n}(Z)]\|_2 \lesssim \frac{\lambda_{j_n}}{j_n^{\gamma_\beta}} \quad \text{(A.37)}$$

and $\mathrm{s}_n \asymp \lambda_{j_n}$ – i.e. s_n is proportional to the reciprocal of the sieve measure of ill-posedness (Chen and Pouzo, 2012). As a result, if $\lambda_j \asymp j^{-\gamma_\lambda}$ and $\gamma_\beta > \max\{5/2, 3 - \gamma_\lambda\}$, then Theorem A.2.5 may be applied to couple $I_n(R)$ by setting $j_n \asymp n^\kappa$ with $(2(\gamma_\lambda + \gamma_\beta))^{-1} < \kappa < \min\{(5 + 2\gamma_\lambda)^{-1}, 1/6\}$, while coupling $I_n(R) - I_n(\Theta)$ additionally requires $\gamma_\beta > 3/2 + \gamma_\lambda$ and $\kappa < (3 + 4\gamma_\lambda)^{-1}$. In contrast, in the severely ill-posed case in which $\lambda_j \asymp \exp\{-\gamma_\lambda j\}$, the conditions of Theorem A.2.5 for coupling $I_n(R) - I_n(\Theta)$ are not satisfied. However, the conditions for coupling $I_n(R)$ can still be met provided $\gamma_\beta > 4$ by setting $j_n = (\log(n) - \kappa(\log(\log(n))))/2\gamma_\lambda$ with $7 < \kappa < 2\gamma_\beta - 1$. Thus, while in the ill-posed case the rate of convergence is too slow to apply Theorem A.2.5(ii), Theorem A.2.5(i) is still able to deliver a coupling upper bound for suitable ℓ_n. ■

References

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis – A Hitchhiker's Guide*. Springer-Verlag, Berlin.

Belloni, A., Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, **186** 345–366.

Blundell, R., Chen, X. and Kristensen, D. (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, **75** 1613–1669.

Blundell, R., Horowitz, J. and Parey, M. (2017). Nonparametric estimation of a nonseparable demand function under the slutsky inequality restriction. *Review of Economics and Statistics*, **99** 291–304.

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.

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. and Christensen, T. M. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric iv regression. *Quantitative Economics*, **9** 39–84.

Chen, X. and Pouzo, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, **80** 277–321.

Chernozhukov, V. and Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, **73** 245–261.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50** 891–916.

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. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, **79** 147–168.

Yurinskii, V. V. (1977). On the error of the gaussian approximation for convolutions. *Theory of Probability and Its Applications*, **2** 236–247.

Zhai, A. (2018). A high-dimensional clt in w_2 distance with near optimal convergence rate. *Probability Theory and Related Fields*, **170** 821–845.

Supplemental Appendix II
Not Intended For Publication

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.L.A.
andres@econ.ucla.edu

April, 2022

This Supplemental Appendix to "Constrained Conditional Moment Restriction Models" contains the proofs for all results. Section S.1 derives rate of convergence results that are employed in our strong and bootstrap approximations. In Section S.2 we establish Theorem 3.1, while the proofs for all remaining results concerning our bootstrap approximation and test are contained in Section S.3. Section S.4 includes the proofs of the results stated in Section 4 and the examples discussed in Supplemental Appendix I. Finally, Sections S.5, S.6, and S.7 develop results that may be of independent interest, and include the analysis of the local parameter space, empirical process coupling results based on Koltchinskii (1994), and bootstrap coupling results.

*Research supported by NSF Grant 1757140.
†Research supported by NSF Grant SES-1426882.

S.1 Rate of Convergence

This section contains consistency and rate of convergence results for $\hat{\Theta}^{\mathrm{r}}_n$. The assumptions in the main text, which are designed to deliver a strong approximation, are stronger than needed for deriving the results in this section. We therefore next introduce a weaker set of assumptions that suffice for obtaining a rate of convergence. To this end, we set

$$Q_P(\theta) \equiv \|E_P[\rho(X,\theta) * q^{k_n}(Z)]\|_{\Sigma_P,p}; \tag{S.1}$$

i.e. Q_P is the population analogue to the criterion function Q_n. In addition, we define

$$\begin{aligned}\overrightarrow{d}_H(A,B,\|\cdot\|_{\mathbf{E}}) &\equiv \sup_{a\in A}\inf_{b\in B}\|a-b\|_{\mathbf{E}}\\ d_H(A,B,\|\cdot\|_{\mathbf{E}}) &\equiv \max\{\overrightarrow{d}_H(A,B,\|\cdot\|_{\mathbf{E}}),\overrightarrow{d}_H(B,A,\|\cdot\|_{\mathbf{E}})\},\end{aligned}$$

which constitute the directed Hausdorff and the Hausdorff distance (under $\|\cdot\|_{\mathbf{E}}$) between two sets A and B. Given these definition, we introduce the following requirements:

Assumption S.1.1. *(i) There are $k_n \times k_n$ matrices $\Sigma_P > 0$ with $\|\hat{\Sigma}_n - \Sigma_P\|_{o,p} = o_P(1)$ uniformly in $P \in \mathbf{P}$; (ii) $\|\Sigma_P\|_{o,p} \vee \|\Sigma_P^{-1}\|_{o,p}$ is uniformly bounded in k_n and $P \in \mathbf{P}$.*

Assumption S.1.2. *Define the sequence $\eta_n \equiv J_n B_n k_n^{1/p}\sqrt{\log(1+k_n)/n}$. Then: (i) $\sup_{\theta\in\Theta^{\mathrm{r}}_{0n}} Q_P(\theta) \times \|\hat{\Sigma}_n - \Sigma_P\|_{o,p} = O_P(\eta_n)$ uniformly in $P \in \mathbf{P}_0$; (ii) $\sup_{\theta\in\Theta^{\mathrm{r}}_{0n}} Q_P(\theta) = \inf_{\theta\in\Theta_n\cap R} Q_P(\theta) + O(\eta_n)$ uniformly in $P \in \mathbf{P}_0$.*

Assumption S.1.3. *There are sets $\mathcal{V}_n(P) \subseteq \Theta_n \cap R$ and a sequence $\{\nu_n\}_{n=1}^{\infty}$ with $\nu_n^{-1} = O(1)$, such that $\hat{\Theta}^{\mathrm{r}}_n \subseteq \mathcal{V}_n(P)$ with probability tending to one uniformly in $P \in \mathbf{P}_0$ and for any $\theta \in \mathcal{V}_n(P)$ and $\eta_n \equiv J_n B_n k_n^{1/p}\sqrt{\log(1+k_n)/n}$ it follows that*

$$\nu_n^{-1}\overrightarrow{d}_H(\{\theta\},\Theta^{\mathrm{r}}_{0n},\|\cdot\|_{\mathbf{E}}) \leq \{Q_P(\theta) - \inf_{\tilde{\theta}\in\Theta_n\cap R} Q_P(\tilde{\theta})\} + O(\eta_n).$$

In particular, note Assumption S.1.1 is implied by Assumption 3.7. Similarly, Assumption S.1.2 follows from Assumptions 3.7(i) and 3.6(ii), while Assumption S.1.3 will be verified by relying on Assumptions 3.4(i), 3.4(ii) or 3.12(iii) (depending on the choice of τ_n), and 3.6(ii). Given these assumptions, we next establish a consistency (Lemma S.1.1) and rate of convergence results (Theorem S.1.1) for $\hat{\Theta}^{\mathrm{r}}_n$.

Lemma S.1.1. *Let Assumptions 3.1(i), 3.2(i)(iii), S.1.1, S.1.2(i), $\|\cdot\|_{\mathbf{A}}$ be a norm on $\mathbf{B}_n$ and for $\epsilon > 0$ let $\mathcal{V}_n(P) \equiv \{\theta \in \Theta_n \cap R : \overrightarrow{d}_H(\{\theta\},\Theta^{\mathrm{r}}_{0n},\|\cdot\|_{\mathbf{A}}) \leq \epsilon\}$, and define*

$$S_n(\epsilon) \equiv \inf_{P\in\mathbf{P}_0}\{\inf_{\theta\in(\Theta_n\cap R)\setminus\mathcal{V}_n(P)} Q_P(\theta) - \inf_{\theta\in\Theta_n\cap R} Q_P(\theta)\}.$$

(i) If $\eta_n \vee \tau_n = o(S_n(\epsilon))$ for $\eta_n \equiv k_n^{1/p}\sqrt{\log(1+k_n)}J_n B_n/\sqrt{n}$, then $\hat{\Theta}^{\mathrm{r}}_n \subseteq \mathcal{V}_n(P)$ with

probability tending to one uniformly in $P \in \mathbf{P}_0$. (ii) If Assumption S.1.2(ii) holds and $\eta_n = o(\tau_n)$, then $\Theta_{0n}^{\mathrm{r}} \subseteq \hat{\Theta}_n^{\mathrm{r}}$ with probability tending to one uniformly in $P \in \mathbf{P}_0$.

PROOF: For a given $\epsilon > 0$ first notice that by definition of $\hat{\Theta}_n^{\mathrm{r}}$ and $\mathcal{V}_n(P)$ we have

$$P(\overrightarrow{d}_H(\hat{\Theta}_n^{\mathrm{r}}, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{A}}) > \epsilon) \leq P(\inf_{\theta \in (\Theta_n \cap R)\setminus \mathcal{V}_n(P)} Q_n(\theta) \leq \inf_{\theta \in \Theta_n \cap R} Q_n(\theta) + \tau_n) \tag{S.2}$$

for all $P \in \mathbf{P}_0$. Setting $\hat{Q}_P(\theta) \equiv \|E_P[\rho(X,\theta) * q^{k_n}(Z)]\|_{\hat{\Sigma}_n,p}$ then note that Lemma S.1.2 and $\|\hat{\Sigma}_n\|_{o,p} = O_P(1)$ uniformly in $P \in \mathbf{P}_0$ by Lemma S.1.4 allow us to conclude

$$\inf_{\theta \in (\Theta_n \cap R)\setminus \mathcal{V}_n(P)} \hat{Q}_P(\theta) \leq \inf_{\theta \in (\Theta_n \cap R)\setminus \mathcal{V}_n(P)} Q_n(\theta) + O_P(\eta_n) \tag{S.3}$$

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

$$\inf_{\theta \in \Theta_n \cap R} Q_n(\theta) \leq \inf_{\theta \in \Theta_n \cap R} \hat{Q}_P(\theta) + O_P(\eta_n). \tag{S.4}$$

Next note that for any $a \in \mathbf{R}^{k_n}$ we have $\|\Sigma_P a\|_p \leq \|\Sigma_P \hat{\Sigma}_n^{-1}\|_{o,p}\|\hat{\Sigma}_n a\|_p$, and therefore

$$\begin{aligned}\inf_{\theta \in (\Theta_n \cap R)\setminus \mathcal{V}_n(P)} \hat{Q}_P(\theta) &\geq \|\Sigma_P \hat{\Sigma}_n^{-1}\|_{o,p}^{-1} \inf_{\theta \in (\Theta_n \cap R)\setminus \mathcal{V}_n(P)} Q_P(\theta) \\ &\geq \|\Sigma_P \hat{\Sigma}_n^{-1}\|_{o,p}^{-1}\{S_n(\epsilon) + \inf_{\theta \in \Theta_n \cap R} Q_P(\theta)\}\end{aligned} \tag{S.5}$$

by definition of $S_n(\epsilon)$. Similarly, employing that $\|\hat{\Sigma}_n a\|_p \leq \|\hat{\Sigma}_n \Sigma_P^{-1}\|_{o,p}\|\Sigma_P a\|_p$ yields

$$\begin{aligned}&\inf_{\theta \in \Theta_n \cap R} \hat{Q}_P(\theta) - \|\Sigma_P \hat{\Sigma}_n^{-1}\|_{o,p}^{-1} \inf_{\theta \in \Theta_n \cap R} Q_P(\theta) \\ &\qquad \leq \{\|\hat{\Sigma}_n \Sigma_P^{-1}\|_{o,p} - \|\Sigma_P \hat{\Sigma}_n^{-1}\|_{o,p}^{-1}\} \inf_{\theta \in \Theta_n \cap R} Q_P(\theta).\end{aligned} \tag{S.6}$$

For I_{k_n} the $k_n \times k_n$ identity matrix, then note that $\|I_{k_n}\|_{o,p} = 1$ implies the bound

$$\begin{aligned}|\|\Sigma_P \hat{\Sigma}_n^{-1}\|_{o,p} - 1| = |\|\Sigma_P \hat{\Sigma}_n^{-1}\|_{o,p} - \|I_{k_n}\|_{o,p}| &\leq \|(\Sigma_P - \hat{\Sigma}_n)\hat{\Sigma}_n^{-1}\|_{o,p} \\ &\leq \|\hat{\Sigma}_n^{-1}\|_{o,p}\|\Sigma_P - \hat{\Sigma}_n\|_{o,p} = O_P(\|\Sigma_P - \hat{\Sigma}_n\|_{o,p}),\end{aligned} \tag{S.7}$$

where the final equality holds uniformly in $P \in \mathbf{P}_0$ by Lemma S.1.4. By identical arguments it follows that $|\|\hat{\Sigma}_n \Sigma_P^{-1}\|_{o,p} - 1| = O_P(\|\hat{\Sigma}_n - \Sigma_P\|_{o,p})$ uniformly in $P \in \mathbf{P}_0$, and therefore (S.6), $\Theta_{0n}^{\mathrm{r}} \subseteq \Theta_n \cap R$, and Assumption S.1.2(i) imply that

$$\inf_{\theta \in \Theta_n \cap R} \hat{Q}_P(\theta) - \|\Sigma_P \hat{\Sigma}_n^{-1}\|_{o,p}^{-1} \inf_{\theta \in \Theta_n \cap R} Q_P(\theta) \leq O_P(\eta_n) \tag{S.8}$$

uniformly in $P \in \mathbf{P}_0$. Therefore, (S.2), (S.3), (S.4), (S.5), and (S.8) yield that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\overrightarrow{d}_H(\hat{\Theta}^{\mathrm{r}}_n, \Theta^{\mathrm{r}}_{0n}, \|\cdot\|_{\mathbf{A}}) > \epsilon)$$
$$\leq \limsup_{M\uparrow\infty} \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(S_n(\epsilon) \leq \|\Sigma_P\|_{o,p}\|\hat{\Sigma}_n^{-1}\|_{o,p} M(\eta_n + \tau_n)) = 0,$$

where the equality follows from Lemma S.1.4, Assumption S.1.1(ii), and $\eta_n \vee \tau_n = o(S_n(\epsilon))$ by hypothesis. Part (i) of the lemma then follows by definition of $\mathcal{V}_n(P)$.

In order to establish part (ii) of the lemma, note that the definition of $\hat{\Theta}^{\mathrm{r}}_n$ implies

$$P(\Theta^{\mathrm{r}}_{0n} \subseteq \hat{\Theta}^{\mathrm{r}}_n) \geq P(\sup_{\theta\in\Theta^{\mathrm{r}}_{0n}} Q_n(\theta) \leq \inf_{\theta\in\Theta_n\cap R} Q_n(\theta) + \tau_n) \tag{S.9}$$

for all $P \in \mathbf{P}_0$. Moreover, applying Lemmas S.1.2 and S.1.4 together with $\|\hat{\Sigma}_n a\|_p \leq \|\hat{\Sigma}_n \Sigma_P^{-1}\|_{o,p}\|\Sigma_P a\|_p$ for any $a \in \mathbf{R}^{k_n}$ implies that uniformly in $P \in \mathbf{P}_0$

$$\sup_{\theta\in\Theta^{\mathrm{r}}_{0n}} Q_n(\theta) \leq \sup_{\theta\in\Theta^{\mathrm{r}}_{0n}} \hat{Q}_P(\theta) + O_P(\eta_n)$$
$$\leq \|\hat{\Sigma}_n\Sigma_P^{-1}\|_{o,p} \sup_{\theta\in\Theta^{\mathrm{r}}_{0n}} Q_P(\theta) + O_P(\eta_n) = \inf_{\theta\in\Theta_n\cap R} Q_P(\theta) + O_P(\eta_n), \tag{S.10}$$

where the final equality follows from Assumption S.1.2(ii), identical arguments to those in (S.7) implying $|\|\hat{\Sigma}_n\Sigma_P^{-1}\|_{o,p} - 1| = O_P(\|\hat{\Sigma}_n - \Sigma_P\|_{o,p})$ uniformly in $P \in \mathbf{P}$, and Assumption S.1.2(i). Similarly, Lemmas S.1.2 and S.1.4, $\|\Sigma_P a\|_p \leq \|\Sigma_P\hat{\Sigma}_n^{-1}\|_{o,p}\|\hat{\Sigma}_n a\|_p$ for any $a \in \mathbf{R}^{k_n}$, Assumption S.1.2(i), and result (S.7) imply that uniformly in $P \in \mathbf{P}_0$

$$\inf_{\theta\in\Theta_n\cap R} Q_n(\theta) \geq \inf_{\theta\in\Theta_n\cap R} \hat{Q}_P(\theta) - O_P(\eta_n)$$
$$\geq \|\Sigma_P\hat{\Sigma}_n^{-1}\|_{o,p}^{-1} \inf_{\theta\in\Theta_n\cap R} Q_P(\theta) - O_P(\eta_n) = \inf_{\theta\in\Theta_n\cap R} Q_P(\theta) - O_P(\eta_n). \tag{S.11}$$

Part (ii) of the lemma thus follows from (S.9), (S.10), (S.11), and $\eta_n = o(\tau_n)$. ∎

Theorem S.1.1. *Let Assumptions 3.1(i), 3.2(i)(iii), S.1.1, S.1.2, S.1.3 hold, and*

$$\mathcal{R}_n \equiv \nu_n\{\frac{k_n^{1/p}\sqrt{\log(1+k_n)}J_n B_n}{\sqrt{n}}\}. \tag{S.12}$$

Then uniformly in $P \in \mathbf{P}_0$: (i) $\overrightarrow{d}_H(\hat{\Theta}^{\mathrm{r}}_n, \Theta^{\mathrm{r}}_{0n}, \|\cdot\|_{\mathbf{E}}) = O_P(\mathcal{R}_n + \nu_n\tau_n)$; and (ii) $d_H(\hat{\Theta}^{\mathrm{r}}_n, \Theta^{\mathrm{r}}_{0n}, \|\cdot\|_{\mathbf{E}}) = O_P(\nu_n\tau_n)$ provided $J_n B_n k_n^{1/p}\sqrt{\log(1+k_n)/n} = o(\tau_n)$.

PROOF: Let $\eta_n \equiv k_n^{1/p}\sqrt{\log(1+k_n)}J_n B_n/\sqrt{n}$, $\delta_n^{-1} \equiv \nu_n(\eta_n + \tau_n)$, and $Q_P(\theta) \equiv$

$\|E_P[\rho(X,\theta) * q^{k_n}(Z)]\|_{\Sigma_P,p}$. In addition, we define $A_n \equiv A_{n1} \cap A_{n2} \cap A_{n3}$ where

$$\begin{aligned}
&A_{n1} \equiv \{\hat{\Theta}_n^{\mathrm{r}} \subseteq \mathcal{V}_n(P)\} \\
&A_{n2} \equiv \{\hat{\Sigma}_n^{-1} \text{ exists and } \|\hat{\Sigma}_n^{-1}\|_{o,p} \vee \|\hat{\Sigma}_n\|_{o,p} \vee \|\Sigma_P^{-1}\|_{o,p} \vee \|\Sigma_P\|_{o,p} < B\} \\
&A_{n3} \equiv \{\sup_{\theta \in \Theta_{0n}^{\mathrm{r}}} Q_P(\theta) \times \|\hat{\Sigma}_n - \Sigma_P\|_{o,p} \leq B\eta_n \text{ and } \|\hat{\Sigma}_n - \Sigma_P\|_{o,p} \leq \frac{1}{2B}\}.
\end{aligned} \tag{S.13}$$

Moreover, note that for any $\epsilon > 0$ and B sufficiently large we can conclude that

$$\limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(A_n^c) < \epsilon \tag{S.14}$$

due to Lemma S.1.4 and Assumptions S.1.1(i), S.1.2(i), and S.1.3. Therefore, we obtain

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(\delta_n \overrightarrow{d}_H(\hat{\Theta}_n^{\mathrm{r}}, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{E}}) > 2^M) \\
&\qquad \leq \limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(\delta_n \overrightarrow{d}_H(\hat{\Theta}_n^{\mathrm{r}}, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{E}}) > 2^M;\ A_n) + \epsilon
\end{aligned} \tag{S.15}$$

for any M. For each $P \in \mathbf{P}_0$, next partition $\mathcal{V}_n(P)$ into subsets $S_{n,j}(P)$ defined by

$$S_{n,j}(P) \equiv \{\theta \in \mathcal{V}_n(P) : 2^{j-1} < \delta_n \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{E}}) \leq 2^j\}.$$

Since $\hat{\Theta}_n^{\mathrm{r}} \subseteq \mathcal{V}_n(P)$ under A_n, it follows from the definition of $\hat{\Theta}_n^{\mathrm{r}}$, and (S.15) that

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(\delta_n \overrightarrow{d}_H(\hat{\Theta}_n^{\mathrm{r}}, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{E}}) > 2^M) \\
&\qquad \leq \limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} \sum_{j \geq M}^{\infty} P(\inf_{\theta \in S_{n,j}(P)} Q_n(\theta) \leq \inf_{\theta \in \Theta_n \cap R} Q_n(\theta) + \tau_n;\ A_n) + \epsilon.
\end{aligned} \tag{S.16}$$

Letting $\hat{Q}_P(\theta) \equiv \|E_P[\rho(X,\theta) * q^{k_n}(Z)]\|_{\hat{\Sigma}_n,p}$, we then obtain from Lemma S.1.2 that

$$\inf_{\theta \in \Theta_n \cap R} Q_n(\theta) \leq \inf_{\theta \in \Theta_n \cap R} \hat{Q}_P(\theta) + \|\hat{\Sigma}_n\|_{o,p} \mathcal{Z}_{n,P} \leq \inf_{\theta \in \Theta_n \cap R} \hat{Q}_P(\theta) + B\mathcal{Z}_{n,P} \tag{S.17}$$

where the final inequality holds under the event A_n by (S.13). Moreover, since for any $a \in \mathbf{R}^{k_n}$ we have $\|\hat{\Sigma}_n a\|_p \leq \|\hat{\Sigma}_n \Sigma_P^{-1}\|_{o,p} \|\Sigma_P a\|_p$, we obtain from $\Theta_{0n}^{\mathrm{r}} \subseteq \Theta_n \cap R$ and the inequality $\|\hat{\Sigma}_n \Sigma_P^{-1}\|_{o,p} \leq \|\{\hat{\Sigma}_n - \Sigma_P\}\Sigma_P^{-1}\|_{o,p} + 1$ that under the event A_n we have

$$\begin{aligned}
&\inf_{\theta \in \Theta_n \cap R} \hat{Q}_P(\theta) \leq \|\hat{\Sigma}_n \Sigma_P^{-1}\|_{o,p} \inf_{\theta \in \Theta_n \cap R} Q_P(\theta) \\
&\qquad \leq \{1 + \|\Sigma_P^{-1}\|_{o,p} \|\hat{\Sigma}_n - \Sigma_P\|_{o,p}\} \inf_{\theta \in \Theta_n \cap R} Q_P(\theta) \leq \inf_{\theta \in \Theta_n \cap R} Q_P(\theta) + B^2 \eta_n.
\end{aligned} \tag{S.18}$$

In addition, note that by similar arguments we also obtain from Lemma S.1.2 and

$\|\Sigma_P a\|_p \leq \|\Sigma_P \hat{\Sigma}_n^{-1}\|_{o,p} \|\hat{\Sigma}_n a\|_p$ that under the event A_n we must have

$$\inf_{\theta \in S_{n,j}(P)} Q_n(\theta) \geq \inf_{\theta \in S_{n,j}(P)} \hat{Q}_P(\theta) - \|\hat{\Sigma}_n\|_{o,p} \mathcal{Z}_{n,P}$$
$$\geq \|\Sigma_P \hat{\Sigma}_n^{-1}\|_{o,p}^{-1} \inf_{\theta \in S_{n,j}(P)} Q_P(\theta) - B\mathcal{Z}_{n,P}. \quad \text{(S.19)}$$

Next, we note the triangle inequality, $\|(\Sigma_P - \hat{\Sigma}_n)\hat{\Sigma}_n^{-1}\|_{o,p} \leq \|\hat{\Sigma}_n^{-1}\|_{o,p} \|\hat{\Sigma}_n - \Sigma_P\|_{o,p}$, and $\|\hat{\Sigma}_n^{-1}\|_{o,p} \leq B$ under the event A_n by (S.13) yield the inequality

$$\|\Sigma_P \hat{\Sigma}_n^{-1}\|_{o,p}^{-1} - 1 \geq (\|(\Sigma_P - \hat{\Sigma}_n)\hat{\Sigma}_n^{-1}\|_{o,p} + 1)^{-1} - 1$$
$$\geq -\|(\Sigma_P - \hat{\Sigma}_n)\hat{\Sigma}_n^{-1}\|_{o,p} \geq -B\|\hat{\Sigma}_n - \Sigma_P\|_{o,p}. \quad \text{(S.20)}$$

Therefore, combining results (S.19) and (S.20), together with Assumption S.1.3 and the definition of $S_{n,j}(P)$ we obtain for B sufficiently large that under the event A_n we have

$$\begin{aligned}\inf_{\theta \in S_{n,j}(P)} Q_n(\theta) &\geq (1 - B\|\hat{\Sigma}_n - \Sigma_P\|_{o,p}) \times \inf_{\theta \in S_{n,j}(P)} Q_P(\theta) - B\mathcal{Z}_{n,P} \\ &\geq (1 - B\|\hat{\Sigma}_n - \Sigma_P\|_{o,p})(\inf_{\theta \in \Theta_n \cap R} Q_P(\theta) + \frac{2^{j-1}}{\nu_n \delta_n} - B\eta_n) - B\mathcal{Z}_{n,P} \\ &\geq \inf_{\theta \in \Theta_n \cap R} Q_P(\theta) + \frac{2^{j-2}}{\nu_n \delta_n} - B(\mathcal{Z}_{n,P} + 2B\eta_n), \end{aligned} \quad \text{(S.21)}$$

where the final inequality follows from $\Theta_{0n}^{\mathrm{r}} \subseteq \Theta_n \cap R$ and the definition of the evnet A_n in (S.13). Hence, results (S.16), (S.17), (S.18), and (S.21) yield

$$\begin{aligned}&\limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(\delta_n \overrightarrow{d}_H(\hat{\Theta}_n^{\mathrm{r}}, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{E}}) > 2^M) \\ &\quad \leq \limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} \sum_{j \geq M}^{\infty} P(\frac{2^{j-2}}{\nu_n \delta_n} \leq 3B(B\eta_n + \mathcal{Z}_{n,P}) + \tau_n;\ A_n) + \epsilon \\ &\quad \leq \limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} \sum_{j \geq M}^{\infty} P(2^{(j-3)}(\eta_n + \tau_n) \leq 3B\mathcal{Z}_{n,P}) + \epsilon, \end{aligned} \quad \text{(S.22)}$$

where in the final inequality we employed that we had defined $\delta_n^{-1} \equiv \nu_n(\eta_n + \tau_n)$. Therefore, $\mathcal{Z}_{n,P} \in \mathbf{R}_+$, Lemma S.1.2, $\tau_n \geq 0$, and Markov's inequality yield

$$\limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} \sum_{j \geq M}^{\infty} P(2^{(j-3)}(\eta_n + \tau_n) \leq 3B\mathcal{Z}_{n,P})$$
$$\lesssim \limsup_{M \uparrow \infty} \limsup_{n \to \infty} \sum_{j \geq M} 2^{-j} \times \frac{1}{\eta_n} \frac{k_n^{1/p} \sqrt{\log(1 + k_n)} J_n B_n}{\sqrt{n}} = 0, \quad \text{(S.23)}$$

where in the final result we employed that $\eta_n \equiv k_n^{1/p} \sqrt{\log(1 + k_n)} J_n B_n / \sqrt{n}$. The first claim of the theorem therefore follows from (S.22), (S.23), and ϵ being arbitrary.

To establish the second claim of the theorem, next define the event $A_{n4} \equiv \{\Theta^{\mathrm{r}}_{0n} \subseteq \hat{\Theta}^{\mathrm{r}}_n\}$. Since $\overrightarrow{d}_H(\Theta^{\mathrm{r}}_{0n}, \hat{\Theta}^{\mathrm{r}}_n, \|\cdot\|_{\mathbf{E}}) = 0$ whenever the event A_{n4} occurs, we can conclude from Lemma S.1.1(ii) and part (i) of this theorem that

$$\limsup_{M\uparrow\infty} \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\delta_n d_H(\hat{\Theta}^{\mathrm{r}}_n, \Theta^{\mathrm{r}}_{0n}, \|\cdot\|_{\mathbf{E}}) > 2^M)$$
$$= \limsup_{M\uparrow\infty} \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\delta_n \overrightarrow{d}_H(\hat{\Theta}^{\mathrm{r}}_n, \Theta^{\mathrm{r}}_{0n}, \|\cdot\|_{\mathbf{E}}) > 2^M) = 0, \quad \text{(S.24)}$$

and thus the theorem follows from $\delta_n^{-1} = \nu_n(\eta_n + \tau_n)$ and $\eta_n = o(\tau_n)$. ■

Corollary S.1.1. *If Assumptions 3.1(i), 3.2(i)(iii), 3.3(i), 3.4, 3.6(ii), and 3.7 hold, then $\overrightarrow{d}_H(\hat{\theta}_n, \Theta^{\mathrm{r}}_{0n}, \|\cdot\|_{\mathbf{E}}) = O_P(\mathcal{R}_n)$ uniformly in $P \in \mathbf{P}_0$.*

Proof: Follows from Theorem S.1.1(i) applied with $\tau_n \equiv a_n/\sqrt{n}$ after noting that $a_n = o(1)$ (by Assumption 3.3(i)) implies $\nu_n a_n/\sqrt{n} = o(\mathcal{R}_n)$ and: (i) Assumption S.1.1 holds by Assumption 3.7; (ii) Assumption S.1.2(i) holds by Assumptions 3.6(ii) and 3.7(i); (iii) Assumption S.1.2(ii) holds by $Q_P(\theta) \geq 0$ and Assumption 3.6(ii); and (iv) Assumption S.1.3 holds with $\tau_n \equiv a_n/\sqrt{n}$ by Assumptions 3.4 and 3.6(ii), the triangle inequality, and $\inf_{\theta\in\Theta_n\cap R} Q_P(\theta) \leq \sup_{\theta\in\Theta^{\mathrm{r}}_{0n}} Q_P(\theta)$ due to $\Theta^{\mathrm{r}}_{0n} \subseteq \Theta_n \cap R$. ■

Corollary S.1.2. *Let Assumptions 3.1(i), 3.2(i)(iii), 3.3(i), 3.4(i), 3.6(ii), 3.7, and 3.12(iii) hold. Then uniformly in $P \in \mathbf{P}_0$: (i) $\overrightarrow{d}_H(\hat{\Theta}^{\mathrm{r}}_n, \Theta^{\mathrm{r}}_{0n}, \|\cdot\|_{\mathbf{E}}) = O_P(\mathcal{R}_n + \nu_n\tau_n)$; and (ii) $d_H(\hat{\Theta}^{\mathrm{r}}_n, \Theta^{\mathrm{r}}_{0n}, \|\cdot\|_{\mathbf{E}}) = O_P(\nu_n\tau_n)$ provided $J_nB_nk_n^{1/p}\sqrt{\log(1+k_n)/n} = o(\tau_n)$.*

Proof: Follows from Theorem S.1.1 after noting that $a_n = o(1)$ (by Assumption 3.3(i)) implies: (i) Assumption S.1.1 holds by Assumption 3.7; (ii) Assumption S.1.2(i) holds by Assumptions 3.6(ii) and 3.7(i); (iii) Assumption S.1.2(ii) holds by $Q_P(\theta) \geq 0$ and Assumption 3.6(ii); and (iv) Assumption S.1.3 holds by Assumptions 3.4(i), 3.6(ii), 3.12(iii), the triangle inequality, and $\Theta^{\mathrm{r}}_{0n} \subseteq \Theta_n \cap R$. ■

Lemma S.1.2. *Let $\hat{Q}_P(\theta) \equiv \|E_P[\rho(X,\theta) * q^{k_n}(Z)]\|_{\hat{\Sigma}_n,p}$, and Assumptions 3.1(i), 3.2(i), and 3.2(iii) hold. Then, for each $P \in \mathbf{P}$ there are random $\mathcal{Z}_{n,P} \in \mathbf{R}_+$ with*

$$|Q_n(\theta) - \hat{Q}_P(\theta)| \leq \|\hat{\Sigma}_n\|_{o,p} \times \mathcal{Z}_{n,P},$$

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

Proof: Let $\mathcal{G}_n \equiv \{fq_{k,\jmath} : f \in \mathcal{F}_n,\ 1 \leq \jmath \leq \mathcal{J} \text{ and } 1 \leq k \leq k_{n,\jmath}\}$. Note that by Assumption 3.2(i), $\|q_{k,\jmath}\|_\infty \leq B_n$ 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 3.2(iii), it follows that $G_n \equiv B_nF_n$ is an envelope for $\mathcal{G}_n$ satisfying $\sup_{P\in\mathbf{P}} E_P[G_n^2(V)] < \infty$. Thus, we obtain

$$\sup_{P\in\mathbf{P}} E_P[\sup_{g\in\mathcal{G}_n} |\frac{1}{\sqrt{n}}\sum_{i=1}^n (g(V_i) - E_P[g(V)])|] \lesssim \sup_{P\in\mathbf{P}} J_{[\,]}(\|G_n\|_{P,2}, \mathcal{G}_n, \|\cdot\|_{P,2}) \quad \text{(S.25)}$$

by Theorem 2.14.2 in van der Vaart and Wellner (1996). Moreover, also notice that Lemma S.1.3, the change of variables $u = \epsilon/B_n$, and $B_n \geq 1$ imply

$$\sup_{P\in\mathbf{P}} J_{[\,]}(\|G_n\|_{P,2}, \mathcal{G}_n, \|\cdot\|_{P,2}) \leq \sup_{P\in\mathbf{P}} \int_0^{\|G_n\|_{P,2}} \sqrt{1+\log(k_n N_{[\,]}(\epsilon/B_n, \mathcal{F}_n, \|\cdot\|_{P,2}))}d\epsilon$$
$$\leq (1+\sqrt{\log(k_n)})B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}(\|F_n\|_{P,2}, \mathcal{F}_n, \|\cdot\|_{P,2}) = O(\sqrt{\log(1+k_n)}B_n J_n), \quad \text{(S.26)}$$

where the final equality follows from Assumption 3.2(iii). Next define $\mathcal{Z}_{n,P} \in \mathbf{R}_+$ by

$$\mathcal{Z}_{n,P} \equiv \frac{k_n^{1/p}}{\sqrt{n}} \times \sup_{g\in\mathcal{G}_n} |\frac{1}{\sqrt{n}}\sum_{i=1}^n (g(V_i) - E_P[g(V)])|$$

and note (S.25) and (S.26) imply $\sup_{P\in\mathbf{P}} E_P[\mathcal{Z}_{n,P}] = O(k_n^{1/p}\sqrt{\log(1+k_n)}B_n J_n/\sqrt{n})$ as desired. Moreover, for any $\theta \in \Theta_n \cap R$, the definitions of $\mathbb{G}_n(\theta)$, $\mathcal{G}_n$, and $\mathcal{Z}_{n,P}$ yield

$$|Q_n(\theta) - \hat{Q}_P(\theta)| \leq \frac{\|\hat{\Sigma}_n\|_{o,p}}{\sqrt{n}} \times \|\mathbb{G}_n(\theta)\|_p$$
$$\leq \|\hat{\Sigma}_n\|_{o,p} \times \frac{k_n^{1/p}}{\sqrt{n}} \times \sup_{g\in\mathcal{G}_n} |\frac{1}{\sqrt{n}}\sum_{i=1}^n (g(V_i) - E_P[g(V)])| \equiv \|\hat{\Sigma}_n\|_{o,p} \times \mathcal{Z}_{n,P},$$

which establishes the claim of the lemma. ■

Lemma S.1.3. *Let $\{g_j\}_{j=1}^J$ be functions satisfying $\max_{1\leq j\leq J}\|g_j\|_\infty \leq C < \infty$ and define $\mathcal{G}_n \equiv \{fg_j : f \in \mathcal{F}_n,\ 1 \leq j \leq J\}$. Then for any P it follows that*

$$N_{[\,]}(\epsilon, \mathcal{G}_n, \|\cdot\|_{P,2}) \leq J \times N_{[\,]}(\epsilon/C, \mathcal{F}_n, \|\cdot\|_{P,2}).$$

Proof: First define $g_j^+ \equiv g_j \vee 0$ and $g_j^- \equiv g_j \wedge 0$, where $\vee$ and $\wedge$ denote the pointwise maximums and minimums. If $\{[f_{i,l}, f_{i,u}]\}_i$ is a collection of brackets for $\mathcal{F}_n$ satisfying

$$\int (f_{i,u} - f_{i,l})^2 dP \leq \epsilon^2 \quad \text{(S.27)}$$

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

$$\{[g_j^+ f_{i,l} + g_j^- f_{i,u},\ g_j^- f_{i,l} + g_j^+ f_{i,u}]\}_{i,j}. \quad \text{(S.28)}$$

Moreover, since $|g_j| = g_j^+ - g_j^-$ by construction, we also obtain from result (S.27) that

$$\int (g_j^+ f_{i,u} + g_j^- f_{i,l} - g_j^+ f_{i,l} - g_j^- f_{i,u})^2 dP = \int (f_{i,u} - f_{i,l})^2 |g_j|^2 dP \leq \epsilon^2 C^2. \quad \text{(S.29)}$$

Since there are $J \times N_{[\,]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2})$ brackets in (S.28), we conclude from (S.29) that $N_{[\,]}(\epsilon, \mathcal{G}_n, \|\cdot\|_{P,2}) \leq J \times N_{[\,]}(\epsilon/C, \mathcal{F}_n, \|\cdot\|_{P,2})$, which establishes the lemma. ■

Lemma S.1.4. *If Assumption S.1.1 holds, then there is a constant $B < \infty$ such that*

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

Proof: Recall that $\hat{\Sigma}_n$ and Σ_P are $k_n \times k_n$ matrices, though the dependence on k_n was suppressed from the notation. Then note that by Assumption S.1.1(ii) there exists a constant $B < \infty$ such that for all $P \in \mathbf{P}$ and k_n we have that

$$\max\{\|\Sigma_P\|_{o,p}, \|\Sigma_P^{-1}\|_{o,p}\} < \frac{B}{2}. \tag{S.30}$$

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

$$\hat{\Sigma}_n = \Sigma_P\{I_{k_n} - \Sigma_P^{-1}(\Sigma_P - \hat{\Sigma}_n)\}. \tag{S.31}$$

By Theorem 2.9 in Kress (1999), the matrix $\{I_{k_n} - \Sigma_P^{-1}(\Sigma_P - \hat{\Sigma}_n)\}$ is invertible and the operator norm of its inverse is bounded by two when $\|\Sigma_P^{-1}(\Sigma_P - \hat{\Sigma}_n)\|_{o,p} < 1/2$. Since Σ_P is invertible by Assumption S.1.1(i), result (S.31) implies that $\hat{\Sigma}_n$ is invertible if and only if $\{I_{k_n} - \Sigma_P^{-1}(\Sigma_P - \hat{\Sigma}_n)\}$ is invertible, which yields that

$$\begin{aligned} &P(\hat{\Sigma}_n^{-1} \text{ exists and } \|\{I_{k_n} - \Sigma_P^{-1}(\Sigma_P - \hat{\Sigma}_n)\}^{-1}\|_{o,p} < 2) \\ &\qquad \geq P(\|\Sigma_P^{-1}(\hat{\Sigma}_n - \Sigma_P)\|_{o,p} < \frac{1}{2}) \geq P(\|\hat{\Sigma}_n - \Sigma_P\|_{o,p} < \frac{1}{B}), \end{aligned} \tag{S.32}$$

where we employed $\|\Sigma_P^{-1}(\hat{\Sigma}_n - \Sigma_P)\|_{o,p} \leq \|\Sigma_P^{-1}\|_{o,p}\|\hat{\Sigma}_n - \Sigma_P\|_{o,p}$ and (S.30). Hence, since (S.31) implies $\hat{\Sigma}_n^{-1} = \{I_{k_n} - \Sigma_P^{-1}(\Sigma_P - \hat{\Sigma}_n)\}^{-1}\Sigma_P^{-1}$ whenever $\{I_{k_n} - \Sigma_P^{-1}(\Sigma_P - \hat{\Sigma}_n)\}^{-1}$ exists, the bound $\|\Sigma_P^{-1}\|_{o,p} < B/2$ and result (S.32) allow us to conclude

$$P(\hat{\Sigma}_n^{-1} \text{ exists and } \|\hat{\Sigma}_n^{-1}\|_{o,p} < B) \geq P(\|\hat{\Sigma}_n - \Sigma_P\|_{o,p} < \frac{1}{B}). \tag{S.33}$$

Finally, since $\|\hat{\Sigma}_n\|_{o,p} \leq B/2 + \|\hat{\Sigma}_n - \Sigma_P\|_{o,p}$ by (S.30), result (S.33) implies that

$$\begin{aligned} &\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P(\hat{\Sigma}_n^{-1} \text{ exists and } \max\{\|\hat{\Sigma}_n\|_{o,p}, \|\hat{\Sigma}_n^{-1}\|_{o,p}\} < B) \\ &\qquad \geq \liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P(\|\hat{\Sigma}_n - \Sigma_P\|_{o,p} < \min\{\frac{B}{2}, \frac{1}{B}\}) = 1, \end{aligned}$$

where the equality, and hence the lemma, follows from Assumption S.1.1(i). ■

Corollary S.1.3. *If Assumption 3.7 holds, then for some $B < \infty$ it follows that:*

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

Proof: Follows from Lemma S.1.4 and Assumption 3.7 together with $a_n = o(1)$, which is imposed by Assumption 3.3(i) (or 3.11), implying Assumption S.1.1 holds. ■

Lemma S.1.5. *If $a \in \mathbf{R}^d$, then $\|a\|_{\tilde{p}} \le d^{(\frac{1}{\tilde{p}}-\frac{1}{p})_+}\|a\|_p$ for any $\tilde{p}, p \in [2,\infty]$.*

PROOF: The case $p \le \tilde{p}$ trivially follows from $\|a\|_{\tilde{p}} \le \|a\|_p$ for all $a \in \mathbf{R}^d$. For the case $p > \tilde{p}$, let $a = (a_1, \ldots, a_d)'$ and note that by Hölder's inequality we obtain

$$\|a\|_{\tilde{p}}^{\tilde{p}} = \sum_{i=1}^{d}\{|a_i|^{\tilde{p}} \times 1\} \le \{\sum_{i=1}^{d}(|a_i|^{\tilde{p}})^{\frac{p}{\tilde{p}}}\}^{\frac{\tilde{p}}{p}}\{\sum_{i=1}^{d} 1^{\frac{p}{p-\tilde{p}}}\}^{1-\frac{\tilde{p}}{p}} = \{\sum_{i=1}^{d}|a_i|^{p}\}^{\frac{\tilde{p}}{p}} d^{1-\frac{\tilde{p}}{p}}. \quad (S.34)$$

Thus, the claim of the lemma for $p > \tilde{p}$ follows from taking the $1/\tilde{p}$ power in (S.34). ■

S.2 Strong Approximation

This Section contains the proof of Theorem 3.1 and supporting results.

PROOF OF THEOREM 3.1: First note that by Assumption 3.7(ii) there is a constant $C_0 < \infty$ such that $\|\Sigma_P\|_{o,p} \le C_0$ for all $P \in \mathbf{P}_0$. Hence, Assumption 3.6(ii) and Lemma S.1.5 imply that for all $P \in \mathbf{P}_0$, $\theta \in \Theta_{0n}^{\mathrm{r}}$, and $h \in V_n(\theta, R|\ell_n)$ we have

$$\begin{aligned}&\|\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}}) * q^{k_n}(Z)] - \mathbb{D}_P(\theta)[h]\|_{\Sigma_P,p} \\ &\qquad \le C_0\|\sqrt{n}E_P[(\rho(X,\theta+\frac{h}{\sqrt{n}}) - \rho(X,\theta)) * q^{k_n}(Z)] - \mathbb{D}_P(\theta)[h]\|_2 + o(a_n). \quad (S.35)\end{aligned}$$

Moreover, Lemma S.2.5 and the maps $m_{P,\jmath}$ satisfying Assumption 3.5(i) imply that

$$\begin{aligned}&\sum_{\jmath=1}^{\mathcal{J}}\sum_{k=1}^{k_{n,\jmath}}\langle\sqrt{n}\{m_{P,\jmath}(\theta+\frac{h}{\sqrt{n}}) - m_{P,\jmath}(\theta)\} - \nabla m_{P,\jmath}(\theta)[h], q_{k,\jmath}\rangle_{L_P^2}^2 \\ &\qquad \le \sum_{\jmath=1}^{\mathcal{J}} C_1\|\sqrt{n}\{m_{P,\jmath}(\theta+\frac{h}{\sqrt{n}}) - m_{P,\jmath}(\theta) - \nabla m_{P,\jmath}(\theta)[\frac{h}{\sqrt{n}}]\}\|_{P,2}^2 \\ &\qquad \le \sum_{\jmath=1}^{\mathcal{J}} C_1 K_m^2 \times n \times \|\frac{h}{\sqrt{n}}\|_{\mathbf{L}}^2 \times \|\frac{h}{\sqrt{n}}\|_{\mathbf{E}}^2 \quad (S.36)\end{aligned}$$

for some constant $C_1 < \infty$ and all $P \in \mathbf{P}_0$, $\theta \in \Theta_{0n}^{\mathrm{r}}$, and $h \in V_n(\theta, R|\ell_n)$. Therefore, by results (S.35) and (S.36), the law of iterated expectations, the definition of $\mathcal{S}_n(\mathbf{L}, \mathbf{E})$, and $K_m\ell_n^2 \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-1/2})$ by hypothesis, we obtain that

$$\begin{aligned}&\sup_{P\in\mathbf{P}_0}\sup_{\theta\in\Theta_{0n}^{\mathrm{r}}}\sup_{h\in V_n(\theta,R|\ell_n)}\|\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}}) * q^{k_n}(Z)] - \mathbb{D}_P(\theta)[h]\|_{\Sigma_P,p} \\ &\qquad \lesssim K_m \times \sqrt{n}\ell_n^2 \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) + o(a_n) = o(a_n). \quad (S.37)\end{aligned}$$

Next, note that since $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$, Assumption 3.6(i) implies there is a sequence $\tilde{\ell}_n$ satisfying the conditions of Lemma S.2.1 and $\ell_n = o(\tilde{\ell}_n)$. Therefore, applying Lemma S.2.1 we obtain uniformly in $P \in \mathbf{P}_0$

$$I_n(R) = \inf_{\theta\in\Theta_{0n}^{\mathrm{r}}} \inf_{h\in V_n(\theta,R|\tilde{\ell}_n)} \|\mathbb{W}_P(\theta)+\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}})*q^{k_n}(Z)]\|_{\Sigma_P,p}+o_P(a_n). \quad \text{(S.38)}$$

Moreover, since $\ell_n = o(\tilde{\ell}_n)$ implies that $V_n(\theta,R|\tilde{\ell}_n) \subseteq V_n(\theta,R|\ell_n)$ for all $\theta\in\Theta_n\cap R$ for n sufficiently large, we obtain uniformly in $P\in\mathbf{P}_0$ that

$$\begin{aligned}
&\inf_{\theta\in\Theta_{0n}^{\mathrm{r}}} \inf_{h\in V_n(\theta,R|\tilde{\ell}_n)} \|\mathbb{W}_P(\theta)+\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}})*q^{k_n}(Z)]\|_{\Sigma_P,p} \\
&\le \inf_{\theta\in\Theta_{0n}^{\mathrm{r}}} \inf_{h\in V_n(\theta,R|\ell_n)} \|\mathbb{W}_P(\theta)+\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}})*q^{k_n}(Z)]\|_{\Sigma_P,p} \\
&= \inf_{\theta\in\Theta_{0n}^{\mathrm{r}}} \inf_{h\in V_n(\theta,R|\ell_n)} \|\mathbb{W}_P(\theta)+\mathbb{D}_P(\theta)[h]\|_{\Sigma_P,p}+o(a_n), \qquad \text{(S.39)}
\end{aligned}$$

where the final equality following from (S.37), Assumption 3.7(ii) and Lemma S.2.6. Thus, the first claim of the Theorem follows from (S.38) and (S.39), while the second follows by noting that if $K_m\mathcal{R}_n^2 \times \mathcal{S}_n(\mathbf{L},\mathbf{E}) = o(a_n n^{-1/2})$, then we may set ℓ_n to simultaneously satisfy the conditions of Lemma S.2.1 and $K_m\ell_n^2 \times \mathcal{S}_n(\mathbf{L},\mathbf{E}) = o(a_n n^{-1/2})$, which obviates the need to introduce $\tilde{\ell}_n$ in (S.38) and (S.39). ■

Lemma S.2.1. *Let Assumptions 3.1(i), 3.2(i), 3.2(iii), 3.3, 3.4, 3.6, and 3.7 hold. Then, for any sequence $\{\ell_n\}$ satisfying $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$ and $\mathcal{R}_n = o(\ell_n)$, we have uniformly in $P\in\mathbf{P}_0$ that:*

$$I_n(R) = \inf_{\theta\in\Theta_{0n}^{\mathrm{r}}} \inf_{h\in V_n(\theta,R|\ell_n)} \|\mathbb{W}_P(\theta)+\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}})*q^{k_n}(Z)]\|_{\Sigma_P,p}+o_P(a_n).$$

Proof: First note that the required sequence $\{\ell_n\}$ exists by Assumption 3.6(i). Next, note that by Assumption 3.4(ii) and Corollary S.1.1 there is a $\hat{\theta}_n\in\Theta_n\cap R$ satisfying

$$Q_n(\hat{\theta}_n) \le \inf_{\theta\in\Theta_n\cap R} Q_n(\theta) + o(a_n/\sqrt{n}) \qquad \text{(S.40)}$$

and $\overrightarrow{d}_H(\hat{\theta}_n, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{E}}) = O_P(\mathcal{R}_n)$ uniformly in $P\in\mathbf{P}_0$. Hence, defining $(\Theta_{0n}^{\mathrm{r}})^{\ell_n} \equiv \{\theta\in\Theta_n\cap R : \overrightarrow{d}_H(\theta,\Theta_{0n}^{\mathrm{r}},\|\cdot\|_{\mathbf{E}}) \le \ell_n\}$, which implicitly depends on $P\in\mathbf{P}_0$, we obtain

$$I_n(R) = \inf_{\theta\in(\Theta_{0n}^{\mathrm{r}})^{\ell_n}} \sqrt{n}Q_n(\theta) + o_P(a_n) \qquad \text{(S.41)}$$

uniformly in $P\in\mathbf{P}_0$ due to $\mathcal{R}_n = o(\ell_n)$, $\overrightarrow{d}_H(\hat{\theta}_n,\Theta_{0n}^{\mathrm{r}},\|\cdot\|_{\mathbf{E}}) = O_P(\mathcal{R}_n)$, $(\Theta_{0n}^{\mathrm{r}})^{\ell_n} \subseteq \Theta_n\cap R$ by construction, result (S.40), and the definition of $I_n(R)$. Next, note that by

Assumption 3.3(i), Corollary S.1.3, and Lemma S.2.6 it follows that

$$|\inf_{\theta\in(\Theta_{0n}^{\mathrm{r}})^{\ell_n}} \sqrt{n}Q_n(\theta) - \inf_{\theta\in(\Theta_{0n}^{\mathrm{r}})^{\ell_n}} \|\mathbb{W}_P(\theta) + \sqrt{n}E_P[\rho(X,\theta) * q^{k_n}(Z)]\|_{\hat{\Sigma}_n,p}|$$
$$\leq \|\hat{\Sigma}_n\|_{o,p} \times \sup_{\theta\in\Theta_n\cap R} \|\mathbb{G}_n(\theta) - \mathbb{W}_P(\theta)\|_p = o_P(a_n) \quad \text{(S.42)}$$

uniformly in $P \in \mathbf{P}_0$. Similarly, employing Corollary S.1.3, Lemmas S.2.2, S.2.6, and ℓ_n satisfying $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$ yields

$$\inf_{\theta\in\Theta_{0n}^{\mathrm{r}}} \inf_{h\in V_n(\theta,R|\ell_n)} \|\mathbb{W}_P(\theta + \frac{h}{\sqrt{n}}) + \sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}}) * q^{k_n}(Z)]\|_{\hat{\Sigma}_n,p}$$
$$= \inf_{\theta\in\Theta_{0n}^{\mathrm{r}}} \inf_{h\in V_n(\theta,R|\ell_n)} \|\mathbb{W}_P(\theta) + \sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}}) * q^{k_n}(Z)]\|_{\hat{\Sigma}_n,p} + o_P(a_n)$$

uniformly in $P \in \mathbf{P}_0$, which together with results (S.41) and (S.42), and Lemma S.2.3 establish the claim of the lemma. ■

Lemma S.2.2. *Let Assumptions 3.2(i) and 3.3(ii) hold. If $\{\delta_n\}$ is a sequence satisfying $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}(\delta_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$, then uniformly in $P \in \mathbf{P}$:*

$$\sup_{\theta\in\Theta_{0n}^{\mathrm{r}}} \sup_{h\in V_n(\theta,R|\delta_n)} \|\mathbb{W}_P(\theta+\frac{h}{\sqrt{n}}) - \mathbb{W}_P(\theta)\|_p = o_P(a_n).$$

Proof: Since $\|q_{k,\jmath}\|_\infty \leq B_n$ for all $1 \leq \jmath \leq \mathcal{J}$ and $1 \leq k \leq k_{n,\jmath}$ by Assumption 3.2(i), Assumption 3.3(ii) yields for any $P \in \mathbf{P}$, $\theta \in \Theta_n \cap R$, and $h \in V_n(\theta, R|\delta_n)$ that

$$E_P[\|\rho(X,\theta+\frac{h}{\sqrt{n}}) - \rho(X,\theta)\|_2^2 q_{k,\jmath}^2(Z)] \leq K_\rho^2 B_n^2 \|\frac{h}{\sqrt{n}}\|_{\mathbf{E}}^{2\kappa_\rho} \leq K_\rho^2 B_n^2 \delta_n^{2\kappa_\rho}. \quad \text{(S.43)}$$

Set $\mathcal{G}_n \equiv \{f q_{k,\jmath}$ for some $f \in \mathcal{F}_n,\ 1 \leq \jmath \leq \mathcal{J},\ 1 \leq k \leq k_{n,\jmath}\}$ and let $\mathbb{G}_P$ be a Gaussian process on $\mathcal{G}_n$ satisfying $E[\mathbb{G}_P(g_1)] = 0$ and $E[\mathbb{G}_P(g_1)\mathbb{G}_P(g_2)] = \mathrm{Cov}_P\{g_1(V), g_2(V)\}$ for any $g_1, g_2 \in \mathcal{G}_n$. Since $\|a\|_p \leq k_n^{1/p}\|a\|_\infty$ for any $a \in \mathbf{R}^{k_n}$, result (S.43) yields

$$E_P[\sup_{\theta\in\Theta_{0n}^{\mathrm{r}}} \sup_{h\in V_n(\theta,R|\delta_n)} \|\mathbb{W}_P(\theta+\frac{h}{\sqrt{n}}) - \mathbb{W}_P(\theta)\|_p]$$
$$\leq k_n^{1/p} \times E[\sup_{g_1,g_2\in\mathcal{G}_n:\|g_1-g_2\|_{P,2}\leq K_\rho B_n\delta_n^{\kappa_\rho}} |\mathbb{G}_P(g_1) - \mathbb{G}_P(g_2)|]. \quad \text{(S.44)}$$

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

$$\sup_{P\in\mathbf{P}} E_P[\sup_{g_1,g_2\in\mathcal{G}_n:\|g_1-g_2\|_{P,2}\leq K_\rho B_n\delta_n^{\kappa_\rho}} |\mathbb{G}_P(g_1)-\mathbb{G}_P(g_2)|]$$
$$\lesssim \sup_{P\in\mathbf{P}}\int_0^{K_\rho B_n\delta_n^{\kappa_\rho}} \sqrt{\log N_{[\,]}(\epsilon,\mathcal{G}_n,\|\cdot\|_{P,2})}d\epsilon$$
$$\lesssim \sup_{P\in\mathbf{P}}\sqrt{\log(1+k_n)}B_n\int_0^{K_\rho\delta_n^{\kappa_\rho}} \sqrt{1+\log N_{[\,]}(u,\mathcal{F}_n,\|\cdot\|_{P,2})}du, \qquad \text{(S.45)}$$

where the second inequality follows from Lemma S.1.3 and the change of variables $u=\epsilon/B_n$. However, note that since $N_{[\,]}(u,\mathcal{F}_n,\|\cdot\|_{P,2})$ is decreasing in u, it follows that $J_{[\,]}(K_\rho\delta_n^{\kappa_\rho},\mathcal{F}_n,\|\cdot\|_{P,2})\leq K_\rho J_{[\,]}(\delta_n^{\kappa_\rho},\mathcal{F}_n,\|\cdot\|_{P,2})$. Therefore, the lemma follows from results (S.44) and (S.45), the definition of $J_{[\,]}(\epsilon,\mathcal{F}_n,\|\cdot\|_{P,2})$, and $k_n^{1/p}\sqrt{\log(1+k_n)}B_n\times \sup_{P\in\mathbf{P}} J_{[\,]}(\delta_n^{\kappa_\rho},\mathcal{F}_n,\|\cdot\|_{P,2})=o(a_n)$ by hypothesis. ■

Lemma S.2.3. *Let Assumptions 3.2(i), 3.2(iii), 3.6(ii), and 3.7 hold with $a_n=o(1)$. For any positive sequence δ_n it then follows that uniformly in $P\in\mathbf{P}_0$ we have*

$$\inf_{\theta\in\Theta_{0n}^{\mathrm{r}}}\inf_{h\in V_n(\theta,R|\delta_n)}\|\mathbb{W}_P(\theta)+\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}})*q^{k_n}(Z)]\|_{\Sigma_P,p}$$
$$=\inf_{\theta\in\Theta_{0n}^{\mathrm{r}}}\inf_{h\in V_n(\theta,R|\delta_n)}\|\mathbb{W}_P(\theta)+\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}})*q^{k_n}(Z)]\|_{\hat{\Sigma}_n,p}+o_P(a_n).$$

Proof: First note that by Assumption 3.7(ii) there is a $C_0<\infty$ such that $\|\Sigma_P\|_{o,p}\vee\|\Sigma_P^{-1}\|_{o,p}\leq C_0$ for all $P\in\mathbf{P}$. Since $\|\hat{\Sigma}_n a\|_p\leq\|\hat{\Sigma}_n\Sigma_P^{-1}\|_{o,p}\|\Sigma_P a\|_p$ for any $a\in\mathbf{R}^{k_n}$, and $\|\hat{\Sigma}_n\Sigma_P^{-1}\|_{o,p}\leq\|\Sigma_P^{-1}\|_{o,p}\|\hat{\Sigma}_n-\Sigma_P\|_{o,p}+1$ by the triangle inequality, we obtain

$$\{C_0\|\hat{\Sigma}_n-\Sigma_P\|_{o,p}+1\}\|\mathbb{W}_P(\theta)+\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}})*q^{k_n}(Z)]\|_{\Sigma_P,p}$$
$$\geq\|\mathbb{W}_P(\theta)+\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}})*q^{k_n}(Z)]\|_{\hat{\Sigma}_n,p} \qquad \text{(S.46)}$$

for any $\theta\in\Theta_{0n}^{\mathrm{r}}$ and $h\in V_n(\theta,R|\delta_n)$. Moreover, $\|\Sigma_P\|_{o,p}\leq C_0$, $0\in V_n(\theta,R|\delta_n)$ for any $\theta\in\Theta_n\cap R$, and Assumption 3.6(ii) imply uniformly in $P\in\mathbf{P}$ that

$$\inf_{\theta\in\Theta_{0n}^{\mathrm{r}}}\inf_{h\in V_n(\theta,R|\delta_n)}\|\mathbb{W}_P(\theta)+\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}})*q^{k_n}(Z)]\|_{\Sigma_P,p}$$
$$\lesssim\sup_{\theta\in\Theta_n\cap R}\|\mathbb{W}_P(\theta)\|_p+o(a_n)=O_P(k_n^{1/p}\sqrt{\log(1+k_n)}B_nJ_n)+o(a_n) \qquad \text{(S.47)}$$

where the final equality holds uniformly in $P\in\mathbf{P}_0$ by Lemma S.2.4 and Markov's inequality. Therefore, results (S.46), (S.47), and Assumption 3.7(i) imply

$$\inf_{\theta\in\Theta_{0n}^{\mathrm{r}}}\inf_{h\in V_n(\theta,R|\delta_n)}\|\mathbb{W}_P(\theta)+\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}})*q^{k_n}(Z)]\|_{\Sigma_P,p}+o_P(a_n)$$
$$\geq \inf_{\theta\in\Theta_{0n}^{\mathrm{r}}}\inf_{h\in V_n(\theta,R|\delta_n)}\|\mathbb{W}_P(\theta)+\sqrt{n}E_P[\rho(X,\theta+\frac{h}{\sqrt{n}})*q^{k_n}(Z)]\|_{\hat{\Sigma}_n,p} \quad (S.48)$$

uniformly in $P\in\mathbf{P}_0$. Next, note that Assumption 3.7 implies Assumption S.1.1 and therefore Lemma S.1.4 yields that $\|\hat{\Sigma}_n\|_{o,p}\vee\|\hat{\Sigma}_n^{-1}\|_{o,p}=O_P(1)$ uniformly in $P\in\mathbf{P}$. The lemma then follows from (S.48) and noting that the reverse inequality also holds by identical arguments but relying on $\|\hat{\Sigma}_n\|_{o,p}\vee\|\hat{\Sigma}_n^{-1}\|_{o,p}=O_P(1)$ uniformly in $P\in\mathbf{P}$ rather than on $\|\Sigma_P\|_{o,p}\vee\|\Sigma_P^{-1}\|_{o,p}\leq C_0$. ■

Lemma S.2.4. *If Assumptions 3.2(i) and 3.2(iii) hold, then for some $C<\infty$ we have:*

$$\sup_{P\in\mathbf{P}}E_P[\sup_{\theta\in\Theta_n\cap R}\|\mathbb{W}_P(\theta)\|_p]\leq Ck_n^{1/p}\sqrt{\log(1+k_n)}B_nJ_n.$$

PROOF: Let $\mathcal{G}_n\equiv\{fq_{k,\jmath}:f\in\mathcal{F}_n,\ 1\leq\jmath\leq\mathcal{J},\text{ and }1\leq k\leq k_{n,\jmath}\}$ and $\mathbb{G}_P$ be a Gaussian process on $\mathcal{G}_n$ satisfying $E[\mathbb{G}_P(g_1)]=0$ and $E[\mathbb{G}_P(g_1)\mathbb{G}_P(g_2)]=\mathrm{Cov}_P\{g_1(V),g_2(V)\}$ for any $g_1,g_2\in\mathcal{G}_n$. Then note $\|a\|_p\leq d^{1/p}\|a\|_\infty$ for any $a\in\mathbf{R}^d$ implies that

$$E_P[\sup_{\theta\in\Theta_n\cap R}\|\mathbb{W}_P(\theta)\|_p]\leq k_n^{1/p}E_P[\sup_{g\in\mathcal{G}_n}|\mathbb{G}_P(g)|]$$
$$\leq k_n^{1/p}\{E_P[|\mathbb{G}_P(g_0)|]+C_1\int_0^\infty\sqrt{\log N_{[\,]}(\epsilon,\mathcal{G}_n,\|\cdot\|_{P,2})}d\epsilon\}, \quad (S.49)$$

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 $G_n\equiv B_nF_n$ for F_n as in Assumption 3.2(iii) 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 $2\|G_n\|_{P,2}$ covering $\mathcal{G}_n$, and as a result we obtain

$$\int_0^\infty\sqrt{\log N_{[\,]}(\epsilon,\mathcal{G}_n,\|\cdot\|_{P,2})}d\epsilon$$
$$\leq\int_0^{2\|G_n\|_{P,2}}\sqrt{1+\log N_{[\,]}(\epsilon,\mathcal{G}_n,\|\cdot\|_{P,2})}d\epsilon\leq C_2\sqrt{\log(1+k_n)}B_nJ_n, \quad (S.50)$$

where the final inequality holds for some $C_2<\infty$ by result (S.26) and $N_{[\,]}(u,\mathcal{G}_n,\|\cdot\|_{P,2})$ being decreasing in u. Furthermore, since $E_P[|\mathbb{G}_P(g_0)|]\leq\|g_0\|_{P,2}\leq\|G_n\|_{P,2}$ we have

$$E_P[|\mathbb{G}_P(g_0)|]\leq\|G_n\|_{P,2}\leq\int_0^{\|G_n\|_{P,2}}\sqrt{1+\log N_{[\,]}(u,\mathcal{G}_n,\|\cdot\|_{P,2})}du. \quad (S.51)$$

Thus, the claim of the lemma follows from (S.49), (S.50), and (S.51). ■

Lemma S.2.5. *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

$$\sum_{k=1}^{k_{n,\jmath}} \langle f, q_{k,\jmath}\rangle^2_{L^2_P} \leq C E_P[(E_P[f(V)|Z_\jmath])^2]. \tag{S.52}$$

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

$$J_{\jmath,n}(\{c_k\}) = \sum_{k=1}^{k_{n,\jmath}} c_k q_{k,\jmath}.$$

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

$$\begin{aligned}
\sup_{P\in\mathbf{P}} \sup_{n\geq 1} \|J_{\jmath,n}\|^2_o &= \sup_{P\in\mathbf{P}} \sup_{n\geq 1} \sup_{\{c_k\}:\sum_k c_k^2 = 1} \|J_{\jmath,n}(\{c_k\})\|^2_{P,2} \\
&= \sup_{P\in\mathbf{P}} \sup_{n\geq 1} \sup_{\{c_k\}:\sum_k c_k^2 = 1} E_P[(\sum_{k=1}^{k_{n,\jmath}} c_k q_{k,\jmath}(Z_\jmath))^2] \leq \sup_{\{c_k\}:\sum_k c_k^2 = 1} C \sum_{k=1}^\infty c_k^2 = C
\end{aligned} \tag{S.53}$$

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

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

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

$$\sum_{k=1}^{k_{n,\jmath}} \langle g, q_{k,\jmath}\rangle^2_{L^2_P} = \|J^*_{\jmath,n}(g)\|^2_{\ell^2(\mathbb{N})} \leq \|J^*_{\jmath,n}\|^2_o \|g\|^2_{P,2} = \|J_{\jmath,n}\|^2_o \|g\|^2_{P,2}. \tag{S.54}$$

Therefore, since $E_P[f(V) q_{k,\jmath}(Z_\jmath)] = E_P[E_P[f(V)|Z_\jmath] q_{k,\jmath}(Z_\jmath)]$ for any $f \in L^2_P$, setting $g(Z_\jmath) = E_P[f(V)|Z_\jmath]$ in (S.54) and employing (S.53) yields the lemma. ■

Lemma S.2.6. *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*

$$|\inf_{\lambda\in\Lambda} \|WA(\lambda)\|_p - \inf_{\lambda\in\Lambda} \|WB(\lambda)\|_p| \leq \|W\|_{o,p} \times \sup_{\lambda\in\Lambda} \|A(\lambda) - B(\lambda)\|_p.$$

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

$$\inf_{\lambda\in\Lambda} \|WB(\lambda)\|_p - \inf_{\lambda\in\Lambda} \|WA(\lambda)\|_p \leq \|WB(\lambda_a)\|_p - \|WA(\lambda_a)\|_p + \eta$$
$$\leq \|W(B(\lambda_a) - A(\lambda_a))\|_p + \eta \leq \|W\|_{o,p} \times \sup_{\lambda\in\Lambda} \|A(\lambda) - B(\lambda)\|_p + \eta, \quad \text{(S.55)}$$

where the second result follows from the triangle inequality, and the final result from $\|Wv\|_p \leq \|W\|_{o,p}\|v\|_p$ for any $v \in \mathbf{R}^k$. By identical manipulations we also have

$$\inf_{\lambda\in\Lambda} \|WA(\lambda)\|_p - \inf_{\lambda\in\Lambda} \|WB(\lambda)\|_p \leq \|W\|_{o,p} \times \sup_{\lambda\in\Lambda} \|A(\lambda) - B(\lambda)\|_p + \eta. \quad \text{(S.56)}$$

Thus, since η was arbitrary, the lemma follows from results (S.55) and (S.56). ■

S.3 Bootstrap Approximation

This appendix contains the proof of all results concerning the bootstrap approximation. We first introduce two assumptions that generalize Assumption 3.13 (at the cost of introducing additional notation) and deliver a stronger version of Theorem 3.2.

Assumption S.3.1. *There is an $\epsilon > 0$ and scalars $\mathcal{D}_n(\mathbf{L}, \mathbf{E})$ and $\mathcal{D}_n(\mathbf{B}, \mathbf{E})$ such that for any $P \in \mathbf{P}$, $\theta \in \Theta^{\mathrm{r}}_{0n}$, and $\theta_1 \in \Theta_n \cap R$ satisfying $\|\theta_1 - \theta\|_{\mathbf{E}} \leq \epsilon$, there exists $\tilde\theta \in \Theta^{\mathrm{r}}_{0n}$ such that $\|\theta - \tilde\theta\|_{\mathbf{E}} = 0$, $\|\tilde\theta - \theta_1\|_{\mathbf{L}} \leq \mathcal{D}_n(\mathbf{L}, \mathbf{E})\|\tilde\theta - \theta_1\|_{\mathbf{E}}$, and $\|\tilde\theta - \theta_1\|_{\mathbf{B}} \leq \mathcal{D}_n(\mathbf{B}, \mathbf{E})\|\tilde\theta - \theta_1\|_{\mathbf{E}}$.*

Assumption S.3.2. *(i) Either Υ_F and Υ_G are affine or $(\mathcal{R}_n + \nu_n\tau_n)\mathcal{D}_n(\mathbf{B}, \mathbf{E}) = o(1)$; (ii) $k_n^{1/p}\sqrt{\log(1 + k_n)}B_n \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho} \vee (\nu_n\tau_n)^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$, $K_m\ell_n^2\mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$, $K_m\ell_n(\mathcal{R}_n + \nu_n\tau_n)\mathcal{D}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$, $\ell_n(\ell_n + \{\mathcal{R}_n + \nu_n\tau_n\}\mathcal{D}_n(\mathbf{B}, \mathbf{E}))1\{K_f > 0\} = o(a_n n^{-\frac{1}{2}})$; (iii) $\limsup 1\{K_g > 0\}\ell_n/r_n < 1/2$ and $(\mathcal{R}_n + \nu_n\tau_n)\mathcal{D}_n(\mathbf{B}, \mathbf{E}) = o(r_n)$.*

In particular, note Assumption S.3.1 holds with $\mathcal{D}_n(\mathbf{L}, \mathbf{E}) = \mathcal{S}_n(\mathbf{L}, \mathbf{E})$, $\mathcal{D}_n(\mathbf{B}, \mathbf{E}) = \mathcal{S}_n(\mathbf{B}, \mathbf{E})$, and $\tilde\theta = \theta$. Hence, Assumption 3.13 implies Assumptions S.3.1 and S.3.2. In general, however, $\mathcal{D}_n(\mathbf{L}, \mathbf{E})$ and $\mathcal{D}_n(\mathbf{B}, \mathbf{E})$ can be smaller than $\mathcal{S}_n(\mathbf{L}, \mathbf{E})$ and $\mathcal{S}_n(\mathbf{B}, \mathbf{E})$ while the introduction of a $\tilde\theta \neq \theta$ eases requirements in partially identified models.

Our next theorem consists of two parts. The first part, which replaces Assumption 3.13 with S.3.1 and S.3.2, can by the preceding discussion be seen as a generalization of Theorem 3.2. The second part shows that, under additional restrictions, it is possible to replace the norm $\|\cdot\|_{\mathbf{B}}$ in the definition of $\hat V_n(\theta, R|\ell)$ (as in (21)) with the norm $\|\cdot\|_{\mathbf{E}}$ – an observation that is sometimes helpful in easing rate restrictions.

Theorem S.3.1. *Let Assumptions 3.1, 3.2, 3.3, 3.4(i), 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, 3.12(i)(iii), S.3.1, and S.3.2 hold. Then, the following statements hold:*

(i) If Assumption 3.12(ii) holds, then there is a $\tilde\ell_n \asymp \ell_n$ such that uniformly in $P \in \mathbf{P}_0$

$$\hat U_n(R|\ell_n) \geq U_P^\star(R|\tilde\ell_n) + o_P(a_n).$$

(ii) In addition, suppose for some $\epsilon > 0$ and $\|\cdot\|_{\mathbf{I}}$ satisfying $\|h\|_{\mathbf{E}} \leq \|h\|_{\mathbf{I}}$ for all $h \in \mathbf{B}_n$, we have that for all $P \in \mathbf{P}_0$, $\{\theta \in \mathbf{B}_n : \vec{d}_H(\theta, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{I}}) \leq \epsilon\} \subseteq \Theta_n$ and $P(\{\theta \in \mathbf{B}_n : \vec{d}_H(\theta, \hat{\Theta}_n^{\mathrm{r}}, \|\cdot\|_{\mathbf{I}}) \leq \epsilon\} \subseteq \Theta_n)$ tends to one uniformly in $P \in \mathbf{P}_0$. If Υ_F and Υ_G are affine, then part (i) holds with $\hat{U}_n(R|\ell_n)$ as in (17) but with

$$\hat{V}_n(\theta, R|\ell) \equiv \{h \in \mathbf{B}_n : h \in G_n(\theta),\ \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0,\ \textit{and}\ \|\frac{h}{\sqrt{n}}\|_{\mathbf{I}} \leq \ell\}. \quad \text{(S.57)}$$

Proof: First note Assumptions 3.6(i) and S.3.2(ii) imply $\mathcal{R}_n \vee \nu_n \tau_n = o(1)$. Hence, by Assumption S.3.2(ii) we may apply Lemma S.3.2 to obtain uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R|\ell_n) = \inf_{\theta \in \hat{\Theta}_n^{\mathrm{r}}} \inf_{h \in \hat{V}_n(\theta, R|\ell_n)} \|\mathbb{W}_P^{\star}(\theta) + \mathbb{D}_P(\theta)[h]\|_{\Sigma_P, p} + o_P(a_n). \quad \text{(S.58)}$$

Thus, we may select $\hat{\theta}_n \in \hat{\Theta}_n^{\mathrm{r}}$ and $\hat{h}_n \in \hat{V}_n(\hat{\theta}_n, R|\ell_n)$ so that uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R|\ell_n) = \|\mathbb{W}_P^{\star}(\hat{\theta}_n) + \mathbb{D}_P(\hat{\theta}_n)[\hat{h}_n]\|_{\Sigma_P, p} + o_P(a_n). \quad \text{(S.59)}$$

Next note that by Assumptions 3.6(i), S.3.1, and S.3.2 there is a δ_n so that $\delta_n \mathcal{D}_n(\mathbf{B}, \mathbf{E}) = o(r_n)$, $\delta_n \mathcal{D}_n(\mathbf{B}, \mathbf{E}) = o(1)$ if either Υ_F or Υ_G are not affine, $\mathcal{R}_n + \nu_n \tau_n = o(\delta_n)$, and

$$\ell_n \delta_n \mathcal{D}_n(\mathbf{B}, \mathbf{E}) 1\{K_f > 0\} = o(a_n n^{-\frac{1}{2}}) \quad \text{(S.60)}$$

$$K_m \delta_n \ell_n \mathcal{D}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}}) \quad \text{(S.61)}$$

$$k_n^{1/p} \sqrt{\log(1 + k_n)} B_n \times \sup_{P \in \mathbf{P}} J_{[\,]}(\delta_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n). \quad \text{(S.62)}$$

Next, notice that Corollary S.1.2(i) implies that there exists a $\theta_{0n} \in \Theta_{0n}^{\mathrm{r}}$ such that

$$\|\hat{\theta}_n - \theta_{0n}\|_{\mathbf{E}} = o_P(\delta_n) \quad \text{(S.63)}$$

uniformly in $P \in \mathbf{P}_0$ due to $(\mathcal{R}_n + \nu_n \tau_n) = o(\delta_n)$. Furthermore, by Assumption S.3.1 we can assume without loss of generality that θ_{0n} in addition satisfies

$$\|\hat{\theta}_n - \theta_{0n}\|_{\mathbf{L}} = o_P(\mathcal{D}_n(\mathbf{L}, \mathbf{E})\delta_n) \qquad \|\hat{\theta}_n - \theta_{0n}\|_{\mathbf{B}} = o_P(\mathcal{D}_n(\mathbf{B}, \mathbf{E})\delta_n) \quad \text{(S.64)}$$

uniformly in $P \in \mathbf{P}_0$. In addition note that since $\|q_{k,\jmath}\|_\infty \leq B_n$ for all $1 \leq \jmath \leq \mathcal{J}$ and $1 \leq k \leq k_{n,\jmath}$ by Assumption 3.2(i), we obtain from Assumption 3.3(ii) together with result (S.63) that with probability tending to one uniformly in $P \in \mathbf{P}_0$ we have

$$E_P[\|\rho(X, \hat{\theta}_n) - \rho(X, \theta_{0n})\|_2^2 q_{k,\jmath}^2(Z_\jmath)] \leq B_n^2 K_\rho^2 \delta_n^{2\kappa_\rho}. \quad \text{(S.65)}$$

Set $\mathcal{G}_n \equiv \{f q_{k,\jmath} : f \in \mathcal{F}_n, 1 \leq \jmath \leq \mathcal{J},\ 1 \leq k \leq k_{n,\jmath}\}$ and let $\mathbb{G}_P$ be a Gaussian process on $\mathcal{G}_n$ satisfying $E[\mathbb{G}_P(g_1)\mathbb{G}_P(g_2)] = \mathrm{Cov}_P\{g_1(V), g_2(V)\}$ and $E[\mathbb{G}_P(g_1)] = 0$ for any $g_1, g_2 \in \mathcal{G}_n$. Since (S.65) holds with probability tending to one uniformly in $P \in \mathbf{P}_0$, Assumption 3.7(ii), result (S.45), and δ_n satisfying (S.62) imply for any $\epsilon > 0$ that

$$\begin{aligned}&\limsup_{n\to\infty}\sup_{P\in\mathbf{P}_0} P(\|\mathbb{W}^\star_P(\hat{\theta}_n) - \mathbb{W}^\star_P(\theta_{0n})\|_{\Sigma_P,p} > a_n\epsilon)\\ &\leq \limsup_{n\to\infty}\sup_{P\in\mathbf{P}_0}\frac{1}{a_n\epsilon}k_n^{1/p}E_P[\sup_{g_1,g_2\in\mathcal{G}_n:\|g_1-g_2\|_{P,2}\leq B_nK_\rho\delta_n^{\kappa_\rho}}|\mathbb{G}_P(g_1)-\mathbb{G}_P(g_2)|] = 0. \end{aligned} \quad \text{(S.66)}$$

Similarly, result (S.63) implies $\overrightarrow{d}_H(\hat{\theta}_n, \Theta^{\mathrm{r}}_{0n}, \|\cdot\|_{\mathbf{E}}) \leq \epsilon$ with probability tending to one uniformly in $P\in\mathbf{P}_0$ for any $\epsilon > 0$. Hence, Lemma S.3.4 yields uniformly in $P \in \mathbf{P}_0$

$$\begin{aligned}\|\mathbb{D}_P(\theta_{0n})[\hat{h}_n] - \mathbb{D}_P(\hat{\theta}_n)[\hat{h}_n]\|_{\Sigma_P,p} &\lesssim \|\Sigma_P\|_{o,p}\times K_m\|\hat{\theta}_n-\theta_{0n}\|_{\mathbf{L}}\|\hat{h}_n\|_{\mathbf{E}} + o_P(a_n)\\ &\lesssim \|\Sigma_P\|_{o,p}\times K_m\mathcal{D}_n(\mathbf{L},\mathbf{E})\delta_n\ell_n\sqrt{n} + o_P(a_n) = o_P(a_n),\end{aligned} \quad \text{(S.67)}$$

where the second inequality follows from $\|\hat{h}_n/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ due to $\hat{h}_n/\sqrt{n} \in \hat{V}_n(\hat{\theta}_n, R|\ell_n)$, Assumption 3.12(i), and (S.64). In turn, the final result in (S.67) follows from (S.61) and Assumption 3.7(ii). Next, note the conditions of Theorem S.5.1(i) hold because: Either Υ_F and Υ_G are affine (implying $K_f = K_g = 0$) or $\delta_n\mathcal{D}_n(\mathbf{B},\mathbf{E}) = o(1)$, and $\delta_n\mathcal{D}_n(\mathbf{B},\mathbf{E}) = o(r_n)$ and $\limsup \ell_n/r_n 1\{K_g > 0\} < 1/2$ by Assumption S.3.2(iii) imply

$$r_n \geq 2(\ell_n + \delta_n\mathcal{D}_n(\mathbf{B},\mathbf{E}))1\{K_g > 0\}$$

for n sufficiently large. Hence, Theorem S.5.1(i), Assumption 3.12(ii), and $\|h\|_{\mathbf{E}} \lesssim \|h\|_{\mathbf{B}}$ for all $h \in \mathbf{B}_n$ by Assumption 3.12(i), imply that there is a constant $M < \infty$ for which with probability tending to one uniformly in $P \in \mathbf{P}_0$ we have that

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

It follows from Assumption S.3.2(ii) and (S.60) that there is a $h_{0n} \in V_n(\theta_{0n}, R|M\ell_n)$ such that $\|h_{0n} - \hat{h}_n\|_{\mathbf{B}} = o_P(a_n)$ uniformly in $P \in \mathbf{P}_0$, and hence Assumption 3.7(ii), Lemma S.3.4, and $\|h\|_{\mathbf{E}} \lesssim \|h\|_{\mathbf{B}}$ by Assumption 3.12(i) yield

$$\|\mathbb{D}_P(\theta_{0n})[\hat{h}_n] - \mathbb{D}_P(\theta_{0n})[h_{0n}]\|_{\Sigma_P,p} \lesssim \|\Sigma_P\|_{o,p}\times\|\hat{h} - h_{0n}\|_{\mathbf{E}} = o_P(a_n) \quad \text{(S.68)}$$

uniformly in $P \in \mathbf{P}_0$. Therefore, combining results (S.59), (S.66), (S.67), and (S.68) together with $\theta_{0n} \in \Theta^{\mathrm{r}}_{0n}$ and $h_{0n} \in V_n(\theta_{0n}, R|M\ell_n)$ yields

$$\begin{aligned}\hat{U}_n(R|\ell_n) &= \|\mathbb{W}^\star_P(\theta_{0n}) + \mathbb{D}_P(\theta_{0n})[h_{0n}]\|_{\Sigma_P,p} + o_P(a_n)\\ &\geq \inf_{\theta\in\Theta^{\mathrm{r}}_{0n}}\inf_{h\in V_n(\theta,R|M\ell_n)}\|\mathbb{W}^\star_P(\theta) + \mathbb{D}_P(\theta)[h]\|_{\Sigma_P,p} + o_P(a_n)\end{aligned}$$

uniformly in $P \in \mathbf{P}_0$. The first part of theorem then follows by setting $\tilde{\ell}_n = M\ell_n$.

In order to establish the second part of the theorem, note that the only assumptions that potentially require the norm $\|\cdot\|_{\mathbf{B}}$ to be stronger than $\|\cdot\|_{\mathbf{I}}$ are Assumptions 3.8, 3.9, 3.10 (pertaining to the differentiability of Υ_F and Υ_G) and Assumption 3.12(ii) (since a stronger norm $\|\cdot\|_{\mathbf{B}}$ makes $(\hat{\Theta}_n^{\mathrm{r}})^{\epsilon}$ smaller). We therefore establish part (ii) of the theorem by repeating the arguments employed in showing part (i) while carefully re-examining the role played by the norm $\|\cdot\|_{\mathbf{B}}$. To this end, note that since

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(\{\theta \in \mathbf{B}_n : \overrightarrow{d}_H(\theta, \hat{\Theta}_n^{\mathrm{r}}, \|\cdot\|_{\mathbf{I}}) \leq \epsilon\} \subseteq \Theta_n) = 1, \tag{S.69}$$

we may apply Lemma S.3.2 with $\|\cdot\|_{\mathbf{B}}$ set to equal $\|\cdot\|_{\mathbf{I}}$ to still obtain that

$$\hat{U}_n(R|\ell_n) = \inf_{\theta\in\hat{\Theta}_n^{\mathrm{r}}} \inf_{h\in\hat{V}_n(\theta,R|\ell_n)} \|\mathbb{W}_P^{\star}(\theta) + \mathbb{D}_P(\theta)[h]\|_{\Sigma_P,p} + o_P(a_n). \tag{S.70}$$

Letting $\hat{\theta}_n$ and $\hat{h}_n$ be defined as in (S.59) (but with $\hat{V}_n(\theta, R|\ell)$ as defined in (S.57)), then observe that since results (S.66) and (S.67) do not rely on Assumptions 3.8, 3.9, 3.10 or 3.12(ii), we can conclude from result (S.70) that uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R|\ell_n) = \|\mathbb{W}_P^{\star}(\theta_{0n}) + \mathbb{D}_P(\theta_{0n})[\hat{h}_n]\|_{\Sigma_P,p} + o_P(a_n) \tag{S.71}$$

for some $\theta_{0n} \in \Theta_{0n}^{\mathrm{r}}$. Next, note $\delta_n\mathcal{D}_n(\mathbf{B},\mathbf{E}) = o(r_n)$ and $K_f = K_g = 0$ due to Υ_F and Υ_G being affine, together with Theorem S.5.1(ii) imply that

$$\begin{aligned}
\hat{V}_n(\hat{\theta}_n, R|\ell_n) &\equiv \{h \in \mathbf{B}_n : h \in G_n(\hat{\theta}_n),\ \Upsilon_F(\hat{\theta}_n + \frac{h}{\sqrt{n}}) = 0,\ \|\frac{h}{\sqrt{n}}\|_{\mathbf{I}} \leq \ell_n\} \\
&\subseteq \{h \in \mathbf{B}_n : \Upsilon_G(\theta_{0n} + \frac{h}{\sqrt{n}}) \leq 0,\ \Upsilon_F(\theta_{0n} + \frac{h}{\sqrt{n}}) = 0,\ \|\frac{h}{\sqrt{n}}\|_{\mathbf{I}} \leq \ell_n\} \\
&\subseteq V_n(\theta_{0n}, R|\ell_n),
\end{aligned}$$

with probability tending to one uniformly in $P \in \mathbf{P}_0$, and where the final inequality follows from $\ell_n \downarrow 0$, $\{\theta \in \mathbf{B}_n : \overrightarrow{d}_H(\theta, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{I}}) \leq \epsilon\} \subseteq \Theta_n$ and $\|\cdot\|_{\mathbf{E}} \leq \|\cdot\|_{\mathbf{I}}$ by hypothesis. Therefore, we can conclude that $\hat{h}_n \in V_n(\theta_{0n}, R|\ell_n)$ with probability tending to one uniformly in $P \in \mathbf{P}_0$, which by (S.71) yields

$$\hat{U}_n(R|\ell_n) \geq \inf_{\theta\in\Theta_{0n}^{\mathrm{r}}} \inf_{h\in V_n(\theta,R|\ell_n)} \|\mathbb{W}_P^{\star}(\theta_{0n}) + \mathbb{D}_P(\theta_{0n})[h]\|_{\Sigma_P,p} + o_P(a_n),$$

and hence establishes the second claim of the theorem. ■

Proof of Theorem 3.2: Follows from immediately from Theorem S.3.1(i) and Assumption 3.13 implying Assumptions S.3.1 and S.3.2 are satisfied by setting $\mathcal{D}_n(\mathbf{B},\mathbf{E}) = \mathcal{S}_n(\mathbf{B},\mathbf{E})$, $\mathcal{D}_n(\mathbf{L},\mathbf{E}) = \mathcal{S}_n(\mathbf{L},\mathbf{E})$ and $\theta = \tilde{\theta}$. ■

Proof of Corollary 3.1: We establish the corollary by appealing to Lemmas S.3.5

and S.3.6. To this end, we first note Theorem 3.2 allows us to conclude that

$$\hat{U}_n(R|\ell_n) \geq U_P^\star(R|\tilde{\ell}_n) + o_P(a_n) \tag{S.72}$$

uniformly in $P \in \mathbf{P}_0$ with $\ell_n \asymp \tilde{\ell}_n$, while Assumption 3.13(ii) implies $K_m \tilde{\ell}_n^2 \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-\frac{1}{2}})$ and $k_n^{1/p}\sqrt{\log(1+k_n)} B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}(\tilde{\ell}_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$, and hence

$$I_n(R) \leq U_P(R|\tilde{\ell}_n) + o_P(a_n) \tag{S.73}$$

uniformly in $P \in \mathbf{P}_0$ by Theorem 3.1(i). Moreover, applying Lemma S.3.5 with $B_n = \hat{U}_n(R|\ell_n)$, $D_n = \{V_i\}_{i=1}^n$, and $C_{P,n}^\star = U_P^\star(R|\tilde{\ell}_n)$ yields, for some $\delta_n = o(1)$, that

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(\hat{c}_n + \frac{a_n}{2} > q_{1-\alpha-\delta_n,P}(U_P^\star(R|\tilde{\ell}_n))) = 1, \tag{S.74}$$

where $q_{\tau,P}(U_P^\star(R|\tilde{\ell}_n))$ denotes the τ quantile of $U_P^\star(R|\tilde{\ell}_n)$. Since $U_P^\star(R|\tilde{\ell}_n) \stackrel{d}{=} U_P(R|\tilde{\ell}_n)$, results (S.73), (S.74), and Assumption 3.14 verify the conditions of Lemma S.3.6 (applied with $T_n = I_n(R)$ and $C_{P,n} = U_P(R|\tilde{\ell}_n)$) and therefore the corollary follows. ■

Proof of Corollary 3.2: In what follows, we use a "u" superscript for parameters associated with setting $R = \Theta$ – e.g., $\mathbf{B}_n^{\mathrm{u}}$ denotes the vector subspace generated by Θ_n. First note Theorem 3.1(i) (for R as in (13)) and Theorem 3.1(ii) (for $R = \Theta$) imply that for any $\ell_n, \ell_n^{\mathrm{u}} \downarrow 0$ satisfying $k_n^{1/p}\sqrt{\log(1+k_n)} B_n \times \{\sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) + \sup_{P\in\mathbf{P}} J_{[\,]}((\ell_n^{\mathrm{u}})^{\kappa_\rho}, \mathcal{F}_n^{\mathrm{u}}, \|\cdot\|_{P,2})\} = o(a_n)$, $K_m(\ell_n^{\mathrm{u}})^2 \times \mathcal{S}_n^{\mathrm{u}}(\mathbf{L}, \mathbf{E}) = o(a_n n^{-1/2})$, $K_m \ell_n^2 \times \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-1/2})$ and $\mathcal{R}_n^{\mathrm{u}} = o(\ell_n^{\mathrm{u}})$ it follows uniformly in $P \in \mathbf{P}_0$ that

$$I_n(R) - I_n(\Theta) \leq U_P(R|\ell_n) - U_P(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n). \tag{S.75}$$

Next note that we may apply Theorem 3.2 to obtain that uniformly in $P \in \mathbf{P}_0$ we have

$$\hat{U}_n(R|\ell_n) \geq U_P^\star(R|\tilde{\ell}_n) + o_P(a_n) \tag{S.76}$$

with $\tilde{\ell}_n \asymp \ell_n$. Similarly, also note that Lemma S.3.7 implies uniformly in $P \in \mathbf{P}_0$ that

$$\hat{U}_n(\Theta| + \infty) \leq U_P^\star(\Theta|\tilde{\ell}_n^{\mathrm{u}}) + o_P(a_n), \tag{S.77}$$

for $\tilde{\ell}_n^{\mathrm{u}} \downarrow 0$ satisfying Assumption 3.13(ii) and $\mathcal{R}_n^{\mathrm{u}} = o(\tilde{\ell}_n^{\mathrm{u}})$. In particular, it follows from results (S.76) and (S.77) that uniformly in $P \in \mathbf{P}_0$ we have

$$\hat{U}_n(R|\ell_n) - \hat{U}_n(\Theta| + \infty) \geq U_P^\star(R|\tilde{\ell}_n) - U_P^\star(\Theta|\tilde{\ell}_n^{\mathrm{u}}) + o_P(a_n) \tag{S.78}$$

for sequences $\tilde{\ell}_n, \tilde{\ell}_n^{\mathrm{u}} \downarrow 0$ satisfying the rate requirements needed for (S.75) to hold (i.e. with $\ell_n, \ell_n^{\mathrm{u}}$ replaced by $\tilde{\ell}_n, \tilde{\ell}_n^{\mathrm{u}}$). The corollary then follows by the same arguments as in Corollary 3.1 but employing (S.75) and (S.78) in place of (S.72) and (S.73). ■

Lemma S.3.1. *Suppose there is a $\mathcal{A}_n(P) \subseteq \Theta_n \cap R$ such that $\|h\|_{\mathbf{E}} \leq \nu_n \|\mathbb{D}_P(\theta)[h]\|_p$ for all $\theta \in \mathcal{A}_n(P)$ and $h \in \sqrt{n}\{\mathbf{B}_n \cap R - \theta\}$. If the estimator $\hat{\mathbb{D}}_n(\theta)$ satisfies*

$$\sup_{\theta \in \mathcal{A}_n(P)} \sup_{h \in \sqrt{n}\{\mathbf{B}_n \cap R - \theta\} : \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \geq \ell_n} \frac{\|\hat{\mathbb{D}}_n(\theta)[h] - \mathbb{D}_P(\theta)[h]\|_p}{\|h\|_{\mathbf{E}}} = o_P(\nu_n^{-1}) \tag{S.79}$$

and $\hat{\Theta}_n^{\mathrm{r}} \subseteq \mathcal{A}_n(P)$ with probability tending to one uniformly in $P \in \mathbf{P}_0$, Assumptions 3.2(i)(iii), 3.7, 3.11 hold, and $\mathcal{S}_n(\mathbf{B}, \mathbf{E})\mathcal{R}_n = o(\ell_n)$, then uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(R|\ell_n) = \inf_{\theta \in \hat{\Theta}_n^{\mathrm{r}}} \inf_{h \in \hat{V}_n(\theta, R|+\infty)} \|\hat{\mathbb{W}}_n(\theta) + \hat{\mathbb{D}}_n(\theta)[h]\|_{\hat{\Sigma}_n, p} + o_P(a_n). \tag{S.80}$$

PROOF: In the following arguments, we note that the only requirement on $\hat{\mathbb{D}}_n(\theta)$ is that it satisfy condition (S.79). As a result, the lemma applies to estimators $\hat{\mathbb{D}}_n(\theta)$ besides the numerical derivative examined in the main text.

In order to establish the result, we first let $\hat{\theta}_n \in \hat{\Theta}_n^{\mathrm{r}}$ and $\hat{h}_n \in \hat{V}_n(\hat{\theta}_n, R|+\infty)$ satisfy

$$\inf_{\theta \in \hat{\Theta}_n^{\mathrm{r}}} \inf_{h \in \hat{V}_n(\theta, R|+\infty)} \|\hat{\mathbb{W}}_n(\theta) + \hat{\mathbb{D}}_n(\theta)[h]\|_{\hat{\Sigma}_n, p} = \|\hat{\mathbb{W}}_n(\hat{\theta}_n) + \hat{\mathbb{D}}_n(\hat{\theta}_n)[\hat{h}_n]\|_{\hat{\Sigma}_n, p} + o(a_n).$$

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

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(\|\frac{\hat{h}_n}{\sqrt{n}}\|_{\mathbf{B}} \geq \ell_n) = 0. \tag{S.81}$$

To this end, note $0 \in \hat{V}_n(\theta, R|+\infty)$ for all $\theta \in \Theta_n \cap R$, the triangle inequality, $\|\hat{\Sigma}_n\|_{o,p} = O_P(1)$ uniformly in $P \in \mathbf{P}$ by Corollary S.1.3, and Assumption 3.11 yield

$$\begin{aligned} \|\hat{\mathbb{D}}_n(\hat{\theta}_n)[\hat{h}_n]\|_{\hat{\Sigma}_n, p} &\leq \|\hat{\mathbb{W}}_n(\hat{\theta}_n) + \hat{\mathbb{D}}_n(\hat{\theta}_n)[\hat{h}_n]\|_{\hat{\Sigma}_n, p} + \|\hat{\mathbb{W}}_n(\hat{\theta}_n)\|_{\hat{\Sigma}_n, p} \\ &\leq 2\|\hat{\Sigma}_n\|_{o,p} \|\mathbb{W}_P^\star(\hat{\theta}_n)\|_p + o_P(a_n) \end{aligned} \tag{S.82}$$

uniformly in $P \in \mathbf{P}$. Hence, since $\hat{\theta}_n \in \hat{\Theta}_n^{\mathrm{r}} \subseteq \Theta_n \cap R$ almost surely, we obtain from result (S.82), $\|\hat{\Sigma}_n\|_{o,p} = O_P(1)$ uniformly in $P \in \mathbf{P}$, and Lemma S.2.4 that

$$\|\hat{\mathbb{D}}_n(\hat{\theta}_n)[\hat{h}_n]\|_{\hat{\Sigma}_n, p} \leq 2\|\hat{\Sigma}_n\|_{o,p} \sup_{\theta \in \Theta_n \cap R} \|\mathbb{W}_P^\star(\theta)\|_p + o_P(a_n) = O_P(k_n^{1/p}\sqrt{\log(1+k_n)} B_n J_n) \tag{S.83}$$

uniformly in $P \in \mathbf{P}$. Since $\hat{h}_n \in \hat{V}_n(\hat{\theta}_n, R|+\infty)$ implies $\hat{h}_n \in \sqrt{n}\{\mathbf{B}_n \cap R - \hat{\theta}_n\}$ and $\hat{\theta}_n \in \hat{\Theta}_n^{\mathrm{r}} \subseteq \mathcal{A}_n(P)$ with probability tending to one uniformly in $P \in \mathbf{P}_0$, we obtain from the first hypothesis of the lemma that $\|\hat{h}_n\|_{\mathbf{E}} \leq \nu_n \|\mathbb{D}_P(\hat{\theta}_n)[\hat{h}_n]\|_p$ with probability tending to one uniformly in $P \in \mathbf{P}_0$. Therefore, it follows that

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\ell_n \leq \|\frac{\hat{h}_n}{\sqrt{n}}\|_{\mathbf{B}}) \\
&\qquad = \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\ell_n \leq \|\frac{\hat{h}_n}{\sqrt{n}}\|_{\mathbf{B}} \text{ and } \|\hat{h}_n\|_{\mathbf{E}} \leq \nu_n \|\mathbb{D}_P(\hat{\theta}_n)[\hat{h}_n]\|_p) \\
&\qquad \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\ell_n \leq \|\frac{\hat{h}_n}{\sqrt{n}}\|_{\mathbf{B}} \text{ and } \|\hat{h}_n\|_{\mathbf{E}} \leq 2\nu_n \|\hat{\mathbb{D}}_n(\hat{\theta}_n)[\hat{h}_n]\|_p), \qquad \text{(S.84)}
\end{aligned}$$

where the inequality follows from condition (S.79). Hence, results (S.83) and (S.84), 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

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\ell_n \leq \|\frac{\hat{h}_n}{\sqrt{n}}\|_{\mathbf{B}}) \\
&\qquad \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\ell_n \leq 2\frac{\nu_n}{\sqrt{n}} \mathcal{S}_n(\mathbf{B},\mathbf{E}) \|\hat{\mathbb{D}}_n(\hat{\theta}_n)[\hat{h}_n]\|_p) = 0, \qquad \text{(S.85)}
\end{aligned}$$

which establishes (S.81) and hence the claim of the lemma. ■

Lemma S.3.2. *Let Assumptions 3.1(i), 3.2, 3.3, 3.4(i), 3.5(i), 3.6(ii), 3.7, 3.11, 3.12 hold and $\mathcal{R}_n \vee \nu_n\tau_n = o(1)$. If $\ell_n \downarrow 0$ satisfies $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$ and $K_m \ell_n^2 \mathcal{S}_n(\mathbf{L},\mathbf{E}) = o(a_n n^{-\frac{1}{2}})$, then uniformly in $P \in \mathbf{P}_0$ we have*

$$\hat{U}_n(R|\ell_n) = \inf_{\theta\in\hat{\Theta}_n^{\mathrm{r}}} \inf_{h\in\hat{V}_n(\theta,R|\ell_n)} \|\mathbb{W}_P^\star(\theta) + \mathbb{D}_P(\theta)[h]\|_{\Sigma_P,p} + o_P(a_n).$$

Proof: First note that Corollary S.1.2(i) implies $\vec{d}_H(\hat{\Theta}_n^{\mathrm{r}}, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{E}}) = O_P(\mathcal{R}_n + \nu_n\tau_n)$ uniformly in $P \in \mathbf{P}_0$. Hence, since $\mathcal{R}_n \vee \nu_n\tau_n = o(1)$, for any $\epsilon > 0$ it follows that

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(\hat{\Theta}_n^{\mathrm{r}} \subseteq \{\theta \in \Theta_n \cap R : \vec{d}_H(\theta, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{E}}) \leq \epsilon\}) = 1. \qquad \text{(S.86)}$$

Furthermore, for any $\theta \in \hat{\Theta}_n^{\mathrm{r}}$ and $h \in \hat{V}_n(\theta, R|\ell_n)$ note that $\Upsilon_G(\theta + h/\sqrt{n}) \leq 0$ and $\Upsilon_F(\theta + h/\sqrt{n}) = 0$ by definition of $\hat{V}_n(\theta, R|\ell_n)$. Thus, $\theta + h/\sqrt{n} \in R$ for any $\theta \in \hat{\Theta}_n^{\mathrm{r}}$ and $h \in \hat{V}_n(\theta, R|\ell_n)$, and hence Assumption 3.12(ii) allows us to conclude

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(\theta + \frac{h}{\sqrt{n}} \in \Theta_n \cap R \text{ for all } \theta \in \hat{\Theta}_n^{\mathrm{r}} \text{ and } h \in \hat{V}_n(\theta, R|\ell_n)) \\
&\qquad = \liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(\theta + \frac{h}{\sqrt{n}} \in \Theta_n \text{ for all } \theta \in \hat{\Theta}_n^{\mathrm{r}} \text{ and } h \in \hat{V}_n(\theta, R|\ell_n)) = 1 \qquad \text{(S.87)}
\end{aligned}$$

due to $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n \downarrow 0$ for any $h \in \hat{V}_n(\theta, R|\ell_n)$. In particular, note that result (S.87) and Assumption 3.12(i) imply that for some $M < \infty$ we have $\hat{V}_n(\theta, R|\ell_n) \subseteq V_n(\theta, R|\ell_n/M)$ for all $\theta \in \hat{\Theta}_n^{\mathrm{r}}$ with probability tending to one uniformly in $P \in \mathbf{P}_0$.

Thus, (S.86) and Lemma S.3.3 allow us to conclude that uniformly in $P \in \mathbf{P}_0$ we have

$$\sup_{\theta \in \hat{\Theta}_n^{\mathrm{r}}} \sup_{h \in \hat{V}_n(\theta, R|\ell_n)} \|\hat{\mathbb{D}}_n(\theta)[h] - \mathbb{D}_P(\theta)[h]\|_p = o_P(a_n). \tag{S.88}$$

Moreover, since $\hat{\Theta}_n^{\mathrm{r}} \subseteq \Theta_n \cap R$ almost surely, we also have from Assumption 3.11 that

$$\sup_{\theta \in \hat{\Theta}_n^{\mathrm{r}}} \|\hat{\mathbb{W}}_n(\theta) - \mathbb{W}_P^\star(\theta)\|_p = o_P(a_n) \tag{S.89}$$

uniformly in $P \in \mathbf{P}$. Therefore, since $\|\hat{\Sigma}_n\|_{o,p} = O_P(1)$ uniformly in $P \in \mathbf{P}$ by Corollary S.1.3, we obtain from results (S.88) and (S.89) and Lemma S.2.6 that

$$\hat{U}_n(R|\ell_n) = \inf_{\theta \in \hat{\Theta}_n^{\mathrm{r}}} \inf_{h \in \hat{V}_n(\theta, R|\ell_n)} \|\mathbb{W}_P^\star(\theta) + \mathbb{D}_P(\theta)[h]\|_{\hat{\Sigma}_n, p} + o_P(a_n) \tag{S.90}$$

uniformly in $P \in \mathbf{P}_0$. Next, note that by Assumption 3.7(ii) there exists a constant $C_0 < \infty$ such that $\|\Sigma_P^{-1}\|_{o,p} \le C_0$ for all $P \in \mathbf{P}$. Thus, using that $\|\hat{\Sigma}_n a\|_p \le \|\hat{\Sigma}_n \Sigma_P^{-1}\|_{o,p} \|\Sigma_P a\|_p$ for any $a \in \mathbf{R}^{k_n}$ and the triangle inequality we obtain

$$\|\mathbb{W}_P^\star(\theta) + \mathbb{D}_P(\theta)[h]\|_{\hat{\Sigma}_n, p} \le \{C_0 \|\hat{\Sigma}_n - \Sigma_P\|_{o,p} + 1\} \|\mathbb{W}_P^\star(\theta) + \mathbb{D}_P(\theta)[h]\|_{\Sigma_P, p} \tag{S.91}$$

for any $\theta \in \Theta_n \cap R$, $h \in \mathbf{B}_n$, and $P \in \mathbf{P}$. In particular, since $0 \in \hat{V}_n(\theta, R|\ell_n)$ for any $\theta \in \Theta_n \cap R$, Assumption 3.7, Markov's inequality, and Lemma S.2.4 yield

$$\begin{aligned} \|\hat{\Sigma}_n - \Sigma_P\|_{o,p} \times \inf_{\theta \in \hat{\Theta}_n^{\mathrm{r}}} \inf_{h \in \hat{V}_n(\theta, R|\ell_n)} &\|\mathbb{W}_P^\star(\theta) + \mathbb{D}_P(\theta)[h]\|_{\Sigma_P, p} \\ &\le \|\hat{\Sigma}_n - \Sigma_P\|_{o,p} \times \sup_{\theta \in \Theta_n \cap R} \|\mathbb{W}_P^\star(\theta)\|_{\Sigma_P, p} = o_P(a_n) \end{aligned} \tag{S.92}$$

uniformly in $P \in \mathbf{P}$. It then follows from (S.91) and (S.92) that uniformly in $P \in \mathbf{P}$

$$\begin{aligned} \inf_{\theta \in \hat{\Theta}_n^{\mathrm{r}}} \inf_{h \in \hat{V}_n(\theta, R|\ell_n)} &\|\mathbb{W}_P^\star(\theta) + \mathbb{D}_P(\theta)[h]\|_{\hat{\Sigma}_n, p} \\ &\le \inf_{\theta \in \hat{\Theta}_n^{\mathrm{r}}} \inf_{h \in \hat{V}_n(\theta, R|\ell_n)} \|\mathbb{W}_P^\star(\theta) + \mathbb{D}_P(\theta)[h]\|_{\Sigma_P, p} + o_P(a_n). \end{aligned} \tag{S.93}$$

The reverse inequality to (S.93) can be obtained by identical arguments but employing $\max\{\|\hat{\Sigma}_n\|_{o,p}, \|\hat{\Sigma}_n^{-1}\|_{o,p}\} = O_P(1)$ uniformly in $P \in \mathbf{P}$ by Corollary S.1.3 instead of $\|\Sigma_P\|_{o,p} \vee \|\Sigma_P^{-1}\|_{o,p}$ being bouded uniformly in $P \in \mathbf{P}$. The claim of the Lemma then follows from (S.90) and (S.93) (and its reverse inequality). ■

Lemma S.3.3. *Let Assumptions 3.2(i)(ii), 3.3, and 3.5(i) hold, and define the sets*

$$V_n(\theta, R|\ell_n) \equiv \{h \in \mathbf{B}_n : \theta + \frac{h}{\sqrt{n}} \in \Theta_n \cap R \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{E}} \le \ell_n\}. \tag{S.94}$$

If $\ell_n \downarrow 0$ satisfies $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$ 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 $P \in \mathbf{P}_0$:

$$\sup_{\theta\in\Theta_n\cap R:\overrightarrow{d}_H(\theta,\Theta_{0n}^{\mathrm{r}},\|\cdot\|_{\mathbf{E}})\le\epsilon}\ \sup_{h\in V_n(\theta,R|\ell_n)} \|\hat{\mathbb{D}}_n(\theta)[h] - \mathbb{D}_P(\theta)[h]\|_p = o_P(a_n). \tag{S.95}$$

Proof: By definition of $V_n(\theta, R|\ell_n)$, we have $\theta + h/\sqrt{n} \in \Theta_n \cap R$ for any $\theta \in \Theta_n \cap R$, $h \in V_n(\theta, R|\ell_n)$. Therefore, since $\|h/\sqrt{n}\|_{\mathbf{E}} \le \ell_n$ for all $h \in V_n(\theta, R|\ell_n)$ we obtain

$$\begin{aligned}\sup_{\theta\in\Theta_n\cap R}\ \sup_{h\in V_n(\theta,R|\ell_n)} &\|\hat{\mathbb{D}}_n(\theta)[h] - \sqrt{n}E_P[(\rho(X,\theta+\frac{h}{\sqrt{n}}) - \rho(X,\theta)) * q^{k_n}(Z)]\|_p \\ &\le \sup_{\theta_1,\theta_2\in\Theta_n\cap R:\|\theta_1-\theta_2\|_{\mathbf{E}}\le\ell_n} \|\mathbb{G}_n(\theta_1) - \mathbb{G}_n(\theta_2)\|_p \\ &\le \sup_{\theta_1,\theta_2\in\Theta_n\cap R:\|\theta_1-\theta_2\|_{\mathbf{E}}\le\ell_n} \|\mathbb{W}_P(\theta_1) - \mathbb{W}_P(\theta_2)\|_p + o_P(a_n)\end{aligned} \tag{S.96}$$

uniformly in $P \in \mathbf{P}$ by Assumption 3.3(i). Next note Assumptions 3.2(i) and 3.3(ii) imply that for any $1 \le \jmath \le \mathcal{J}$ and $1 \le k \le k_{n,\jmath}$ we must have

$$\sup_{P\in\mathbf{P}}\ \sup_{\theta_1,\theta_2\in\Theta_n\cap R:\|\theta_1-\theta_2\|_{\mathbf{E}}\le\ell_n} E_P[\|\rho(X,\theta_1) - \rho(X,\theta_2)\|_2^2 q_{k,\jmath}^2(Z_\jmath)] \le B_n^2 K_\rho^2 \ell_n^{2\kappa_\rho}. \tag{S.97}$$

Define $\mathcal{G}_n \equiv \{f q_{k,\jmath} : f \in \mathcal{F}_n, 1 \le \jmath \le \mathcal{J} \text{ and } 1 \le k \le k_{n,\jmath}\}$ and let $\mathbb{G}_P$ be a Gaussian process on $\mathcal{G}_n$ satisfying $E[\mathbb{G}_P(g_1)\mathbb{G}_P(g_2)] = \mathrm{Cov}_P\{g_1(V), g_2(V)\}$ and $E[\mathbb{G}_P(g_1)] = 0$ for any $g_1, g_2 \in \mathcal{G}_n$. By result (S.97) and $\|a\|_p \le k_n^{1/p}\|a\|_\infty$ for any $a \in \mathbf{R}^{k_n}$ we obtain

$$\begin{aligned}E[\sup_{\theta_1,\theta_2\in\Theta_n\cap R:\|\theta_1-\theta_2\|_{\mathbf{E}}\le\ell_n} &\|\mathbb{W}_P(\theta_1) - \mathbb{W}_P(\theta_2)\|_p] \\ &\le k_n^{1/p} \times E[\sup_{g_1,g_2\in\mathcal{G}_n:\|g_1-g_2\|_{P,2}\le B_nK_\rho\ell_n^{\kappa_\rho}} |\mathbb{G}_P(g_1) - \mathbb{G}_P(g_2)|].\end{aligned} \tag{S.98}$$

Therefore, the calculations in (S.45), Markov's inequality, and $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$ by hypothesis, yield that

$$\sup_{\theta\in\Theta_n\cap R}\ \sup_{h\in V_n(\theta,R|\ell_n)} \|\hat{\mathbb{D}}_n(\theta)[h] - \sqrt{n}E_P[(\rho(X,\theta+\frac{h}{\sqrt{n}}) - \rho(X,\theta)) * q^{k_n}(Z)]\|_p = o_P(a_n) \tag{S.99}$$

uniformly in $P \in \mathbf{P}$. Next, let $\epsilon > 0$ be sufficiently small for Assumption 3.5(i) to hold and define the neighborhood $\mathcal{N}_n \equiv \{\theta \in \Theta_n \cap R : \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{E}}) \le \epsilon\}$. We can then conclude from Lemmas S.1.5 and S.2.5, and Assumption 3.5(i) that

$$\sup_{\theta\in\mathcal{N}_n}\sup_{h\in V_n(\theta,R|\ell_n)}\|\sqrt{n}E_P[(\rho(X,\theta+\frac{h}{\sqrt{n}})-\rho(X,\theta))*q^{k_n}(Z)]-\mathbb{D}_P(\theta)[h]\|_p$$
$$\lesssim \sup_{\theta\in\mathcal{N}_n}\sup_{h\in V_n(\theta,R|\ell_n)}\{K_m\times\sqrt{n}\|\frac{h}{\sqrt{n}}\|_{\mathbf{E}}\|\frac{h}{\sqrt{n}}\|_{\mathbf{L}}\}=o(a_n), \quad \text{(S.100)}$$

where the final equality follows from $K_m\ell_n^2\times\mathcal{S}_n(\mathbf{L},\mathbf{E})=o(a_n n^{-1/2})$ by hypothesis. Hence, the Lemma follows from results (S.99) and (S.100). ∎

Lemma S.3.4. *Let Assumptions 3.2(ii) and 3.5(ii)(iii) hold. Then there are constants $\epsilon>0$ and $C<\infty$ such that for all n, $P\in\mathbf{P}$, $\theta_0\in\Theta_{0n}^{\mathrm{r}}$, $\theta_1\in\Theta_n\cap R$ satisfying $\vec{d}_H(\theta_1,\Theta_{0n}^{\mathrm{r}},\|\cdot\|_{\mathbf{E}})\leq\epsilon$, and $h_0,h_1\in\mathbf{B}_n$ it follows that*

$$\|\mathbb{D}_P(\theta_0)[h_0]-\mathbb{D}_P(\theta_1)[h_1]\|_p\leq C\{\|h_0-h_1\|_{\mathbf{E}}+K_m\|\theta_0-\theta_1\|_{\mathbf{L}}\|h_1\|_{\mathbf{E}}\}.$$

Proof: We first fix $\epsilon>0$ such that Assumptions 3.5(ii)(iii) are satisfied. Then note that by Lemmas S.1.5 and S.2.5 it follows that there is a constant $C_0<\infty$ with

$$\|\mathbb{D}_P(\theta_0)[h_0]-\mathbb{D}_P(\theta_1)[h_1]\|_p\leq\{\sum_{\jmath=1}^{\mathcal{J}}C_0\|\nabla m_{P,\jmath}(\theta_0)[h_0]-\nabla m_{P,\jmath}(\theta_1)[h_1]\|_{P,2}^2\}^{1/2}.$$

Moreover, since $(h_0-h_1)\in\mathbf{B}_n$, we can also conclude from Assumptions 3.5(ii)(iii) that

$$\begin{aligned}&\|\nabla m_{P,\jmath}(\theta_0)[h_0]-\nabla m_{P,\jmath}(\theta_1)[h_1]\|_{P,2}\\&\quad\leq\|\nabla m_{P,\jmath}(\theta_0)[h_0-h_1]\|_{P,2}+\|\nabla m_{P,\jmath}(\theta_0)[h_1]-\nabla m_{P,\jmath}(\theta_1)[h_1]\|_{P,2}\\&\quad\leq M\|h_0-h_1\|_{\mathbf{E}}+K_m\|\theta_1-\theta_0\|_{\mathbf{L}}\|h_1\|_{\mathbf{E}}\end{aligned}$$

for some $M<\infty$, and therefore the claim of the lemma follows. ∎

Lemma S.3.5. *Let B_n and D_n be observable random variables, $C_{P,n}^\star$ be a potentially unobservable random variable depending on $P\in\mathbf{P}$, and for any $\alpha\in(0,1)$ define*

$$\hat{q}_\alpha\equiv\inf\{u:P(B_n\leq u|D_n)\geq\alpha\}\qquad q_{\alpha,P}\equiv\inf\{u:P(C_{P,n}^\star\leq u)\geq\alpha\}.$$

If $B_n\geq C_{P,n}^\star+o_P(a_n)$ (with $a_n>0$) uniformly in $P\in\mathbf{P}$ and $C_{P,n}^\star$ is independent of D_n, then there exists a $\delta_n\downarrow 0$ such that $\liminf_{n\to\infty}\inf_{P\in\mathbf{P}}P(\hat{q}_\alpha+a_n\geq q_{\alpha-\delta_n,P})=1$.

Proof: In the statement of the lemma, $\mathbf{P}$ and a_n represent a generic set of distributions and positive sequence – i.e. they need not be the same as in the main text. To establish the result, note Markov's inequality and the law of iterated expectations yield

$$\limsup_{n\to\infty}\sup_{P\in\mathbf{P}}P(P(C_{P,n}^\star>B_n+a_n|D_n)>\epsilon)\leq\limsup_{n\to\infty}\sup_{P\in\mathbf{P}}\frac{1}{\epsilon}P(C_{P,n}^\star>B_n+a_n)=0,$$

where the final equality follows from $B_n \geq C^{\star}_{P,n} + o_P(a_n)$ uniformly in $P \in \mathbf{P}$ by hypothesis. Thus, we conclude there exists some sequence $\delta_n \downarrow 0$ such that the event

$$\Omega_n(P) \equiv \{D_n | P(C^{\star}_{P,n} > B_n + a_n | D_n) \leq \delta_n\}$$

satisfies $P(\Omega_n(P)^c) = o(1)$ uniformly in $P \in \mathbf{P}$. Hence, for any $t \in \mathbf{R}$ we obtain that

$$\begin{aligned} P(B_n \leq t | D_n) 1\{D_n \in \Omega_n(P)\} &\leq P(B_n \leq t \text{ and } C^{\star}_{P,n} \leq B_n + a_n | D_n) + \delta_n \\ &\leq P(C^{\star}_{P,n} \leq t + a_n) + \delta_n, \end{aligned} \tag{S.101}$$

where in the final inequality we employed that $C^{\star}_{P,n}$ is independent of D_n. Therefore, setting $t = \hat{q}_\alpha$ in (S.101) implies that, under $\Omega_n(P)$, we have $\hat{q}_\alpha + a_n \geq q_{\alpha - \delta_n, P}$. Since $\sup_{P \in \mathbf{P}} P(\Omega_n(P)^c) = o(1)$, the claim of the lemma follows. ■

Lemma S.3.6. *Let $T_n \leq C_{P,n} + o_P(a_n)$ uniformly in $P \in \mathbf{P}_0$ with $0 < a_n = o(1)$, define $q_{\alpha,P} \equiv \inf\{u : P(C_{P,n} \leq u) \geq \alpha\}$, and suppose that, for some $\delta_n \downarrow 0$, $\hat{c}_n + a_n/2 \geq q_{1-\alpha-\delta_n,P}$ with probability tending to one uniformly in $P \in \mathbf{P}_0$. If for some $\eta_n \geq 0$*

$$\limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(T_n > \hat{c}_n) = \limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(T_n > \hat{c}_n \vee \eta_n) \tag{S.102}$$

and for some sequence ϱ_n satisfying $\varrho_n a_n = o(1)$ we have $\sup_{P \in \mathbf{P}_0} P(|C_{P,n} - t| \leq \epsilon) \leq \varrho_n(\epsilon \wedge 1) + o(1)$ for all $t \in (\eta_n - a_n, +\infty)$, then it follows that

$$\limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(T_n > \hat{c}_n) \leq \alpha.$$

Proof: First note that by condition (S.102), $T_n \leq C_{P,n} + o_P(a_n)$ uniformly in $P \in \mathbf{P}_0$ and the maintained hypothesis on $\hat{c}_n$ we can conclude that

$$\begin{aligned} \limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(T_n > \hat{c}_n) &= \limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(T_n > \hat{c}_n \vee \eta_n) \\ &\leq \limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(C_{P,n} + \frac{a_n}{2} > (q_{1-\alpha-\delta_n,P} - \frac{a_n}{2}) \vee \eta_n) \\ &\leq \limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} P(C_{P,n} + a_n > q_{1-\alpha-\delta_n,P} \vee \eta_n). \end{aligned} \tag{S.103}$$

Next observe that by direct calculation we also have the following inequalities

$$\begin{aligned} &P(C_{P,n} + a_n > q_{1-\alpha-\delta_n,P} \vee \eta_n) - P(C_{P,n} > q_{1-\alpha-\delta_n,P}) \\ &\qquad \leq \begin{cases} 0 & \text{if } \eta_n - a_n \geq q_{1-\alpha-\delta_n,P} \\ P(|C_{P,n} - q_{1-\alpha-\delta_n,P}| \leq a_n) & \text{if } \eta_n - a_n < q_{1-\alpha-\delta_n,P} \end{cases}. \end{aligned} \tag{S.104}$$

Therefore, combining results (S.103) and (S.104) together with $\sup_{P \in \mathbf{P}_0} P(|C_{P,n} - t| \leq$

$\epsilon) \leq \varrho_n(\epsilon \wedge 1) + o(1)$ for all $t \in (\eta_n - a_n, +\infty)$ implies that

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(T_n > \hat{c}_n) \\
&\qquad \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(C_{P,n} > q_{1-\alpha-\delta_n,P}) + \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} \sup_{t>\eta_n - a_n} P(|C_{P,n} - t| \leq a_n) \\
&\qquad \leq \alpha + \delta_n + \varrho_n(a_n \wedge 1). && \text{(S.105)}
\end{aligned}$$

The claim of the lemma therefore follows from $\delta_n = o(1)$ and $\varrho_n a_n = o(1)$. ■

Lemma S.3.7. *Let the conditions of Theorems 3.1(ii) and 3.2 hold with $R = \Theta$ and suppose that ℓ_n^{u} satisfies $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}((\ell_n^{\mathrm{u}})^{\kappa_\rho} \vee (\nu_n^{\mathrm{u}}\tau_n^{\mathrm{u}})^{\kappa_\rho}, \mathcal{F}_n^{\mathrm{u}}, \|\cdot\|_{P,2}) = o(a_n)$, $K_m \ell_n^{\mathrm{u}}(\ell_n^{\mathrm{u}} + \mathcal{R}_n^{\mathrm{u}} + \nu_n^{\mathrm{u}}\tau_n^{\mathrm{u}}) \times \mathcal{S}_n^{\mathrm{u}}(\mathbf{L}, \mathbf{E}) = o(a_n n^{-1/2})$, and $\mathcal{R}_n^{\mathrm{u}} = o(\ell_n^{\mathrm{u}})$. (i) If $\tau_n^{\mathrm{u}} \downarrow 0$ satisfies $J_n^{\mathrm{u}} B_n k_n^{1/p}\sqrt{\log(1+k_n)/n} = o(\tau_n^{\mathrm{u}})$ and $\nu_n^{\mathrm{u}}\tau_n^{\mathrm{u}} \times \mathcal{S}_n^{\mathrm{u}}(\mathbf{B}, \mathbf{E}) = o(1)$, then*

$$\hat{U}_n(\Theta| + \infty) \leq U_P^\star(\Theta|\ell_n^{\mathrm{u}}) + o_P(a_n)$$

uniformly in $P \in \mathbf{P}_0$. (ii) If $\mathcal{S}_n^{\mathrm{u}}(\mathbf{B}, \mathbf{E}) \times \mathcal{R}_n^{\mathrm{u}} = o(1)$ and Θ_{0n}^{u} is a singleton for all $P \in \mathbf{P}_0$ and n sufficiently large, then part (i) of the lemma continues to hold if $\tau_n^{\mathrm{u}} = 0$.

Proof: First note that since we required $J_n^{\mathrm{u}} B_n k_n^{1/p}\sqrt{\log(1+k_n)/n} = o(\tau_n^{\mathrm{u}})$ and we assumed all other conditions of Corollary S.1.2(ii) are satisfied when $\Theta = R$, it follows

$$d_H(\hat{\Theta}_n^{\mathrm{u}}, \Theta_{0n}^{\mathrm{u}}, \|\cdot\|_{\mathbf{E}}) = O_P(\nu_n^{\mathrm{u}}\tau_n^{\mathrm{u}}) \tag{S.106}$$

uniformly in $P \in \mathbf{P}_0$. Therefore, Lemma S.3.3 yields, uniformly in $P \in \mathbf{P}_0$, that

$$\sup_{\theta\in\hat{\Theta}_n^{\mathrm{u}}} \sup_{h\in V_n(\theta,\Theta|\ell_n^{\mathrm{u}})} \|\hat{\mathbb{D}}_n(\theta)[h] - \mathbb{D}_P(\theta)[h]\|_p = o_P(a_n). \tag{S.107}$$

We further note that since $\hat{\Theta}_n^{\mathrm{u}} \subseteq \Theta_n$, Assumption 3.11 holding with $R = \Theta$ implies

$$\sup_{\theta\in\hat{\Theta}_n^{\mathrm{u}}} \|\hat{\mathbb{W}}_n(\theta) - \mathbb{W}_P^\star(\theta)\|_p = o_P(a_n) \tag{S.108}$$

uniformly in $P \in \mathbf{P}_0$. Hence, by results (S.107) and (S.108), $\|\hat{\Sigma}_n\|_{o,p} = O_P(1)$ uniformly in $P \in \mathbf{P}_0$ by Corollary S.1.3, and $V_n(\theta, \Theta|\ell_n^{\mathrm{u}}) \subseteq \hat{V}_n(\theta, \Theta| + \infty)$ imply that

$$\begin{aligned}
\hat{U}_n(\Theta| + \infty) &\leq \inf_{\theta\in\hat{\Theta}_n^{\mathrm{u}}} \inf_{h\in V_n(\theta,\Theta|\ell_n^{\mathrm{u}})} \|\mathbb{W}_P^\star(\theta) + \mathbb{D}_P(\theta)[h]\|_{\hat{\Sigma}_n,p} + o_P(a_n) \\
&= \inf_{\theta\in\hat{\Theta}_n^{\mathrm{u}}} \inf_{h\in V_n(\theta,\Theta|\ell_n^{\mathrm{u}})} \|\mathbb{W}_P^\star(\theta) + \mathbb{D}_P(\theta)[h]\|_{\Sigma_P,p} + o_P(a_n) && \text{(S.109)}
\end{aligned}$$

uniformly in $P \in \mathbf{P}_0$, and where the equality can be established by employing identical arguments to those used in Lemma S.3.2 (see, in particular, (S.91)-(S.93)). Also note that, by hypothesis, there is an $\eta_n \downarrow 0$ satisfying $\nu_n^{\rm u}\tau_n^{\rm u} \times \mathcal{S}_n^{\rm u}(\mathbf{B},\mathbf{E}) = o(\eta_n)$ and define

$$\mathcal{E}_n(\theta) \equiv V_n(\theta,\Theta|\ell_n^{\rm u}) \cap \{h \in \mathbf{B}_n^{\rm u} : \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \eta_n\}.$$

Next, select $\theta_{0n} \in \Theta_{0n}^{\rm u}$ and $h_{0n} \in \mathcal{E}_n(\theta_{0n})$ so that the following equality is satisfied

$$\inf_{\theta\in\Theta_{0n}^{\rm u}} \inf_{h\in\mathcal{E}_n(\theta)} \|\mathbb{W}_P^\star(\theta)+\mathbb{D}_P(\theta)[h]\|_{\Sigma_P,p} = \|\mathbb{W}_P^\star(\theta_{0n})+\mathbb{D}_P(\theta_{0n})[h_{0n}]\|_{\Sigma_P,p}+o(a_n). \quad (S.110)$$

Assumption 3.13 holding with $R = \Theta$ implies $K_m\ell_n^{\rm u}(\nu_n^{\rm u}\tau_n^{\rm u})\mathcal{S}_n^{\rm u}(\mathbf{L},\mathbf{E}) = o(a_n n^{-1/2})$ and $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \sup_{P\in\mathbf{P}} J_{[\,]}((\nu_n^{\rm u}\tau_n^{\rm u})^{\kappa_\rho},\mathcal{F}_n^{\rm u},\|\cdot\|_{P,2}) = o(a_n)$. Hence, there is δ_n with

$$K_m\delta_n\ell_n^{\rm u}\mathcal{S}_n^{\rm u}(\mathbf{L},\mathbf{E}) = o(a_n n^{-1/2}) \quad (S.111)$$

$$k_n^{1/p}\sqrt{\log(1+k_n)}B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}(\delta_n^{\kappa_\rho},\mathcal{F}_n^{\rm u},\|\cdot\|_{P,2}) = o(a_n), \quad (S.112)$$

and $\nu_n^{\rm u}\tau_n^{\rm u} = o(\delta_n)$. Moreover, note result (S.106) implies there is a $\hat\theta_{0n}$ in $\hat\Theta_n^{\rm u}$ such that

$$\|\theta_{0n} - \hat\theta_{0n}\|_{\mathbf{E}} = O_P(\nu_n^{\rm u}\tau_n^{\rm u})$$

uniformly in $P \in \mathbf{P}_0$. Thus, $\nu_n^{\rm u}\tau_n^{\rm u} = o(\delta_n)$ and $\hat\theta_{0n} \in \hat\Theta_n^{\rm u} \subseteq \Theta_n$ implies that $\sqrt{n}(\hat\theta_{0n} - \theta_{0n}) \in V_n(\theta_{0n},\Theta|\delta_n)$ with probability tending to one uniformly in $P \in \mathbf{P}_0$. Hence, applying Lemma S.2.2 with $\Theta_{0n}^{\rm u}$ and $V_n(\theta,\Theta|\delta_n)$ in place of $\Theta_{0n}^{\rm r}$ and $V_n(\theta,R|\delta_n)$, yields

$$\|\mathbb{W}_P^\star(\hat\theta_{0n}) - \mathbb{W}_P^\star(\theta_{0n})\|_p = o_P(a_n) \quad (S.113)$$

uniformly in $P \in \mathbf{P}_0$. Furthermore, Lemma S.3.4, $h_{0n} \in \mathcal{E}_n(\theta_{0n})$ and result (S.111) imply that with probability tending to one uniformly in $P \in \mathbf{P}_0$ we must have

$$\|\mathbb{D}_P(\hat\theta_{0n})[h_{0n}] - \mathbb{D}_P(\theta_{0n})[h_{0n}]\|_p \leq K_m\mathcal{S}_n^{\rm u}(\mathbf{L},\mathbf{E})\delta_n\ell_n^{\rm u}\sqrt{n} = o(a_n). \quad (S.114)$$

Therefore, Assumption 3.7(ii), $\hat\theta_{0n} \in \hat\Theta_n^{\rm u}$, $h_{0n} \in \mathcal{E}_n(\theta_{0n})$, $\mathcal{E}_n(\theta_{0n}) \subseteq V_n(\hat\theta_{0n},\Theta|\ell_n^{\rm u})$ by Assumption 3.12(ii), and results (S.110), (S.113), and (S.114) yield that

$$\inf_{\theta\in\Theta_{0n}^{\rm u}} \inf_{h\in\mathcal{E}_n(\theta)} \|\mathbb{W}_P^\star(\theta)+\mathbb{D}_P(\theta)[h]\|_{\Sigma_P,p}$$
$$\geq \inf_{\theta\in\hat\Theta_n^{\rm u}} \inf_{h\in V_n(\theta,\Theta|\ell_n^{\rm u})} \|\mathbb{W}_P^\star(\theta)+\mathbb{D}_P(\theta)[h]\|_{\Sigma_P,p} + o_P(a_n) \quad (S.115)$$

uniformly in $P \in \mathbf{P}_0$. To conclude, note that Assumption 3.4 holding with $R = \Theta$, Corollary S.1.1, and $\mathcal{R}_n^{\rm u}\times\mathcal{S}_n^{\rm u}(\mathbf{B},\mathbf{E}) = o(\eta_n)$ due to $\mathcal{R}_n^{\rm u} = o(\tau_n^{\rm u}\nu_n^{\rm u})$ and $\nu_n^{\rm u}\tau_n^{\rm u}\times\mathcal{S}_n^{\rm u}(\mathbf{B},\mathbf{E}) =$

$o(\eta_n)$ allow us to conclude that uniformly in $P \in \mathbf{P}_0$ we have

$$
\begin{aligned}
I_n(\Theta) &= \inf_{\theta \in \Theta^{\mathrm{u}}_{0n}} \inf_{h \in \mathcal{E}_n(\theta)} \sqrt{n} Q_n(\theta + \frac{h}{\sqrt{n}}) + o_P(a_n) \\
&= \inf_{\theta \in \Theta^{\mathrm{u}}_{0n}} \inf_{h \in \mathcal{E}_n(\theta)} \|\mathbb{W}_P(\theta) + \mathbb{D}_P(\theta)[h]\|_{\Sigma_P, p} + o_P(a_n),
\end{aligned} \tag{S.116}
$$

where the second equality follows by identical arguments to those employed in Theorem 3.1(ii). Combining result (S.116) with Theorem 3.1(ii) and employing the fact that $\mathbb{W}^\star_P$ and $\mathbb{W}_P$ share the same distribution we thus obtain, uniformly in $P \in \mathbf{P}_0$, that

$$
\begin{aligned}
&\inf_{\theta \in \Theta^{\mathrm{u}}_{0n}} \inf_{h \in V_n(\theta, \Theta | \ell^{\mathrm{u}}_n)} \|\mathbb{W}^\star_P(\theta) + \mathbb{D}_P(\theta)[h]\|_{\Sigma_P, p} \\
&\qquad = \inf_{\theta \in \Theta^{\mathrm{u}}_{0n}} \inf_{h \in \mathcal{E}_n(\theta)} \|\mathbb{W}^\star_P(\theta) + \mathbb{D}_P(\theta)[h]\|_{\Sigma_P, p} + o_P(a_n).
\end{aligned} \tag{S.117}
$$

The claim of part (i) of the lemma therefore follows (S.109), (S.115), and (S.117). To establish part (ii) note that if Θ^{u}_{0n} is a singleton, then $\overrightarrow{d}_H(\hat{\Theta}^{\mathrm{u}}_n, \Theta^{\mathrm{u}}_{0n}, \|\cdot\|_{\mathbf{E}}) = d_H(\hat{\Theta}^{\mathrm{u}}_n, \Theta^{\mathrm{u}}_{0n}, \|\cdot\|_{\mathbf{E}})$ and therefore Corollary S.1.2(i) implies $d_H(\hat{\Theta}^{\mathrm{u}}_n, \Theta^{\mathrm{u}}_{0n}, \|\cdot\|_{\mathbf{E}}) = O_P(\mathcal{R}^{\mathrm{u}}_n)$ uniformly in $P \in \mathbf{P}_0$. Part (ii) of the lemma can then be established by replacing $\nu^{\mathrm{u}}_n \tau^{\mathrm{u}}_n$ with $\mathcal{R}^{\mathrm{u}}_n$ in the arguments employed in establishing part (i). ■

Corollary S.3.1. *Suppose that* $I_n(R) \leq U_P(R|\tilde{\ell}_n) + o_P(a_n)$ *and* $\hat{U}_n(R|\ell_n) \geq U^\star_P(R|\tilde{\ell}_n) + o_P(a_n)$ *uniformly in* $P \in \mathbf{P}_0$ *with* $0 < a_n = o(1)$, $U_P(R|\tilde{\ell}_n) \overset{d}{=} U^\star_P(R|\tilde{\ell}_n)$, *and* $U^\star_P(R|\tilde{\ell}_n)$ *independent of* $\{V_i\}_{i=1}^n$. *Then for any constant* $\eta \in (0, \alpha)$ *it follows that*

$$
\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(I_n(R) > \hat{q}_{1-\alpha+\eta}(\hat{U}_n(R|\ell_n)) + \eta) \leq \alpha.
$$

Proof: Since $I_n(R) \leq U_P(R|\tilde{\ell}_n) + o_P(a_n)$ uniformly in $P \in \mathbf{P}_0$ by hypothesis, we obtain

$$
\begin{aligned}
&\limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(I_n(R) > \hat{q}_{1-\alpha+\eta}(\hat{U}_n(R|\ell_n)) + \eta) \\
&\qquad \leq \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(U_P(R|\tilde{\ell}_n) + a_n > \hat{q}_{1-\alpha+\eta}(\hat{U}_n(R|\ell_n)) + \eta) \\
&\qquad \leq \limsup_{n \to \infty} \sup_{P \in \mathbf{P}_0} P(U_P(R|\tilde{\ell}_n) > q_{1-\alpha+\eta-\delta_n, P}(U^\star_P(R|\tilde{\ell}_n)) + \eta - 2a_n) \\
&\qquad \leq \alpha,
\end{aligned} \tag{S.118}
$$

where the second inequality holds for $q_{1-\alpha+\eta-\delta_n, P}(U^\star_P(R|\tilde{\ell}_n))$ the $1-\alpha+\eta-\delta_n$ quantile of $U^\star_P(R|\tilde{\ell}_n)$ and some $\delta_n = o(1)$ by Lemma S.3.5 applied with $B_n = \hat{U}_n(R|\ell_n)$, $C^\star_{P,n} = U^\star_P(R|\tilde{\ell}_n)$, and $D_n = \{V_i\}_{i=1}^n$. In turn, the final inequality in (S.118) follows from $\eta > 0$, $a_n = o(1)$, $\delta_n = o(1)$, and $U_P(R|\tilde{\ell}_n) \overset{d}{=} U^\star_P(R|\tilde{\ell}_n)$. ■

S.4 Illustrative Examples

In this Section, we include the proofs for all the examples discussed in the main text and Supplemental Appendix I – i.e., the results stated in Section 4 of the main text and in Section A.2 of Supplemental Appendix I.

S.4.1 Proofs for Section 4

Proof of Theorem 4.1: We establish the claim of the theorem by verifying the conditions of Theorem 3.1(ii) for both R as in (29) (to couple $I_n(R)$) and $R = \Theta$ (to couple $I_n(\Theta)$). To this end, note that Assumption 3.1(i) is imposed in Assumption 4.1(i), Assumption 3.2(i) holds with $B_n \asymp \sqrt{k_n}$ by Assumption 4.2(i), Assumption 3.2(ii) is directly imposed in Assumption 4.2(ii), and Assumption 3.2(iii) is satisfied with $J_n \asymp \sqrt{j_n \log(1 + j_n)}$ by Lemma S.4.2 and $\|f\|_\infty \leq 3$ for any $f \in \mathcal{F}_n$. The coupling requirement of Assumption 3.3(i) is satisfied for $R = \Theta$, and hence also for R as in (29), with $a_n = (\log(n))^{-1/2}$ by Lemma S.4.4 and Assumption 4.2(iv). Moreover, Assumptions 3.3(ii), 3.4, and 3.5 also hold by Lemmas S.4.1 and S.4.3. To verify Assumption 3.6, we first note that Assumption 3.6(ii) is implied by Assumptions 4.1(iv) and 4.3(ii). Furthermore, as argued, $B_n \asymp \sqrt{k_n}$, $J_n \asymp \sqrt{j_n \log(1 + j_n)}$, and $\nu_n \asymp 1$ by Lemma S.4.1, which yields that $\mathcal{R}_n \lesssim k_n \sqrt{j_n} \log(1 + k_n)/\sqrt{n}$ since $k_n \geq j_n$ by Assumption 4.2(iii). Thus, $\kappa_\rho = 1$ by Lemma S.4.3 and Lemma S.4.2 imply that Assumption 3.6(i) holds by Assumption 4.2(iv). By similar arguments, it also follows that Assumption 3.7 is implied by Assumption 4.3, and that the requirements $k_n^{1/p}\sqrt{\log(1 + k_n)} B_n \sup_{P \in \mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$ and $\mathcal{R}_n = o(\ell_n)$ are implied by $k_n\sqrt{j_n}\log^2(n)\ell_n = o(1)$ and $k_n\sqrt{j_n}\log(n)/\sqrt{n} = o(\ell_n)$. Since $K_m = 0$ in this application, it follows all the conditions of Theorem 3.1(ii) hold for both $R = \Theta$ and R as in (29), and hence the theorem follows. ■

Proof of Lemma 4.1: The result essentially follows from Theorem 1 in Walkup and Wets (1969). To map our problem into their setting, note that since $\{\delta_s\}_{s=1}^{s_n}$ are orthogonal, every $\mu \in \mathcal{M}_n$ can be identified with a unique $(\alpha_1, \ldots, \alpha_{s_n}) \equiv \alpha \in \mathbf{R}^{s_n}$ through the relation $\mu = \sum_{s=1}^{s_n} \alpha_s \delta_s$ – e.g., by $\alpha_s = \mu(S_s)$ for S_s the support of δ_s. With some abuse of notation, for the remaining of the proof we therefore employ α and μ interchangeably. Further note that, for any $\theta = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu)$, the restrictions $\Upsilon_G(\theta) \leq 0$, $\Upsilon_F^{(\mu)}(\theta) = 0$, and $\Upsilon_F^{(\mathrm{s})}(\theta) = 0$ depend only on μ and define a closed convex polyhedron on $\mathbf{R}^{s_n}$, which we denote by K_n. Next, define the map $\Lambda_n : \mathbf{R}^{s_n} \to \mathbf{R}^{\mathcal{J}\mathcal{L}}$ to be given by

$$\Lambda_n(\alpha) = \{\sum_{s=1}^{s_n} \alpha_s(\int 1\{g(w_l, \eta) \leq c_\jmath\}\delta_s(d\eta))\}_{1 \leq \jmath \leq \mathcal{J}, 1 \leq l \leq \mathcal{L}} \qquad \text{(S.119)}$$

and note that for any $\theta = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu) = \theta \in \Theta_n \cap R$, it follows by (35) that

$$\Gamma_n(\theta) = K_n \cap \Lambda_n^{-1}(\{F(c_\jmath|w_l)\}_{1\leq \jmath \leq \mathcal{J}, 1\leq l\leq \mathcal{L}}). \tag{S.120}$$

Let d_n denote the dimension of the null space of Λ_n, and note that if $d_n = s_n$, then $\Gamma_n(\theta_1) = \Gamma_n(\theta_2)$ for any $\theta_1, \theta_2 \in \Theta_n \cap R$ by result (S.120), and hence the conclusion of the lemma is immediate. On the other hand, if $1 \leq d_n \leq s_n - 1$, then Theorem 1 in Walkup and Wets (1969) implies there is a C_n such that for any $\theta_1, \theta_2 \in \Theta_n \cap R$ we have

$$\begin{aligned} d_H(\Gamma_n(\theta_1), \Gamma_n(\theta_2), \|\cdot\|_2) &\leq C_n\{\sum_{\jmath=1}^{\mathcal{J}}\sum_{l=1}^{\mathcal{L}}(F_1(c_\jmath|w_l) - F_2(c_\jmath|w_l))^2\}^{1/2} \\ &\lesssim C_n \sum_{\jmath=1}^{\mathcal{J}} \|F_1(c_\jmath|\cdot) - F_2(c_\jmath|\cdot)\|_\infty, \end{aligned} \tag{S.121}$$

and where the norm $\|\cdot\|_2$ on $\Gamma_n(\theta)$ is understood as the usual Euclidean norm on the corresponding $\alpha \in \mathbf{R}^{s_n}$. Similarly, we note that if $d_n = 0$, then Λ_n is invertible and (S.121) holds with $C_n = \|\Lambda_n^{-1}\|_o$. Also note that for any $\mu = \sum_{s=1}^{s_n}\alpha_s\delta_s$ and $\tilde{\mu} = \sum_{s=1}^{s_n}\tilde{\alpha}_s\delta_s$ we have $\|\mu - \tilde{\mu}\|_{TV} = \|\alpha - \tilde{\alpha}\|_1$ due to the measures $\{\delta_s\}_{s=1}^{s_n}$ being orthogonal. Hence, since $\|a\|_1 \leq \sqrt{s_n}\|a\|_2$ for any $a \in \mathbf{R}^{s_n}$, result (S.121) yields

$$d_H(\Gamma_n(\theta_1), \Gamma_n(\theta_2), \|\cdot\|_{TV}) \lesssim \sqrt{s_n} C_n \sum_{\jmath=1}^{\mathcal{J}} \|F_1(c_\jmath|\cdot) - F_2(c_\jmath|\cdot)\|_\infty,$$

which establishes the claim of the lemma by setting $\zeta_n \asymp C_n\sqrt{s_n}$. ■

Proof of Theorem 4.2: Let $\hat{V}_n(\theta, R|\ell) \equiv \hat{V}_n(\theta, R| + \infty) \cap \{h \in \mathbf{B}_n : \|h/\sqrt{n}\|_{\mathbf{E}} \leq \ell\}$, recall $\|\theta\|_{\mathbf{E}} = \sum_{\jmath=1}^{\mathcal{J}} \sup_{P\in\mathbf{P}} \|F(c_\jmath|\cdot)\|_{P,2}$ for any $\theta = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu)$, define

$$\hat{E}_n(R|\ell_n) = \inf_{\theta\in\hat{\Theta}_n^{\mathrm{r}}} \inf_{h\in\hat{V}_n(\theta,R|\ell_n)} \{\sum_{\jmath=1}^{\mathcal{J}} \|\hat{\mathbb{W}}_{\jmath,n}(\theta) + \hat{\mathbb{D}}_{\jmath,n}[h]\|_{\hat{\Sigma}_{\jmath,n},2}\}^{1/2},$$

and note that for any ℓ_n satisfying the conditions of the theorem, Assumption 4.4(iii) and Lemma S.4.9 imply $\hat{U}_n(R| + \infty) = \hat{E}_n(R|\ell_n) + o_P(a_n)$ uniformly in $P \in \mathbf{P}_0$. Hence, to establish the theorem it suffices to show there are $\ell_n \asymp \tilde{\ell}_n$ and $\ell_n^{\mathrm{u}} \asymp \tilde{\ell}_n^{\mathrm{u}}$ such that

$$\begin{aligned} \hat{E}_n(R|\ell_n) &\geq U_P^\star(R|\tilde{\ell}_n) + o_P(a_n) \\ \hat{E}_n(R|\ell_n) - \hat{U}_n(\Theta| + \infty) &\geq U_P^\star(R|\tilde{\ell}_n) - U_P^\star(\Theta|\tilde{\ell}_n^{\mathrm{u}}) + o_P(a_n) \end{aligned} \tag{S.122}$$

uniformly in $P \in \mathbf{P}_0$. To this end, we rely on Theorem S.3.1(ii) (for $\hat{E}_n(R|\ell_n)$) and Lemma S.3.7. Also note that in the proof of Theorem 4.1 we showed Assumptions 4.1, 4.2, and 4.3 imply Assumptions 3.1-3.7 hold with $B_n \asymp \sqrt{k_n}$, $J_n \asymp \sqrt{j_n \log(1 + j_n)}$, $\nu_n \asymp 1$, $\mathcal{R}_n \asymp k_n\sqrt{j_n \log(1 + k_n)\log(1 + j_n)/n}$, $a_n = (\log(n))^{-1/2}$, $\kappa_\rho = 1$, $\|\theta\|_{\mathbf{L}} =$

$\|\theta\|_{\mathbf{E}} = \sum_{\jmath=1}^{\mathcal{J}} \sup_{P\in\mathbf{P}} \|F(c_\jmath|\cdot)\|_{P,2}$, and $\|\theta\|_{\mathbf{B}} = \sum_{\jmath=1}^{\mathcal{J}} \|F(c_\jmath|\cdot)\|_\infty + \|\mu\|_{TV}$ for $R = \Theta$ and R as in (29).

In order to apply Theorem S.3.1(ii), we set $\|\theta\|_{\mathbf{I}} = \max_{1\leq\jmath\leq\mathcal{J}} \mathcal{J}\|F(c_\jmath|\cdot)\|_\infty$ for any $\theta = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu) \in \mathbf{B}_n$ and note Assumption 4.4(i) and Lemma S.4.5 verify Assumptions 3.8, 3.9, and 3.10 are satisfied with $K_g = 0$ and $K_f = 0$. Also note Assumption 4.4(iii) and Lemma S.4.8 verify Assumption 3.11 and Assumptions 3.12(i)(iii) are immediate given the definitions of $\|\cdot\|_{\mathbf{E}}$ and $\|\cdot\|_{\mathbf{B}}$ and $\mathcal{V}_n(P) = \Theta_n \cap R$ by Lemma S.4.1. Also note $\{\theta \in \mathbf{B}_n : \vec{d}_H(\theta, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{I}}) \leq 1/2\} \subseteq \Theta_n$ for n sufficiently large by Assumption 4.4(iv) and the definitions of Θ_n and $\|\cdot\|_{\mathbf{I}}$. Moreover, Assumptions 4.1(ii)(iii) imply

$$\sup_{h\in\mathbf{B}_n} \frac{\|h\|_{\mathbf{I}}}{\|h\|_{\mathbf{E}}} = \sup_{h\in\mathbf{B}_n} \frac{\max_{1\leq\jmath\leq\mathcal{J}} \mathcal{J}\|F(c_\jmath|\cdot)\|_\infty}{\sum_{\jmath=1}^{\mathcal{J}} \sup_{P\in\mathbf{P}} \|F(c_\jmath|\cdot)\|_{P,2}} \lesssim \sqrt{j_n}, \tag{S.123}$$

and hence Corollary S.1.2(i), $\nu_n \asymp 1$, and $\mathcal{R}_n \asymp k_n\sqrt{j_n \log(1+k_n)\log(1+j_n)/n}$ yield

$$\vec{d}_H(\hat{\Theta}_n^{\mathrm{r}}, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{I}}) \lesssim \sqrt{j_n}\,\vec{d}_H(\hat{\Theta}_n^{\mathrm{r}}, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{E}}) = O_P(\frac{k_n j_n \log(n)}{\sqrt{n}} + \sqrt{j_n}\tau_n)$$

uniformly in $P \in \mathbf{P}_0$. In particular, Assumptions 4.4(iii)(v) imply $\vec{d}_H(\hat{\Theta}_n^{\mathrm{r}}, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{I}}) = o_P(1)$ uniformly in $P \in \mathbf{P}_0$, and therefore since, as argued, we have $\{\theta \in \mathbf{B}_n : \vec{d}_H(\theta, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{I}}) \leq 1/2\} \subseteq \Theta_n$ for n sufficiently large, we obtain

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P(\{\theta \in \mathbf{B}_n : \vec{d}_H(\{\theta\}, \hat{\Theta}_n^{\mathrm{r}}, \|\cdot\|_{\mathbf{I}}) \leq 1/4\} \subseteq \Theta_n) = 1. \tag{S.124}$$

Next, observe Lemma 4.1, Assumption 4.1(ii) and the definitions of $\|\cdot\|_{\mathbf{E}}$, $\|\cdot\|_{\mathbf{L}}$, and $\|\cdot\|_{\mathbf{B}}$ imply Assumption S.3.1 holds with $\mathcal{D}_n(\mathbf{B},\mathbf{E}) \asymp \zeta_n\sqrt{j_n}$ and $\mathcal{D}_n(\mathbf{L},\mathbf{E}) = 1$. Since $K_m = K_g = K_f = 0$ and Υ_F and Υ_G are affine, the only requirements imposed by Assumption S.3.2 are that $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho} \vee (\nu_n\tau_n)^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$ and $(\mathcal{R}_n + \nu_n\tau_n)\mathcal{D}_n(\mathbf{B},\mathbf{E}) = o(r_n)$, which are implied by Assumption 4.4(v), Lemma S.4.2, and $k_n\sqrt{j_n}\log^2(n)\ell_n = o(1)$ by hypothesis. Hence, all the conditions of Theorem S.3.1(ii) hold, which implies there is a $\tilde{\ell}_n \asymp \ell_n$ such that uniformly in $P \in \mathbf{P}_0$

$$\hat{E}_n(R|\ell_n) \geq U_P^\star(R|\tilde{\ell}_n) + o_P(a_n). \tag{S.125}$$

Finally, to apply Lemma S.3.7 to $\hat{U}_n(\Theta| + \infty)$, note that we can set the norm $\|\cdot\|_{\mathbf{B}}$ to equal $\|\theta\|_{\mathbf{B}} = \max_{1\leq\jmath\leq\mathcal{J}} \|F(c_\jmath|\cdot)\|_\infty$ and interpret Υ_G and Υ_F as satisfying $\Upsilon_G(\theta) = \Upsilon_F(\theta) = 0$ for all $\theta \in \mathbf{B}$ (since $R = \Theta$). Hence, Assumptions 3.8, 3.9, and 3.10, 3.12(i) are immediate, while Assumption 3.11 is satisfied by Assumption 4.4(iii) and Lemma S.4.8. Further note since Θ_0 is an equivalence class under $\|\cdot\|_{\mathbf{E}}$ and $\|\cdot\|_{\mathbf{B}}$, when studying the unconstrained statistic we can treat the model as identified. As a result, we may set $\tau_n^{\mathrm{u}} = 0$ and Assumption 3.12(ii) holds by the same arguments employed in (S.124), while Assumption 3.12(iii) is immediate since $\mathcal{V}_n(P) = \Theta_n \cap R$. In order to apply Lemma S.3.7(ii), it therefore only remains to verify that $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\mathrm{u}}, \mathcal{F}_n^{\mathrm{u}}, \|\cdot\|_{P,2}) = o(a_n)$, $\mathcal{R}_n^{\mathrm{u}} = o(\ell_n^{\mathrm{u}})$, and $\mathcal{S}_n^{\mathrm{u}}(\mathbf{B},\mathbf{E})\times\mathcal{R}_n^{\mathrm{u}} = o(1)$, which are implied by $k_n\sqrt{j_n}\log^2(n)\ell_n^{\mathrm{u}} = o(1)$, $k_n\sqrt{j_n}\log(n)/\sqrt{n} = o(\ell_n^{\mathrm{u}})$, and Assumption 4.4(iii) respectively. Thus, (S.125) and Lemma S.3.7(ii) verify (S.122) with $\tilde{\ell}_n^{\mathrm{u}} = \ell_n^{\mathrm{u}}$ and $\tilde{\ell}_n \asymp \ell_n$, which in turn establishes the theorem. ■

Lemma S.4.1. *If Assumptions 4.1(iii), 4.2(iii), and 4.3(ii) hold, then Assumption 3.4 holds with $R = \Theta$ and R as in (29), $\mathcal{V}_n(P) = \Theta_n \cap R$, $\|\theta\|_{\mathbf{E}} = \sum_{\jmath=1}^{\mathcal{J}} \sup_{P\in\mathbf{P}} \|F(c_\jmath|\cdot)\|_{P,2}$ for any $\theta = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu) \in \mathbf{B}$, and $\nu_n^{-1} \asymp 1$.*

Proof: First note that since we are setting $\mathcal{V}_n(P) = \Theta_n \cap R$, Assumption 3.4(ii) is immediate. To verify Assumption 3.4(i), let $\|\theta\|_{\mathbf{E}} = \sum_{\jmath=1}^{\mathcal{J}} \sup_{P\in\mathbf{P}} \|F(c_\jmath|\cdot)\|_{P,2}$ for any $\theta = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu) \in \mathbf{B}$. Then note that any $(\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu) = \theta \in \Theta_n$ must be such that $F(c_\jmath|\cdot) = p^{j_n\prime}\beta_{\jmath,\theta}$ for some $\beta_{\jmath,\theta} \in \mathbf{R}^{j_n}$ and, similarly, $\Pi_n\theta_0 = (\{F_n(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu_n)$ must satisfy $F_n(c_\jmath|\cdot) = p^{j_n\prime}\beta_{\jmath,n}$. The Cauchy Schwarz inequality, and Assumptions 4.1(iii) and 4.2(iii) then yield that uniformly in $P \in \mathbf{P}_0$ we must have

$$\|\theta - \Pi_n\theta_0\|_{\mathbf{E}} \lesssim \sum_{\jmath=1}^{\mathcal{J}} \|\beta_{\jmath,\theta} - \beta_{\jmath,n}\|_2 \lesssim \sum_{\jmath=1}^{\mathcal{J}} \|E_P[q^{k_n}(W)p^{j_n}(W)'(\beta_{\jmath,\theta} - \beta_{\jmath,n})]\|_2$$
$$\lesssim \{\sum_{\jmath=1}^{\mathcal{J}} \|E_P[(F(c_\jmath|W) - F_n(c_\jmath|W))q^{k_n}(W)]\|^2_{\Sigma_{\jmath,P},2}\}^{1/2}, \quad \text{(S.126)}$$

where the final inequality holds due to $\|\Sigma_{\jmath,P}^{-1}\|_{o,2}$ being uniformly bounded by Assumption 4.3(ii) and $\sum_{\jmath=1}^{\mathcal{J}} |a^{(\jmath)}| \le \sqrt{\mathcal{J}}\|a\|_2$ for any $(a^{(1)}, \ldots, a^{(\mathcal{J})}) = a \in \mathbf{R}^{\mathcal{J}}$. Result (S.126) and the definition of $\rho_\jmath(X,\theta)$ in (28) verify Assumption 3.4(i) holds with $\nu_n^{-1} \asymp 1$. ■

Lemma S.4.2. *Define the class $\mathcal{F}_n \equiv \{f : f(v) = (1\{y \le c_\jmath\} - p^{j_n}(w)'\beta)$ for some $1 \le \jmath \le \mathcal{J}$ and $\|p^{j_n\prime}\beta\|_\infty \le 2\}$ and suppose that Assumptions 4.1(ii)(iii) hold. Then, it follows that $\sup_{P\in\mathbf{P}} N_{[\,]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2}) \lesssim (1 \vee (\sqrt{j_n}K/\epsilon)^{j_n})$ for some $K < \infty$, and in addition $\sup_{P\in\mathbf{P}} J_{[\,]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2}) \lesssim \epsilon\sqrt{j_n}(1 + \sqrt{\log(1 \vee (\sqrt{j_n}/\epsilon))})$.*

Proof: First note that for any $p^{j_n\prime}\beta_1$ and $p^{j_n\prime}\beta_2$, the Cauchy-Schwarz inequality yields

$$|p^{j_n}(w)'\beta_1 - p^{j_n}(w)'\beta_2| \le \sup_w \|p^{j_n}(w)\|_2\|\beta_1 - \beta_2\|_2 \lesssim \sqrt{j_n}\|\beta_1 - \beta_2\|_2,$$

where in the final inequality we employed Assumption 4.1(ii). Hence, Theorem 2.7.11 in van der Vaart and Wellner (1996), $\|\beta\|_2 \asymp \sup_{P\in\mathbf{P}} \|p^{j_n\prime}\beta\|_{P,2}$ by Assumption 4.1(iii), and $\sup_{P\in\mathbf{P}} \|p^{j_n\prime}\beta\|_{P,2} \le \|p^{j_n\prime}\beta\|_\infty \le 2$ for any $p^{j_n\prime}\beta \in \Theta_n$ imply

$$\sup_{P\in\mathbf{P}} N_{[\,]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2}) \lesssim 1 \vee (\frac{K\sqrt{j_n}}{\epsilon})^{j_n}, \quad \text{(S.127)}$$

for some $K < \infty$, which establishes the first claim of the lemma. For the second claim of the lemma, we employ (S.127) and the change of variables $v = u/\epsilon$ to obtain

$$\sup_{P\in\mathbf{P}} J_{[]}(\epsilon,\mathcal{F}_n,\|\cdot\|_{P,2}) \lesssim \epsilon + \int_0^\epsilon (\log(1\vee(\frac{K\sqrt{j_n}}{u})^{j_n}))^{1/2}du$$
$$= \epsilon(1+\sqrt{j_n}\int_0^1 (\log(1\vee(\frac{K\sqrt{j_n}}{v\epsilon})))^{1/2}dv) \lesssim \sqrt{j_n}\epsilon(1+\sqrt{\log(1\vee(\sqrt{j_n}/\epsilon))}),$$

where the final inequality follows from $(1\vee ab) \leq (1\vee a)(1\vee b)$ for any $a,b\in\mathbf{R}_+$. ■

Lemma S.4.3. *Let $\rho_\jmath : \mathbf{R}\times\mathbf{W}\times\Theta$ be as defined in (28). It then follows Assumptions 3.3(ii) and 3.5 hold with $\kappa_\rho = 1$, $K_\rho = 1$, $K_m = 0$, $M < \infty$, $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_{\mathbf{E}}$, and $\|\theta\|_{\mathbf{E}} = \sum_{\jmath=1}^{\mathcal{J}}\sup_{P\in\mathbf{P}}\|F(c_\jmath|\cdot)\|_{P,2}$ for any $\theta = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}},\mu)\in\mathbf{B}$.*

Proof: First note that for any $(\{F_1(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}},\mu_1) = \theta_1\in\mathbf{B}$ and $(\{F_2(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}},\mu_2) = \theta_2\in\mathbf{B}$, we obtain from (28) and the definition of $\|\cdot\|_{\mathbf{E}}$ that for all $P\in\mathbf{P}$

$$E_P[\|\rho(X,\theta_1)-\rho(X,\theta_2)\|_2^2] = \sum_{\jmath=1}^{\mathcal{J}} E_P[(F_1(c_\jmath|W)-F_2(c_\jmath|W))^2] \leq \|\theta_1-\theta_2\|_{\mathbf{E}}^2,$$

which verifies Assumption 3.3(ii) holds with $\kappa_\rho = 1$ and $K_\rho = 1$. Next, for any $P\in\mathbf{P}$ define $\nabla m_{P,\jmath}(\theta)[h] = -F_h(c_\jmath|W)$ for all $\theta\in\mathbf{B}$ and $(\{F_h(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}},\mu_h) = h\in\mathbf{B}$. Since $m_{P,\jmath}(\theta) = P(Y\leq c_\jmath|W) - F(c_\jmath|W)$ for any $\theta = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}},\mu)\in\mathbf{B}$, direct calculation verifies Assumption 3.5 holds with $K_m = 0$, $M = 1$, and $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_{\mathbf{E}}$. ■

Lemma S.4.4. *If $k_n^3 j_n^2\log^2(n) = o(n)$, Assumptions 4.1(i)-(iii) and 4.2(i) hold, then Assumption 3.3(i) holds with $R = \Theta$ for any a_n with $k_n^3 j_n^2\log^2(n)/n = o(a_n^2)$.*

Proof: We establish the result by applying Lemma S.4.6. To this end, we let $\tilde{j}_n = \mathcal{J}+j_n$ set $\{r_j\}_{j=1}^{\tilde{j}_n} = \{1\{y\leq c_\jmath\}\}_{\jmath=1}^{\mathcal{J}}\cup\{p_j\}_{j=1}^{j_n}$ and let $r^{\tilde{j}_n}(x) \equiv (r_1(x),\ldots,r_{\tilde{j}_n}(x))'$. Next note that any $f\in\mathcal{F}_n$ may be written as $r^{\tilde{j}_n\prime}\beta$ for some $\beta\in\mathbf{R}^{\tilde{j}_n}$. Moreover, since $\sup_{P\in\mathbf{P}}\max_{1\leq\jmath\leq\mathcal{J}}\|F(c_\jmath|\cdot)\|_{P,2} \leq \max_{1\leq\jmath\leq\mathcal{J}}\|F(c_\jmath|\cdot)\|_\infty \leq 2$ for any $(\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}},\mu) = \theta\in\Theta_n$, Assumption 4.1(iii) implies that there exists a $C_0 < \infty$ (independent of $\tilde{j}_n$) such that $\|\beta\|_2 \leq C_0$ whenever $r^{\tilde{j}_n\prime}\beta\in\mathcal{F}_n$. Hence, by Assumptions 4.1(ii) and 4.2(i), we may apply Lemma S.4.6 with $b_{1n}\asymp\sqrt{j_n}$, $b_{2n}\asymp k_n$, and $C_n = O(1)$, from which the claim of the present lemma immediately follows. ■

Lemma S.4.5. *Let $\mathbf{B} = (\bigotimes_{\jmath=1}^{\mathcal{J}} C_B(\mathbf{W}))\times\mathcal{M}$ and Θ, Υ_G, and Υ_F be as defined in (27), (30), and (31). If $\Psi(g,\cdot)$ is bounded on Ω, then Assumptions 3.8, 3.9, and 3.10 are satisfied with $K_g = 0$, $\nabla\Upsilon_G(\theta)[h] = \Upsilon_G(h)$, $K_f = 0$, and $\nabla\Upsilon_F(\theta)[h]$ equal to*

$$\nabla\Upsilon_F(\theta)[h] = (\Upsilon_F^{(\mathrm{e})}(h), \Upsilon_F^{(\mu)}(h)+1, \Upsilon_F^{(\mathrm{s})}(h)+\lambda). \tag{S.128}$$

Proof: For any measure $\mu \in \mathcal{M}$ let $\mu = \mu^+ - \mu^-$ denote its Jordan decomposition, $|\mu| = \mu^+ + \mu^-$, and recall the total variation of μ equals $\|\mu\|_{TV} = |\mu|(\Omega)$. Since $\Upsilon_G : \mathbf{B} \to \ell^\infty(\mathcal{B})$ is linear, in order to verify Assumption 3.8 we need only show that Υ_G is continuous. To this end, recall that for any $(\{F(c_j|\cdot)\}_{j=1}^{\mathcal{J}}, \mu) = \theta \in \mathbf{B}$ we had defined $\|\theta\|_{\mathbf{B}} = \sum_{j=1}^{\mathcal{J}} \|F(c_j|\cdot)\|_\infty + \|\mu\|_{TV}$. Hence, employing the definition of Υ_G we obtain

$$\|\Upsilon_G\|_o = \sup_{\|\theta\|_{\mathbf{B}}=1} \|\Upsilon_G(\theta)\|_\infty = \sup_{\mu:\|\mu\|_{TV}=1} \sup_{B\in\mathcal{B}} |\mu(B)| \leq \sup_{\mu:\|\mu\|_{TV}=1} |\mu|(\Omega) = 1,$$

which, by linearity of Υ_G, implies Assumption 3.8 holds with $\nabla\Upsilon_G = \Upsilon_G$ and $K_g = 0$. By similar arguments, note that $\Upsilon_F^{(\mathrm{e})} : \mathbf{B} \to \mathbf{R}^{\mathcal{J}\mathcal{L}}$, as defined in (31), is linear and

$$\begin{aligned}\|\Upsilon_F^{(\mathrm{e})}\|_o^2 &= \sup_{\|\theta\|_{\mathbf{B}}=1} \sum_{j=1}^{\mathcal{J}}\sum_{l=1}^{\mathcal{L}} (F(c_j|w_l) - \int 1\{g(w_l,\eta) \leq c_j\}\mu(d\eta))^2 \\ &\leq \sum_{j=1}^{\mathcal{J}}\sum_{l=1}^{\mathcal{L}} \{2 \sup_{\|F(c_j|\cdot)\|_\infty=1} (F(c_j|w_l))^2 + 2\sup_{\|\mu\|_{TV}=1} (|\mu|(\Omega))^2\} = 4\mathcal{J}\mathcal{L}. \quad \text{(S.129)}\end{aligned}$$

Moreover, note that for any bounded $f : \Omega \to \mathbf{R}$ and $\mu_1, \mu_2 \in \mathcal{M}$ it follows that

$$\int_\Omega f(\eta)(\mu_1(d\eta) - \mu_2(d\eta)) \leq \|f\|_\infty |\mu_1 - \mu_2|(\Omega) = \|f\|_\infty \|\mu_1 - \mu_2\|_{TV},$$

which implies $\Upsilon_F^{(\mu)}$ and $\Upsilon_F^{(\mathrm{s})}$ are Fréchet differentiable with $\nabla\Upsilon_F^{(\mu)} = \Upsilon_F^{(\mu)} + 1$, $\nabla\Upsilon_F^{(\mathrm{s})} = \Upsilon_F^{(\mathrm{s})} + \lambda$, $\|\nabla\Upsilon_F^{(\mu)}\|_o \leq 1$, and $\|\nabla\Upsilon_F^{(\mathrm{s})}\|_o \leq \|\Psi(g,\cdot)\|_\infty$. By (S.129) we may therefore conclude Assumptions 3.9(i)(ii)(iii) are satisfied with $\nabla\Upsilon_F$ as in (S.128) and $K_f = 0$. Furthermore, note that (provided $\Theta_n \cap R \neq \emptyset$) there is a $\theta^\star \in \mathbf{B}_n$ such that $\Upsilon_F(\theta^\star) = 0$, which together with (S.128) implies the range of $\nabla\Upsilon_F$ equals $\mathbf{F}_n$ and hence Assumption 3.9(iv) holds. Finally, we note Assumption 3.10 is immediate due to Υ_F being affine. ■

Lemma S.4.6. *Let $\{r_j\}_{j=1}^{j_n}$ be functions of X, $r^{j_n}(x) = (r_1(x), \ldots, r_{j_n}(x))'$, define the class $\mathcal{G}_n = \{r^{j_n\prime}\beta$ for some β with $\|\beta\|_2 \leq C_n\}$, and suppose $b_{1n} \equiv \sup_x \|r^{j_n}(x)\|_2$ and $b_{2n} \equiv \sup_z \|q^{k_n}(z)\|_2$ are finite. If $\{X_i, Z_i\}_{i=1}^n$ is i.i.d. with $(X,Z) \sim P \in \mathbf{P}$, then there is an isonormal Gaussian process $\mathbb{G}_P$ such that uniformly in $P \in \mathbf{P}$*

$$\sup_{g\in\mathcal{G}_n} \|\frac{1}{\sqrt{n}}\sum_{i=1}^n (g(X_i)q^{k_n}(Z_i) - E_P[g(X)q^{k_n}(Z)]) - \mathbb{G}_P(gq^{k_n})\|_2 = O_P(\frac{C_n\sqrt{k_n j_n}b_{1n}b_{2n}\log(n)}{\sqrt{n}}). \quad \text{(S.130)}$$

Proof: For notational simplicity, we first define a $k_n \times j_n$ matrix $\mathbb{E}_n^{(1)}$ to be given by

$$\mathbb{E}_n^{(1)} \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \{q^{k_n}(Z_i)r^{j_n}(X_i)' - E_P[q^{k_n}(Z)r^{j_n}(X)']\}.$$

For any matrix A let $\text{vec}\{A\}$ denote a column vector consisting of the unique elements of A and set $\mathbb{E}_n \equiv \text{vec}\{\mathbb{E}_n^{(1)}\}$, noting that $\mathbb{E}_n$ has dimension (at most) $j_n k_n$. Our first step is to couple $\mathbb{E}_n$ to a normal vector $\mathbb{N}_P$. To this end, we note that

$$\sup_{z,x} \|\text{vec}\{q^{k_n}(z)r^{j_n}(x)' - E_P[q^{k_n}(Z)r^{j_n}(X)']\}\|_2^2$$
$$\leq \sup_{z,x} 4\text{trace}\{q^{k_n}(z)r^{j_n}(x)'r^{j_n}(x)q^{k_n}(z)'\} \leq 4b_{1n}^2 b_{2n}^2$$

by definition of b_{1n} and b_{2n}. Since the dimension of $\mathbb{E}_n$ is at most $j_n k_n$, Theorem 1.1 in Zhai (2018) and Markov's inequality imply, provided the underlying probability space is suitably rich, that there is a Gaussian vector $\mathbb{N}_P$ such that

$$\|\mathbb{E}_n - \mathbb{N}_P\|_2 = O_P(\frac{\sqrt{k_n j_n} b_{1n} b_{2n} \log(n)}{\sqrt{n}}) \tag{S.131}$$

uniformly in $P \in \mathbf{P}$. Next observe that for any $g \in \mathcal{G}_n$ there exists a $\beta \in \mathbf{R}^{j_n}$ such that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (g(X_i)q^{k_n}(Z_i) - E_P[g(X)q^{k_n}(Z)]) = \mathbb{E}_n^{(1)}\beta.$$

Hence, letting $\mathbb{N}_P^{(1)}$ denote the $k_n \times j_n$ matrix built from the corresponding entries of the normal vector $\mathbb{N}_P$, we define the Gaussian process $\mathbb{G}_P$ by setting

$$\mathbb{G}_P(gq^{k_n}) = \mathbb{N}_P^{(1)}\beta$$

for any $r^{j_n\prime}\beta = g \in \mathcal{G}_n$. Therefore, since $\|\beta\|_2 \leq C_n$ by definition of $\mathcal{G}_n$, and the operator norm is bounded by the Frobenius norm, we obtain from result (S.131) that

$$\sup_{g \in \mathcal{G}_n} \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (g(X_i)q^{k_n}(Z_i) - E_P[g(X)q^{k_n}(Z)]) - \mathbb{G}_P(gq^{k_n})\|_2$$
$$\leq \|\mathbb{E}_n^{(1)} - \mathbb{N}_P^{(1)}\|_{o,2} C_n = O_P(\frac{C_n \sqrt{k_n j_n} b_{1n} b_{2n} \log(n)}{\sqrt{n}})$$

uniformly in $P \in \mathbf{P}$, and hence the claim of the lemma follows. ■

Lemma S.4.7. *Let $\{r_j\}_{j=1}^{j_n}$ be a set of functions of X, $r^{j_n}(x) \equiv (r_1(x), \ldots, r_{j_n}(x))'$, and suppose $\sup_x \|r^{j_n}(x)\|_2 \lesssim b_{1n}$, $\sup_z \|q^{k_n}(z)\|_2 \lesssim b_{2n}$, and $E_P[q^{k_n}(Z)q^{k_n}(Z)']$ and $E_P[r^{j_n}(X)r^{j_n}(X)']$ have eigenvalues bounded uniformly in $P \in \mathbf{P}$, j_n, k_n. If $\{X_i, Z_i\}_{i=1}^{n}$ is i.i.d. with $(X, Z) \sim P \in \mathbf{P}$, then there is a $K < \infty$ such that for all $\delta \geq 0$*

$$\sup_{P \in \mathbf{P}} P(\|\frac{1}{n} \sum_{i=1}^{n} q^{k_n}(Z_i)r^{j_n}(X_i)' - E_P[q^{k_n}(Z)r^{j_n}(X)']\|_{o,2} > \delta)$$
$$\leq (j_n + k_n) \exp\{-\frac{n\delta^2 K}{b_{1n}^2 \vee b_{2n}^2 + \delta b_{1n} b_{2n}}\}.$$

Proof: We first define a $k_n \times j_n$ random matrix $\mathbb{M}_{i,n}$ satisfying $E_P[\mathbb{M}_{i,n}] = 0$ by

$$\mathbb{M}_{i,n} \equiv \frac{1}{n}\{q^{k_n}(Z_i)r^{j_n}(X_i)' - E_P[q^{k_n}(Z)r^{j_n}(X)']\}.$$

Since for any random matrix A we have $\|E[A]\|_o \leq E[\|A\|_o]$ by Jensen's inequality, $\|A\|_o^2 \leq \text{trace}\{A'A\}$, $\sup_x \|r^{j_n}(x)\|_2 \lesssim b_{1n}$, and $\sup_z \|q^{k_n}(z)\|_2 \lesssim b_{2n}$ imply

$$\|\mathbb{M}_{i,n}\|_o^2 \lesssim \|\frac{1}{n}q^{k_n}(Z_i)r^{j_n}(X_i)'\|_o^2 + E_P[\|\frac{1}{n}q^{k_n}(Z)r^{j_n}(X)'\|_o^2]$$
$$\lesssim \frac{\sup_z \|q^{k_n}(z)\|_2^2 \times \sup_x \|r^{j_n}(x)\|_2^2}{n^2} \lesssim \frac{b_{1n}^2 b_{2n}^2}{n^2}. \quad \text{(S.132)}$$

Moreover, since the eigenvalues of $E_P[q^{k_n}(Z)q^{k_n}(Z)']$ are bounded uniformly in $P \in \mathbf{P}$ by assumption and $\sup_x \|r^{j_n}(x)\|_2 \lesssim b_{1n}$ it additionally follows that

$$\sup_{P\in\mathbf{P}} \|\sum_{i=1}^n E_P[\mathbb{M}_{i,n}\mathbb{M}_{i,n}']\|_o \leq \sup_{P\in\mathbf{P}} \frac{2}{n}\|E_P[q^{k_n}(Z)q^{k_n}(Z)'\|r^{j_n}(X)\|_2^2]\|_o \lesssim \frac{b_{1n}^2}{n}. \quad \text{(S.133)}$$

Identical arguments but relying on the eigenvalues of $E_P[r^{j_n}(X)r^{j_n}(X)']$ being bounded uniformly in $P \in \mathbf{P}$ and $\sup_x \|q^{k_n}(x)\|_2 \lesssim b_{2n}$ by hypothesis further yield that

$$\sup_{P\in\mathbf{P}} \|\sum_{i=1}^n E_P[\mathbb{M}_{i,n}'\mathbb{M}_{i,n}]\|_o \lesssim \frac{b_{2n}^2}{n}. \quad \text{(S.134)}$$

The claim of the lemma then follows from results (S.132), (S.133), and (S.134) allowing us to apply Theorem 1.6 in Tropp (2012) with $\sigma^2 \asymp (b_{1n}^2 \vee b_{2n}^2)/n$ and $R \asymp b_{1n}b_{2n}/n$. ■

Lemma S.4.8. *If Assumptions 4.1(i)-(iii), 4.2(i)(ii) hold, and* $j_n^3 k_n^2 \log(1 + j_n k_n) = o(n)$*, then it follows that Assumption 3.11 holds with* $R = \Theta$ *for any sequence* a_n *satisfying* $k_n^{1/p}(k_n^2 j_n^5 \log^3(1 + k_n j_n)/n)^{1/4} = o(a_n)$*.*

Proof: Let $\mathcal{G}_n \equiv \{g : g(x) = 1\{y \leq c_\jmath\} - p^{j_n}(w)'\beta$ for some $1 \leq \jmath \leq \mathcal{J}$ and $\|p^{j_n\prime}\beta\|_\infty \leq 2\}$ and $\tilde{\mathcal{F}}_n \equiv \{gq_k : g \in \mathcal{G}_n$ and $1 \leq k \leq k_n\}$. Further let $\mathbb{G}_P^\star$ be a Gaussian process on $\tilde{\mathcal{F}}_n$ independent of $\{V_i\}_{i=1}^n$, satisfying $E[\mathbb{G}_P^\star(f_1)] = 0$ and $E[\mathbb{G}_P^\star(f_1)\mathbb{G}_P^\star(f_2)] = \text{Cov}_P\{f_1, f_2\}$ for any $f_1, f_2 \in \tilde{\mathcal{F}}_n$, and for any $f \in \tilde{\mathcal{F}}_n$ define $\hat{\mathbb{G}}_n(f)$ to be given by

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

where $\{\omega_i\}_{i=1}^n$ are the same weights used in building $\hat{\mathbb{W}}_n$. Then note that when $R = \Theta$ and for $\mathbb{W}_P^\star(\theta) \equiv (\mathbb{G}_P^\star(\rho_1(\cdot,\theta)q^{k_n})', \ldots \mathbb{G}_P^\star(\rho_{\mathcal{J}}(\cdot,\theta)q^{k_n})')'$, we obtain

$$\sup_{\theta\in\Theta_n} \|\hat{\mathbb{W}}_n(\theta) - \mathbb{W}_P^\star(\theta)\|_p \lesssim k_n^{1/p} \sup_{f\in\tilde{\mathcal{F}}_n} |\hat{\mathbb{G}}_n(f) - \mathbb{G}_P^\star(f)|. \quad \text{(S.135)}$$

We will therefore establish the lemma by employing (S.135) and applying Theorem S.7.1(i) to the class $\tilde{\mathcal{F}}_n$. To this end, define $f^{d_n}(V)$ to be given by

$$f^{d_n}(V) \equiv g^{d_n}(V) - E_P[g^{d_n}(V)] \qquad g^{d_n}(V) \equiv q^{k_n}(Z) \otimes \begin{pmatrix} p^{j_n}(W) \\ 1\{Y \leq c_1\} \\ \vdots \\ 1\{Y \leq c_{\mathcal{J}}\} \end{pmatrix} \tag{S.136}$$

and note $d_n = k_n(j_n + \mathcal{J})$. Next observe that applying Lemma S.4.22 with $D_1 \equiv (p^{j_n}(W)', 1\{Y \leq c_1\}, \ldots, 1\{Y \leq c_{\mathcal{J}}\})'$ and $D_2 = q^{k_n}(Z)$ allows us to conclude

$$\sup_{P\in\mathbf{P}} \overline{\text{eig}}\{E_P[g^{d_n}(V)g^{d_n}(V)']\} \leq \sup_{P\in\mathbf{P}}(\|\overline{\text{eig}}\{D_1D_1'\}\|_{P,\infty} \times \overline{\text{eig}}\{E_P[D_2D_2']\}) \lesssim j_n, \tag{S.137}$$

where the final inequality holds by Assumptions 4.1(ii) and 4.2(ii). Hence, since in addition $\overline{\text{eig}}\{E_P[g^{d_n}(V)]E[g^{d_n}(V)']\} \leq \overline{\text{eig}}\{E_P[g^{d_n}(V)g^{d_n}(V)']\}$, results (S.136) and (S.137) imply Assumption S.7.1(i) holds with $C_n \asymp j_n$. Next note Assumption S.7.1(ii) is satisfied with $K_n \asymp \sqrt{k_n j_n}$ by Assumptions 4.1(ii) and 4.2(i). By Assumptions 4.1(iii) it also follows that $\|\beta\|_2 \asymp \sup_{P\in\mathbf{P}} \|p^{j_n\prime}\beta\|_{P,2} \leq \|p^{j_n\prime}\beta\|_\infty$. Hence, by definition of $\tilde{\mathcal{F}}_n$, there is a $C_0 < \infty$ such that any $f \in \tilde{\mathcal{F}}$ satisfies $f(V) - E_P[f(V)] = f^{d_n}(V)'\beta$ for some β in

$$\mathcal{B}_n \equiv \{\beta \in \mathbf{R}^{d_n} : \beta = e_k \otimes \gamma \text{ for some } \gamma \in \mathbf{R}^{j_n+\mathcal{J}} \text{ with } \|\gamma\|_2 \leq C_0\},$$

where $e_k \in \mathbf{R}^{k_n}$ has its k^{th} coordinate equal to one and all other coordinates equal to zero. In particular, it follows that Assumption S.7.2(i) is immediate with $G_{n,P}$ equal to the zero function and $J_{1n} = 0$. Moreover, setting $\mathcal{C}_n \equiv \{\gamma \in \mathbf{R}^{j_n+\mathcal{J}} : \|\gamma\|_2 \leq C_0\}$, we can then conclude from the definition of $\mathcal{B}_n$ and $N(\epsilon, \mathcal{C}_n, \|\cdot\|_2) \lesssim 1 \vee (C_0/\epsilon)^{j_n}$ that

$$\int_0^\infty \sqrt{\log(N(\epsilon, \mathcal{B}_n, \|\cdot\|_2))}d\epsilon$$
$$\lesssim \int_0^{C_0} \sqrt{\log(k_n) + \log(N(\epsilon, \mathcal{C}_n, \|\cdot\|_2))}d\epsilon \lesssim \sqrt{\log(k_n)} + \sqrt{j_n},$$

which verifies Assumption S.7.2(ii) is satisfied with $J_{2n} \asymp \sqrt{\log(k_n)} + \sqrt{j_n}$. Thus, applying Theorem S.7.1(i) with $K_n \asymp \sqrt{k_n, j_n}$, $C_n \asymp j_n$, $d_n \lesssim k_n j_n$, $J_{1n} = 0$, and $J_{2n} \asymp \sqrt{\log(k_n)} + \sqrt{j_n}$ implies that uniformly in $P \in \mathbf{P}$ we have

$$\sup_{f\in\tilde{\mathcal{F}}_n} |\hat{\mathbb{G}}_n(f) - \mathbb{G}_P^\star(f)| = O_P(\{\frac{k_n^2 j_n^5 \log^3(1 + k_n j_n)}{n}\}^{1/4}) \tag{S.138}$$

provided that $j_n^3 k_n^2 \log(1 + j_n k_n) = o(n)$. Since the latter condition is satisfied by hypothesis, the claim of the lemma then follows from (S.135) and (S.138). ■

Lemma S.4.9. *Define* $\|\theta\|_{\mathbf{E}} = \sum_{\jmath=1}^{\mathcal{J}} \sup_{P\in\mathbf{P}} \|F(c_\jmath|\cdot)\|_{P,2}$ *and for* $\hat{V}_n(\theta, R| + \infty)$ *as in*

(34) *let* $\hat{V}_n(\theta, R|\ell_n) = \hat{V}_n(\theta, R|+\infty) \cap \{h : \|h/\sqrt{n}\|_{\mathbf{E}} \leq \ell_n\}$*. If Assumptions 4.1, 4.2, and 4.3 hold, then for any* $a_n = o(1)$ *and* $\ell_n = o(1)$ *satisfying* $k_n^4 j_n^5 \log^3(1 + k_n j_n)/n = o(a_n^4)$ *and* $k_n\sqrt{j_n}\log(n)/\sqrt{n} = o(\ell_n)$ *it follows uniformly in* $P \in \mathbf{P}_0$ *that*

$$\hat{U}_n(R|+\infty) = \inf_{\theta \in \hat{\Theta}_n^{\mathrm{r}}} \inf_{h \in \hat{V}_n(\theta, R|\ell_n)} \{\sum_{\jmath=1}^{\mathcal{J}} \|\hat{\mathbb{W}}_{\jmath,n}(\theta) + \hat{\mathbb{D}}_{\jmath,n}[h]\|_{\hat{\Sigma}_{\jmath,n},2}\}^{1/2} + o_P(a_n).$$

Proof: We establish the claim of the lemma by verifying the conditions of Lemma S.3.1. To this end, recall that in the proof of Theorem 4.1 we argued that Assumptions 3.2(i)(iii) and 3.7 hold with $B_n \asymp \sqrt{k_n}$ and $J_n \asymp \sqrt{j_n \log(1 + j_n)}$. Moreover, Assumption 4.1(iii) implies that for any $(\{p^{j_n\prime}\beta_{\jmath,h}\}_{\jmath=1}^{\mathcal{J}}, \mu) = h \in \mathbf{B}_n$ we have

$$\|h\|_{\mathbf{E}} \lesssim \sum_{\jmath=1}^{\mathcal{J}} \|\beta_{\jmath,h}\|_2 \lesssim \{\sum_{\jmath=1}^{\mathcal{J}} \|\mathbb{D}_{\jmath,P}[h]\|_2^2\}^{1/2} = \|\mathbb{D}_P[h]\|_2, \tag{S.139}$$

where the second inequality follows from $\mathbb{D}_{\jmath,P}[h] = -E_P[q^{k_n}(Z)p^{j_n}(W)'\beta_{\jmath,h}]$ and the smallest singular values of $E_P[q^{k_n}(Z)p^{j_n}(W)']$ being bounded away from zero uniformly in $P \in \mathbf{P}$ by Assumption 4.2(iii). Since $\nu_n \asymp 1$ by Lemma S.4.1 and the derivative $\mathbb{D}_P(\theta)$ does not depend on θ, we conclude $\|h\|_{\mathbf{E}} \leq \nu_n \|\mathbb{D}_P[h]\|_2$ for all $h \in \mathbf{B}_n$ – i.e., in verifying the conditions of Lemma S.3.1 we may set $\mathcal{A}_n(P) = \Theta_n \cap R$. In order to verify condition (S.79) of Lemma S.3.1 we note that since $\|h\|_{\mathbf{E}} \asymp \sum_{\jmath=1}^{\mathcal{J}} \|\beta_{\jmath,h}\|_2$ by Assumption 4.1(iii), the definitions of the operator norm $\|\cdot\|_{o,2}$, $\hat{\mathbb{D}}_{\jmath,n}$, and $\mathbb{D}_{\jmath,P}$ imply that

$$\sup_{h \in \mathbf{B}_n} \frac{\|\hat{\mathbb{D}}_n[h] - \mathbb{D}_P[h]\|_2}{\|h\|_{\mathbf{E}}} \lesssim \|\frac{1}{n}\sum_{i=1}^{n} q^{k_n}(Z_i)p^{j_n}(W_i)' - E_P[q^{k_n}(Z)p^{j_n}(W)']\|_{o,2} = o_P(1),$$

where the final equality holds uniformly in $P \in \mathbf{P}$ by applying Lemma S.4.7 with $b_{1n} = \sqrt{j_n}$, $b_{2n} = k_n$ (by Assumptions 4.1(ii) and 4.2(i)) and employing that $k_n \geq j_n$ and $k_n^2 \log(k_n)/n = o(1)$ by Assumptions 4.2(iii)(iv). Finally, we note that $j_n^5 k_n^4 \log^3(1 + j_n k_n)/n = o(a_n^4)$ by hypothesis, and employing Lemma S.4.8 with $p = 2$ yields that Assumption 3.11 holds for $R = \Theta$, and hence also for R as in (29). The only condition of Lemma S.3.1 that remains to be verified is that $\mathcal{S}_n(\mathbf{B}, \mathbf{E})\mathcal{R}_n = o(\ell_n)$. To this end, we observe that since $\hat{V}_n(\theta, R|\ell_n)$ is defined through the constraint $\|h\|_{\mathbf{E}} \leq \ell_n$ (instead of $\|\cdot\|_{\mathbf{B}} \leq \ell_n$), it suffices to verify $\mathcal{R}_n = o(\ell_n)$ – i.e. for the purposes of this lemma we may set $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{\mathbf{E}}$. However, since as argued $J_n \asymp \sqrt{j_n \log(1 + j_n)}$, $B_n = \sqrt{k_n}$, and $\nu_n \asymp 1$, we have $\mathcal{R}_n \asymp k_n\sqrt{j_n \log(1 + k_n)\log(1 + j_n)}/\sqrt{n}$, and the requirement $\mathcal{R}_n = o(\ell_n)$ is implied by $k_n\sqrt{j_n}\log(n)/\sqrt{n} = o(\ell_n)$. Thus, the claim of the lemma follow from Lemma S.3.1. ■

S.4.2 Proofs for Section A.2.1

Proof of Theorem A.2.1: We establish the theorem by simply applying Theorem 3.1(ii) to both R as in (A.5) (to couple $I_n(R)$) and to $R = \Theta$ (to couple $I_n(\Theta)$). To this end, note that as discussed Assumption 3.1(ii)(iii) holds, while Assumption 3.1(i) is directly imposed in A.2.1(i). Since $q^{k_n}(Z)$ equals the vector $(1, \ldots, 1)' \in \mathbf{R}^{\mathcal{J}}$, it further follows Assumption 3.2(i) holds with $B_n = 1$, while Assumption 3.2(ii) is automatically satisfied. We further note that Assumption 3.2(iii) holds for $R = \Theta$ (and hence also for R as in (A.5)) with $J_n = C_0$ for some $C_0 < \infty$ by Assumption A.2.2(ii) and Lemma S.4.11. Also note Assumption 3.3(i) is satisfied for $R = \Theta$, and hence also for R as in (A.5), by Lemma S.4.12. Additionally, since Θ is convex by Assumption A.2.1(iii), the mean value theorem and Assumption A.2.2(ii) imply that

$$E_P[\|\rho(X, \theta_1) - \rho(X, \theta_2)\|_2^2] \leq E_P[\sup_{\theta \in \Theta} \|\nabla_\theta \rho(X, \theta)\|_{o,2}^2]\|\theta_1 - \theta_2\|_2^2$$

for all $\theta_1, \theta_2 \in \Theta$, which verifies Assumption 3.3(ii) holds with $\kappa_\rho = 1$ and $\|\cdot\|_{\mathbf{E}} = \|\cdot\|_2$. Lemma S.4.10 additionally verifies that Assumption 3.4 holds with $\|\cdot\|_{\mathbf{E}} = \|\cdot\|_2$ and $\nu_n^{-1} = \eta$ for some $\eta > 0$ when $R = \Theta$ and hence also when R is as in (A.5). Furthermore, we note that in this problem $\mathcal{R}_n \asymp n^{-1/2}$ because $\nu_n \asymp 1$, $J_n = O(1)$, $k_n = \mathcal{J}$, and $B_n = 1$. To verify Assumption 3.5, note that in this application $\nabla m_{P,\jmath}(\theta) = E_P[\nabla_\theta \rho_\jmath(X, \theta)]$. Hence, Assumptions 3.5(i)(ii) hold with $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_2$ due to $E_P[\sup_{\theta \in \Theta} \|\nabla_\theta^2 \rho_\jmath(X, \theta)\|_{o,2}]$ being bounded in $P \in \mathbf{P}$ by Assumption A.2.2(ii). Similarly, Assumption 3.5(iii) is satisfied due to $E_P[\sup_{\theta \in \Theta} \|\nabla_\theta \rho(X, \theta)\|_{o,2}]$ being bounded by Assumption A.2.2(ii). Finally, we note that since $\mathcal{R}_n \asymp n^{-1/2}$ and $\kappa_\rho = 1$, Lemma S.4.11 verifies Assumption 3.6(i). Assumption 3.6(ii) is immediate since $E_P[\rho(X, \theta_0)] = 0$, while Assumption 3.7 holds by Assumption A.2.4. To conclude, simply note that the condition $k_n^{1/p}\sqrt{\log(1 + k_n)}B_n \times \sup_{P \in \mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$ is implied by $\ell_n\sqrt{\log(1/\ell_n)} = o(a_n)$ by Lemma S.4.11, and $K_m \mathcal{R}_n^2 = o(a_n/\sqrt{n})$ is implied by $n^{-1/2} = o(a_n)$. ■

Proof of Theorem A.2.2: We first define a variable $\hat{E}_n(R|\ell_n)$ to be given by

$$\hat{E}_n(R|\ell_n) \equiv \inf_{h \in \hat{V}_n(\hat{\theta}_n, R|\ell_n)} \|\hat{\mathbb{W}}_n(\hat{\theta}_n) + \hat{\mathbb{D}}_n(\hat{\theta}_n)[h]\|_{\hat{\Sigma}_n,2}$$

and note Lemma S.4.15 implies $\hat{U}_n(R| + \infty) = \hat{E}_n(R|\ell_n) + o_P(a_n)$ uniformly in $P \in \mathbf{P}_0$ for any $\ell_n \downarrow 0$ satisfying the conditions of the theorem. Therefore, to establish the theorem it suffices to show that uniformly in $P \in \mathbf{P}_0$ we have

$$\hat{E}_n(R|\ell_n) \geq U_P^\star(R|\tilde{\ell}_n) + o_P(a_n)$$
$$\hat{E}_n(R|\ell_n) - \hat{U}_n(\Theta| + \infty) \geq U_P^\star(R|\tilde{\ell}_n) - U_P^\star(\Theta|\tilde{\ell}_n^{\mathrm{u}}) + o_P(a_n).$$

with $\ell_n \asymp \tilde{\ell}_n$ and $\tilde{\ell}_n^{\mathrm{u}}$ satisfying the conditions of the theorem. To this end we rely on Theorem 3.2 (for $\hat{E}_n(R|\ell_n)$) and Lemma S.3.7. Next note that in the proof of Theorem A.2.1 we established that Assumptions A.2.1, A.2.2, A.2.3, and A.2.4, imply Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, and 3.7 hold with $\mathcal{R}_n \asymp n^{-1/2}$, $\nu_n \asymp 1$, $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{\mathbf{E}} = \|\cdot\|_{\mathbf{L}} = \|\cdot\|_2$, $\kappa_\rho = 1$, and $a_n = \sqrt{\log(n)}/n^{\frac{1}{10+5d_\theta}}$ for $R = \Theta$ and R as in (A.5). We thus avoid repeating the arguments, and verify only that Assumptions 3.8, 3.9, 3.10, 3.11, 3.12, and 3.13 hold for $R = \Theta$ and R as in (A.5).

Next note Lemma S.4.14 implies Assumptions 3.8, 3.9, and 3.10 are satisfied, while Lemma S.4.13 verifies Assumption 3.11 with $a_n = \sqrt{\log(n)}/n^{\frac{1}{10+5d_\theta}}$ for $R = \Theta$, and hence also for R as in (A.5). Assumption 3.12(i) is immediate since $\|\cdot\|_{\mathbf{E}} = \|\cdot\|_{\mathbf{B}} = \|\cdot\|_2$, while Assumptions 3.12(ii)(iii) are implied by Assumption A.2.5(i), $\|\hat{\theta}_n - \theta_0\|_2 = o_P(1)$ uniformly in $P \in \mathbf{P}_0$ (which we showed in establishing Theorem A.2.1), and $\mathcal{V}_n(P) \equiv \{\theta \in \Theta : \|\theta - \theta_0\|_2 \le \epsilon\}$ for some $\epsilon > 0$ by Lemma S.4.10. Assumption 3.13(i) is immediate since $\mathcal{S}_n(\mathbf{B}, \mathbf{E}) = 1$ and the choices of $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$ correspond to setting $\tau_n = o(n^{-1/2})$. Similarly, Lemma S.4.11, $\mathcal{S}_n(\mathbf{L}, \mathbf{E}) = 1$, and $n^{-1/2} = o(\ell_n)$ imply that the condition $\ell_n^2\sqrt{\log(1/\ell_n)} = o(a_n n^{-\frac{1}{2}})$ verifies Assumption 3.13(ii). Moreover, since $\ell_n = o(r_n)$ and $n^{-1/2} = o(r_n)$ Assumption 3.13(iii) holds. Hence, Theorem 3.2 implies

$$\hat{E}_n(R|\ell_n) \ge U_P^\star(R|\tilde{\ell}_n) + o_P(a_n) \tag{S.140}$$

uniformly in $P \in \mathbf{P}_0$ for some $\ell_n \asymp \tilde{\ell}_n$. Similarly, since $\mathcal{R}_n^{\mathrm{u}} \asymp n^{-1/2}$, the conditions of Lemma S.3.7(ii) are immediate and hence by (S.140) there are $\ell_n \asymp \tilde{\ell}_n$ and $\ell_n^{\mathrm{u}} \asymp \tilde{\ell}_n^{\mathrm{u}}$ with

$$\hat{E}_n(R|\ell_n) - \hat{U}_n(\Theta| + \infty) \ge U_P^\star(R|\tilde{\ell}_n) - U_P^\star(\Theta|\tilde{\ell}_n^{\mathrm{u}}) + o_P(a_n). \tag{S.141}$$

The theorem therefore follows from (S.140), (S.141) and Lemma S.4.15. ■

Lemma S.4.10. *If Assumptions A.2.1, A.2.2, A.2.3, and A.2.4(ii) hold, then Assumption 3.4 is satisfied with $R = \Theta$ and R as in* (A.5)*, $\|\cdot\|_{\mathbf{E}} = \|\cdot\|_2$, $\nu_n^{-1} = \eta$ for some $\eta > 0$, and $\mathcal{V}_n(P) \equiv \{\theta \in \Theta : \|\theta - \theta_0\|_2 \le \epsilon\}$ for some $\epsilon > 0$.*

Proof: To verify Assumption 3.4(ii), note Assumptions A.2.1(i), A.2.2(ii), A.2.4, and A.2.3(i) and Lemma S.4.11 allow us to apply Lemma S.1.1(i) with $\|\cdot\|_{\mathbf{A}} = \|\cdot\|_2$, $J_n = O(1)$ and $S_n(\epsilon) > 0$ to conclude $\hat{\theta}_n \in \mathcal{V}_n(P) \equiv \{\theta \in \Theta_n : \|\theta - \theta_0\|_2 \le \epsilon\}$ with probability tending to one uniformly in $P \in \mathbf{P}_0$ for any $\epsilon > 0$ and for both $R = \Theta$ and R as in (A.5). In order to verify Assumption 3.4(i), next note that Θ being convex and Assumption A.2.2(ii) imply that for some $C_0 < \infty$ we have

$$\|E_P[\rho(X,\theta)] - E_P[\rho(X,\theta_0)] - E_P[\nabla_\theta \rho(X,\theta_0)](\theta - \theta_0)\|_2 \le C_0\|\theta - \theta_0\|_2^2$$

for all $\theta \in \Theta$. Hence, since the smallest singular value of $E_P[\nabla_\theta \rho(X,\theta_0)]$ is bounded away from zero uniformly in $P \in \mathbf{P}_0$ by Assumption A.2.3(ii), we obtain for some $C_1 < \infty$

$$\begin{aligned}\|\theta - \theta_0\|_2 &\leq C_1 \|E_P[\nabla_\theta \rho(X, \theta_0)](\theta - \theta_0)\|_2 \\ &\leq C_1\{\|E_P[\rho(X,\theta)] - E_P[\rho(X,\theta_0)]\|_2 + C_0\|\theta - \theta_0\|_2^2\}\end{aligned} \tag{S.142}$$

for all $\theta \in \Theta$ and $P \in \mathbf{P}_0$. Therefore, provided $\epsilon > 0$ is set sufficiently small in defining $\mathcal{V}_n(P) \equiv \{\theta \in \Theta_n : \|\theta - \theta_0\|_2 \leq \epsilon\}$, it follows that Assumption 3.4(i) holds with $\|\cdot\|_{\mathbf{E}} = \|\cdot\|_2$ and $\nu_n^{-1} = \eta$ for some $\eta > 0$ due to (S.142) and Assumption A.2.4(ii). ■

Lemma S.4.11. *Let $\mathcal{F} \equiv \{\rho_\jmath(\cdot, \theta) :$ for some $\theta \in \Theta$ and $1 \leq \jmath \leq \mathcal{J}\}$. If Assumptions A.2.1(iii) and A.2.2 hold, then it follows that $\sup_{P\in\mathbf{P}} N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{P,2}) \lesssim 1 \vee \epsilon^{-d_\theta}$ and $\sup_{P\in\mathbf{P}} J_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{P,2}) \lesssim \epsilon(1 + \sqrt{\log(1 \vee \epsilon^{-1})})$.*

Proof: Since Θ is convex by Assumption A.2.1(iii), the mean value theorem and Assumption A.2.2(i) imply for any $\theta_1, \theta_2 \in \Theta$ and $1 \leq \jmath \leq \mathcal{J}$ that

$$|\rho_\jmath(x, \theta_1) - \rho_\jmath(x, \theta_2)| \leq \sup_{\theta \in \Theta} \|\nabla_\theta \rho(x, \theta)\|_{o,2} \|\theta_1 - \theta_2\|_2. \tag{S.143}$$

Setting $D(x) \equiv \sup_{\theta\in\Theta} \|\nabla_\theta \rho(x, \theta)\|_{o,2}$, then note that Theorem 2.7.11 in van der Vaart and Wellner (1996) and the right hand side of (S.143) not depending on $\jmath$ imply

$$N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{P,2}) \leq \mathcal{J} \times N(\frac{\epsilon}{2\|D\|_{P,2}}, \Theta, \|\cdot\|_2) \lesssim 1 \vee \epsilon^{-d_\theta}, \tag{S.144}$$

where we employed that $N(\epsilon, \Theta, \|\cdot\|_2) \lesssim 1 \vee \epsilon^{-d_\theta}$ due to Θ being bounded by Assumption A.2.1(iii) and $\sup_{P\in\mathbf{P}} \|D\|_{P,2} < \infty$ by Assumption A.2.2(ii).

For the second claim of the Lemma we employ the bound in (S.144) to obtain

$$\begin{aligned}\sup_{P\in\mathbf{P}} J_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{P,2}) &\lesssim \int_0^\epsilon (1 + \log(1 \vee u^{-d_\theta}))^{1/2} du \\ &= \epsilon \int_0^1 (1 + \log(1 \vee (\epsilon v)^{-d_\theta}))^{1/2} dv \lesssim \epsilon(1 + \sqrt{\log(1 \vee \epsilon^{-1})}),\end{aligned}$$

where the first equality follows from the change of variables $v = u/\epsilon$ and the final inequality is implied by the inequality $1 \vee (ab) \leq (1 \vee a)(1 \vee b)$. ■

Lemma S.4.12. *If Assumptions A.2.1(i)(iii) and A.2.2 hold, then it follows that Assumption 3.3(i) is satisfied with $R = \Theta$ and $a_n = \sqrt{\log(n)}/n^{\frac{1}{6+5d_\theta}}$.*

Proof: Let $\epsilon_n = \sqrt{\log(n)}/n^{\frac{1}{6+5d_\theta}}$ and set $\delta_n \equiv 1 \wedge (\epsilon_n^2 \sqrt{n})^{-\frac{2}{2+5d_\theta}}$, which note satisfies $1 \geq \delta_n = o(1)$. Further define $N_n \equiv N(\delta_n, \Theta, \|\cdot\|_2)$ and set $\{\theta_k\}_{k=1}^{N_n}$ to be the center of the N_n balls covering Θ. For notational simplicity, we also let

$$r_{n,P}(x) \equiv ((\rho(x, \theta_1) - E_P[\rho(X, \theta_1)])', \ldots, (\rho(x, \theta_{N_n}) - E_P[\rho(X, \theta_{N_n})])')'$$

and note $r_{n,P}(x) \in \mathbf{R}^{\mathcal{J}N_n}$. For any $P \in \mathbf{P}$ and $\eta > 0$ further define $C_{n,P}(\eta)$ to equal

$$C_{n,P}(\eta) \equiv \frac{(\mathcal{J}N_n)E_P[\|r_{n,P}(X)\|_2^3]}{\eta^3\epsilon_n^3\sqrt{n}}. \tag{S.145}$$

It then follows by Yurinskii's coupling (see, e.g., Theorem 10.10 in Pollard (2002)) that there exists a Gaussian vector $\mathbb{N}_{n,P} \in \mathbf{R}^{\mathcal{J}N_n}$ and universal constant K_0 such that

$$P(\|\frac{1}{\sqrt{n}}\sum_{i=1}^n r_{n,P}(X_i) - \mathbb{N}_{n,P}\|_2 > 3\eta\epsilon_n) \leq K_0 C_{n,P}(\eta)(1 + \frac{|\log(1/C_{n,P}(\eta))|}{\mathcal{J}N_n}). \tag{S.146}$$

Next note Assumption A.2.2(ii), Jensen's inequality, and the convexity of $u \mapsto |u|^{\frac{3}{2}}$ yield

$$\sup_{P\in\mathbf{P}} E_P[\|r_{n,P}(X)\|_2^3] \lesssim (\mathcal{J}N_n)^{\frac{3}{2}} \times \sup_{P\in\mathbf{P}} \frac{1}{\mathcal{J}N_n}\sum_{k=1}^{N_n}\sum_{\jmath=1}^{\mathcal{J}} E_P[|\rho_\jmath(X,\theta_k)|^3] \lesssim N_n^{\frac{3}{2}}. \tag{S.147}$$

In particular, since $N(\epsilon,\Theta,\|\cdot\|_2) \lesssim 1 \vee \epsilon^{-d_\theta}$, it follows from $\delta_n \leq 1$ that $N_n \lesssim \delta_n^{-d_\theta}$, and hence by (S.147) and the definition of $C_{n,P}(\eta)$ in (S.145) we obtain

$$\sup_{P\in\mathbf{P}} C_{n,P}(\eta) \lesssim \frac{N_n^{\frac{5}{2}}}{\eta^3\epsilon_n^3\sqrt{n}} \lesssim \frac{1}{\eta^3\epsilon_n^3(n\delta_n^{5d_\theta})^{\frac{1}{2}}}. \tag{S.148}$$

Moreover, since the function $u \mapsto u(1 + |\log(1/u)|/A)$ with $A \geq 1$ is increasing in u on the interval $(0,1]$ and $\epsilon_n^3(n\delta_n^{5d_\theta})^{\frac{1}{2}} \to \infty$, we obtain from results (S.146) and (S.148) that

$$\begin{aligned}\limsup_{n\to\infty}\sup_{P\in\mathbf{P}} P(\|\frac{1}{\sqrt{n}}\sum_{i=1}^n(r_{n,P}(X_i) - \mathbb{N}_{n,P}\|_2 > 3\eta\epsilon_n) \\ \lesssim \limsup_{n\to\infty}\frac{1}{\eta^3\epsilon_n^3(n\delta_n^{5d_\theta})^{\frac{1}{2}}}(1 + \frac{|\log(\eta^3\epsilon_n^3(n\delta_n^{5d_\theta})^{\frac{1}{2}})|}{\mathcal{J}N_n}) = 0,\end{aligned} \tag{S.149}$$

where the final result follows by direct calculation. Letting $\mathbb{S}_{n,P}$ denote the linear span of $r_{n,P}$ in L_P^2 we then employ $\mathbb{N}_{n,P}$ to define a Gaussian process $\mathbb{G}_P^{(1)}$ on $\mathbb{S}_{n,P}$ by setting

$$\mathbb{G}_P^{(1)}(\sum_{k=1}^{N_n}\lambda_k'\rho(\cdot,\theta_k)) \equiv (\lambda_1',\ldots,\lambda_{N_n}')\mathbb{N}_{n,P} \tag{S.150}$$

for any $\{\lambda_k\}_{k=1}^{N_n}$ with $\lambda_k \in \mathbf{R}^{\mathcal{J}}$. Letting $\text{Proj}\{f|\mathbb{S}_{n,P}\}$ denote the projection of f onto $\mathbb{S}_{n,P}$ under $\|\cdot\|_{P,2}$, and assuming the probability space is suitably large to carry an isonormal process $\mathbb{G}_P^{(2)}$ on $\{(f - \int f dP) - \text{Proj}\{f - \int f dP|\mathbb{S}_{n,P}\} : f \in \mathcal{F}\}$ that is independent of $\mathbb{G}_P^{(1)}$, we then define the isonormal process $\mathbb{G}_P$ to be given by

$$\mathbb{G}_P(f) \equiv \mathbb{G}_P^{(1)}(\text{Proj}\{f|\mathbb{S}_{n,P}\}) + \mathbb{G}_P^{(2)}(f - \text{Proj}\{f|\mathbb{S}_{n,P}\}). \tag{S.151}$$

Next let $\Pi_n\theta$ denote the projection of any $\theta \in \Theta$ onto $\{\theta_k\}_{k=1}^{N_n}$ under $\|\cdot\|_2$ and define

$$\mathcal{G}_{n,P} \equiv \{(\rho_\jmath(\cdot,\theta)-\rho_\jmath(\cdot,\Pi_n\theta))-E_P[(\rho_\jmath(X,\theta)-\rho_\jmath(X,\Pi_n\theta))] : \theta \in \Theta, 1 \leq \jmath \leq \mathcal{J}\}. \quad \text{(S.152)}$$

By the mean value theorem, Θ being convex by Assumption A.2.1(iii), and $\|\theta - \Pi_n\theta\|_2 \leq \delta_n$ for every $\theta \in \Theta$ due to δ_n-balls around $\{\theta_k\}_{k=1}^{N_n}$ covering Θ, it follows that

$$\begin{aligned}\sup_{\theta\in\Theta} &|(\rho_\jmath(x,\theta) - \rho_\jmath(x,\Pi_n\theta)) - E_P[(\rho_\jmath(X,\theta) - \rho_\jmath(X,\Pi_n\theta))]| \\ &\leq \{\sup_{\theta\in\Theta}\|\nabla_\theta\rho(x,\theta)\|_{o,2} + \sup_{P\in\mathbf{P}} E_P[\sup_{\theta\in\Theta}\|\nabla_\theta\rho(X,\theta)\|_{o,2}]\} \times \delta_n.\end{aligned}$$

Hence, setting $G(x) \equiv 1 \vee \{\sup_{\theta\in\Theta}\|\nabla_\theta\rho(x,\theta)\|_{o,2} + \sup_{P\in\mathbf{P}} E_P[\sup_{\theta\in\Theta}\|\nabla_\theta\rho(X,\theta)\|_{o,2}]\}$ it follows that $G\delta_n$ is an envelope for $\mathcal{G}_{n,P}$, which by Assumption A.2.2(ii) satisfies $\sup_{P\in\mathbf{P}}\|G\delta_n\|_{P,2} \lesssim \delta_n$. Further note that if $[f_l, f_u]$ is a bracket containing a function f, then $[f_l - E_P[f_u(X)], f_u - E_P[f_l(X)]]$ contains $f - E_P[f(X)]$ and satisfies

$$\|f_u - f_l - E_P[f_l(X) - f_u(X)]\|_{P,2} \leq 2\|f_u - f_l\|_{P,2}$$

by Jensen's inequality and the triangle inequality. Therefore, Lemma S.4.11 implies

$$\sup_{P\in\mathbf{P}} N_{[\,]}(\epsilon, \mathcal{G}_{n,P}, \|\cdot\|_{P,2}) \lesssim N_n \times (1 \vee \epsilon^{-d_\theta}),$$

and hence Theorem 2.14.2 in van der Vaart and Wellner (1996) together with $\mathcal{G}_{n,P}$ having envelope $\delta_n G$ with $G \geq 1$, $\sup_{P\in\mathbf{P}}\|G\|_{P,2} < \infty$, and $N_n \lesssim \delta_n^{-d_\theta}$ yield

$$\begin{aligned}\sup_{P\in\mathbf{P}} &E_P[\sup_{g\in\mathcal{G}_{n,P}}|\frac{1}{\sqrt{n}}\sum_{i=1}^n(g(X_i) - E_P[g(X)])|] \\ &\lesssim \sup_{P\in\mathbf{P}}\{\delta_n\|G\|_{P,2}\int_0^1(1 + \log N_{[\,]}(\epsilon\delta_n\|G\|_{P,2}, \mathcal{G}_{n,P}, \|\cdot\|_{P,2}))^{\frac{1}{2}}d\epsilon\} \\ &\lesssim \delta_n\int_0^1(1 + \log(N_n) + \log(1 \vee (\epsilon\delta_n)^{-d_\theta}))^{1/2}d\epsilon \\ &\lesssim \delta_n(1 + \log(\delta_n^{-d_\theta}))^{1/2}. \end{aligned} \quad \text{(S.153)}$$

Therefore, the definitions of δ_n and ϵ_n, result (S.153) and Markov's inequality imply

$$\begin{aligned}\limsup_{n\to\infty}\sup_{P\in\mathbf{P}} &P(\sup_{g\in\mathcal{G}_{n,P}}|\frac{1}{\sqrt{n}}\sum_{i=1}^n(g(X_i) - E_P[g(X)])| > \eta\epsilon_n) \\ &\lesssim \limsup_{n\to\infty}\frac{\delta_n(1 + \log(\delta_n^{-d_\theta}))^{1/2}}{\eta\epsilon_n} = 0. \end{aligned} \quad \text{(S.154)}$$

Similarly, since $\mathbb{G}_P$ is Gaussian and $0 \in \mathcal{G}_{n,P}$, Corollary 2.2.8 in van der Vaart and Wellner

(1996) and packing numbers being bounded by bracketing numbers imply

$$\sup_{P\in\mathbf{P}} E_P[\sup_{g\in\mathcal{G}_{n,P}} |\mathbb{G}_P(g)|] \lesssim \sup_{P\in\mathbf{P}} \int_0^\infty (\log N_{[\,]}(\epsilon,\mathcal{G}_{n,P},\|\cdot\|_{P,2}))^{\frac{1}{2}} d\epsilon$$
$$\lesssim \sup_{P\in\mathbf{P}} \int_0^{2\delta_n\|G\|_{P,2}} (\log N_{[\,]}(\epsilon,\mathcal{G}_{n,P},\|\cdot\|_{P,2}))^{\frac{1}{2}} d\epsilon \lesssim \delta_n(1+\log(\delta_n^{-d_\theta}))^{1/2}, \quad (S.155)$$

where in the second inequality we employed that the bracket $[-\delta_n G, \delta_n G]$ covers $\mathcal{G}_{n,P}$ due to $\delta_n G$ being an envelope for $\mathcal{G}_{n,P}$, and the final inequality follows from the change of variables $u = \epsilon/(2\delta_n\|G\|_{P,2})$ and the same manipulations as in (S.153). Hence,

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} P(\sup_{g\in\mathcal{G}_{n,P}} |\mathbb{G}_P(g)| > \eta\epsilon_n) \lesssim \limsup_{n\to\infty} \frac{\delta_n(1+\log(\delta_n^{-d_\theta}))^{1/2}}{\eta\epsilon_n} = 0, \quad (S.156)$$

by result (S.155) and Markov's inequality. To conclude, for any $\theta\in\Theta$ set $\mathbb{W}_P(\theta)$ to be

$$\mathbb{W}_P(\theta) \equiv (\mathbb{G}_P(\rho_1(\cdot,\theta)),\ldots,\mathbb{G}_P(\rho_{\mathcal{J}}(\cdot,\theta))'$$

and note that the definitions of $\mathbb{G}_P$ in (S.150) and (S.151), and of $\mathcal{G}_{n,P}$ in (S.152), yield

$$\sup_{\theta\in\Theta} \|\mathbb{G}_n(\theta) - \mathbb{W}_P(\theta)\|_2 \le \|\frac{1}{\sqrt{n}}\sum_{i=1}^n r_{n,P}(X_i) - \mathbb{N}_{n,P}\|_2$$
$$+ \sup_{g\in\mathcal{G}_{n,P}} \sqrt{\mathcal{J}}|\frac{1}{\sqrt{n}}\sum_{i=1}^n (g(X_i) - E_P[g(X)])| + \sup_{g\in\mathcal{G}_{n,P}} \sqrt{\mathcal{J}}|\mathbb{G}_P(g)|.$$

Thus the lemma follows from (S.149), (S.154), and (S.156). ■

Lemma S.4.13. *If Assumptions A.2.1(i)(iii) and A.2.2 hold, then it follows that Assumption 3.11 is satisfied with $R = \Theta$ and $a_n = \log^{3/4}(n)/n^{\frac{1}{12+2d_\theta}}$.*

Proof: We establish the lemma by relying on Theorem S.7.1(i) in Section S.7. To this end set $\zeta_n = n^{-\frac{1}{2(6+d_\theta)}}$, $M_n = n^{\frac{1}{6+d_\theta}}$, and $N_n \equiv N(\zeta_n, \Theta, \|\cdot\|_2)$. By Assumption A.2.2(ii) the function $F(x) \equiv (1 + \sup_{\theta\in\Theta}\|\rho(x,\theta)\|_2)$ is integrable, and for any $\theta\in\Theta$ we let

$$\tilde\rho(x,\theta) \equiv (\rho_1(x,\theta)1\{F(x)\le M_n\},\ldots,\rho_{\mathcal{J}}(x,\theta)1\{F(x)\le M_n\})'.$$

Defining $d_n = \mathcal{J}N_n$ and $\{\theta_k\}_{k=1}^{N_n}$ to be the centers of the ζ_n-balls covering Θ we then let

$$f_n^{d_n}(X) \equiv (\tilde\rho(X,\theta_1)' - E_P[\tilde\rho(X,\theta_1)'],\ldots,\tilde\rho(X,\theta_{N_n})' - E_P[\tilde\rho(X,\theta_{N_n})'])'.$$

Next note that since each entry of the matrix $f_n^{d_n}(X)f_n^{d_n}(X)'$ is almost surely bounded by $2M_n^2$ it follows that $\|E_P[f_n^{d_n}(X)f_n^{d_n}(X)']\|_{o,2} \le 2d_nM_n^2$, and hence Assumption S.7.1 in Section S.7 holds with $C_n \asymp d_nM_n^2$ and $K_n \asymp M_n$. For every $\theta\in\Theta$ let $\Pi_n\theta$ denote its projection (under $\|\cdot\|_2$) onto $\{\theta_k\}_{k=1}^{N_n}$ and define the class $\mathcal{G}_{n,P} \equiv \{(\rho_j(\cdot,\theta) - \tilde\rho_j(\cdot,\Pi_n\theta)) -$

$E_P[\rho_\jmath(X,\theta) - \tilde{\rho}_\jmath(X,\Pi_n\theta)] : \theta \in \Theta$ and $1 \leq \jmath \leq \mathcal{J}\}$. Further observe that

$$\begin{aligned}
&\sup_{g \in \mathcal{G}_{n,P}} |g(x)| \\
&\leq \max_{1\leq \jmath \leq \mathcal{J}} \sup_{\theta\in\Theta} 2|\rho_\jmath(x,\theta) - \rho_\jmath(x,\Pi_n\theta)| + F(x)1\{F(x) > M_n\} + E_P[F(X)1\{F(X) > M_n\}] \\
&\leq \sup_{\theta\in\Theta} 2\|\nabla_\theta \rho(x,\theta)\|_{o,2}\|\theta - \Pi_n\theta\|_2 + F(x)1\{F(x) > M_n\} + E_P[F(X)1\{F(X) > M_n\}]
\end{aligned} \tag{S.157}$$

where in the second inequality we employed the mean value theorem and Θ being convex by Assumption A.2.1(iii). In particular, since the ζ_n-balls centered around $\{\theta_k\}_{k=1}^{N_n}$ cover Θ and $\zeta_n \leq 1$, result (S.157) implies that the function

$$G(x) \equiv 2 \sup_{\theta\in\Theta} \|\nabla_\theta\rho(x,\theta)\|_{o,2} + F(x) + \sup_{P\in\mathbf{P}} E_P[F(X)]$$

is an envelope for $\mathcal{G}_{n,P}$, while Assumption A.2.2(ii) implies $\sup_{P\in\mathbf{P}} E_P[G^2(X)] < \infty$. Moreover, result (S.157) and Markov's, Jensen's, and Holder's inequalities yield that

$$\begin{aligned}
\sup_{g\in\mathcal{G}_{n,P}} \|g\|_{P,2} &\leq \zeta_n\|G\|_{P,2} + 2\{E_P[F^3(X)]\}^{\frac{2}{3}}\{P(F(X) > M_n)\}^{\frac{1}{3}} \\
&\leq (\zeta_n + M_n^{-1/2} \times 2\sup_{P\in\mathbf{P}}(E_P[F^3(X)])^{1/2}) \times \|G\|_{P,2},
\end{aligned} \tag{S.158}$$

where in the final equality we employed that $\|G\|_{P,2} \geq 1$ because $F(X) \geq 1$. Thus, by result (S.158) and Assumption A.2.2(ii), we may set $\delta_n \equiv C(\zeta_n + M_n^{-1/2})$ and obtain $\|g\|_{P,2} \leq \delta_n\|G\|_{P,2}$ for all $g \in \mathcal{G}_{n,P}$ and $P \in \mathbf{P}$ provided C is chosen large enough. Next note that since Θ being bounded by Assumption A.2.1(iii) implies $N_n \lesssim \zeta_n^{-d_\theta}$, we obtain

$$\sup_{P\in\mathbf{P}} N_{[\,]}(\epsilon, \mathcal{G}_{n,P}, \|\cdot\|_{P,2}) \lesssim N_n \times (1\vee\epsilon)^{-d_\theta} \lesssim \zeta_n^{-d_\theta} \times (1\vee\epsilon)^{-d_\theta} \tag{S.159}$$

due to Lemma S.4.11. Hence, the change of variables $u = \epsilon/(\delta_n\|G\|_{P,2})$ implies that

$$\begin{aligned}
&\sup_{P\in\mathbf{P}} \int_0^{\delta_n\|G\|_{P,2}} (1+\log N_{[\,]}(\epsilon,\mathcal{G}_{n,P},\|\cdot\|_{P,2}))^{1/2} d\epsilon \\
&\qquad \lesssim \sup_{P\in\mathbf{P}} \delta_n\|G\|_{P,2} \int_0^1 (1+\log(\zeta_n^{-d_\theta}) + \log(1\vee(u\delta_n\|G\|_{P,2})^{-d_\theta}))^{1/2} du \\
&\qquad \lesssim \delta_n(1+\log(\zeta_n^{-1}))^{1/2}
\end{aligned} \tag{S.160}$$

where in the inequalities we employed result (S.159), $\zeta_n \lesssim \delta_n$, and $\sup_{P\in\mathbf{P}}\|G\|_{P,2} < \infty$. In particular, results (S.159) and (S.160) together with Lemma S.7.3 imply that Assumption S.7.2(i) in Section S.7 is satisfied with $J_{1n} \lesssim \delta_n(1+\log(\zeta_n^{-1}))^{1/2}$. Similarly, note that in this application, the set $\mathcal{B}_n$ in Assumption S.7.2(ii) consists of $0 \in \mathbf{R}^{d_n}$ and the set of vectors in $\mathbf{R}^{d_n}$ with one coordinate equal to one and all other coordinates equal to zero. Thus, Assumption S.7.2(ii) holds with $J_{2n} = (\log(1+d_n))^{1/2} \lesssim (1+\log(\zeta_n^{-1}))^{1/2}$.

We have so far verified Assumptions S.7.1 and S.7.2 in Section S.7 hold with $d_n \lesssim \zeta_n^{-d_\theta}$, $K_n = M_n$, $C_n \lesssim M_n^2 \zeta_n^{-d_\theta}$, $J_{1n} \lesssim \delta_n(1+\log(\zeta_n^{-1}))^{1/2}$, and $J_{2n} \lesssim (1+\log(\zeta_n^{-1}))^{1/2}$. Since we had set $\zeta_n = n^{-1/(2(6+d_\theta))}$, $M_n = n^{1/(6+d_\theta)}$, and $\delta_n = C(\zeta_n + M_n^{-1/2})$ the requirement that $d_n \log(1+d_n) K_n^2 = o(n)$ imposed by Theorem S.7.1(i) holds as well. Therefore, Assumption A.2.1(i) and Theorem S.7.1(i) finally enable us to conclude that there exists a process $\mathbb{W}_P^\star$ that is independent of the data $\{X_i\}_{i=1}^n$ and such that

$$\sup_{\theta \in \Theta} \|\hat{\mathbb{W}}_n(\theta) - \mathbb{W}_P^\star(\theta)\|_2 = O_P(\log^{3/4}(n) n^{-1/(12+2d_\theta)})$$

uniformly in $P \in \mathbf{P}$, which conclude the proof of the lemma. ■

Lemma S.4.14. *If Assumption A.2.1(ii), A.2.5(ii)-(vi), and A.2.6 hold, then it follows that Assumptions 3.8, 3.9, and 3.10 are satisfied.*

PROOF: Recall that in this setting $\mathbf{G} = \mathbf{R}^{d_G}$ and $\|\cdot\|_{\mathbf{G}} = \|\cdot\|_\infty$. For ϵ and B^ϵ as in Assumption A.2.5, let $N_\epsilon(\theta_0) \equiv \{\theta \in \Theta : \|\theta - \theta_0\|_2 \leq \epsilon\}$ noting that $N_\epsilon(\theta_0) \subseteq B^\epsilon$ and that $N_\epsilon(\theta_0)$ implicitly depends on P through θ_0 (which depends on P through (A.4)). For any $\theta_1, \theta_2 \in N_\epsilon(\theta_0)$, $N_\epsilon(\theta_0) \subseteq B^\epsilon$ and Proposition 7.3.3 in Luenberger (1969) imply

$$\|\Upsilon_G(\theta_1) - \Upsilon_G(\theta_2) - \nabla\Upsilon_G(\theta_1)[\theta_1 - \theta_2]\|_{\mathbf{G}} \leq \{\sup_{\theta \in B^\epsilon} \max_{1 \leq j \leq d_G} \|\nabla^2 \Upsilon_{G,j}(\theta)\|_{o,2}\} \frac{\|\theta_1 - \theta_2\|_2^2}{2}. \tag{S.161}$$

Similarly, for any $\theta_1, \theta_2 \in N_\epsilon(\theta_0)$, Proposition 7.3.2 in Luenberger (1969) yields

$$\begin{aligned} \|\nabla\Upsilon_G(\theta_1) - \nabla\Upsilon_G(\theta_2)\|_o &= \sup_{\|h\|_2 = 1} \max_{1 \leq j \leq d_G} |(\nabla\Upsilon_{G,j}(\theta_1) - \nabla\Upsilon_{G,j}(\theta_2))[h]| \\ &\leq \{\sup_{\theta \in B^\epsilon} \max_{1 \leq j \leq d_G} \|\nabla^2\Upsilon_{G,j}(\theta)\|_{o,2}\} \|\theta_1 - \theta_2\|_2. \end{aligned} \tag{S.162}$$

Since $\|\nabla^2\Upsilon_{G,j}(\theta)\|_{o,2}$ is uniformly bounded on B^ϵ by Assumption A.2.5(v), it follows from results (S.161) and (S.162) that Assumptions 3.8(i)(ii) are satisfied with

$$K_g \equiv \sup_{\theta \in B^\epsilon} \max_{1 \leq j \leq d_G} \|\nabla^2 \Upsilon_{G,j}(\theta)\|_{o,2}.$$

Assumption A.2.5(iii) additionally implies $\sup_{\theta \in B^\epsilon} \|\nabla\Upsilon_G(\theta)\|_{o,2} < \infty$, and hence verifies Assumption 3.8(iii). By identical arguments, but recalling $\mathbf{F} = \mathbf{R}^{d_F}$ and $\|\cdot\|_{\mathbf{F}} = \|\cdot\|_2$, it follows Assumptions A.2.5(iii)-(iv) imply Assumptions 3.9(i)-(iii) hold with

$$K_f \equiv \sqrt{d_F} \sup_{\theta \in B^\epsilon} \max_{1 \leq j \leq d_F} \|\nabla^2 \Upsilon_{F,j}(\theta)\|_{o,2}. \tag{S.163}$$

To conclude, note that since Assumption A.2.5(vi) implies the range of $\nabla\Upsilon_F(\theta)$ equals $\mathbf{R}^{d_F}$ for all $\theta \in B^\epsilon$, it follows that $\nabla\Upsilon_F(\theta)$ admits a right inverse. Moreover, if Υ_F is affine, then $K_f = 0$ and hence Assumption 3.9(iv) holds. On the other hand, if Υ_F is nonlinear, then note $\nabla\Upsilon_F(\theta)^- = \nabla\Upsilon_F(\theta)'(\nabla\Upsilon_F(\theta)\nabla\Upsilon_F(\theta)')^{-1}$ and therefore $\|\nabla\Upsilon_F(\theta)^-\|_{o,2}$ is bounded for all $\theta \in B^\epsilon$ due to $\|\nabla\Upsilon_F(\theta)\|_{o,2}$ being bounded on B^ϵ by Assumption A.2.5(ii), and the smallest singular value of $\nabla\Upsilon_F(\theta)'$ being bounded away from zero on B^ϵ by Assumption A.2.6(ii). It follows Assumption 3.9(iv) holds as well. Since Assumption A.2.6 directly implies Assumption 3.10, the lemma follows. ■

Lemma S.4.15. *Let Assumptions A.2.1, A.2.2, A.2.3, and A.2.4 hold, and set* $a_n = \sqrt{\log(n)}/n^{\frac{1}{10+5d_\theta}}$. *For any* ℓ_n *with* $n^{-1/2} = o(\ell_n)$, *it follows uniformly in* $P \in \mathbf{P}_0$ *that*

$$\hat{U}_n(R|+\infty) = \inf_{h\in\hat{V}_n(\hat{\theta}_n,R|\ell_n)} \|\hat{\mathbb{W}}_n(\hat{\theta}_n) + \hat{\mathbb{D}}_n(\hat{\theta}_n)[h]\|_{\hat{\Sigma}_n,2} + o_P(a_n).$$

Proof: We establish the lemma by relying on Lemma S.3.1. To this end note that in the proof of Theorem A.2.1, Assumptions 3.1(i), 3.2(i)(iii), and 3.7 were verified and $\hat{\theta}_n$ was shown to be consistent for θ_0 uniformly in $P \in \mathbf{P}_0$ with $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{\mathbf{E}} = \|\cdot\|_2$, $\mathcal{R}_n = n^{-1/2}$, and $\nu_n \asymp 1$ for both R as in (A.5). Next, note Lemma S.4.13 verifies Assumption 3.11 holds with $a_n = \sqrt{\log(n)}/n^{\frac{1}{10+5d_\theta}}$ for $R = \Theta$ and hence also for R as in (A.5). Moreover, the mean value theorem and Θ being convex imply that

$$|\frac{\partial}{\partial\theta_k}\rho_\jmath(x,\theta_1) - \frac{\partial}{\partial\theta_k}\rho_\jmath(x,\theta_2)| \le \max_{1\le\jmath\le\mathcal{J}}\sup_{\theta\in\Theta}\|\nabla^2_\theta\rho_\jmath(x,\theta)\|_{o,2}\|\theta_1-\theta_2\|_2 \tag{S.164}$$

for any $\theta_1, \theta_2 \in \Theta$, $1 \le \jmath \le \mathcal{J}$, and $1 \le k \le d_\theta$. Thus, Assumption A.2.2(ii) implies there exists a $C_0 < \infty$ such that for all $P \in \mathbf{P}$ and $\theta_1, \theta_2 \in \Theta$ it follows that

$$\|E_P[\nabla_\theta\rho(X,\theta_1) - \nabla_\theta\rho(X,\theta_2)]\|_{o,2} \le C_0\|\theta_1-\theta_2\|_2.$$

In particular, the function $\theta \mapsto E_P[\nabla_\theta\rho(X,\theta)]$ is uniformly continuous in θ and $P \in \mathbf{P}$, which implies by Assumption A.2.3(ii) that there is an $\epsilon_0 > 0$ such that the smallest singular value of $E_P[\nabla_\theta\rho(X,\theta)]$ is bounded away from zero on $\{\theta \in \Theta : \|\theta-\theta_0\|_2 \le \epsilon_0$ for some $P \in \mathbf{P}_0\}$ (where recall θ_0 implicitly depends on P through (A.4)). Since $\|\mathbb{D}_P(\theta)[h]\|_2 \equiv \|E_P[\nabla_\theta\rho(X,\theta)h]\|_2$, $\nu_n \asymp 1$, $p = 2$, and $\|\cdot\|_{\mathbf{E}} = \|\cdot\|_2$, the Lemma S.3.1 requirement that $\|h\|_{\mathbf{E}} \le \nu_n\|\mathbb{D}_P(\theta)[h]\|_2$ for all $\theta \in \mathcal{A}_n(P)$, $P \in \mathbf{P}_0$, and $h \in \sqrt{n}\{\mathbf{B}_n \cap R - \theta\}$ holds with $\mathcal{A}_n(P) = (\theta_0)^{\epsilon_0}$ and $R = \Theta$ (and hence also for R as in (A.5)). Moreover, by uniform consistency (in $P \in \mathbf{P}_0$) of $\hat{\theta}_n$ it follows that $\hat{\theta}_n \in \mathcal{A}_n(P)$ with probability tending to one uniformly in $P \in \mathbf{P}_0$.

To conclude, define $\mathcal{F} \equiv \{\frac{\partial}{\partial\theta_k}\rho_\jmath(\cdot,\theta) :$ for some $\theta \in \Theta,\ 1 \le \jmath \le \mathcal{J},\ 1 \le k \le d_\theta\}$ and let $F(x) \equiv \max_{1\le\jmath\le\mathcal{J}}\sup_{\theta\in\Theta}\|\nabla^2_\theta\rho_\jmath(x,\theta)\|_{o,2}$. Then note that if ϵ-balls around $\{\theta_i\}_{i=1}^{N_\epsilon}$ cover Θ, then result (S.164) implies that the brackets $[\frac{\partial}{\partial\theta_k}\rho_\jmath(\cdot,\theta_i) - \epsilon F, \frac{\partial}{\partial\theta_k}\rho_\jmath(\cdot,\theta_i) + \epsilon F]$ cover $\mathcal{F}$. Writing these brackets as $\{[f_{l,k}, f_{u,k}]\}_{k=1}^{K_\epsilon}$ for conciseness, further note that $K_\epsilon = \mathcal{J}d_\theta N_\epsilon < \infty$ since $N_\epsilon < \infty$ due to Θ being compact, and $C_1 \equiv \sup_{P\in\mathbf{P}}\|F\|_{P,1} < \infty$

by Assumption A.2.2(ii). Moreover, by definition of $[f_{l,k}, f_{u,k}]$ it further follows that

$$E_P[f_{u,k}(X) - f_{l,k}(X)] = \|f_{u,k} - f_{l,k}\|_{P,1} \leq 2\epsilon C_1 \tag{S.165}$$

for all $P \in \mathbf{P}$. Hence, employing the bound $f(x) - E_P[f(X)] \leq f_{u,k}(x) - E_P[f_{l,k}(X)]$ for $[f_{l,k}, f_{u,k}]$ the bracket containing f, we obtain from result (S.165) that

$$\begin{aligned}
&\sup_{f\in\mathcal{F}}\{\frac{1}{n}\sum_{i=1}^{n} f(X_i) - E_P[f(X)]\} \\
&\qquad \leq \max_{1\leq k\leq K_\epsilon} |\frac{1}{n}\sum_{i=1}^{n} f_{u,k}(X_i) - E_P[f_{u,k}(X)]| + 2\epsilon C_1 = 2\epsilon C_1 + o_P(1),
\end{aligned} \tag{S.166}$$

where the equality holds uniformly in $P \in \mathbf{P}$ by Assumption A.2.2(ii), $K_\epsilon < \infty$, and Theorem 2.8.1 in van der Vaart and Wellner (1996). By identical arguments, we have

$$\begin{aligned}
&\inf_{f\in\mathcal{F}}\{\frac{1}{n}\sum_{i=1}^{n} f(X_i) - E_P[f(X)]\} \\
&\qquad \geq -\max_{1\leq k\leq K_\epsilon} |\frac{1}{n}\sum_{i=1}^{n} f_{l,k}(X_i) - E_P[f_{l,k}(X)]| - 2\epsilon C_1 = -2\epsilon C_1 + o_P(1),
\end{aligned} \tag{S.167}$$

uniformly in $P \in \mathbf{P}$. We thus conclude from results (S.166) and (S.167) that $\mathcal{F}$ is Glivenko-Cantelli uniformly in $P \in \mathbf{P}$. Since by Assumption A.2.5(i) there exists an $\epsilon > 0$ such that $\{\theta : \|\theta - \theta_0\|_2 \leq \epsilon\} \subseteq \Theta$ for all $P \in \mathbf{P}_0$, we can conclude

$$\begin{aligned}
&\sup_{\theta:\|\theta-\theta_0\|_2\leq\epsilon} \sup_{h\in\mathbf{R}^{d_\theta}:\|\frac{h}{\sqrt{n}}\|_2\geq\ell_n} \frac{\|\hat{\mathbb{D}}_n(\theta)[h] - \mathbb{D}_P(\theta)[h]\|_2}{\|h\|_2} \\
&\qquad \leq \sup_{\theta\in\Theta} \|\frac{1}{n}\sum_{i=1}^{n} \nabla_\theta \rho(X_i, \theta) - E_P[\nabla_\theta \rho(X, \theta)]\|_{o,2} = o_P(1)
\end{aligned} \tag{S.168}$$

uniformly in $P \in \mathbf{P}$, and where the equality follows from $\mathcal{F}$ being Glivenko-Cantelli uniformly in $P \in \mathbf{P}$. Since $\nu_n \asymp 1$, result (S.168) verifies condition (S.79) in Lemma S.3.1. This concludes verifying the requirements of Lemma S.3.1 and hence the present Lemma follows for any ℓ_n satisfying $\mathcal{S}_n(\mathbf{B}, \mathbf{E})\mathcal{R}_n = o(\ell_n)$, which in this application is equivalent to $n^{-1/2} = o(\ell_n)$ due to $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{\mathbf{E}} = \|\cdot\|_2$ and $\mathcal{R}_n \asymp n^{-1/2}$. ■

S.4.3 Proofs for Section A.2.2

Proof of Theorem A.2.3: We establish the theorem by simply verifying the conditions of Theorem 3.1(ii) for both R as corresponding to (A.13) and (A.14) (to couple $I_n(R)$) and to $R = \Theta$ (to couple $I_n(\Theta)$). To this end, note Assumption 3.1(i) is imposed in Assumption A.2.7(i), Assumption 3.2(i) holds with $B_n \asymp \sqrt{k_n}$ by Assumption

A.2.9(i), and Assumption 3.2(ii) is satisfied by Assumption A.2.9(ii). Further note that for $R = \Theta$, the class $\mathcal{F}_n$ has bounded envelope F_n by Assumption A.2.7(iii) and $\|g\|_\infty \le C_0$ for any $(g, \gamma) \in \Theta$. Hence, by Lemma S.4.17 it follows that Assumption 3.2(iii) holds with $J_n \asymp \sqrt{j_n \log(1 + j_n)}$ when $R = \Theta$, and therefore also for R as corresponding to (A.13) and (A.14). Next, we note Lemma S.4.18 and $j_n^2 k_n^3 \log^3(n)/n = o(1)$ by Assumption A.2.9(iv) imply Assumption 3.3(i) for both specifications of R under consideration and any a_n satisfying $a_n = O((\log(n))^{-1})$. To verify Assumption 3.3(ii), we observe that for any $(g_1, \gamma_1) \in \Theta_n$ and $(g_2, \gamma_2) \in \Theta_n$, Assumption A.2.7(iii) implies

$$E_P[((Q - g_1(S, Y) - W'\gamma_1) - (Q - g_2(S, Y) - W'\gamma_2))^2] \lesssim \sup_{P \in \mathbf{P}} \|g_1 - g_2\|_{P,2}^2 + \|\gamma_1 - \gamma_2\|_2^2. \quad \text{(S.169)}$$

Hence, since $\|\cdot\|_{\mathbf{E}} \equiv \sup_{P \in \mathbf{P}} \|\cdot\|_{P,2} + \|\cdot\|_2$ it follows Assumption 3.3(ii) holds with $\kappa_\rho = 1$ and some $K_\rho < \infty$. Lemma S.4.16 additionally verifies that Assumption 3.4 holds with $\|\cdot\|_{\mathbf{E}} \equiv \sup_{P \in \mathbf{P}} \|\cdot\|_{P,2} + \|\cdot\|_2$, $\mathcal{V}_n(P) = \Theta_n \cap R$, and $\nu_n^{-1} \asymp \mathrm{s}_n$ for both $R = \Theta$ and R as corresponding to (A.13) and (A.14). Further note that in this application

$$\nabla m_P(\theta)[h] = -E_P[g_h(S, Y) + W'\gamma_h | Z]$$

for any $\theta \in \mathbf{B}$ and $(g_h, \gamma_h) = h \in \mathbf{B}$. By direct calculation it then follows Assumptions 3.5(i)(ii) hold with $K_m = 0$ for $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_{\mathbf{E}}$, and Assumption 3.5(iii) is satisfied for some $M < \infty$ by result (S.169) and Jensen's inequality. To verify Assumption 3.6(i), note that since as argued $\nu_n \asymp 1/\mathrm{s}_n$, $p = 2$, $J_n \asymp \sqrt{j_n \log(1 + j_n)}$, and $B_n \asymp \sqrt{k_n}$, it follows $\mathcal{R}_n \asymp k_n \sqrt{j_n} \log(1 + k_n)/\mathrm{s}_n \sqrt{n}$ due to $j_n \le k_n$ by Assumption A.2.9(iii). Therefore, since $\kappa_\rho = 1$, Lemma S.4.17 implies Assumption 3.6(i) demands $k_n \sqrt{j_n \log(1 + k_n)} \mathcal{R}_n (1 + \sqrt{\log(1 \vee (\sqrt{j_n}/\mathcal{R}_n))}) = o(a_n)$, which is satisfied with $a_n = 1/\sqrt{\log(n)}$ by Assumption A.2.9(iv). In turn, Assumption 3.6(ii) holds with $a_n = (\log(n))^{-1/2}$ by Assumptions A.2.8(iv) and A.2.10(ii). Finally, we note that Assumption A.2.10 implies Assumption 3.7 due $B_n \lesssim \sqrt{k_n}$, $p = 2$, $J_n \asymp \sqrt{j_n \log(1 + j_n)}$, and $a_n = (\log(n))^{-1/2}$. To conclude, note that since $K_m = 0$, the condition $K_m \mathcal{R}_n^2 \mathcal{S}_n(\mathbf{L}, \mathbf{E}) = o(a_n n^{-1/2})$ is automatically satisfied, the requirement $\mathcal{R}_n = o(\ell_n)$ is equivalent to $k_n \sqrt{j_n} \log(1 + k_n)/\mathrm{s}_n \sqrt{n} = o(\ell_n)$, and the condition $k_n^{1/p} \sqrt{\log(1 + k_n)} B_n \times \sup_{P \in \mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$ is implied by $k_n \sqrt{j_n \log(1 + k_n)} \ell_n \sqrt{\log(\sqrt{j_n}/\ell_n)} = o((\log(n))^{-1/2})$. Thus, all the conditions of Theorem 3.1(ii) hold for both $R = \Theta$ and R corresponding to (A.13) and (A.14), and thus the claim of the theorem follows. ■

Proof of Theorem A.2.4: We first define the variable $\hat{E}_n(R|\ell_n)$ to be given by

$$\hat{E}_n(R|\ell_n) \equiv \inf_{h \in \hat{V}_n(\hat{\theta}_n, R|\ell_n)} \|\hat{\mathbb{W}}_n(\hat{\theta}_n) + \hat{\mathbb{D}}_n[h]\|_{\hat{\Sigma}_n, 2}$$

and note that for any sequence ℓ_n satisfying the conditions of the theorem, Lemma

S.4.21 implies $\hat{U}_n(R|+\infty) = \hat{E}_n(R|\ell_n) + o_P(a_n)$ uniformly in $P \in \mathbf{P}_0$. To establish the theorem, it therefore suffices to show that uniformly in $P \in \mathbf{P}_0$

$$\hat{E}_n(R|\ell_n) \geq U_P^\star(R|\tilde{\ell}_n) + o_P(a_n)$$
$$\hat{E}_n(R|\ell_n) - \hat{U}_n(\Theta|+\infty) \geq U_P^\star(R|\tilde{\ell}_n) - U_P^\star(\Theta|\tilde{\ell}_n^{\mathrm{u}}) + o_P(a_n)$$

with $\ell_n \asymp \tilde{\ell}_n$ and $\tilde{\ell}_n^{\mathrm{u}}$ satisfying the requirements of the theorem. To this end we rely on Theorem 3.2 (for $\hat{E}_n(R|\ell_n)$) and Lemma S.3.7(ii). We note that in the proof of Theorem A.2.3 we established that Assumptions A.2.7, A.2.8, A.2.9, and A.2.10 imply Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7 hold with $\mathcal{R}_n \asymp k_n\sqrt{j_n}\log(1+k_n)/\mathrm{s}_n\sqrt{n}$, $B_n \asymp \sqrt{k_n}$, $\nu_n \asymp 1/\mathrm{s}_n$, $\|\theta\|_{\mathbf{B}} = \|g\|_{1,\infty} \vee \|\gamma\|_2$ and $\|\theta\|_{\mathbf{L}} = \|\theta\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}}\|g\|_{P,2} + \|\gamma\|_2$ for $\theta = (g,\gamma)$, $\kappa_\rho = 1$, and $a_n = (\log(n))^{-1/2}$ for $R = \Theta$ and R as corresponding to (A.13) and (A.14). We thus only verify Assumptions 3.8, 3.9, 3.10, 3.11, 3.12, and 3.13 for $R = \Theta$ and R as corresponding to (A.13) and (A.14).

Next note that Lemma S.4.20 implies Assumptions 3.8, 3.9, and 3.10 are satisfied with $K_g = 2$ and $K_f = 0$, while Lemma S.4.19 and Assumption A.2.11(iv) imply Assumption 3.11 holds with $R = \Theta$ (and hence for R corresponding to (A.13) and (A.14)) with $a_n = (\log(n))^{-1/2}$. Further note that since $\sup_{P\in\mathbf{P}}\|g\|_{P,2}+\|\gamma\|_2 \leq 2(\|g\|_{1,\infty}\vee\|\gamma\|_2)$ for any $(g,\gamma) \in C^1_B(\Omega)\times\mathbf{R}^{d_w}$, it follows that Assumption 3.12(i) holds with $M = 2$. To verify Assumption 3.12(ii) note that by the definitions of $\|\cdot\|_{\mathbf{B}}$ and $\|\cdot\|_{\mathbf{E}}$ in this application and the eigenvalues of $E_P[p^{j_n}(S,Y)p^{j_n}(S,Y)']$ being bounded away from zero uniformly in $P \in \mathbf{P}$ by Assumption A.2.8(iii) we obtain

$$\mathcal{S}_n(\mathbf{B},\mathbf{E}) = \sup_{(\beta,\gamma)} \frac{\|p^{j_n\prime}\beta\|_{1,\infty}\vee\|\gamma\|_2}{\sup_{P\in\mathbf{P}}\|p^{j_n\prime}\beta\|_{P,2}+\|\gamma\|_2}$$
$$\leq 1\vee\sup_\beta\frac{\|p^{j_n\prime}\beta\|_{1,\infty}}{\sup_{P\in\mathbf{P}}\|p^{j_n\prime}\beta\|_{P,2}} \lesssim 1\vee\sup_\beta\frac{\|p^{j_n\prime}\beta\|_{1,\infty}}{\|\beta\|_2} \lesssim j_n^{3/2} \quad \text{(S.170)}$$

where the final equality follows from Assumptions A.2.8(i)(ii). In particular, note that result (S.170), $\mathcal{R}_n \asymp k_n\sqrt{j_n}\log(1+k_n)/\mathrm{s}_n\sqrt{n}$, and Assumption A.2.9(iv) imply that $\mathcal{R}_n\mathcal{S}_n(\mathbf{B},\mathbf{E}) = o(1)$. Thus, since setting $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$ to be the minimizers of Q_n (respectively over $\Theta_n\cap R$ and Θ_n) corresponds to setting $\tau_n = 0$, Assumption A.2.11(ii) and $\mathcal{R}_n\mathcal{S}_n(\mathbf{B},\mathbf{E}) = o(1)$ implies Assumption 3.12(ii) holds. We also note Assumption 3.12(iii) is immediate since $\mathcal{V}_n(P) = \Theta_n\cap R$ by Lemma S.4.16. To conclude, note Assumption 3.13(i) holds since we showed $\mathcal{R}_n\mathcal{S}_n(\mathbf{B},\mathbf{E}) = o(1)$. Moreover, since $B_n \asymp \sqrt{k_n}$ and $K_f = K_m = 0$, Lemma S.4.17 implies Assumption 3.13(ii) holds for any $\ell_n, \ell_n^{\mathrm{u}}$ satisfying $k_n\sqrt{j_n\log(1+k_n)}(\ell_n\vee\ell_n^{\mathrm{u}})(1+\sqrt{\log(\sqrt{j_n}/(\ell_n\vee\ell_n^{\mathrm{u}}))}) = o(a_n)$. Similarly, we obtain that Assumption 3.13(iii) is satisfied provided $\ell_n = o(r_n)$ (imposed in the theorem) and $k_nj_n^2\log(1+k_n)/\mathrm{s}_n\sqrt{n} = o(r_n)$ (implied by Assumption A.2.11(iii)), while the requirement $\mathcal{R}_n^{\mathrm{u}} = o(\ell_n^{\mathrm{u}})$ is implied by $k_nj_n^2\log(1+k_n)/\mathrm{s}_n\sqrt{n} = o(\ell_n^{\mathrm{u}})$. Hence, the conditions of Theorem 3.2 and Lemma S.3.7(ii) hold, and the theorem follows. ■

Lemma S.4.16. *If Assumptions A.2.7(ii), A.2.8(iii), A.2.9(iii), and A.2.10(ii) hold, then Assumption 3.4 is satisfied with both $R = \Theta$ and R corresponding to* (A.13) *and* (A.14)*, $\|\cdot\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}} \|\cdot\|_{P,2} + \|\cdot\|_2$, $\mathcal{V}_n(P) = \Theta_n \cap R$, and $\nu_n^{-1} \asymp \mathrm{s}_n$.*

Proof: By Assumption A.2.7(ii) there is a unique $\theta_0 \equiv (g_0, \gamma_0) \in \Theta \cap R$ for which (A.9) holds, and we let $\Pi_n\theta_0 = (g_n, \gamma_0) \in \Theta_n \cap R$ for $g_n = p^{j_n\prime}\beta_n$. To verify Assumption 3.4(i) is satisfied we set $\|\theta\|_{\mathbf{E}} \equiv \sup_{P\in\mathbf{P}} \|g\|_{P,2} + \|\gamma\|_2$ for any $(g,\gamma) = \theta \in \mathbf{B}$. Since the eigenvalues of $E_P[p^{j_n}(S,Y)p^{j_n}(S,Y)']$ are bounded uniformly in j_n and $P \in \mathbf{P}$ by Assumption A.2.8(iii) we can conclude for any $\theta = (p^{j_n\prime}\beta, \gamma)$ that

$$\begin{aligned} \mathrm{s}_n\|\theta - \Pi_n\theta_0\|_{\mathbf{E}} &\lesssim \mathrm{s}_n\{\|\beta - \beta_n\|_2 + \|\gamma - \gamma_0\|_2\} \\ &\lesssim \|E_P[(p^{j_n}(S,Y)'(\beta - \beta_n) + W'(\gamma - \gamma_0))q^{k_n}(Z)]\|_{\Sigma_P,2} \\ &= \|E_P[(\rho(X,\theta) - \rho(X,\Pi_n\theta_0))q^{k_n}(Z)]\|_{\Sigma_P,2} \end{aligned} \tag{S.171}$$

where the second inequality holds by Assumptions A.2.9(iii) and A.2.10(ii), while the final equality holds by definition of $\rho(X,\theta)$ (see (A.11)). Thus, we conclude from (S.171) that Assumption 3.4(i) holds with $\nu_n^{-1} \asymp \mathrm{s}_n$ and $\mathcal{V}_n(P) = \Theta_n \cap R$. Finally, note Assumption 3.4(ii) is immediate since $\mathcal{V}_n(P) = \Theta_n \cap R$. ■

Lemma S.4.17. *Define the class $\mathcal{F}_n \equiv \{f : f(v) = (q - g(s,y) - w'\gamma)$ for some $(g,\gamma) \in \Theta_n\}$ and suppose that Assumptions A.2.7(iii) and A.2.8(i)(iii) hold. Then, it follows that $\sup_{P\in\mathbf{P}} N_{[\,]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2}) \lesssim 1 \vee (\sqrt{j_n}K/\epsilon)^{j_n+d_w}$ for some $K < \infty$, and in addition $\sup_{P\in\mathbf{P}} J_{[\,]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2}) \lesssim \epsilon\sqrt{j_n}(1 + \sqrt{\log(1 \vee (\sqrt{j_n}/\epsilon))})$.*

Proof: Define the classes $\mathcal{F}_{1n} \equiv \{f : f(v) = q - w'\gamma$ with $\|\gamma\|_2 \le C_0\}$ and $\mathcal{F}_{2n} \equiv \{p^{j_n\prime}\beta : \|p^{j_n\prime}\beta\|_{1,\infty} \le C_0\}$, and then note that by definition of $\mathcal{F}_n$ we have

$$\sup_{P\in\mathbf{P}} N_{[\,]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2}) \le \sup_{P\in\mathbf{P}} N_{[\,]}(\frac{\epsilon}{2}, \mathcal{F}_{1n}, \|\cdot\|_{P,2}) \times \sup_{P\in\mathbf{P}} N_{[\,]}(\frac{\epsilon}{2}, \mathcal{F}_{2n}, \|\cdot\|_{P,2}). \tag{S.172}$$

Next observe that since the support of W is bounded uniformly in $P \in \mathbf{P}$ by Assumption A.2.7(iii), the Cauchy-Schwarz inequality, the covering numbers of $\{\gamma \in \mathbf{R}^{d_w} : \|\gamma\|_2 \le C_0\}$ being bounded (up to a multiplicative constant) by $1 \vee \epsilon^{-d_w}$, and Theorem 2.7.11 in van der Vaart and Wellner (1996) allow us to conclude that

$$\sup_{P\in\mathbf{P}} N_{[\,]}(\frac{\epsilon}{2}, \mathcal{F}_{1n}, \|\cdot\|_{P,2}) \lesssim 1 \vee \epsilon^{-d_w}. \tag{S.173}$$

Similarly, for any $p^{j_n\prime}\beta_1, p^{j_n\prime}\beta_2 \in \mathcal{F}_{2n}$, the Cauchy-Schwarz inequality implies that

$$|p^{j_n}(s,y)'\beta_1 - p^{j_n}(s,y)'\beta_2| \le \sup_{(s,y)} \|p^{j_n}(s,y)\|_2\|\beta_1 - \beta_2\|_2 \lesssim \sqrt{j_n}\|\beta_1 - \beta_2\|_2,$$

where in the final inequality we employed Assumption A.2.8(i). Hence, Theorem 2.7.11

in van der Vaart and Wellner (1996), $\|\beta\|_2 \asymp \|p^{j_n\prime}\beta\|_{P,2}$ uniformly in $P \in \mathbf{P}$ by Assumption A.2.8(iii), and $\|p^{j_n\prime}\beta\|_{P,2} \leq \|p^{j_n\prime}\beta\|_\infty \leq C_0$ for any $p^{j_n\prime}\beta \in \Theta_n$ imply that

$$\sup_{P\in\mathbf{P}} N_{[\,]}(\frac{\epsilon}{2}, \mathcal{F}_{2n}, \|\cdot\|_{P,2}) \lesssim 1 \vee (\frac{K\sqrt{j_n}}{\epsilon})^{j_n} \tag{S.174}$$

for some $K < \infty$. Thus, the first claim of the lemma follows from results (S.172), (S.173), and (S.174). For the second claim of the lemma, we employ the first claim of the lemma and the change of variables $v = u/\epsilon$ to obtain the bound

$$\begin{aligned}\sup_{P\in\mathbf{P}} J_{[\,]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2}) &\lesssim \epsilon + \int_0^\epsilon (\log(1 \vee (\frac{K\sqrt{j_n}}{u})^{j_n+d_w}))^{1/2} du \\ &= \epsilon(1 + \sqrt{j_n + d_w}\int_0^1 (\log(1 \vee (\frac{K\sqrt{j_n}}{v\epsilon})))^{1/2} dv) \lesssim \sqrt{j_n}\epsilon(1 + \sqrt{\log(1 \vee (\sqrt{j_n}/\epsilon))}),\end{aligned}$$

where we used that $(1 \vee ab) \leq (1 \vee a)(1 \vee b)$ whenever a and b are positive. ■

Lemma S.4.18. *Let Assumptions A.2.7(i)(iii), A.2.8(i)(iii), and A.2.9(i) hold. If $a_n \downarrow 0$ and $k_n^3 j_n^2 \log^2(n)/n = o(a_n)$, then Assumption 3.3(i) holds with $R = \Theta$.*

PROOF: We establish the claim of the lemma by relying on Lemma S.4.6. To this end, let $\tilde{j}_n = (1 + d_w) + j_n$, set $r^{j_n}(x) \equiv (q, w', p^{j_n}(x)')'$, and observe any $f \in \mathcal{F}_n$ can be written as $f = r^{j_n\prime}\delta$ for some $\delta \in \mathbf{R}^{\tilde{j}_n}$. Moreover, by Assumption A.2.8(iii) and definition of Θ_n, it follows that there exists an $M < \infty$ such that $\mathcal{F}_n \subseteq \{r^{\tilde{j}_n\prime}\delta : \|\delta\|_2 \leq M\}$, while Assumptions A.2.7(iii), A.2.8(i), and A.2.9(i) imply $\sup_x \|r^{j_n}(x)\|_2 \lesssim \sqrt{j_n}$ and $\sup_z \|q^{k_n}(z)\|_2 \leq \sqrt{k_n}\max_{1\leq k\leq k_n} \|q_k\|_\infty \lesssim k_n$. The claim of the lemma therefore follows from applying Lemma S.4.6 with $b_{1n} \asymp \sqrt{j_n}$, $b_{2n} \asymp k_n$, and $C_n = M$. ■

Lemma S.4.19. *Suppose Assumptions A.2.7(i)(iii), A.2.8(i)(iii), A.2.9(i)(ii) hold and let $\mathcal{C}_n \equiv \{\beta \in \mathbf{R}^{j_n} : \|p^{j_n\prime}\beta\|_{1,\infty} \leq C_0\}$ and $\mathcal{E}_n \equiv \int_0^\infty \sqrt{\log(N(\epsilon, \mathcal{C}_n, \|\cdot\|_2))}d\epsilon$. If $j_n^2 k_n^2 \log(1 + k_n j_n) = o(n)$, then it follows that Assumption 3.11 holds with $R = \Theta$ for any sequence a_n satisfying $k_n^{1/p}(\sqrt{\log(k_n)} + \mathcal{E}_n) j_n^{3/4} k_n^{1/2} \log^{1/4}(1 + j_n k_n)/n^{1/4} = o(a_n)$.*

PROOF: Recall that in this application $X \equiv (Q, S, Y, W')'$ and, when $R = \Theta$, we have $\mathcal{F}_n \equiv \{f : f(x) = (q - g(s, y) - w'\gamma)$ for some $(g, \gamma) \in \Theta_n\}$. Also define $\tilde{\mathcal{F}}_n \equiv \{f q_k : f \in \mathcal{F}_n$ and $1 \leq k \leq k_n\}$, for $\{\omega_i\}_{i=1}^n$ the weights used in building $\hat{\mathbb{W}}_n$ set

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

for any $f \in \tilde{\mathcal{F}}_n$, and let $\mathbb{G}_P^\star$ denote an isonormal process on $\mathcal{F}_n$ independent of $\{V_i\}_{i=1}^n$. Setting $\mathbb{W}_P^\star(\theta) \equiv (\mathbb{G}_P^\star(\rho(\cdot,\theta)q_1), \ldots, \mathbb{G}_P^\star(\rho(\cdot,\theta)q_{k_n}))'$, then note that

$$\sup_{\theta\in\Theta} \|\hat{\mathbb{W}}_n(\theta) - \mathbb{W}_P^\star(\theta)\|_p \leq k_n^{1/p} \sup_{f\in\tilde{\mathcal{F}}_n} |\hat{\mathbb{G}}_n(f) - \mathbb{G}_P^\star(f)|. \tag{S.175}$$

We will establish the lemma by relying on (S.175) and applying Theorem S.7.1 to couple $\hat{\mathbb{W}}_n$ and $\mathbb{G}_P^\star$ on $\tilde{\mathcal{F}}_n$. To this end, define $d_n = k_n(j_n + d_w + 1)$ and let

$$f_n^{d_n}(V) \equiv g^{d_n}(V) - E_P[g^{d_n}(V)] \qquad g^{d_n}(V) \equiv q^{k_n}(Z) \otimes (p^{j_n}(S,Y)', Q, W')'. \quad \text{(S.176)}$$

Next, we set $D_1 \equiv (Q, W', p^{j_n}(S,Y)')'$ and $D_2 = q^{k_n}(Z)$, and for $\overline{\text{eig}}\{D_1 D_1'\}$ the largest eigenvalue of the matrix $D_1 D_1'$, then note that we must have

$$\sup_{P\in\mathbf{P}} \|\overline{\text{eig}}\{D_1 D_1'\}\|_{P,\infty} \leq \sup_{P\in\mathbf{P}} \|\text{trace}\{D_1 D_1'\}\|_{P,\infty} \lesssim j_n, \quad \text{(S.177)}$$

where the final inequality follows from Assumptions A.2.7(iii) and A.2.8(i). Hence, since $\overline{\text{eig}}\{E_P[q^{k_n}(Z)q^{k_n}(Z)']\}$ is bounded uniformly in $P \in \mathbf{P}$ by Assumption A.2.9(ii), result (S.177) and Lemma S.4.22 imply $\overline{\text{eig}}\{E_P[g^{d_n}(V)g^{d_n}(V)'] \lesssim j_n$. It thus follows from $\overline{\text{eig}}\{E_P[g^{d_n}(V)]E_P[g^{d_n}(V)']\} \leq \overline{\text{eig}}\{E_P[g^{d_n}(V)g^{d_n}(V)']\}$ and definition (S.176) that Assumption S.7.1(i) is satisfied with $C_n \asymp j_n$. Similarly, note that Assumptions A.2.7(iii), A.2.8(i), and A.2.9(i) imply Assumption S.7.1(ii) holds with $K_n \asymp \sqrt{j_n k_n}$. Moreover, Assumption S.7.2(i) is immediate with $G_{n,P}$ equal to the zero function and $J_{1n} = 0$. Finally, note that any function $f \in \tilde{\mathcal{F}}_n$ has the structure

$$f(v) = q_k(z)(q - p^{j_n}(s,y)'\beta - w'\gamma) \text{ for some } (p^{j_n\prime}\beta, \gamma) \in \Theta_n. \quad \text{(S.178)}$$

Therefore, for $\mathcal{B}_n$ as defined in Assumption S.7.2(ii), $\mathcal{C}_n \equiv \{\beta \in \mathbf{R}^{j_n} : \|p^{j_n\prime}\beta\|_{1,\infty} \leq C_0\}$, and $\mathcal{G}_n \equiv \{\gamma \in \mathbf{R}^{d_w} : \|\gamma\|_2 \leq C_0\}$, we can conclude that

$$\begin{aligned} N(\epsilon, \mathcal{B}_n, \|\cdot\|_2) &\leq k_n \times N(\frac{\epsilon}{2}, \mathcal{G}_n, \|\cdot\|_2) \times N(\frac{\epsilon}{2}, \mathcal{C}_n, \|\cdot\|_2) \\ &\lesssim k_n \times ((\frac{1}{\epsilon})^{d_w} \vee 1) \times N(\frac{\epsilon}{2}, \mathcal{C}_n, \|\cdot\|_2), \end{aligned} \quad \text{(S.179)}$$

where in the second inequality we employed that $N(\epsilon, \mathcal{G}_n, \|\cdot\|_2) \lesssim (1/\epsilon)^{d_w} \vee 1$. Furthermore, note that Assumption A.2.8(iii) implies that $\|\beta\|_2 \asymp \|p^{j_n\prime}\beta\|_{P,2}$ uniformly in j_n and $P \in \mathbf{P}$, and hence since $\|p^{j_n\prime}\beta\|_{P,2} \leq \|p^{j_n\prime}\beta\|_{1,\infty}$, the definition of Θ_n and (S.178) implies that the radius of $\mathcal{B}_n$ under $\|\cdot\|_2$ is uniformly bounded in n. Thus, the bound in (S.179) yields that for some $M < \infty$ we must have

$$\begin{aligned} &\int_0^\infty \sqrt{\log(N(\epsilon, \mathcal{B}_n, \|\cdot\|_2))}d\epsilon \\ &\quad \lesssim \int_0^M \sqrt{\log(k_n)}d\epsilon + \int_0^1 \sqrt{\log(1/\epsilon)}d\epsilon + \int_0^M \sqrt{\log(N(\epsilon/2, \mathcal{C}_n, \|\cdot\|_2))}d\epsilon \\ &\quad \lesssim \sqrt{\log(k_n)} + \int_0^\infty \sqrt{\log(N(u, \mathcal{C}_n, \|\cdot\|_2))}du, \end{aligned}$$

where the final inequality follows from $N(\epsilon, \mathcal{C}_n, \|\cdot\|_2)$ being (weakly) larger than one for all ϵ and the change of variables $u = \epsilon/2$. Hence, Assumption S.7.2(ii) holds with

$J_{2n} = \sqrt{\log(k_n)} + \mathcal{E}_n$, and as a result Theorem S.7.1 implies uniformly in $P \in \mathbf{P}$

$$\sup_{f\in\tilde{\mathcal{F}}_n} |\hat{\mathbb{G}}_n(f) - \mathbb{G}^\star_P(f)| = O_P((\sqrt{\log(k_n)} + \mathcal{E}_n)\{\frac{j_n^3 k_n^2 \log(1 + j_n k_n)}{n}\}^{1/4}). \tag{S.180}$$

The claim of the lemma therefore follows from (S.175) and (S.180). ∎

Lemma S.4.20. *If $\mathbf{B} = C^1_B(\Omega) \times \mathbf{R}^{d_w}$ and Υ_G, Υ_F, and Θ are as defined in (A.13), (A.14), and (A.15), then it follows that Assumptions 3.8, 3.9, and 3.10 are satisfied with $K_g = 2$, $K_f = 0$, and for any $\theta = (g, \gamma)$ and $h = (g_h, \gamma_h)$, $\nabla\Upsilon_G(\theta)[h]$ equals*

$$\nabla\Upsilon_G(\theta)[h](s, y) = \frac{\partial}{\partial s} g_h(s, y) + g(s, y)\frac{\partial}{\partial y} g_h(s, y) + g_h(s, y)\frac{\partial}{\partial y} g(s, y).$$

PROOF: Recall that in this application $\mathbf{G} = C^0_B(\Omega)$ and $\|\theta\|_{\mathbf{B}} = \max\{\|g\|_{1,\infty}, \|\gamma\|_2\}$. Hence, for any $\theta_1 = (g_1, \gamma_1) \in \mathbf{B}$ and $\theta_2 = (g_2, \gamma_2) \in \mathbf{B}$ we obtain that

$$\begin{aligned}&\|\Upsilon_G(\theta_1) - \Upsilon_G(\theta_2) - \nabla\Upsilon_G(\theta_1)[\theta_1 - \theta_2]\|_{\mathbf{G}}\\ &\qquad \le \sup_{(s,y)\in\Omega} |g_1(s, y) - g_2(s, y)| \times \sup_{(s,y)\in\Omega} |\frac{\partial}{\partial y}(g_1(s, y) - g_2(s, y))| \le \|g_1 - g_2\|^2_{1,\infty},\end{aligned}$$

which verifies Assumption 3.8(i) holds with $K_g = 2$. Similarly, we additionally conclude

$$\begin{aligned}&\|\nabla\Upsilon_G(\theta_1) - \nabla\Upsilon_G(\theta_2)\|_o\\ &\quad = \sup_{g_h:\|g_h\|_{1,\infty}\le 1} \sup_{(s,y)\in\Omega} |(g_1(s, y) - g_2(s, y))\frac{\partial}{\partial y} g_h(s, y) + g_h(s, y)\frac{\partial}{\partial y}(g_1(s, y) - g_2(s, y))|\\ &\quad \le 2\|g_1 - g_2\|_{1,\infty},\end{aligned} \tag{S.181}$$

which verifies Assumption 3.8(ii) holds with $K_g = 2$ as well. Moreover, note that since any $\theta = (g, \gamma) \in \Theta$ satisfies $\|g\|_{1,\infty} \le C_0$, it follows that $\|\tilde{g}\|_{1,\infty} \le C_0 + \epsilon$ for any $\tilde{g} \in \Theta^\epsilon$. Thus, by identical arguments to those in (S.181) we obtain

$$\|\nabla\Upsilon_G(\theta)\|_o \le 2\|g\|_{1,\infty} \le 2(C_0 + \epsilon),$$

which thus verifies Assumption 3.8(iii) holds with $M = 2(C_0 + \epsilon)$.

Next note $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ is affine and continuous, and hence $\nabla\Upsilon_F(\theta)[h] = \Upsilon_F(h) - c_0$ for all $\theta, h \in \mathbf{B}$. Therefore, Assumptions 3.9(i)(ii) hold with $K_f = 0$, while

$$\sup_{g_h:\|g_h\|_{1,\infty}\le 1} |g_h(s_0, y_0)| \le 1$$

implies Assumption 3.9(iii) is satisfied with $M = 1$. Since Υ_F being affine and $K_f = 0$ further imply that Assumptions 3.9(iv) and 3.10 hold, the lemma follows. ∎

Lemma S.4.21. *Let* $a_n = (\log(n))^{-1/2}$ *and Assumptions A.2.7, A.2.8, A.2.9, A.2.10 hold. If* ℓ_n *satisfies* $k_n j_n^2 \log(1+k_n)/\mathrm{s}_n\sqrt{n} = o(\ell_n)$*, then uniformly in* $P \in \mathbf{P}_0$*:*

$$\hat{U}_n(R|+\infty) = \inf_{h\in \hat{V}_n(\hat{\theta}_n,R|\ell_n)} \|\hat{\mathbb{W}}_n(\hat{\theta}_n) + \hat{\mathbb{D}}_n[h]\|_{\hat{\Sigma}_n,2} + o_P(a_n).$$

Proof: We establish the lemma by applying Lemma S.3.1. To this end, recall that in the proof of Theorem A.2.3, Assumptions 3.2(i)(iii) and 3.7 were verified to hold with $B_n \asymp \sqrt{k_n}$ and $J_n \asymp \sqrt{j_n \log(1+j_n)}$. Since the eigenvalues of $E_P[p^{j_n}(S,Y)p^{j_n}(S,Y)']$ are bounded uniformly in $P \in \mathbf{P}$ by Assumption A.2.8(iii) and $\|\theta\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}} \|g\|_{P,2} + \|\gamma\|_2$ for any $\theta = (g,\gamma)$, it also follows that for any $h = (p^{j_n\prime}\beta_h, \gamma_h)$ we have

$$\begin{aligned}\|h\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}} \|p^{j_n\prime}\beta_h\|_{P,2} + \|\gamma\|_2 &\lesssim \|\beta_h\|_2 + \|\gamma_h\|_2 \\ &\lesssim \frac{1}{\mathrm{s}_n}\|E_P[q^{k_n}(Z)(p^{j_n}(S,Y)'\beta_h + W'\gamma_h)]\|_2 = \frac{1}{\mathrm{s}_n}\|\mathbb{D}_P[h]\|_2,\end{aligned}$$

where the second inequality holds by Assumption A.2.9(iii) and the final equality follows from the definition of $\mathbb{D}_P[h]$. Hence, since $\nu_n \asymp 1/\mathrm{s}_n$ by Lemma S.4.16 and $p=2$, we conclude the Lemma S.3.1 requirement that $\|h\|_{\mathbf{E}} \leq \nu_n \|\mathbb{D}_P(\theta)[h]\|_p$ for all $\theta \in \mathcal{A}_n(P)$ holds with $\mathcal{A}_n(P) = \Theta_n \cap R$. Next, define the $k_n \times (j_n + d_w)$ matrix

$$\mathbb{M}_{i,n} \equiv \frac{1}{n}\{q^{k_n}(Z_i)(p^{j_n}(S_i,Y_i)'\ W_i') - E_P[q^{k_n}(Z)(p^{j_n}(S,Y)'\ W')]\}, \tag{S.182}$$

which satisfies $E_P[\mathbb{M}_{i,n}] = 0$. Since $\|(p^{j_n\prime}\beta,\gamma)\|_{\mathbf{E}} \asymp \|\beta\|_2 + \|\gamma\|_2$ by Assumption A.2.8(iii), we then conclude from (S.182) that for some $C_1 < \infty$ we must have

$$\begin{aligned}\sup_{P\in\mathbf{P}} P(\sup_{h\in\mathbf{B}_n} \frac{\|\hat{\mathbb{D}}_n[h] - \mathbb{D}_P[h]\|_2}{\|h\|_{\mathbf{E}}} > \mathrm{s}_n) &\leq \sup_{P\in\mathbf{P}} P(\|\frac{1}{n}\sum_{i=1}^n \mathbb{M}_{i,n}\|_{o,2} > C_1\mathrm{s}_n) \\ &\leq (j_n + d_w + k_n)\exp\{-\frac{n\mathrm{s}_n^2 C_2}{(k_n^2 \vee j_n) + \mathrm{s}_n k_n\sqrt{j_n}}\} = o(1), \end{aligned} \tag{S.183}$$

where the final inequality follows by applying Lemma S.4.7 with $b_{1n} = \sqrt{j_n}$ (by Assumptions A.2.7(iii) and A.2.8(i)) and $b_{2n} = k_n$ (by Assumption A.2.9(i)), while the final equality results from $\log(k_n)k_n^2/\mathrm{s}_n^2 n = o(1)$ by Assumption A.2.9(iv) and $k_n \geq j_n$ by Assumption A.2.8(iii). Hence, $\nu_n \asymp 1/\mathrm{s}_n$ and (S.183) imply condition (S.79) in Lemma S.3.1 holds. Finally, we note that by Assumption A.2.11(iv), we may apply Lemma S.4.19 with $p=2$ to conclude that Assumption 3.11 holds with $R = \Theta$ (and hence for R as corresponding to (A.13) and (A.14)) with $a_n = (\log(n))^{-1/2}$. This concludes verifying the requirements of Lemma S.3.1 and therefore the present Lemma follows for any ℓ_n satisfying $\mathcal{S}_n(\mathbf{B},\mathbf{E})\mathcal{R}_n = o(\ell_n)$, which in this application is equivalent to $k_n j_n^2 \log(1+k_n)/\mathrm{s}_n\sqrt{n} = o(\ell_n)$ due to $\mathcal{S}_n(\mathbf{B},\mathbf{E}) \lesssim j_n^{3/2}$ and $\mathcal{R}_n \asymp k_n\sqrt{j_n}\log(1+k_n)/\mathrm{s}_n\sqrt{n}$. ■

Lemma S.4.22. *Let $D_1 \in \mathbf{R}^{d_1}$, $D_2 \in \mathbf{R}^{d_2}$ be distributed according to Q, and for any matrix A let $\overline{\text{eig}}\{A\}$ denote its largest eigenvalue. Then it follows that*

$$\overline{\text{eig}}\{E_Q[(D_1 \otimes D_2)(D_1 \otimes D_2)']\} \leq \|\overline{\text{eig}}\{D_1 D_1'\}\|_{Q,\infty} \times \overline{\text{eig}}\{E_Q[D_2 D_2']\}.$$

Proof: Let $\mathcal{A} \equiv \{\{a_i\}_{i=1}^{d_1} : a_i \in \mathbf{R}^{d_2}$ and $\sum_{i=1}^{d_1} \|a_i\|_2^2 \leq 1\}$, set $(D_1^{(1)}, \ldots, D_1^{(d_1)}) = D_1 \in \mathbf{R}^{d_1}$, and then note that by direct calculation we obtain that

$$\begin{aligned}
&\overline{\text{eig}}\{E_Q[(D_1 \otimes D_2)(D_1 \otimes D_2)']\} \\
&\quad = \sup_{\{a_i\}_{i=1}^{d_1} \in \mathcal{A}} (a_1', \ldots, a_{d_1}') E_Q[(D_1 \otimes D_2)(D_1 \otimes D_2)'](a_1', \ldots, a_{d_1}')' \\
&\quad = \sup_{\{a_i\}_{i=1}^{d_1} \in \mathcal{A}} E_Q[(\sum_{i=1}^{d_1} (a_i' D_2) D_1^{(i)})^2] \\
&\quad \leq \|\overline{\text{eig}}\{D_1 D_1'\}\|_{Q,\infty} \sup_{\{a_i\}_{i=1}^{d_1} \in \mathcal{A}} \sum_{i=1}^{d_1} E_Q[(a_i' D_2)^2].
\end{aligned}$$

However, since $\sum_{i=1}^{d_1} \|a_i\|_2^2 \leq 1$ for all $\{a_i\}_{i=1}^{d_1} \in \mathcal{A}$, we additionally have the inequality

$$\sup_{\{a_i\}_{i=1}^{d_1} \in \mathcal{A}} \sum_{i=1}^{d_1} E_Q[(a_i' D_2)^2] \leq \sup_{\{a_i\}_{i=1}^{d_1} \in \mathcal{A}} \sum_{i=1}^{d_1} \overline{\text{eig}}\{E_Q[D_2 D_2']\} \|a_i\|_2^2 = \overline{\text{eig}}\{E_Q[D_2 D_2']\},$$

and therefore the claim of the lemma follows. ■

Lemma S.4.23. *Let λ be the Lebesgue measure, $\{B_b^{(1)}\}_{b=1}^{j_{1n}}$ and $\{B_b^{(2)}\}_{b=1}^{j_{2n}}$ be B-splines on $[0,1]$ of order $r \geq 3$ with no interior knot multiplicity, mesh ratio bounded in n, and $\|\cdot\|_{\lambda,2}$ normalized to have norm one. If $\{p_j\}_{j=1}^{j_n}$ is the tensor product of $\{B_b^{(1)}\}_{b=1}^{j_{1n}}$ and $\{B_b^{(2)}\}_{b=1}^{j_{2n}}$ and $\mathcal{C}_n \equiv \{\beta \in \mathbf{R}^{j_n} : \|p^{j_n\prime}\beta\|_{1,\infty} \leq C_0\}$, then it follows that*

$$\int_0^\infty \sqrt{\log(N(\epsilon, \mathcal{C}_n, \|\cdot\|_2))} d\epsilon \lesssim \sqrt{j_{1n} \wedge j_{2n}} \log(j_n + 1).$$

Proof: We rely heavily on Chapter 5 in DeVore and Lorentz (1993), and note that B_j corresponds to $N_j/\|N_j\|_{\lambda,2}$ in their notation. Throughout, for two sequences a_n and b_n we employ $a_n \asymp b_n$ to mean that there exist constants $\underline{c}$ and $\bar{c}$ such that $\underline{c} a_n \leq b_n \leq \bar{c} a_n$ for all n. In what follows it will also prove convenient to index the elements of $\beta \in \mathbf{R}^{j_n}$ by β_{b_1,b_2} with $1 \leq b_1 \leq j_{1n}$ and $1 \leq b_2 \leq j_{2n}$. Then note that the mesh ratios corresponding to $\{B_b^{(1)}\}_{b=1}^{j_{1n}}$ and $\{B_b^{(2)}\}_{b=1}^{j_{2n}}$ being bounded uniformly in j_{1n}, j_{2n} and two applications of Theorem 5.4.2 in DeVore and Lorentz (1993) imply that

$$\|p^{j_n\prime}\beta\|_\infty = \sup_{u_1\in[0,1]}\sup_{u_2\in[0,1]}|\sum_{b_2=1}^{j_{2n}} B^{(2)}_{b_2}(u_2)\sum_{b_1=1}^{j_{1n}}\beta_{b_1,b_2}B^{(1)}_{b_1}(u_1)|$$
$$\asymp \sup_{u_1\in[0,1]}\max_{1\le b_2\le j_{2n}}\sqrt{j_{2n}}|\sum_{b_1=1}^{j_{1n}}\beta_{b_1,b_2}B^{(1)}_{b_1}(u_1)| \asymp \sqrt{j_{1n}j_{2n}}\|\beta\|_\infty \quad (S.184)$$

uniformly in $\beta\in\mathbf{R}^{j_n}$. By similar arguments we also obtain uniformly in $\beta\in\mathbf{R}^{j_n}$ that

$$\sup_{u_1\in(0,1)}\sup_{u_2\in[0,1]}|\sum_{b_2=1}^{j_{2n}} B^{(2)}_{b_2}(u_2)\sum_{b_1=1}^{j_{1n}}\frac{\partial}{\partial u_1}\{\beta_{b_1,b_2}B^{(1)}_{b_1}(u_1)\}|$$
$$\asymp \sup_{u_1\in(0,1)}\max_{1\le b_2\le j_{2n}}\sqrt{j_{2n}}|\sum_{b_1=1}^{j_{1n}}\frac{\partial}{\partial u_1}\{\beta_{b_1,b_2}B^{(1)}_{b_1}(u_1)\}|$$
$$\asymp \max_{1\le b_2\le j_{2n}}\max_{2\le b_1\le j_{1n}}\sqrt{j_{2n}}j_{1n}^{3/2}|\beta_{b_1,b_2}-\beta_{b_1-1,b_2}|, \quad (S.185)$$

where the second result follows by employing equation (3.11) and Theorem 5.4.2 in Chapter 5 of DeVore and Lorentz (1993) and the mesh ratio of $\{B^{(1)}_b\}_{b=1}^{j_{1n}}$ being bounded. Since by identical arguments we can also derive the symmetric (to (S.185)) relationship

$$\sup_{u_1\in[0,1]}\sup_{u_2\in(0,1)}|\sum_{b_1=1}^{j_{1n}} B^{(1)}_{b_1}(u_1)\sum_{b_2=1}^{j_{2n}}\frac{\partial}{\partial u_2}\{\beta_{b_1,b_2}B^{(2)}_{b_2}(u_2)\}|$$
$$\asymp \max_{1\le b_1\le j_{1n}}\max_{2\le b_2\le j_{2n}}\sqrt{j_{1n}}j_{2n}^{3/2}|\beta_{b_1,b_2}-\beta_{b_1,b_2-1}|, \quad (S.186)$$

it follows from results (S.184), (S.185), and (S.186) that there is an $M_0<\infty$ such that

$$\max_{1\le b_1\le j_{1n}}\max_{1\le b_2\le j_{2n}}|\beta_{b_1,b_2}|\le M_0/\sqrt{j_n}$$
$$\max_{1\le b_2\le j_{2n}}\max_{2\le b_1\le j_{1n}}|\beta_{b_1,b_2}-\beta_{b_1-1,b_2}|\le M_0/(j_{1n}\sqrt{j_n})$$
$$\max_{1\le b_1\le j_{1n}}\max_{2\le b_2\le j_{2n}}|\beta_{b_1,b_2}-\beta_{b_1,b_2-1}|\le M_0/(j_{2n}\sqrt{j_n}) \quad (S.187)$$

for all $\beta\in\mathcal{C}_n$. Hence, in order to establish the claim of the lemma, it suffices to bound the covering numbers for the set defined by (S.187).

We proceed by combining two bounds, one for "small" ϵ and one for "large" ϵ. First, assume without loss of generality $j_{1n}\ge j_{2n}$, let $c_n\equiv\lceil\log(j_{1n}+1)\rceil$, and define the sets

$$\frac{\epsilon}{3\sqrt{j_n}}k_1\le\beta_{b_1,b_2}\le\frac{\epsilon}{3\sqrt{j_n}}(k_1+1) \text{ for all } b_1=mc_n+1 \text{ with } 0\le m\le\lceil j_{1n}/c_n\rceil-1$$
$$\frac{\epsilon}{3c_n\sqrt{j_n}}k_2\le\beta_{b_1,b_2}-\beta_{b_1-1,b_2}\le\frac{\epsilon}{3c_n\sqrt{j_n}}(k_2+1) \text{ otherwise} \quad (S.188)$$

where k_1, k_2 are non-zero integers – i.e. the sets (in $\mathbf{R}^{j_n}$) defined in (S.188) consist of "chains" along the b_1 dimension that reset every c_n integers. To compute the diameter of the sets in (S.188), then note that since all "chains" have the same structure

$$\begin{aligned}
&\sup \|\beta - \tilde{\beta}\|_2^2 \text{ s.t. } \beta, \tilde{\beta} \text{ satisfying (S.188)} \\
&\qquad \leq \sup j_{2n} \lceil \frac{j_{1n}}{c_n} \rceil \sum_{b_1=1}^{c_n} (\beta_{b_1, j_{2n}} - \tilde{\beta}_{b_1, j_{2n}})^2 \text{ s.t. } \beta, \tilde{\beta} \text{ satisfying (S.188)} \\
&\qquad \leq j_{2n} \lceil \frac{j_{1n}}{c_n} \rceil \sum_{b_1=1}^{c_n} \frac{\epsilon^2}{9 j_n} \{1 + \frac{2(b_1 - 1)}{c_n}\}^2, \qquad \text{(S.189)}
\end{aligned}$$

where the final inequality follows from (S.188). Since $\lceil j_{1n}/c_n \rceil c_n \leq (j_{1n} + c_n) \leq 2 j_{1n}$ due to $\lceil j_{1n}/c_n \rceil \leq 1 + j_{1n}/c_n$ and $c_n \leq j_{1n}$, it follows from $j_n = j_{1n} j_{2n}$ that every set in (S.188) is contained in a ball of radius ϵ. Moreover, by (S.187) the total number of sets with the structure in (S.188) needed to cover the set $\mathcal{C}_n$ is bounded by

$$(\lceil \frac{6M_0}{\epsilon} \rceil)^{j_{2n} \lceil \frac{j_{1n}}{c_n} \rceil} (\lceil \frac{6 M_0 c_n}{\epsilon j_{1n}} \rceil)^{j_{2n} \lceil \frac{j_{1n}}{c_n} \rceil c_n}. \qquad \text{(S.190)}$$

Next, we employ again the bound $\lceil j_{1n}/c_n \rceil c_n \leq 2 j_{1n}$ and $\lceil a \rceil \leq 2a$ whenever $a \geq 1$, to obtain from (S.189) and (S.190) that whenever $\epsilon \leq 6 M_0 c_n / j_{1n}$ we have

$$\begin{aligned}
N(\epsilon, \mathcal{C}_n, \|\cdot\|_2) &\leq (\frac{12 M_0}{\epsilon})^{\frac{2 j_n}{c_n}} (\frac{12 M_0 c_n}{\epsilon j_{1n}})^{2 j_n} \\
&= (\frac{12 M_0 c_n}{\epsilon j_{1n}} (\frac{j_{1n}}{c_n})^{\frac{1}{c_n + 1}})^{\frac{2 j_n (c_n+1)}{c_n}} \leq (\frac{M_1 \log(1 + j_{1n})}{\epsilon j_{1n}})^{4 j_n}, \qquad \text{(S.191)}
\end{aligned}$$

where the final equality holds for some $M_1 < \infty$ due to $(c_n+1)/c_n \leq 2$ and $(j_{1n}/c_n)^{\frac{1}{c_n+1}} \leq j_{1n}^{\frac{1}{\log(1+j_{1n})}} = O(1)$ because $c_n = \lceil \log(1 + j_{1n}) \rceil$.

The bound in (S.191) is valid only for $\epsilon \leq 6 M_0 c_n / j_{1n}$. To obtain a bound for $\epsilon \geq 6 M_0 c_n / j_{1n}$, let $\{\mathbb{Z}_{b_1, b_2}\}_{b_1, b_2}$ be independent standard normal random variables. By Sudakov's inquality (see, e.g., Proposition A.2.5 in van der Vaart and Wellner (1996)), it then follows that for some $M_2 < \infty$ independent of n we have that

$$\sqrt{\log(N(\epsilon, \mathcal{C}_n, \|\cdot\|_2))} \leq \frac{M_2}{\epsilon} E[\sup_{\beta \in \mathcal{C}_n} \sum_{b_1=1}^{j_{1n}} \sum_{b_2=1}^{j_{2n}} \beta_{b_1, b_2} \mathbb{Z}_{b_1, b_2}]. \qquad \text{(S.192)}$$

Next, for notational convenience define $\Delta_{b_1} \beta_{b_1, b_2} = (\beta_{b_1, b_2} - \beta_{b_1 - 1, b_2})$ and $\Delta_{b_2} \beta_{b_1, b_2} =$

$(\beta_{b_1,b_2} - \beta_{b_1,b_2-1})$, and then note that by (S.187) it follows that

$$\sup_{\beta \in \mathcal{C}_n} \sum_{b_1=1}^{j_{1n}} \sum_{b_2=1}^{j_{2n}} \beta_{b_1,b_2} \mathbb{Z}_{b_1,b_2}$$
$$= \sup_{\beta \in \mathcal{C}_n} \sum_{b_2=1}^{j_{2n}} \sum_{b_1=2}^{j_{1n}} \Delta_{b_1} \beta_{b_1,b_2} \sum_{\tilde{b}_1=b_1}^{j_{1n}} \mathbb{Z}_{\tilde{b}_1,b_2} + \sum_{b_2=2}^{j_{2n}} \Delta_{b_2} \beta_{1,b_2} \sum_{b_1=1}^{j_{1n}} \sum_{\tilde{b}_2=b_2}^{j_{2n}} \mathbb{Z}_{b_1,\tilde{b}_2} + \beta_{1,1} \sum_{b_1=1}^{j_{1n}} \sum_{b_2=1}^{j_{2n}} \mathbb{Z}_{b_1,b_2}$$
$$\leq \sum_{b_2=1}^{j_{2n}} \sum_{b_1=2}^{j_{1n}} \frac{M_0}{j_{1n}\sqrt{j_n}} |\sum_{\tilde{b}_1=b_1}^{j_{1n}} \mathbb{Z}_{\tilde{b}_1,b_2}| + \sum_{b_2=2}^{j_{2n}} \frac{M_0}{j_{2n}\sqrt{j_n}} |\sum_{b_1=1}^{j_{1n}} \sum_{\tilde{b}_2=b_2}^{j_{2n}} \mathbb{Z}_{b_1,\tilde{b}_2}| + \frac{M_0}{\sqrt{j_n}} |\sum_{b_1=1}^{j_{1n}} \sum_{b_2=1}^{j_{2n}} \mathbb{Z}_{b_1,b_2}|.$$

Hence, employing that if $\mathbb{W} \sim N(0, \sigma^2)$ then $E[|\mathbb{W}|] \lesssim \sigma$, we can conclude that

$$E[\sup_{\beta \in \mathcal{C}_n} \sum_{b_1=1}^{j_{1n}} \sum_{b_2=1}^{j_{2n}} \beta_{b_1,b_2} \mathbb{Z}_{b_1,b_2}] \lesssim \sum_{b_2=1}^{j_{2n}} \sum_{b_1=2}^{j_{1n}} \frac{\sqrt{j_{1n} - b_1}}{j_{1n}\sqrt{j_n}} + \sum_{b_2=2}^{j_{2n}} \frac{\sqrt{j_{1n}(j_{2n} - b_2)}}{j_{2n}\sqrt{j_n}} + 1$$
$$\leq \frac{j_n\sqrt{j_{1n}}}{j_{1n}\sqrt{j_n}} + \frac{j_{2n}\sqrt{j_{1n}j_{2n}}}{j_{2n}\sqrt{j_n}} + 1 \leq 3\sqrt{j_{2n}}$$

where in the final inequality we employed that $j_n = j_{1n}j_{2n}$. Hence, by (S.192) we have

$$\sqrt{\log(N(\epsilon, \mathcal{C}_n, \|\cdot\|_2))} \lesssim \frac{\sqrt{j_{2n}}}{\epsilon}. \tag{S.193}$$

To conclude the proof, we combine the bounds in (S.191) and (S.193). In particular, setting $\delta_n \equiv 6M_0\lceil\log(j_{1n}+1)\rceil/j_{1n}$ and observing that $\|\beta\|_2 \asymp \|p^{j_n\prime}\beta\|_{\lambda,2} \leq C_0$ for all $\beta \in \mathcal{C}_n$ allows us to conclude that for some $M_2 < \infty$ we must have

$$\int_0^\infty \sqrt{\log(N(\epsilon, \mathcal{C}_n, \|\cdot\|_2))} d\epsilon \lesssim \int_{\delta_n}^{M_2} \frac{\sqrt{j_{2n}}}{\epsilon} + \sqrt{j_n} \int_0^{\delta_n} (\log(\frac{M_1 \log(j_{1n})}{\epsilon j_{1n}}))^{1/2} d\epsilon$$
$$\lesssim \sqrt{j_{2n}} \log(1 + j_{1n}) + \frac{\sqrt{j_n}\log(1+j_{1n})}{j_{1n}} \int_0^1 (\log(\frac{1}{u}))^{1/2} du \lesssim \sqrt{j_{2n}} \log(1 + j_{1n})$$

where the second inequality follows from the change of variables $u = \epsilon/\delta_n$ and the final inequality employed that $j_n = j_{1n}j_{2n}$ and $j_{2n} \leq j_{1n}$. Substituting $j_{2n} = j_{1n} \wedge j_{2n}$ and employing $j_{1n} \leq j_n$ establishes the lemma. ■

S.4.4 Proofs for Section A.2.3

Proof of Theorem A.2.5: We establish the result by applying Theorem 3.1 to both $R = \Theta$ and R corresponding to (A.26) and (A.27). To this end, note that Assumption 3.1(i) is directly imposed in Assumption A.2.12(i), Assumption 3.2(i) holds with $B_n = O(1)$ by Assumption A.2.15(i), Assumption 3.2(ii) is directly imposed by Assumption A.2.15(iii), and Assumption 3.2(iii) holds with $F_n = 1$ and $J_n = O(1)$ by

Lemma S.4.27. Next, we apply Lemma S.4.28 with $\pi_{0n} = O(1)$ and $\pi_{1n} = O(k_n)$ (which is possible by Assumptions A.2.15(i)(ii)) to obtain that Assumption 3.3(i) holds for both $R = \Theta$ and R corresponding to (A.26) and (A.27) for any a_n satisfying $k_n^{1/p}\sqrt{j_n}\log(n)(n^{1/6} \vee k_n)/n^{1/3} = o(a_n)$ and in particular it holds for $a_n \asymp (\log(n))^{-1/2}$ by Assumption A.2.15(v). We also note Assumptions 3.3(ii), 3.4, and 3.5 hold with $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_\infty$, $\|\cdot\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}}\|\cdot\|_{P,2}$, and $\kappa_\rho = 1/2$ by Lemmas S.4.25 and S.4.29. To verify Assumption 3.6, note $J_n = O(1)$, $B_n = O(1)$, and $\nu_n \asymp \sqrt{k_n}/\mathrm{s}_n k_n^{1/p}$ by Lemma S.4.25 imply in this application we have $\mathcal{R}_n \asymp \sqrt{k_n\log(1+k_n)}/\mathrm{s}_n\sqrt{n}$. Thus, Lemma S.4.27 and Assumption A.2.15(v) verify Assumption 3.6(i), while Assumption A.2.13(iii) implies Assumption 3.6(ii), and Assumption A.2.16 implies Assumption 3.7. Finally, since $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_\infty$ by Lemma S.4.29, Assumptions A.2.13(i)(ii) yield

$$\sup_\beta \frac{\|p^{j_n\prime}\beta\|_\infty}{\sup_{P\in\mathbf{P}}\|p^{j_n\prime}\beta\|_{P,2}} \lesssim \frac{\sqrt{j_n}\|\beta\|_2}{\|\beta\|_2} = \sqrt{j_n}. \tag{S.194}$$

Therefore, the condition $K_m\ell_n^2 \times \mathcal{S}_n(\mathbf{L},\mathbf{E}) = o(a_n n^{-1/2})$ is equivalent to $\ell_n^2\sqrt{nj_n\log(n)} = o(1)$. Moreover, by Lemma S.4.27, Assumption A.2.13(ii), $\kappa_\rho = 1/2$, and $B_n = O(1)$ the condition $k_n^{1/p}\sqrt{\log(1+k_n)}B_n \times \sup_{P\in\mathbf{P}} J_{[\,]}(\ell_n^{\kappa_\rho}, \mathcal{F}_n, \|\cdot\|_{P,2}) = o(a_n)$ is implied by the restriction $k_n^{1/p}\sqrt{j_n\ell_n\log(1+k_n)\log(1/\ell_n)} = o((\log(n))^{-1/2})$. Thus, the first claim of the theorem follows from Theorem 3.1(i) applied to $I_n(R)$.

The second claim of the Theorem follows from applying Theorem 3.1(ii) to $I_n(\Theta)$. To this end, note that the only remaining condition to verify is that $\mathcal{R}_n^2 \times \mathcal{S}_n(\mathbf{L},\mathbf{E}) = o(a_n n^{-1/2})$. Using that, as already argued, $\mathcal{R}_n \asymp \sqrt{k_n\log(1+k_n)}/\mathrm{s}_n\sqrt{n}$ and result (S.194) we note that a sufficient condition for this final requirement is that $k_n\log(1+k_n)\sqrt{j_n\log(n)}/\mathrm{s}_n^2\sqrt{n} = o(1)$ and therefore the theorem follows. ■

PROOF OF THEOREM A.2.6: We proceed by relying on Theorem 3.2 and Lemma S.3.7(ii). To this end, we note that, for both $R = \Theta$ and R corresponding to (A.26) and (A.27), the proof of Theorem A.2.5 established that Assumptions 3.1(i), 3.2, 3.4, 3.3, 3.5, 3.6, and 3.7 are satisfied with $B_n = O(1)$, $J_n = O(1)$, $\|\cdot\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}}\|\cdot\|_{P,2}$, $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_\infty$, $\nu_n \asymp \sqrt{k_n}/\mathrm{s}_n k_n^{1/p}$, $\mathcal{R}_n \asymp \sqrt{k_n\log(1+k_n)}/\mathrm{s}_n\sqrt{n}$, $\kappa_\rho = 1/2$, and $\mathcal{S}_n(\mathbf{L},\mathbf{E}) \lesssim \sqrt{j_n}$.

Next, note that Assumptions 3.8, 3.9, and 3.10 are satisfied by Lemma S.4.30 with $K_g = 0$ and $K_f > 0$. To verify Assumption 3.11 we apply Lemma S.4.31 with $\pi_{0n} = O(1)$ and $\pi_{1n} \lesssim k_n$, which is possible by Assumptions A.2.15(i)(ii). In particular, Lemma S.4.31 evaluated at $d_n \asymp (nk_n)^{3/13}$ and Assumption A.2.17(iii) yield

$$\sup_{\theta\in\Theta_n}\|\hat{\mathbb{W}}_n(\theta) - \mathbb{W}_P^\star(\theta)\|_p = o_P((\log(n))^{-1/2}) \tag{S.195}$$

uniformly in $P \in \mathbf{P}$, which implies Assumption 3.11 is satisfied for both $R = \Theta$ and R corresponding to (A.26) and (A.27). Next note that Assumption 3.12(i) is immediate since $\|\cdot\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}_0}\|\cdot\|_{P,2}$ and $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{2,\infty}$, while Assumption 3.12(iii) follows from

Lemma S.4.24 and $\mathcal{V}_n(P) \equiv \{\theta \in \Theta_n : \|\theta - \Pi_n\theta_0\|_{1,\infty} \leq \epsilon\}$ for some $\epsilon > 0$ by Lemma S.4.25. In order to verify Assumption 3.12(ii) and the rate conditions of Assumption 3.13, note that the eigenvalues of $E_P[p^{j_n}(D)p^{j_n}(D)']$ being bounded away from zero uniformly in $P \in \mathbf{P}$ by Assumption A.2.13(ii), Assumptions A.2.13(i) and A.2.17(iv) together with $\|\cdot\|_{\mathbf{B}} = \|\cdot\|_{2,\infty}$ and $\|\cdot\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}}\|\cdot\|_{P,2}$ deliver the bound

$$\mathcal{S}_n(\mathbf{B},\mathbf{E}) = \sup_{\beta} \frac{\|p^{j_n\prime}\beta\|_{2,\infty}}{\sup_{P\in\mathbf{P}}\|p^{j_n\prime}\beta\|_{P,2}} \lesssim \frac{j_n^{5/2}\|\beta\|_2}{\|\beta\|_2} = j_n^{5/2}. \tag{S.196}$$

Thus, we note Assumption 3.13(i) follows from Assumption A.2.17(v), result (S.196), and $\mathcal{R}_n \asymp \sqrt{k_n \log(1+k_n)}/\mathrm{s}_n\sqrt{n}$ implying $\mathcal{R}_n\mathcal{S}_n(\mathbf{B},\mathbf{E}) = o(1)$. Furthermore, we note $\mathcal{R}_n\mathcal{S}_n(\mathbf{B},\mathbf{E}) = o(1)$, Assumption A.2.17(ii), and the definition of Θ in (A.28) imply assumption 3.12(ii) is satisfied as well. Assumption 3.13(ii) similarly follows from Assumption A.2.17(v), result (S.196), $\mathcal{R}_n \asymp \sqrt{k_n \log(1+k_n)}/\mathrm{s}_n\sqrt{n}$, $\kappa_\rho = 1/2$, and Lemma S.4.27. Finally, Assumption 3.13(iii) also follows by Assumption A.2.17(v), result (S.196), $\mathcal{R}_n \asymp \sqrt{k_n(\log(1+k_n))}/\mathrm{s}_n\sqrt{n}$, and $K_g = 0$. We have thus verified the conditions of Theorem 3.2 for R corresponding to (A.26) and (A.27), and hence

$$\hat{U}_n(R|\ell_n) \geq U_P^\star(R|\tilde{\ell}_n) + o_P(a_n)$$

uniformly in $P \in \mathbf{P}_0$ for some $\tilde{\ell}_n \asymp \ell_n$. Similarly, under the additional conditions imposed on the second part of this theorem, Lemma S.3.7(ii) implies that for any $\tilde{\ell}_n^{\mathrm{u}}$ satisfying the conditions of Theorem A.2.5(ii) it follows that uniformly in $P \in \mathbf{P}_0$

$$\hat{U}_n(\Theta|+\infty) \leq U_P^\star(\Theta|\tilde{\ell}_n^{\mathrm{u}}) + o_P(a_n),$$

which implies the second claim of the theorem also holds. ■

Lemma S.4.24. *Let Assumptions A.2.12(i)(iii), A.2.13(ii)(iii), A.2.14(i), A.2.16, and A.2.15(i)(iii)(iv) hold, $k_n \log(1+k_n)/n = o(1)$, and suppose that*

$$Q_n(\hat{\theta}_n) \leq \inf_{\theta\in\Theta_n\cap R} Q_n(\theta) + o(n^{-1/2}) \qquad Q_n(\hat{\theta}_n^{\mathrm{u}}) \leq \inf_{\theta\in\Theta_n} Q_n(\theta) + o(n^{-1/2}).$$

Then: $\|\hat{\theta}_n - \Pi_n\theta_0\|_{1,\infty} \vee \|\hat{\theta}_n^{\mathrm{u}} - \Pi_n\theta_0\|_{1,\infty} = o_P(1)$ uniformly in $P \in \mathbf{P}_0$.

Proof: We establish the result by verifying the conditions of Lemma S.1.1 with $\tau_n = o(n^{-1/2})$. First note that, for both $R = \Theta$ and R corresponding to (A.26), Assumption 3.1(i) is directly imposed in Assumption A.2.12(i), Assumption 3.2(i) holds with $B_n = O(1)$ by Assumption A.2.15(i), Assumption 3.2(iii) holds with $F_n = 1$ and $J_n = O(1)$ by Lemma S.4.27, Assumption S.1.1 is implied by Assumption A.2.16, and Assumption S.1.2(i) follows from Assumption A.2.13(iii). Next, define the following neighborhood

$$\mathcal{V}_n(P) \equiv \{\theta \in \Theta_n : \|\theta - \Pi_n\theta_0\|_{1,\infty} \leq \epsilon\} \tag{S.197}$$

for any $\epsilon > 0$ and $P \in \mathbf{P}_0$ (where recall θ_0 is implicitly a function of P through (A.24)). Further set $Q_P(\theta) \equiv \|E_P[\rho(X,\theta)q^{k_n}(Z)]\|_{\Sigma_P,p}$ and note that since for any $a \in \mathbf{R}^{k_n}$ we have $\|a\|_p \leq \|\Sigma_P^{-1}\|_{o,p}\|\Sigma_P a\|_p$ and $\|\Sigma_P^{-1}\|_{o,p}$ is bounded uniformly in k_n and $P \in \mathbf{P}$ by Assumption A.2.16(ii), we obtain from Lemma S.1.5 and Assumption A.2.13(iii) that

$$\begin{aligned} S_n(\epsilon) &\equiv \inf_{P\in\mathbf{P}_0}\{\inf_{\theta\in(\Theta_n\cap R)\setminus\mathcal{V}_n(P)} Q_P(\theta) - \inf_{\theta\in\Theta_n\cap R} Q_P(\theta)\} \\ &\gtrsim \inf_{P\in\mathbf{P}_0}\inf_{\theta\in(\Theta_n\cap R)\setminus\mathcal{V}_n(P)} \frac{k_n^{1/p}}{\sqrt{k_n}}\|E_P[q^{k_n}(Z)\rho(X,\theta)]\|_2 + O((n\log(n))^{-1/2}). \quad \text{(S.198)} \end{aligned}$$

We further note that the eigenvalues of $E_P[q^{k_n}(Z)q^{k_n}(Z)']$ being bounded uniformly in k_n and $P \in \mathbf{P}$ together with Lemma S.2.5 and Assumption A.2.15(iv) yield

$$\begin{aligned} &\sup_{P\in\mathbf{P}}\sup_{\theta\in\Theta}\|E_P[q^{k_n}(Z)(E_P[\rho(X,\theta)|Z] - q^{k_n}(Z)'\pi_n(\theta))]\|_2^2 \\ &\qquad\qquad \lesssim \sup_{P\in\mathbf{P}}\sup_{\theta\in\Theta} E_P[(E_P[\rho(X,\theta)|Z] - q^{k_n}(Z)'\pi_n(\theta))^2] = o(1). \quad \text{(S.199)} \end{aligned}$$

Therefore, since the eigenvalues of $E_P[q^{k_n}(Z)q^{k_n}(Z)']$ are bounded away from zero by Assumption A.2.15(iii), we obtain from result (S.199) that

$$\begin{aligned} &\inf_{P\in\mathbf{P}_0}\inf_{\theta\in(\Theta_n\cap R)\setminus\mathcal{V}_n(P)}\|E_P[q^{k_n}(Z)\rho(X,\theta)]\|_2 \\ &\qquad \geq \inf_{P\in\mathbf{P}_0}\inf_{\theta\in(\Theta_n\cap R)\setminus\mathcal{V}_n(P)}\|E_P[q^{k_n}(Z)q^{k_n}(Z)'\pi_n(\theta)]\|_2 + o(1) \\ &\qquad \gtrsim \inf_{P\in\mathbf{P}_0}\inf_{\theta\in(\Theta_n\cap R)\setminus\mathcal{V}_n(P)}(E_P[(q^{k_n}(Z)'\pi_n(\theta))^2])^{1/2} + o(1). \quad \text{(S.200)} \end{aligned}$$

Also note that Assumption A.2.13(iii) implies that for n sufficiently large, we have the set inclusion $\Theta_n \cap R \setminus \mathcal{V}_n(P) \subseteq \{\theta \in \Theta : \|\theta - \Pi_n\theta_0\|_{1,\infty} \geq \epsilon\} \subseteq \{\theta \in \Theta : \|\theta - \theta_0\|_{1,\infty} \geq \epsilon/2\}$ holding for all $P \in \mathbf{P}_0$. Hence, (S.198), (S.200), and Assumption A.2.15(iv) yield

$$S_n(\epsilon) \gtrsim \inf_{P\in\mathbf{P}_0}\inf_{\theta\in\Theta:\|\theta-\theta_0\|_{1,\infty}\geq\epsilon}\frac{k_n^{1/p}}{\sqrt{k_n}}(E_P[(P(Y\leq\theta(D)|Z)-\tau)^2])^{1/2} + o(\frac{k_n^{1/p}}{\sqrt{k_n}}). \quad \text{(S.201)}$$

Since $J_n \asymp 1$ and $B_n \asymp 1$, Assumption A.2.14(i), result (S.201) and $k_n\log(1+k_n)/n = o(1)$ by hypothesis imply that $k_n^{1/p}\sqrt{\log(1+k_n)}J_nB_n/\sqrt{n} = o(S_n(\epsilon))$ as required by lemma S.1.1. The preceding arguments apply for both $R = \Theta$ and R corresponding to (A.26) and (A.27), and therefore the claim of the lemma follows. ■

Lemma S.4.25. *Let $k_n\log(1+k_n)/n = o(1)$, and Assumptions A.2.12(i)(iii), A.2.14, A.2.13(ii)(iii), A.2.16, and A.2.15(i)(iii)(iv) hold. For both $R = \Theta$ and R corresponding to (A.26) and (A.27), it follows that Assumption 3.4 holds with $\|\cdot\|_{\mathbf{E}} \equiv \sup_{P\in\mathbf{P}}\|\cdot\|_{P,2}$, $\nu_n \asymp \sqrt{k_n}/\mathrm{s}_nk_n^{1/p}$, and $\mathcal{V}_n(P) \equiv \{\theta \in \Theta_n : \|\theta - \Pi_n\theta_0\|_{1,\infty} \leq \epsilon\}$ for some $\epsilon > 0$.*

Proof: For either $R = \Theta$ or R corresponding to (A.26) and (A.27) set $\mathcal{V}_n(P) \equiv \{\theta \in$

$\Theta_n \cap R : \|\theta - \Pi_n\theta_0\|_{1,\infty} \leq \epsilon\}$ and note that Assumption 3.4(ii) is then satisfied by Lemma S.4.24. Further observe that since $\Pi_n\theta_0 \in \Theta_n$, there exists a β_{0n} such that $\Pi_n\theta_0 = p^{j_n\prime}\beta_{0n}$. For any $\theta = p^{j_n\prime}\beta \in \mathcal{V}_n(P)$, it then follows by Assumptions A.2.13(ii) and A.2.14(ii) that

$$\sup_{P\in\mathbf{P}} \|p^{j_n\prime}(\beta - \beta_{0n})\|_{P,2} \lesssim \|\beta - \beta_{0n}\|_2$$
$$\leq \frac{1}{\mathrm{s}_n} \times \inf_{P\in\mathbf{P}_0} \inf_{\|\theta-\Pi_n\theta_0\|_{1,\infty}\leq\epsilon} \|E_P[f_{Y|DZ,P}(\theta(D)|D,Z)q^{k_n}(Z)p^{j_n}(D)'(\beta_n - \beta_{0n})]\|_2.$$

Hence, since $\|\cdot\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}} \|\cdot\|_{P,2}$, the mean value theorem allows us to conclude that

$$\frac{\mathrm{s}_n k_n^{1/p}}{\sqrt{k_n}} \|p^{j_n\prime}(\beta - \beta_{0n})\|_{\mathbf{E}}$$
$$\lesssim \frac{k_n^{1/p}}{\sqrt{k_n}} \|E_P[(P(Y \leq p^{j_n}(D)'\beta|D,Z) - P(Y \leq p^{j_n}(D)'\beta_{0n}|D,Z))q^{k_n}(Z)]\|_2$$
$$\lesssim \|E_P[(P(Y \leq p^{j_n}(D)'\beta|D,Z) - P(Y \leq p^{j_n}(D)'\beta_{0n}|D,Z))q^{k_n}(Z)]\|_{\Sigma_P,p} \qquad \text{(S.202)}$$

for any $\theta = p^{j_n\prime}\beta \in \mathcal{V}_n(P)$ and $P \in \mathbf{P}_0$, and where the final inequality follows from Lemma S.1.5, $\|a\|_p \leq \|\Sigma_P^{-1}\|_{o,p}\|\Sigma_P a\|_{o,p}$ for any $a \in \mathbf{R}^{k_n}$, and Assumption A.2.16(ii). Since the preceding arguments apply to both $R = \Theta$ and R corresponding to (A.26) and (A.27), result (S.202) verifies Assumption 3.4(i) is satisfied with $\nu_n \asymp \sqrt{k_n}/\mathrm{s}_n k_n^{1/p}$ for both choices of R and therefore the claim of the lemma follows. ■

Lemma S.4.26. *Let Assumption A.2.12(iii) hold. If $f(y,d) = 1\{y \leq \theta(d)\} - \tau$ for some $\theta \in \Theta$ (Θ as in (A.28)) and $z \mapsto q(z)$ is differentiable with bounded level and derivative, then there exists a $K < \infty$ independent of f, such that for all $P \in \mathbf{P}$ we have*

$$\varpi(fq, h, P) \leq K \times \{\|q\|_\infty\sqrt{h} + \|q\|_{1,\infty}h\}.$$

PROOF: First note that since $\|f\|_\infty \leq 1$, we can obtain by direct calculation and the definition of the integral modulus of continuity in (S.283) the upper bound

$$\varpi^2(fq, h, P) \leq 2\|q\|_\infty^2\varpi^2(f, h, P) + 2\varpi^2(q, h, P). \qquad \text{(S.203)}$$

For $\Omega_Z(P)$ the support of Z under P, the mean value theorem then implies that

$$\varpi^2(q, h, P) \equiv \sup_{\|s\|_2\leq h} E_P[(q(Z+s) - q(Z))^2 1\{Z + s \in \Omega_Z(P)\}] \leq \|q\|_{1,\infty}^2 h^2. \qquad \text{(S.204)}$$

Furthermore, for any $(s_y, s_d) \in \mathbf{R}^2$ and $d \in [0,1]$ such that $d + s_d \in [0,1]$, we also note that the mean value theorem and Assumption A.2.12(iii) imply that

$$\begin{aligned} E_P[(1\{Y+s_y \le \theta(D+s_d)\} - 1\{Y \le \theta(D)\})^2 | D=d] \\ = |P(Y \le \theta(D+s_d) - s_y | D=d) - P(Y \le \theta(D)|D=d)| \\ \lesssim |\theta(d+s_d) - s_y - \theta(d)|. \end{aligned} \tag{S.205}$$

Hence, by the law of iterated expectations, a second application of the mean value theorem, and employing that $\|\theta\|_{1,\infty} \le C_0$ by definition of Θ, we can conclude

$$\begin{aligned} \varpi^2(f,h,P) \le \sup_{\|(s_y,s_d)\|_2 \le h} E_P[(1\{Y+s_y \le \theta(D+s_d)\} - 1\{Y \le \theta(D)\})^2 1\{D+s_d \in [0,1]\}] \\ \lesssim \sup_{\|(s_y,s_d)\|_2 \le h} E_P[|\theta(D+s_d) - s_y - \theta(D)| 1\{D+s_d \in [0,1]\}] \lesssim h. \end{aligned} \tag{S.206}$$

The claim of the lemma then follows from (S.203), (S.204), and (S.206). ■

Lemma S.4.27. *Define the class $\mathcal{F}_n \equiv \{f : f(v) = 1\{y \le \theta(d)\} - \tau$ for some $\theta \in \Theta_n\}$ for Θ_n as in (A.29), and suppose that Assumption A.2.12(iii) and A.2.13(ii) hold. For $\zeta_{j_n} \ge (1 \wedge \sup_{d \in [0,1]} \|p^{j_n}(d)\|_2)$, it then follows that for all $\epsilon \le 1$ and some $1 \le K < \infty$*

$$\sup_{P \in \mathbf{P}} N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2}) \le \exp\{\frac{K}{\epsilon}\} \wedge (\frac{K\sqrt{\zeta_{j_n}}}{\epsilon})^{2j_n},$$

and $\sup_{P \in \mathbf{P}} J_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2}) \lesssim \sqrt{1 \wedge \epsilon} \wedge \sqrt{j_n(\log(\zeta_{j_n}) + \log(1 \vee \epsilon^{-1}))}(1 \wedge \epsilon)$.

PROOF: We first note that if $\theta_l(d) \le \theta(d) \le \theta_u(d)$, then it immediately follows that

$$1\{y \le \theta_l(d)\} - \tau \le 1\{y \le \theta(d)\} - \tau \le 1\{y \le \theta_u(d)\}, \tag{S.207}$$

which implies brackets for Θ_n readily yield brackets in $\mathcal{F}_n$. Moreover, by the mean value theorem and Assumption A.2.12(iii) we can in addition conclude that

$$\begin{aligned} E_P[(1\{Y \le \theta_l(D)\} - 1\{Y \le \theta_u(D)\})^2] \\ = E_P[P(Y \le \theta_u(D)|D) - P(Y \le \theta_l(D)|D)] \lesssim E_P[|\theta_u(D) - \theta_l(D)|]. \end{aligned} \tag{S.208}$$

Hence, combining results (S.207) and (S.208) it follows that for some $M_0 < \infty$ we have

$$N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2}) \le N_{[]}(\frac{\epsilon^2}{M_0}, \Theta_n, \|\cdot\|_{P,1}). \tag{S.209}$$

On the other hand, since $\Theta_n \subseteq \Theta$, we also obtain by Corollary 2.7.2 in van der Vaart and Wellner (1996), $\|\cdot\|_{P,2} \le \|\cdot\|_\infty$ and inequality (S.209) that

$$\sup_{P \in \mathbf{P}} N_{[]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2}) \le N_{[]}(\frac{\epsilon^2}{M_0}, \Theta_n, \|\cdot\|_\infty) \le \exp\{\frac{M_1}{\epsilon}\}. \tag{S.210}$$

In addition, the Cauchy-Schwarz inequality implies $\sup_{P\in\mathbf{P}}\|p^{j_n\prime}(\beta_1-\beta_2)\|_{P,1}\leq\zeta_{j_n}\|\beta_1-\beta_2\|_2$ for any $\beta_1,\beta_2\in\mathbf{R}^{j_n}$. Therefore, defining $\mathcal{B}_n\equiv\{\beta\in\mathbf{R}^{j_n}:p^{j_n\prime}\beta\in\Theta_n\}$ Theorem 2.7.11 in van der Vaart and Wellner (1996) allows us to conclude

$$\sup_{P\in\mathbf{P}}N_{[\,]}(\epsilon,\mathcal{F}_n,\|\cdot\|_{P,1})\leq N(\frac{\epsilon^2}{2M_0\zeta_{j_n}},\mathcal{B}_n,\|\cdot\|_2). \tag{S.211}$$

Further note that by Assumption A.2.13(ii), we have $\|p^{j_n\prime}\beta\|_{P,2}\asymp\|\beta\|_2$ uniformly in $P\in\mathbf{P}$ and n, and hence since $\sup_{P\in\mathbf{P}}\|p^{j_n\prime}\beta\|_{P,2}\leq\|p^{j_n\prime}\beta\|_\infty\leq C_0$, it follows

$$\sup_{P\in\mathbf{P}}N_{[\,]}(\epsilon,\mathcal{F}_n,\|\cdot\|_{P,1})\leq(\frac{M_2\sqrt{\zeta_{j_n}}}{\epsilon})^{2j_n} \tag{S.212}$$

for some $M_2<\infty$ due to result (S.211). The first claim of the lemma therefore follows from (S.210) and (S.212). Moreover, noting $N_{[\,]}(1,\mathcal{F}_n,\|\cdot\|_\infty)=1$ we also obtain

$$\begin{aligned}\sup_{P\in\mathbf{P}}J_{[\,]}(\epsilon,\mathcal{F}_n,\|\cdot\|_{P,2})&\leq(\int_0^{1\wedge\epsilon}(1+\frac{K}{u})^{1/2}du)\wedge(\int_0^{1\wedge\epsilon}(1+2j_n\log(\frac{K\sqrt{\zeta_{j_n}}}{u}))^{1/2}du)\\&\lesssim\sqrt{1\wedge\epsilon}\wedge\{\sqrt{j_n\log(\zeta_{j_n})}+\int_0^1(j_n\log(\frac{1}{v(1\wedge\epsilon)}))^{1/2}dv\}(1\wedge\epsilon),\end{aligned} \tag{S.213}$$

where the second inequality follows by the change of variables $v=u/(1\wedge\epsilon)$. The claim of the lemma thus follows from (S.213) and direct calculation. ■

Lemma S.4.28. *Let Assumptions A.2.12(i)(iii)(iv) and A.2.13(ii) hold, Θ_n be as in (A.29), set $\pi_{0n}\equiv\max_{1\leq k\leq k_n}\|q_k\|_\infty$ and $\pi_{1n}\equiv\max_{1\leq k\leq k_n}\|q_k\|_{1,\infty}$, and suppose $\log(k_n\vee\pi_{0n}\vee\sup_d\|p^{j_n}(d)\|_2)=O(\log(n))$. If $j_n/n=o(1)$, then Assumption 3.3(i) holds with $R=\Theta$ for any a_n with $(\pi_{0n}k_n^{1/p}\log(n)\sqrt{j_n}/\sqrt{n})(\sqrt{j_n}+\pi_{0n}n^{1/3}+\pi_{1n}n^{1/6})=o(a_n)$.*

PROOF: We establish the lemma by applying Theorem S.6.1. To this end, define the class $\tilde{\mathcal{F}}_n\equiv\{fq_k$ for some $f\in\mathcal{F}_n,1\leq k\leq k_n\}$ and let $\mathbb{G}_P$ be a Gaussian process on $\tilde{\mathcal{F}}_n$ satisfying $E[\mathbb{G}_P(f_1)]=0$ and $E[\mathbb{G}_P(f_1)\mathbb{G}_P(f_2)]=\mathrm{Cov}_P\{f_1(V),f_2(V)\}$ for any $f_1,f_2\in\tilde{\mathcal{F}}_n$. For any $\theta\in\Theta_n$, set $\mathbb{W}_P(\theta)\equiv(\mathbb{G}_P(\rho(\cdot,\theta)q_1),\ldots,\mathbb{G}_P(\rho(\cdot,\theta)q_{k_n})'$ and note

$$\sup_{\theta\in\Theta_n}\|\mathbb{G}_n(\theta)-\mathbb{W}_P(\theta)\|_p\leq\sup_{f\in\tilde{\mathcal{F}}_n}k_n^{1/p}|\frac{1}{\sqrt{n}}\sum_{i=1}^n(f(V_i)-E_P[f(V)])-\mathbb{G}_P(f)|. \tag{S.214}$$

We proceed by applying Theorem S.6.1 to the class $\tilde{\mathcal{F}}_n$ with $\delta_n\asymp\sqrt{j_n/n}$. Note that Assumptions S.6.1 and S.6.2 are directly imposed in Assumption A.2.12(iv), while Assumption S.6.3(i) is satisfied by Lemma S.4.26, and Assumption S.6.3(ii) holds with $K_n=\pi_{0n}$ since $\mathcal{F}$ has envelope 1. Furthermore, for S_n as in (S.284), we have

$$S_n^2\lesssim\sum_{i=0}^{\lceil\log_2 n\rceil}2^i\{\frac{\pi_{0n}^2}{2^{\frac{i}{3}}}+\frac{\pi_{1n}^2}{2^{\frac{2i}{3}}}\}\lesssim\pi_{0n}^2n^{2/3}+\pi_{1n}^2n^{1/3}$$

due to Lemma S.4.26. Also note that Lemmas S.1.3 and S.4.27 together imply

$$\sup_{P\in\mathbf{P}} \log(N_{[]}(\delta_n, \tilde{\mathcal{F}}_n, \|\cdot\|_{P,2})) \leq \sup_{P\in\mathbf{P}} \log(k_n N_{[]}(\frac{\delta_n}{\pi_{0n}}, \mathcal{F}_n, \|\cdot\|_{P,2})) \lesssim j_n \log(n),$$

where we employed that $\log(k_n \vee \pi_{0n} \vee \sup_d \|p^{j_n}(d)\|_2) = O(\log(n))$ and $\delta_n = \sqrt{j_n/n}$. Similarly, Lemmas S.1.3 and S.4.27 and the change of variables $v = u/\pi_{0n}$ yield

$$\begin{aligned}\sup_{P\in\mathbf{P}} J_{[]}(\delta_n, \tilde{\mathcal{F}}_n, \|\cdot\|_{P,2}) &\leq \sup_{P\in\mathbf{P}} \int_0^{\delta_n} (1 + \log(k_n) + \log(N_{[]}(\frac{u}{\pi_{0n}}, \mathcal{F}_n, \|\cdot\|_{P,2})))^{1/2} du \\ &\leq \delta_n \sqrt{\log(k_n)} + \sup_{P\in\mathbf{P}} J_{[]}(\frac{\delta_n}{\pi_{0n}}, \mathcal{F}_n, \|\cdot\|_{P,2})\pi_{0n} \lesssim \frac{j_n\sqrt{\log(n)}}{\sqrt{n}},\end{aligned}$$

where the final inequality follow from $\log(\pi_{0n}) = O(\log(n))$, $\log(k_n) = O(\log(n))$, and $\log(\sup_d \|p^{j_n}(d)\|_2) = O(\log(n))$. Hence, Theorem S.6.1 implies that

$$\sup_{f\in\tilde{\mathcal{F}}_n} |\frac{1}{\sqrt{n}}\sum_{i=1}^n (f(V_i) - E_P[f(V)]) - \mathbb{G}_P(f)| = O_P(\frac{\pi_{0n}\log(n)\sqrt{j_n}}{\sqrt{n}}\{\sqrt{j_n} + \pi_{0n}n^{1/3} + \pi_{1n}n^{1/6}\})$$

uniformly in $P \in \mathbf{P}$, which together with (S.214) establishes the Lemma. ■

Lemma S.4.29. *If Assumption A.2.12(iii) holds, then it follows that Assumptions 3.3(ii) and 3.5 are satisfied when $R = \Theta$ (Θ as in (A.28)) with $\|\cdot\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}} \|\cdot\|_{P,2}$, $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_\infty$, $\kappa_\rho = 1/2$, $m_P(\theta)(Z) \equiv P(Y \leq \theta(D)|Z)$, and*

$$\nabla m_P(\theta)[h](Z) \equiv E_P[f_{Y|DZ,P}(\theta(D)|D,Z)h(D)|Z]. \tag{S.215}$$

Proof: For $\theta_1 \vee \theta_2$ and $\theta_1 \wedge \theta_2$ the pointwise minimum and maximum of θ_1 and θ_2, note that the conditional density $f_{Y|DZ,P}$ being bounded in (D, Z) and $P \in \mathbf{P}$ by Assumption A.2.12(iii) together with the mean value theorem imply that

$$\begin{aligned}E_P[(\rho(X,\theta_1) - \rho(X,\theta_2))^2] &= E_P[P(Y \leq \theta_1(D) \vee \theta_2(D)|D) - P(Y \leq \theta_1(D) \wedge \theta_2(D)|D)] \\ &\lesssim E_P[\theta_1(D) \vee \theta_2(D) - \theta_1(D) \wedge \theta_2(D)] \leq \sup_{P\in\mathbf{P}} \|\theta_1 - \theta_2\|_{P,2},\end{aligned}$$

where in the final inequality we employed Jensen's inequality and that $\theta_1(d) \vee \theta_2(d) - \theta_1(d) \wedge \theta_2(d) = |\theta_1(d) - \theta_2(d)|$. It thus follows Assumption 3.3(ii) holds with $\|\cdot\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}} \|\cdot\|_{P,2}$ and $\kappa_\rho = 1/2$. Moreover, Jensen's inequality and the mean value theorem imply for some $\bar{\theta}$ such that $\bar{\theta}(d)$ is a convex combination of $\theta_1(d)$ and $\theta_2(d)$ that

$$\begin{aligned}&E_P[(P(Y \leq \theta_1(D)|Z) - P(Y \leq \theta_2(D)|Z) - \nabla m_P(\theta_2)[\theta_1 - \theta_2](Z))^2] \\ &\leq E_P[(\{f_{Y|DZ,P}(\bar{\theta}(D)|D,Z) - f_{Y|DZ,P}(\theta_2(D)|D,Z)\}\{\theta_1(D) - \theta_2(D)\})^2] \\ &\lesssim \|\theta_1 - \theta_2\|_\infty^2 \times \sup_{P\in\mathbf{P}} E_P[(\theta_1(D) - \theta_2(D))^2],\end{aligned}$$

where the final inequality follows from $f_{Y|DZ,P}$ being Lipschitz uniformly in (D,Z) and $P \in \mathbf{P}$. Hence, we may conclude Assumption 3.5(i) is satisfied with $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_\infty$ and $\|\cdot\|_{\mathbf{E}} = \sup_{P\in\mathbf{P}} \|\cdot\|_{P,2}$. Furthermore, once again employing Jensen's inequality and that $f_{Y|DZ,P}$ is Lipschitz uniformly in (D,Z) and $P \in \mathbf{P}$ yields

$$E_P[(E_P[\{f_{Y|DZ,P}(\theta_1(D)|D,Z) - f_{Y|DZ,P}(\theta_2(D)|D,Z)\}h(D)|Z])^2] \lesssim \|\theta_1 - \theta_2\|_\infty^2 \times \sup_{P\in\mathbf{P}} \|h\|_{P,2}^2 \quad \text{(S.216)}$$

which implies Assumption 3.5(ii) is also satisfied under the stated choices of $\|\cdot\|_{\mathbf{L}}$ and $\|\cdot\|_{\mathbf{E}}$. Finally, we note Assumption 3.5(iii) is immediate due to Jensen's inequality and $f_{Y|DZ,P}$ being bounded uniformly in (D,Z) and $P \in \mathbf{P}$. ■

Lemma S.4.30. *If Assumption A.2.17(i) holds, $\mathbf{B} = C_B^2([0,1])$ and Υ_G, Υ_F, and Θ are as defined in (A.26), (A.27) (with $\lambda \neq 0$), and (A.28), then it follows that Assumptions 3.8, 3.9, and 3.10 are satisfied with $K_g = 0$, $\nabla\Upsilon_G(\theta)[h] = -\nabla^2 h$, and*

$$\nabla\Upsilon_F(\theta)[h] = 2\int_0^1 \theta(u)h(u)du - 2(\int_0^1 \theta(u)du)(\int_0^1 h(u)du). \quad \text{(S.217)}$$

Proof: Note that since Υ_G is linear and continuous, it immediately follows that Assumptions 3.8(i) and 3.8(ii) hold with $\nabla\Upsilon_G = \Upsilon_G$ and $K_g = 0$. It further follows from $\nabla\Upsilon_G = \Upsilon_G$ and the definitions of the operator norm $\|\cdot\|_o$ and $\|\cdot\|_{m,\infty}$ that

$$\|\nabla\Upsilon_G(\theta)\|_o = \sup_{\|h\|_{2,\infty}=1} \|-\nabla^2 h\|_\infty \leq 1, \quad \text{(S.218)}$$

which implies Assumption 3.8(iii) holds with $M = 1$. Moreover, by direct calculation

$$|\Upsilon_F(\theta_1) - \Upsilon_F(\theta_2) - \nabla\Upsilon_F(\theta_1)[\theta_1 - \theta_2]| = |\int_0^1 (\theta_1(u) - \theta_2(u))^2 du - (\int_0^1 (\theta_1(u) - \theta_2(u))du)^2| \leq \|\theta_1 - \theta_2\|_{2,\infty}^2, \quad \text{(S.219)}$$

which implies Υ_F is indeed Fréchet differentiable and its derivative is equal to $\nabla\Upsilon_F$ as defined in (S.217). In addition, by (S.217) and Jensen's inequality we have

$$\|\nabla\Upsilon_F(\theta_1) - \nabla\Upsilon_F(\theta_2)\|_o = \sup_{\|h\|_{2,\infty}=1} 2|\int_0^1 (\theta_1(u) - \theta_2(u))(h(u) - \int_0^1 h(\tilde{u})d\tilde{u})du| \leq 2\|\theta_1 - \theta_2\|_{2,\infty}, \quad \text{(S.220)}$$

which together with (S.219) implies Assumptions 3.9(i) and 3.9(ii) hold with $K_f = 2$. Next, note that since $\lambda \neq 0$ it follows that $\mathbf{F}_n = \mathbf{R}$. For any $\theta \in \mathbf{B}_n$ such that $\Upsilon_F(\theta) \neq 0$, we then define $\nabla\Upsilon_F(\theta)^- : \mathbf{F}_n \to \mathbf{B}_n$ to be given (for any $c \in \mathbf{R}$) by

$$\nabla\Upsilon_F(\theta)^-[c](d) \equiv c \times \frac{\theta(d) - \int_0^1 \theta(u)du}{2\Upsilon_F(\theta)}, \tag{S.221}$$

and note that since $\theta \in \mathbf{B}_n$ and the constant function is in $\mathbf{B}_n$ by Assumption A.2.17(i), it follows that $\nabla\Upsilon_F(\theta)^-[c] \in \mathbf{B}_n$. Moreover, by direct calculation we obtain

$$\nabla\Upsilon_F(\theta)\nabla\Upsilon_F(\theta)^-[c] = 2\int_0^1 \theta(u)\{c \times \frac{\theta(u) - \int_0^1 \theta(\tilde{u})d\tilde{u}}{2\Upsilon_F(\theta)}\}du = c \times \frac{2\Upsilon_F(\theta)}{2\Upsilon_F(\theta)} = c, \tag{S.222}$$

which verifies $\nabla\Upsilon_F(\theta)^-$ is indeed the right inverse of $\nabla\Upsilon_F(\theta)$. In addition note that

$$\|\nabla\Upsilon_F(\theta)^-\|_o = \sup_{|c|=1} \|c \times \frac{\theta - \int_0^1 \theta(u)du}{2\Upsilon_F(\theta)}\|_{2,\infty} \le \frac{\|\theta\|_{2,\infty}}{|\Upsilon_F(\theta)|}, \tag{S.223}$$

and hence, since $\|\theta\|_{2,\infty} \le C_0$ and $\Upsilon_F(\theta) = \lambda$ for any $\theta \in \Theta_{0n}^{\mathrm{r}}$, it follows that we may select an $\epsilon > 0$ such that Assumption 3.9(iv) holds with $M = 4C_0/\lambda$.

Next, let θ_2 be the function given by $\theta_2(d) = d^2$ and note that by Assumption A.2.17(i) it follows that $\theta_2 \in \mathbf{B}_n$. For any $\theta \in \Theta_{0n}^{\mathrm{r}}$ we may then set h to equal

$$h \equiv \frac{2\lambda}{C_0}\theta_2 - \frac{\nabla\Upsilon_F(\theta)[\theta_2]}{C_0}\theta, \tag{S.224}$$

which belongs to $\mathbf{B}_n$ since $\theta_2, \theta \in \mathbf{B}_n$. Further observe $\nabla\Upsilon_F(\theta)[\theta] = 2\Upsilon_F(\theta) = 2\lambda$ due to $\theta \in R$, and hence by linearity of $\nabla\Upsilon_F(\theta)$ and (S.224) we can conclude that $h \in \mathbf{B}_n \cap \mathcal{N}(\nabla\Upsilon_F(\theta))$. In addition, it also follows from $\Upsilon_G = \nabla\Upsilon_G$ that

$$\Upsilon_G(\theta)(u) + \nabla\Upsilon_G(\theta)[h](u) = -\nabla^2\theta(u)(1 - \frac{\nabla\Upsilon_F(\theta)[\theta_2]}{C_0}) - \frac{4\lambda}{C_0} \le -\frac{4\lambda}{C_0}, \tag{S.225}$$

where the inequality results from $-\nabla^2\theta(u) \le 0$ due to $\theta \in \Theta_{0n}^{\mathrm{r}}$, $\theta_2(d) = d^2$, and $|\nabla\Upsilon_F(\theta)[\theta_2]| \le C_0$ because $\|\theta\|_{2,\infty} \le C_0$ since $\theta \in \Theta_{0n}^{\mathrm{r}} \subseteq \Theta_n$. By similar arguments and the triangle inequality we also have $\|h\|_{2,\infty} \le 4\lambda/C_0 + C_0$ and hence by (S.225) we conclude Assumption 3.10 is satisfied. ■

Lemma S.4.31. *Suppose Assumptions A.2.12(i)(iii)(iv) and A.2.13(ii) hold, Θ_n be as in (A.29), and let $\pi_{0n} \equiv \max_{1\le k\le k_n} \|q_k\|_\infty$ and $\pi_{1n} \equiv \max_{1\le k\le k_n} \|q_k\|_{1,\infty}$. For any sequence $d_n \uparrow \infty$ such that $d_n^4 \log(1+d_n) = o(n)$ and $\delta_n \asymp d_n^{-1/6} + \pi_{1n}/(\pi_{0n} d_n^{1/3})$ satisfies $\delta_n \log(1+k_n) = o(1)$ it follows that uniformly in $P \in \mathbf{P}$ we have*

$$\sup_{\theta\in\Theta_n} \|\hat{\mathbb{W}}_n(\theta) - \mathbb{W}_P^\star(\theta)\|_p$$
$$= O_P(\frac{k_n^{1/p} d_n^2 \pi_{0n}\sqrt{\log(1+d_n)}}{\sqrt{n}} + \pi_{0n} k_n^{1/p}(\sqrt{\delta_n} + \sqrt{n}\exp\{-n\delta_n^3\})).$$

PROOF: We first define the class $\tilde{\mathcal{F}}_n \equiv \{fq_k \text{ for some } f \in \mathcal{F}_n \text{ and } 1 \leq k \leq k_n\}$, let $\mathbb{G}_P^\star$ be an isonormal Gaussian process on $\tilde{\mathcal{F}}_n$ independent of $\{V_i\}_{i=1}^n$, set $\mathbb{W}_P^\star(\theta) \equiv (\mathbb{G}_P^\star(\rho(\cdot,\theta)q_1),\ldots,\mathbb{G}_P^\star(\rho(\cdot,\theta)q_{k_n}))'$, and for any $f \in \mathcal{F}_n$ define $\hat{\mathbb{G}}_n(f)$ to equal

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

where $\{\omega_i\}_{i=1}^n$ are the same weights employed in $\hat{\mathbb{W}}_n$. These definitions then imply

$$\sup_{\theta\in\Theta_n} \|\hat{\mathbb{W}}_n(\theta) - \mathbb{W}_P^\star(\theta)\|_p \leq \sup_{f\in\tilde{\mathcal{F}}_n} k_n^{1/p}|\hat{\mathbb{G}}_n(f) - \mathbb{G}_P^\star(f)|. \tag{S.226}$$

In what follows, we aim to establish the lemma by applying Theorem S.7.1 to the class $\tilde{\mathcal{F}}_n$ by relying on a Haar basis expansion as in Lemmas S.6.1 and S.6.2. Specifically, note that by Assumption A.2.12(iv) and Lemma S.6.1, there exists a sequence of partitions $\Delta_n(P) = \{\Delta_{d,n}(P) : d = 1,\ldots,d_n\}$ of the support of $V \equiv (Y,D,Z)$ such that $P(\Delta_{d,n}(P)) = 1/d_n$. For any $1 \leq d \leq d_n - 1$ we then set $\{f_{d,n,P}\}_{d=1}^{d_n-1}$ to be given by

$$f_{d,n,P}(V) \equiv \frac{(d_n 1\{V \in \Delta_{d,n}(P)\} - 1)}{\sqrt{d_n - 1}} \tag{S.227}$$

and let $f_{n,P}^{d_n}(v) \equiv (f_{1,n,P}(v),\ldots,f_{d_n-1,n,P}(v))'$. Then note that $E_P[f_{n,P}^{d_n}(V)] = 0$ and

$$E_P[f_{d,n,P}(V)f_{\tilde{d},n,P}(V)] = \begin{cases} 1 & \text{if } d = \tilde{d} \\ -\frac{1}{d_n-1} & \text{if } d \neq \tilde{d} \end{cases}. \tag{S.228}$$

By result (S.228) and direct calculation it follows Assumption S.7.1(i) holds with $C_n \asymp 1$, while (S.227) implies Assumption S.7.1(ii) holds with $K_n \asymp \sqrt{d_n}$. Also note that $\text{Var}_P\{f_{n,P}^{d_n}(V)\} = E_P[f_{n,P}^{d_n}(V)f_{n,P}^{d_n}(V)']$ and therefore using that by (S.228) the smallest eigenvalue of $E_P[f_{n,P}^{d_n}(V)f_{n,P}^{d_n}(V)']$ is of order $1/d_n$, we obtain uniformly in $P \in \mathbf{P}$ that

$$\|\text{Var}_P^{-1}\{f_{n,P}^{d_n}(V)\}\|_{o,2} \lesssim d_n. \tag{S.229}$$

We next aim to verify that Assumption S.7.2 is satisfied by setting $\beta_{n,P}(f)$ to be

$$\beta_{n,P}(f) \equiv \begin{pmatrix} \frac{\sqrt{d_n-1}}{d_n}(E_P[f(V)|V\in\Delta_{1,n}(P)] - E_P[f(V)|V\in\Delta_{d_n,n}(P)]) \\ \vdots \\ \frac{\sqrt{d_n-1}}{d_n}(E_P[f(V)|V\in\Delta_{d_n-1,n}(P)] - E_P[f(V)|V\in\Delta_{d_n,n}(P)]) \end{pmatrix} \tag{S.230}$$

for any $f \in \tilde{\mathcal{F}}_n$. Then observe that, by direct calculation, for any $f \in \tilde{\mathcal{F}}_n$ we have that

$$\begin{aligned}
&f_{n,P}^{d_n}(V)'\beta_{n,P}(f) \\
&= \sum_{d=1}^{d_n-1}(E_P[f(V)|V \in \Delta_{d,n}(P)] - E_P[f(V)|V \in \Delta_{d_n,n}(P)])(1\{V \in \Delta_{d,n}(P)\} - 1/d_n) \\
&= \sum_{d=1}^{d_n} E_P[f(V)|V \in \Delta_{d,n}(P)]1\{V \in \Delta_{d,n}(P)\} - E_P[f(V)], \qquad \text{(S.231)}
\end{aligned}$$

where the final equality follows from $\{\Delta_{d,n}(P)\}_{d=1}^{d_n}$ being a partition of the support of V that satisfies $P(V \in \Delta_{d,n}(P)) = 1/d_n$. Defining $\mathcal{G}_{n,P} \equiv \{(f - \int f dP) - f_{n,P}^{d_n\prime}\beta_{n,P}(f) : f \in \tilde{\mathcal{F}}_n\}$, then observe that since $\mathcal{F}_n$ has envelope 1, it follows the class $\tilde{\mathcal{F}}_n$ has envelope π_{0n} and hence by (S.231) and Jensen's inequality, the class $\mathcal{G}_{n,P}$ has envelope $G_{n,P} \equiv 2\pi_{0n}$. Moreover, by Lemmas S.6.2 and S.4.26 we can in addition conclude that

$$\sup_{P \in \mathbf{P}} \|(f - \int f dP) - f_{n,P}^{d_n\prime}\beta_{n,P}(f)\|_{P,2}^2 \lesssim \frac{\pi_{0n}^2}{d_n^{1/3}} + \frac{\pi_{1n}^2}{d_n^{2/3}},$$

and hence it follows that $\|g\|_{P,2} \leq \delta_n \|G_{n,P}\|_{P,2}$ for all $g \in \mathcal{G}_{n,P}$, $P \in \mathbf{P}$, and δ_n satisfying

$$\delta_n \asymp \frac{1}{d_n^{1/6}} + \frac{\pi_{1n}}{\pi_{0n} d_n^{1/3}}.$$

Next, note that if $\underline{f}(V) \leq f(V) \leq \bar{f}(V)$ almost surely, then result (S.231) yields that

$$\begin{aligned}
\underline{f}(V) - \sum_{d=1}^{d_n} E_P[\bar{f}(V)|V \in \Delta_{d,n}(P)]1\{V \in \Delta_{d,n}(P)\} &\leq (f(V) - \int f dP) - f_n^{d_n}(V)'\beta_{n,P}(f) \\
&\leq \bar{f}(V) - \sum_{d=1}^{d_n} E_P[\underline{f}(V)|V \in \Delta_{d,n}(P)]1\{V \in \Delta_{d,n}(P)\} \qquad \text{(S.232)}
\end{aligned}$$

which implies brackets for $\tilde{\mathcal{F}}_n$ can be employed to obtain brackets for $\mathcal{G}_{n,P}$. Moreover, by the triangle and Jensen's inequality, the width of the brackets built in (S.232) is bounded by $2\|\bar{f} - \underline{f}\|_{P,2}$. Thus, Lemma S.1.3, and $\tilde{\mathcal{F}}_n$ having envelope π_{0n} yields

$$\begin{aligned}
\sup_{P \in \mathbf{P}} \log(N_{[\,]}(\epsilon, \mathcal{G}_{n,P}, \|\cdot\|_{P,2})) &\leq \sup_{P \in \mathbf{P}} \log(N_{[\,]}(\frac{\epsilon}{2}, \tilde{\mathcal{F}}_n, \|\cdot\|_{P,2})) \\
&\leq \log(k_n) + \sup_{P \in \mathbf{P}} \log(N_{[\,]}(\frac{\epsilon}{2\pi_{0n}}, \mathcal{F}_n, \|\cdot\|_{P,2})) \lesssim \log(k_n) + \frac{\pi_{0n}}{\epsilon} 1\{\epsilon \leq 2\pi_{0n}\} \qquad \text{(S.233)}
\end{aligned}$$

where the final inequality follows for any $\epsilon \leq 2\pi_{0n}$ by Lemma S.4.27, and for any $\epsilon > 2\pi_{0n}$ by observing that $\mathcal{F}_n$ is contained in the brackets $[-\tau, 1-\tau]$ which has width 1, and hence $N_{[\,]}(1, \mathcal{F}_n, \|\cdot\|_{P,2}) = 1$. Recalling that $\mathcal{G}_{n,P}$ has envelope $G_{n,P} \equiv 2\pi_{0n}$, we can then use result (S.233) to obtain the following upper bound

$$\sup_{P\in\mathbf{P}} J_{[\,]}(\delta_n\|G_{n,P}\|_{P,2},\mathcal{G}_{n,P},\|\cdot\|_{P,2}) \lesssim \int_0^{2\delta_n\pi_{0n}} \sqrt{1+\log(k_n)+\frac{\pi_{0n}}{\epsilon}1\{\epsilon\le 2\pi_{0n}\}}d\epsilon$$

$$\lesssim \delta_n\pi_{0n}\sqrt{\log(1+k_n)} + \int_0^{2\delta_n\pi_{0n}}\sqrt{\frac{\pi_{0n}}{\epsilon}}d\epsilon \lesssim \pi_{0n}\sqrt{\delta_n}, \quad \text{(S.234)}$$

where in the final inequality we employed that $\delta_n\log(1+k_n)=o(1)$ by hypothesis. Similarly, for $\eta_{n,P}\equiv 1+\log N_{[\,]}(\delta_n\|G_{n,P}\|_{P,2},\mathcal{G}_{n,P},\|\cdot\|_{P,2})$ we can conclude that

$$\sqrt{n}E_P[G_{n,P}(V)\exp\{-\frac{n\delta_n^2\|G_{n,P}\|_{P,2}^2}{G_{n,P}^2(V)\eta_{n,P}}\}] \lesssim \sqrt{n}\pi_{0n}\exp\{-\frac{n\delta_n^2}{1+\log(k_n)+\frac{1}{2\delta_n}1\{\delta_n\le 1\}}\}$$

$$\le \sqrt{n}\pi_{0n}\exp\{-n\delta_n^3\} \quad \text{(S.235)}$$

where the second inequality holds for n sufficiently large due to $\delta_n\log(1+k_n)=o(1)$. Together, results (S.234) and (S.235) verify Assumption S.7.2(i) is satisfied with $J_{1n}\asymp \pi_{0n}(\sqrt{\delta_n}+\sqrt{n}\exp\{-n\delta_n^3\})$. Finally, let $\mathcal{B}_n\equiv\{\beta_{n,P}(f):f\in\tilde{\mathcal{F}}_n,P\in\mathbf{P}\}$ and note (S.230), $P(\Delta_{i,n}(P))=1/d_n$, Jensen's inequality, and $\|f\|_\infty\le\pi_{0n}$ for any $f\in\tilde{\mathcal{F}}_n$ imply $\|\beta_{n,P}(f)\|_2\lesssim\pi_{0n}$ for all $f\in\tilde{\mathcal{F}}_n$ and $P\in\mathbf{P}$. It thus follows that $\mathcal{B}_n$ is contained in a ball of radius $M\pi_{0n}$ for some $M<\infty$, which allows us to conclude

$$\int_0^\infty\sqrt{N(\epsilon,\mathcal{B}_n,\|\cdot\|_2)}d\epsilon \lesssim \int_0^{M\pi_{0n}}\sqrt{d_n\log(\frac{M\pi_{0n}}{\epsilon})}d\epsilon$$

$$=\sqrt{d_n}M\pi_{0n}\int_0^1\sqrt{\log(\frac{1}{u})}du = O(\sqrt{d_n}\pi_{0n}), \quad \text{(S.236)}$$

where the first equality follows from the change of variables $u=\epsilon/M\pi_{0n}$. Result (S.236) verifies Assumption S.7.2(ii) is satisfied with $J_{2n}\asymp\sqrt{d_n}\pi_{0n}$. In summary, since $d_n^4\log(1+d_n)=o(n)$ by hypothesis, it follows that the conditions of Theorem S.7.1(ii) hold with $C_n\asymp 1$, $K_n\asymp\sqrt{d_n}$, $\xi_n\asymp d_n$, $J_{1n}\asymp\pi_{0n}(\sqrt{\delta_n}+\sqrt{n}\exp\{-n\delta_n^3\})$, and $J_{2n}\asymp\sqrt{d_n}\pi_{0n}$. Therefore, Theorem S.7.1(ii) allows us to conclude, uniformly in $P\in\mathbf{P}$, that

$$\sup_{f\in\tilde{\mathcal{F}}_n}|\hat{\mathbb{G}}_n(f)-\mathbb{G}_P^\star(f)| = O_P(\frac{\pi_{0n}d_n^2\sqrt{\log(1+d_n)}}{\sqrt{n}}+\pi_{0n}(\sqrt{\delta_n}+\sqrt{n}\exp\{-n\delta_n^3\})),$$

which together with (S.226) establishes the claim of the Lemma. ■

S.5 Local Parameter Space

This section contains analytical results concerning our approximation to the local parameter space. The main result of this section is the following theorem.

Theorem S.5.1. *Let Assumptions 3.1(ii)(iii), 3.8, 3.9, and 3.10 hold,* $\{\ell_n,\delta_n,r_n\}_{n=1}^\infty$

satisfy $\ell_n \downarrow 0$, $\delta_n 1\{K_f > 0\} \downarrow 0$, $r_n \geq 2(\ell_n + \delta_n)1\{K_g > 0\}$, $r_n/\delta_n \downarrow 0$, *and define*

$$G_n(\theta) \equiv \{h \in \mathbf{B}_n : \Upsilon_G(\theta + \frac{h}{\sqrt{n}}) \leq (\Upsilon_G(\theta) - K_g r_n \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \mathbf{1_G}) \vee (-r_n \mathbf{1_G})\}$$
$$A_n(\theta) \equiv \{h \in \mathbf{B}_n : h \in G_n(\theta),\ \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\}$$
$$T_n(\theta) \equiv \{h \in \mathbf{B}_n : \Upsilon_F(\theta + \frac{h}{\sqrt{n}}) = 0,\ \Upsilon_G(\theta + \frac{h}{\sqrt{n}}) \leq 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq 2\ell_n\}.$$

(i) Then, there exist $M < \infty$, $\epsilon > 0$, *and* $n_0 < \infty$ *such that for all* $n > n_0$, $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$, *and* $\theta_1 \in (\Theta_{0n}^{\mathrm{r}})^{\epsilon} \cap R$ *satisfying* $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$ *we have*

$$\sup_{h_1 \in A_n(\theta_1)} \inf_{h_0 \in T_n(\theta_0)} \|\frac{h_1}{\sqrt{n}} - \frac{h_0}{\sqrt{n}}\|_{\mathbf{B}} \leq M \times \ell_n(\ell_n + \delta_n)1\{K_f > 0\}. \tag{S.237}$$

(ii) If in addition Υ_G *and* Υ_F *are affine, then for any* $\theta_0, \theta_1 \in \mathbf{B}_n \cap R$ *with* $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$

$$\{h \in \mathbf{B}_n : h \in G_n(\theta_1) \text{ and } \Upsilon_F(\theta_1 + \frac{h}{\sqrt{n}}) = 0\}$$
$$\subseteq \{h \in \mathbf{B}_n : \Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) \leq 0 \text{ and } \Upsilon_F(\theta_0 + \frac{h}{\sqrt{n}}) = 0\}.$$

PROOF: We begin by establishing part (ii). First note that if Υ_G is affine, then $K_g = 0$ and since $r_n/\delta_n = o(1)$, Lemma S.5.1(ii) implies that for n sufficiently large

$$G_n(\theta_1) \subseteq \{h \in \mathbf{B}_n : \Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) \leq 0\} \tag{S.238}$$

for any $\theta_0, \theta_1 \in \mathbf{B}_n$ with $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$. Moreover, if Υ_F is affine and continuous, then $\Upsilon_F(\theta) = L(\theta) + c_0$ for some continuous linear map $L : \mathbf{B} \to \mathbf{F}$ and $c_0 \in \mathbf{F}$. It follows that $\nabla\Upsilon_F(\theta)[h] = L(h)$, which does not depend on θ, and since any $\theta \in R$ must satisfy $L(\theta) = -c_0$ (since $\Upsilon_F(\theta) = 0$), we can conclude that $\{h : \Upsilon_F(\theta + h) = 0\} = \{h : L(h) = 0\}$ whenever $\theta \in R$. Therefore part (ii) follows from result (S.238) and $\theta_1, \theta_2 \in R$.

We next turn to the proof of part (i). Throughout, let $\tilde{\epsilon}$ be such that Assumptions 3.8 and 3.9 hold and set $\epsilon = \tilde{\epsilon}/2$. We break up the proof into four steps.

<u>STEP 1:</u> (Decompose $h/\sqrt{n}$). For any $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$, and $h \in \mathbf{B}_n$ set

$$h^{\perp_{\theta_0}} \equiv \nabla\Upsilon_F(\theta_0)^{-}\nabla\Upsilon_F(\theta_0)[h] \qquad\qquad h^{\mathcal{N}_{\theta_0}} \equiv h - h^{\perp_{\theta_0}}, \tag{S.239}$$

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

$$\nabla\Upsilon_F(\theta_0)[h^{\mathcal{N}_{\theta_0}}] = \nabla\Upsilon_F(\theta_0)[h] - \nabla\Upsilon_F(\theta_0)\nabla\Upsilon_F(\theta_0)^{-}\nabla\Upsilon_F(\theta_0)[h] = 0, \tag{S.240}$$

by definition of $h^{\perp\theta_0}$ in (S.239). Next, observe that if $\theta_1 \in (\Theta_{0n}^{\mathrm{r}})^{\epsilon} \cap R$ and $h \in \mathbf{B}_n$ satisfies $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ and $\Upsilon_F(\theta_1 + h/\sqrt{n}) = 0$, then $\theta_1 + h/\sqrt{n} \in (\Theta_{0n}^{\mathrm{r}})^{\tilde{\epsilon}}$ for n sufficiently large, and hence by Assumption 3.9(i) and $\Upsilon_F(\theta_1) = 0$ due to $\theta_1 \in R$

$$\|\nabla\Upsilon_F(\theta_1)[\frac{h}{\sqrt{n}}]\|_{\mathbf{F}} = \|\Upsilon_F(\theta_1 + \frac{h}{\sqrt{n}}) - \Upsilon_F(\theta_1) - \nabla\Upsilon_F(\theta_1)[\frac{h}{\sqrt{n}}]\|_{\mathbf{F}} \leq K_f\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}}^2. \quad \text{(S.241)}$$

Hence, Assumption 3.9(ii), result (S.241), $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$, and $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ imply

$$\begin{aligned}\|\nabla\Upsilon_F(\theta_0)[\frac{h}{\sqrt{n}}]\|_{\mathbf{F}} \\ &\leq \|\nabla\Upsilon_F(\theta_0) - \nabla\Upsilon_F(\theta_1)\|_o\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} + K_f\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}}^2 \leq K_f\ell_n(\delta_n + \ell_n). \quad \text{(S.242)}\end{aligned}$$

Moreover, since $\nabla\Upsilon_F(\theta_0) : \mathbf{F}_n \to \mathbf{B}_n$ satisfies Assumption 3.9(iv), we also have that

$$\begin{aligned}K_f\|h^{\perp\theta_0}\|_{\mathbf{B}} &= K_f\|\nabla\Upsilon_F(\theta_0)^{-}\nabla\Upsilon(\theta_0)[h]\|_{\mathbf{B}} \\ &\leq K_f\|\nabla\Upsilon_F(\theta_0)^{-}\|_o\|\nabla\Upsilon_F(\theta_0)[h]\|_{\mathbf{F}} \leq M_f\|\nabla\Upsilon_F(\theta_0)[h]\|_{\mathbf{F}} \quad \text{(S.243)}\end{aligned}$$

for some $M_f < \infty$. Further note that if $K_f = 0$, then (S.239) and (S.242) imply $h^{\perp\theta_0} = 0$. Thus, combining results (S.242) and (S.243) to handle the case $K_f > 0$ we conclude that for any $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$, $\theta_1 \in (\Theta_{0n}^{\mathrm{r}})^{\epsilon} \cap R$ satisfying $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$ and any $h \in \mathbf{B}_n$ such that $\Upsilon_F(\theta_1 + h/\sqrt{n}) = 0$ and $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ we must have

$$\|\frac{h^{\perp\theta_0}}{\sqrt{n}}\|_{\mathbf{B}} \leq M_f\ell_n(\delta_n + \ell_n)1\{K_f > 0\}. \quad \text{(S.244)}$$

<u>Step 2:</u> (Inequality Constraints). In what follows, it is convenient to define the set

$$S_n(\theta_0, \theta_1) \equiv \{h \in \mathbf{B}_n : \Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) \leq 0,\ \Upsilon_F(\theta_1 + \frac{h}{\sqrt{n}}) = 0,\ \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\}.$$

Then note $r_n \geq 2(\ell_n + \delta_n)1\{K_g > 0\}$, $r_n/\delta_n = o(1)$, and Lemma S.5.1(i) imply that

$$A_n(\theta_1) \subseteq S_n(\theta_0, \theta_1) \quad \text{(S.245)}$$

for n sufficiently large, all $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$, and $\theta_1 \in (\Theta_{0n}^{\mathrm{r}})^{\epsilon}$ satisfying $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$. The proof will proceed by verifying (S.237) holds with $S_n(\theta_0, \theta_1)$ in place of $A_n(\theta_1)$. In particular, if $K_f = 0$, then $\Upsilon_F(\theta_0) = \Upsilon_F(\theta_1)$ due to $\theta_0, \theta_1 \in R$, and Assumptions 3.9(i)(ii) together with (S.245) imply $A_n(\theta_1) \subseteq S_n(\theta_0, \theta_1) \subseteq T_n(\theta_0)$. Hence, result (S.237) holds for the case $K_f = 0$.

For the rest of the proof we therefore assume $K_f > 0$. We further note that Lemma S.5.2 implies that for any $\eta_n \downarrow 0$, there is an $n_0 < \infty$ and $1 \leq C < \infty$ (independent of η_n) such that for all $P \in \mathbf{P}_0$, $n > n_0$, and $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$ there exists a $h_{\theta_0,n} \in \mathbf{B}_n \cap$

$\mathcal{N}(\nabla\Upsilon_F(\theta_0))$ such that for any $\tilde{h} \in \mathbf{B}_n$ for which there exists a $h \in S_n(\theta_0, \theta_1)$ satisfying $\|(\tilde{h} - h)/\sqrt{n}\|_{\mathbf{B}} \leq \eta_n$ the following inequalities hold

$$\Upsilon_G(\theta_0 + \frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{\tilde{h}}{\sqrt{n}}) \leq 0 \qquad \|\frac{h_{\theta_0,n}}{\sqrt{n}}\|_{\mathbf{B}} \leq C\eta_n. \tag{S.246}$$

<u>Step 3:</u> (Equality Constraints). The results in this step allow us to address the challenge that $h \in S_n(\theta_0, \theta_1)$ 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

$$\mathbf{B}_n^{\mathcal{N}_{\theta_0}} \equiv \mathbf{B}_n \cap \mathcal{N}(\nabla\Upsilon_F(\theta_0)) \qquad \mathbf{B}_n^{\perp\theta_0} \equiv \mathcal{R}(\nabla\Upsilon_F(\theta_0)^-\nabla\Upsilon_F(\theta_0)). \tag{S.247}$$

Since $h^{\mathcal{N}_{\theta_0}} \in \mathbf{B}_n^{\mathcal{N}_{\theta_0}}$ by (S.240), the definitions in (S.239) and (S.247) 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

$$\nabla\Upsilon_F(\theta_0)^-\nabla\Upsilon_F(\theta_0)[h] = h \tag{S.248}$$

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 (S.239) 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 (S.248) its inverse is $\nabla\Upsilon_F(\theta_0)^- : \mathbf{F}_n \to \mathbf{B}_n^{\perp\theta_0}$.

Recall Υ_F is Fréchet differentiable on $(\Theta_{0n}^{\mathrm{r}})^{\tilde{\epsilon}}$ by Assumption 3.9(i). The preceding discussion and Assumption 3.9 imply we may apply Lemma S.5.4 with $\mathbf{A}_1 = \mathbf{B}_n^{\mathcal{N}_{\theta_0}}$, $\mathbf{A}_2 = \mathbf{B}_n^{\perp\theta_0}$, and some $K_0 < \infty$ to obtain that for any $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$ 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 is a $h^\star(h^{\mathcal{N}_{\theta_0}}) \in \mathbf{B}_n^{\perp\theta_0}$ such that

$$\Upsilon_F(\theta_0 + h^{\mathcal{N}_{\theta_0}} + h^\star(h^{\mathcal{N}_{\theta_0}})) = 0 \qquad \|h^\star(h^{\mathcal{N}_{\theta_0}})\|_{\mathbf{B}} \leq 2K_0^2\|h^{\mathcal{N}_{\theta_0}}\|_{\mathbf{B}}^2. \tag{S.249}$$

Moreover, for any $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$, $\theta_1 \in (\Theta_{0n}^{\mathrm{r}})^\epsilon \cap R$ with $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$, and any $h \in \mathbf{B}_n$ such that $\Upsilon_F(\theta_1 + h/\sqrt{n}) = 0$ and $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$, result (S.244), the decomposition in (S.239), $\delta_n \downarrow 0$ (since $K_f > 0$), and $\ell_n \downarrow 0$ imply that for n large

$$\|\frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}}\|_{\mathbf{B}} \leq \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} + \|\frac{h^{\perp\theta_0}}{\sqrt{n}}\|_{\mathbf{B}} \leq 2\ell_n. \tag{S.250}$$

Thus, for $h_{\theta_0,n} \in \mathbf{B}_n^{\mathcal{N}_{\theta_0}}$ as in (S.246), $C \geq 1$, and results (S.249) and (S.250) imply that for n sufficiently large we must have for all $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$, $\theta_1 \in \mathbf{B}_n \cap R$ with

$\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$ and $h \in \mathbf{B}_n$ satisfying $\Upsilon_F(\theta_1 + h/\sqrt{n}) = 0$ that

$$\Upsilon_F(\theta_0 + \frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}} + h^\star(\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}})) = 0$$
$$\|h^\star(\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}})\|_{\mathbf{B}} - 16K_0^2C^2(\ell_n^2 + \eta_n^2) \leq 0. \tag{S.251}$$

<u>Step 4:</u> (Build Approximation). In order to employ Steps 2 and 3, we now set η_n to

$$\eta_n = 32(M_f + C^2K_0^2)\ell_n(\ell_n + \delta_n). \tag{S.252}$$

In addition, for any $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$, $\theta_1 \in \mathbf{B}_n \cap R$ satisfying $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$, and any $h \in S_n(\theta_0, \theta_1)$, we let $h^{\mathcal{N}_{\theta_0}}$ be as in (S.239) and define

$$\frac{\hat{h}}{\sqrt{n}} \equiv \frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}} + h^\star(\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}}). \tag{S.253}$$

From Steps 2 and 3 it then follows that for n sufficiently large (independent of $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$, $\theta_1 \in \mathbf{B}_n \cap R$ with $\|\theta_0 - \theta_1\|_{\mathbf{B}} \leq \delta_n$ or $h \in S_n(\theta_0, \theta_1)$) we have

$$\Upsilon_F(\theta_0 + \frac{\hat{h}}{\sqrt{n}}) = 0. \tag{S.254}$$

Moreover, from results (S.251) and (S.252) we also obtain that for n sufficiently large

$$\|h^\star(\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}})\|_{\mathbf{B}} \leq 16C^2K_0^2(\ell_n^2 + \eta_n^2) \leq \frac{\eta_n}{2} + 16C^2K_0^2\eta_n^2 \leq \frac{3}{4}\eta_n. \tag{S.255}$$

Thus, $h = h^{\mathcal{N}_{\theta_0}} + h^{\perp_{\theta_0}}$, (S.244), (S.252), (S.253) and (S.255) imply for large n that $\|(\hat{h} - h - h_{\theta_0,n})/\sqrt{n}\|_{\mathbf{B}} \leq \eta_n$, and employing (S.246) with $\tilde{h} = (\hat{h} - h_{\theta_0,n})/\sqrt{n})$ yields

$$\Upsilon_G(\theta_0 + \frac{\hat{h}}{\sqrt{n}}) \leq 0. \tag{S.256}$$

Since $\|h_{\theta_0,n}/\sqrt{n}\|_{\mathbf{B}} \leq C\eta_n$ by (S.246), results (S.244), (S.251), and $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ for any $h \in S_n(\theta_0, \theta_1)$ imply by (S.252) and $\ell_n \downarrow 0$, $\delta_n \downarrow 0$ that

$$\|\frac{\hat{h}}{\sqrt{n}}\|_{\mathbf{B}} \leq \|\frac{h_{\theta_0,n}}{\sqrt{n}}\|_{\mathbf{B}} + \|h^\star(\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}})\|_{\mathbf{B}} + \|\frac{h^{\perp_{\theta_0}}}{\sqrt{n}}\|_{\mathbf{B}} + \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}}$$
$$\leq C\eta_n + 16C^2K_0^2(\ell_n^2 + \eta_n^2) + M_f\ell_n(\delta_n + \ell_n) + \ell_n \leq 2\ell_n \tag{S.257}$$

for n sufficiently large. Therefore, we conclude from (S.254), (S.256), and (S.257) that

$\hat{h} \in T_n(\theta_0)$. Similarly, (S.244), (S.246), (S.251), and (S.252) yield for some $M < \infty$

$$\|\frac{\hat{h}}{\sqrt{n}} - \frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \|\frac{h_{\theta_0,n}}{\sqrt{n}}\|_{\mathbf{B}} + \|h^\star(\frac{h_{\theta_0,n}}{\sqrt{n}} + \frac{h^{\mathcal{N}_{\theta_0}}}{\sqrt{n}})\|_{\mathbf{B}} + \|\frac{h^{\perp_{\theta_0}}}{\sqrt{n}}\|_{\mathbf{B}}$$
$$\leq C\eta_n + 16C^2K_0^2(\ell_n^2 + \eta_n^2) + M_f\ell_n(\ell_n + \delta_n) \leq M\ell_n(\ell_n + \delta_n),$$

which establishes the (S.237) for the case $K_f > 0$. ■

Lemma S.5.1. *Let Assumptions 3.1(ii)(iii), 3.8 hold, and $\ell_n \downarrow 0$ be given. (i) Then, there are $n_0, M_g < \infty$ and $\epsilon > 0$ such that for all $n > n_0$, $P \in \mathbf{P}_0$, $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$, $\theta_1 \in (\Theta_{0n}^{\mathrm{r}})^\epsilon$:*

$$\{h \in \mathbf{B}_n : \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) \leq (\Upsilon_G(\theta_1) - K_g r\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}}\mathbf{1}_{\mathbf{G}}) \vee (-r\mathbf{1}_{\mathbf{G}}) \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\}$$
$$\subseteq \{h \in \mathbf{B}_n : \Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) \leq 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\}$$

for any $r \geq \{M_g\|\theta_0 - \theta_1\|_{\mathbf{B}} + K_g\|\theta_0 - \theta_1\|_{\mathbf{B}}^2\} \vee 2\{\ell_n + \|\theta_0 - \theta_1\|_{\mathbf{B}}\}1\{K_g > 0\}$. (ii) If in addition Υ_G is affine, then for any n, $\theta_0, \theta_1 \in \mathbf{B}_n$, and $r \geq M_g\|\theta_0 - \theta_1\|_{\mathbf{B}}$ we have

$$\{h \in \mathbf{B}_n : \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) \leq \Upsilon_G(\theta_1) \vee (-r\mathbf{1}_{\mathbf{G}})\} \subseteq \{h \in \mathbf{B}_n : \Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) \leq 0\}.$$

Proof: Let $\tilde{\epsilon} > 0$ be such that Assumption 3.8 holds and set $M_g < \infty$ to satisfy

$$\|\nabla\Upsilon_G(\theta)\|_o \leq M_g \tag{S.258}$$

for any $\theta \in (\Theta_{0n}^{\mathrm{r}})^{\tilde{\epsilon}}$, which is possible by Assumption 3.8(iii). Next, set $\epsilon = \tilde{\epsilon}/2$ and define $N(\delta) \equiv \{\theta \in \mathbf{B}_n : \overrightarrow{d}_H(\{\theta\}, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{B}}) < \delta\}$ for any $\delta > 0$. Then note that for any $\theta_1 \in N(\epsilon)$ and $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ we have $\theta_1 + h/\sqrt{n} \in N(\tilde{\epsilon})$ for n sufficiently large. Therefore, Assumption 3.8(i) allows us to conclude that

$$\|\Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) - \Upsilon_G(\theta_1) - \nabla\Upsilon_G(\theta_1)[\frac{h}{\sqrt{n}}]\|_{\mathbf{G}} \leq K_g\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}}^2. \tag{S.259}$$

Similarly, Assumption 3.8(ii) implies that if $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$ and $\theta_1 \in N(\epsilon)$, then we have

$$\|\nabla\Upsilon_G(\theta_0)[\frac{h}{\sqrt{n}}] - \nabla\Upsilon_G(\theta_1)[\frac{h}{\sqrt{n}}]\|_{\mathbf{G}}$$
$$\leq \|\nabla\Upsilon_G(\theta_0) - \nabla\Upsilon_G(\theta_1)\|_o\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq K_g\|\theta_0 - \theta_1\|_{\mathbf{B}}\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \tag{S.260}$$

for any $h \in \mathbf{B}_n$. Hence, since $\Upsilon_G(\theta_0) \le 0$ due to $\theta_0 \in \Theta_{0n}^{\mathrm{r}} \subseteq \Theta_n \cap R$ we obtain that

$$\begin{aligned}
\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) &+ \{\Upsilon_G(\theta_1) - \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}})\} \\
&\le \{\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) - \Upsilon_G(\theta_0)\} + \{\Upsilon_G(\theta_1) - \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}})\} \\
&\le K_g \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \{2\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} + \|\theta_0 - \theta_1\|_{\mathbf{B}}\}\mathbf{1_G}, \qquad (S.261)
\end{aligned}$$

by (S.259), (S.260), and Lemma S.5.3. Also note for any $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$, $\theta_1 \in N(\epsilon)$, and $h \in \mathbf{B}_n$ with $\|h/\sqrt{n}\|_{\mathbf{B}} \le \ell_n$ we have $\theta_0 + h/\sqrt{n} \in N(\tilde{\epsilon})$ and $\theta_1 + h/\sqrt{n} \in N(\tilde{\epsilon})$ for n sufficiently large. Thus, by Assumptions 3.8(i), result (S.258), and Lemma S.5.3

$$\begin{aligned}
\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) - \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) &\le \nabla\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}})[\theta_0 - \theta_1] + K_g\|\theta_0 - \theta_1\|_{\mathbf{B}}^2 \mathbf{1_G} \\
&\le \{M_g\|\theta_0 - \theta_1\|_{\mathbf{B}} + K_g\|\theta_0 - \theta_1\|_{\mathbf{B}}^2\}\mathbf{1_G}. \qquad (S.262)
\end{aligned}$$

Hence, (S.261) and (S.262) yield for $r \ge \{M_g\|\theta_0 - \theta_1\|_{\mathbf{B}} + K_g\|\theta_0 - \theta_1\|_{\mathbf{B}}^2\} \vee 2\{\ell_n + \|\theta_0 - \theta_1\|_{\mathbf{B}}\}1\{K_g > 0\}$, $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$, $\theta_1 \in N(\epsilon)$, $\|h/\sqrt{n}\|_{\mathbf{B}} \le \ell_n$, and n large

$$\begin{aligned}
\Upsilon_G(\theta_0 + \frac{h}{\sqrt{n}}) &\le \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) + (K_g r\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} - \Upsilon_G(\theta_1))\mathbf{1_G} \wedge r\mathbf{1_G} \\
&= \Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) - (\Upsilon_G(\theta_1) - K_g r\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}})\mathbf{1_G} \vee (-r\mathbf{1_G}) \qquad (S.263)
\end{aligned}$$

where the equality follows from $(-a)\vee(-b) = -(a\wedge b)$ by Theorem 8.6 in Aliprantis and Border (2006). Since $a_1 \le a_2$ and $b_1 \le b_2$ implies $a_1 \wedge b_1 \le a_2 \wedge b_2$ in $\mathbf{G}$ by Corollary 8.7 in Aliprantis and Border (2006), (S.263) implies that for n sufficiently large and any $\theta_0 \in \Theta_{0n}^{\mathrm{r}}$, $\theta_1 \in N(\epsilon)$ and $h \in \mathbf{B}_n$ satisfying $\|h/\sqrt{n}\|_{\mathbf{B}} \le \ell_n$ and

$$\Upsilon_G(\theta_1 + \frac{h}{\sqrt{n}}) \le (\Upsilon_G(\theta_1) - K_g r\|\frac{h}{\sqrt{n}}\|_{\mathbf{B}}\mathbf{1_G}) \vee (-r\mathbf{1_G})$$

we must have $\Upsilon_G(\theta_0 + h/\sqrt{n}) \le 0$, which verifies the first claim of the lemma. For the second claim, just note that if Υ_G is affine, then we may set $K_g = 0$ and $\epsilon = +\infty$ in Assumption 3.8, which leads to the desired simplification. ■

Lemma S.5.2. *If Assumptions 3.1(ii)(iii), 3.8, 3.10(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 $n > n_0$, $P \in \mathbf{P}_0$, and $\theta \in \Theta_{0n}^{\mathrm{r}}$ there is $h_{\theta,n} \in \mathbf{B}_n \cap \mathcal{N}(\nabla\Upsilon_F(\theta))$ with*

$$\Upsilon_G(\theta + \frac{h_{\theta,n}}{\sqrt{n}} + \frac{\tilde{h}}{\sqrt{n}}) \le 0 \qquad \|\frac{h_{\theta,n}}{\sqrt{n}}\|_{\mathbf{B}} \le C\eta_n \qquad (S.264)$$

for all $\tilde{h} \in \mathbf{B}_n$ for which there is a $h \in \mathbf{B}_n$ satisfying $\|(\tilde{h} - h)/\sqrt{n}\|_{\mathbf{B}} \le \eta_n$, $\|h/\sqrt{n}\|_{\mathbf{B}} \le \ell_n$, and the inequality $\Upsilon_G(\theta + h/\sqrt{n}) \le 0$.

PROOF: By Assumption 3.10(ii) there are $\epsilon > 0$ and $M_d < \infty$ such that for every $P \in \mathbf{P}_0$, n, and $\theta \in \Theta^{\mathrm{r}}_{0n}$ there exists a $\bar h_{\theta,n} \in \mathbf{B}_n \cap \mathcal{N}(\nabla \Upsilon_F(\theta))$ satisfying

$$\Upsilon_G(\theta) + \nabla\Upsilon_G(\theta)[\bar h_{\theta,n}] \leq -\epsilon \mathbf{1_G} \qquad \|\bar h_{\theta,n}\|_{\mathbf{B}} \leq M_d. \tag{S.265}$$

Also note Assumption 3.8(iii) and $\ell_n = o(1)$ imply that there is an $M_g < \infty$ such that for n sufficiently large and any $h \in \mathbf{B}_n$ satisfying $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ we must have

$$\|\nabla\Upsilon_G(\theta + \frac{h}{\sqrt{n}})\|_o \leq M_g. \tag{S.266}$$

Moreover, result (S.266), Assumption 3.8(i), Lemma S.5.3, and $\ell_n = o(1)$ imply that for n sufficiently large and any $h \in \mathbf{B}_n$ with $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ we must have

$$\begin{aligned}\Upsilon_G(\theta + \frac{h}{\sqrt{n}}) &\leq \Upsilon_G(\theta) + \nabla\Upsilon_G(\theta)[\frac{h}{\sqrt{n}}] + K_g\|\frac{h}{\sqrt{n}}\|^2_{\mathbf{B}}\mathbf{1_G} \\ &\leq \Upsilon_G(\theta) + \{\|\nabla\Upsilon_G(\theta)\|_o\ell_n + K_g\ell_n^2\}\mathbf{1_G} \leq \Upsilon_G(\theta) + 2M_g\ell_n\mathbf{1_G}. \end{aligned}\tag{S.267}$$

Hence, (S.265) and (S.267) imply for any $h \in \mathbf{B}_n$ with $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ we must have

$$\Upsilon_G(\theta + \frac{h}{\sqrt{n}}) + \nabla\Upsilon_G(\theta)[\bar h_{\theta,n}] \leq \{2M_g\ell_n - \epsilon\}\mathbf{1_G}. \tag{S.268}$$

Next, we let $C_0 > 8M_g/\epsilon$ and aim to show (S.264) holds with $C = C_0 M_d$ by setting

$$\frac{h_{\theta,n}}{\sqrt{n}} \equiv C_0\eta_n \bar h_{\theta,n}. \tag{S.269}$$

To this end, we first note that if $\theta \in \Theta^{\mathrm{r}}_{0n}$, $h \in \mathbf{B}_n$ satisfies $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$ and $\Upsilon_G(\theta + h/\sqrt{n}) \leq 0$, and $\tilde h \in \mathbf{B}_n$ is such that $\|(h - \tilde h)/\sqrt{n}\|_{\mathbf{B}} \leq \eta_n$, then definition (S.269) implies that $\|(h_{\theta,n} + \tilde h)/\sqrt{n}\|_{\mathbf{B}} = o(1)$. Therefore, Assumption 3.8(i), Lemma S.5.3, and $\|(\tilde h - h)/\sqrt{n}\|_{\mathbf{B}} \leq \eta_n$ together allow us to conclude that

$$\begin{aligned}&\Upsilon_G(\theta + \frac{h_{\theta,n}}{\sqrt{n}} + \frac{\tilde h}{\sqrt{n}}) \\ &\quad\leq \Upsilon_G(\theta + \frac{h}{\sqrt{n}}) + \nabla\Upsilon_G(\theta + \frac{h}{\sqrt{n}})[\frac{h_{\theta,n}}{\sqrt{n}} + \frac{(\tilde h - h)}{\sqrt{n}}] + 2K_g(\|\frac{h_{\theta,n}}{\sqrt{n}}\|^2_{\mathbf{B}} + \eta_n^2)\mathbf{1_G} \\ &\quad\leq \Upsilon_G(\theta + \frac{h}{\sqrt{n}}) + \nabla\Upsilon_G(\theta + \frac{h}{\sqrt{n}})[\frac{h_{\theta,n}}{\sqrt{n}}] + \{2K_g\|\frac{h_{\theta,n}}{\sqrt{n}}\|^2_{\mathbf{B}} + 2M_g\eta_n\}\mathbf{1_G}, \end{aligned}\tag{S.270}$$

where the final result follows from result (S.266) and $2K_g\eta_n^2 \leq M_g\eta_n$ for n sufficiently large. Similarly, Assumption 3.8(ii) and Lemma S.5.3 yield

$$\nabla\Upsilon_G(\theta+\frac{h}{\sqrt{n}})[\frac{h_{\theta,n}}{\sqrt{n}}] \le \nabla\Upsilon_G(\theta)[\frac{h_{\theta,n}}{\sqrt{n}}] + \|\nabla\Upsilon_G(\theta+\frac{h}{\sqrt{n}}) - \nabla\Upsilon_G(\theta)\|_o\|\frac{h_{\theta,n}}{\sqrt{n}}\|_{\mathbf{B}}\mathbf{1_G}$$
$$\le \nabla\Upsilon_G(\theta)[\frac{h_{\theta,n}}{\sqrt{n}}] + K_g\ell_n\|\frac{h_{\theta,n}}{\sqrt{n}}\|_{\mathbf{B}}\mathbf{1_G}. \tag{S.271}$$

Hence, results (S.270) and (S.271), $\|h_{\theta,n}/\sqrt{n}\|_{\mathbf{B}} \le C_0 M_d \eta_n$ due to $\|\bar{h}_{\theta,n}\|_{\mathbf{B}} \le M_d$ by (S.265), and $\eta_n \downarrow 0$, $\ell_n \downarrow 0$, imply that for n sufficiently large we have

$$\Upsilon_G(\theta+\frac{h_{\theta,n}}{\sqrt{n}}+\frac{\tilde{h}}{\sqrt{n}}) \le \Upsilon_G(\theta+\frac{h}{\sqrt{n}}) + \nabla\Upsilon_G(\theta)[\frac{h_{\theta,n}}{\sqrt{n}}] + 4M_g\eta_n\mathbf{1_G}. \tag{S.272}$$

In addition, since $C_0\eta_n \downarrow 0$, we have $C_0\eta_n \le 1$ eventually, and hence $\Upsilon_G(\theta+h/\sqrt{n}) \le 0$, $2M_g\ell_n \le \epsilon/2$ for n sufficiently large due to $\ell_n \downarrow 0$, and result (S.268) imply that

$$\Upsilon_G(\theta+\frac{h}{\sqrt{n}}) + C_0\eta_n\nabla\Upsilon_G(\theta)[\bar{h}_{\theta,n}] \le C_0\eta_n\{\Upsilon_G(\theta+\frac{h}{\sqrt{n}}) + \nabla\Upsilon_G(\theta)[\bar{h}_{\theta,n}]\}$$
$$\le C_0\eta_n\{2M_g\ell_n - \epsilon\}\mathbf{1_G} \le -\frac{C_0\eta_n\epsilon}{2}\mathbf{1_G}. \tag{S.273}$$

Thus, we can conclude from results (S.269), (S.272), (S.273), and $C_0 > 8M_g/\epsilon$ that

$$\Upsilon_G(\theta+\frac{h_{\theta,n}}{\sqrt{n}}+\frac{\tilde{h}}{\sqrt{n}}) \le \{4M_g - \frac{C_0\epsilon}{2}\}\eta_n\mathbf{1_G} \le 0,$$

for n sufficiently large, which establishes the claim of the Lemma. ■

Lemma S.5.3. *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 $a_1 \le a_2 + C\mathbf{1_A}$ for any $a_1, a_2 \in \mathbf{A}$ satisfying $\|a_1 - a_2\|_{\mathbf{A}} \le C$.*

PROOF: Since $\mathbf{A}$ is an AM space with unit $\mathbf{1_A}$ we have that $\|a_1 - a_2\|_{\mathbf{A}} \le C$ implies $|a_1 - a_2| \le C\mathbf{1_A}$, and hence the claim follows trivially from $a_1 - a_2 \le |a_1 - a_2|$. ■

Lemma S.5.4. *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 $a \in \mathcal{B}_{\epsilon_0}(a_0) \equiv \{a \in \mathbf{A} : \|a - a_0\|_{\mathbf{A}} \le \epsilon_0\}$.

(ii) $\|F(a+h) - F(a) - \nabla F(a)[h]\|_{\mathbf{C}} \le K_0\|h\|^2_{\mathbf{A}}$ for all $a, a+h \in \mathcal{B}_{\epsilon_0}(a_0)$.

(iii) $\|\nabla F(a_1) - \nabla F(a_2)\|_o \le K_0\|a_1 - a_2\|_{\mathbf{A}}$ for all $a_1, a_2 \in \mathcal{B}_{\epsilon_0}(a_0)$.

(iv) $\nabla F(a_0) : \mathbf{A} \to \mathbf{C}$ has $\|\nabla F(a_0)\|_o \le K_0$.

(v) $\nabla F(a_0) : \mathbf{A}_2 \to \mathbf{C}$ is bijective and $\|\nabla F(a_0)^{-1}\|_o \le K_0$.

Then, for all $h_1 \in \mathbf{A}_1$ with $\|h_1\|_{\mathbf{A}} \le (\epsilon_0/2 \wedge (4K_0^2)^{-1} \wedge 1)^2$ 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 $\|h_2^\star(h_1)\|_{\mathbf{A}} \le 4K_0^2\|h_1\|_{\mathbf{A}}$ for arbitrary $\mathbf{A}_1$, and $\|h_2^\star(h_1)\|_{\mathbf{A}} \le 2K_0^2\|h_1\|^2_{\mathbf{A}}$ 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

$$g(h_1, h_2) \equiv \nabla F(a_0)[h_2] - F(a_0 + h_1 + h_2). \tag{S.274}$$

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

$$h_2 = \nabla F(a_0)^{-1}[g(h_1, h_2)]. \tag{S.275}$$

Letting $T_{h_1} : \mathbf{A}_2 \to \mathbf{A}_2$ be given by $T_{h_1}(h_2) \equiv \nabla F(a_0)^{-1}[g(h_1, h_2)]$, we see from (S.275) that the desired $h_2^\star(h_1)$ must be a fixed point of T_{h_1}. Next, define the set

$$M_0 \equiv \{h_2 \in \mathbf{A}_2 : \|h_2\|_{\mathbf{A}} \leq \delta_0\}$$

for $\delta_0 \equiv (\epsilon_0/2) \wedge (4K_0^2)^{-1} \wedge 1$, and consider an arbitrary $h_1 \in \mathbf{A}_1$ with $\|h_1\|_{\mathbf{A}} \leq \delta_0^2$. 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

$$\begin{aligned} \|g(h_1, h_2) - g(h_1, \tilde{h}_2)\|_{\mathbf{C}} &\leq \sup_{0<\tau<1} \|\nabla_2 g(h_1, h_2 + \tau(\tilde{h}_2 - h_2))\|_o \|h_2 - \tilde{h}_2\|_{\mathbf{A}} \\ &\leq \frac{1}{2K_0} \|h_2 - \tilde{h}_2\|_{\mathbf{A}}, \end{aligned} \tag{S.276}$$

where the final inequality follows by Condition (iii) and $\delta_0^2 \leq \delta_0 \leq (4K_0^2)^{-1}$. Moreover,

$$\begin{aligned} &\|\nabla F(a_0)[h_2] - \nabla F(a_0 + h_1)[h_2]\|_{\mathbf{C}} \\ &\qquad \leq \|\nabla F(a_0) - \nabla F(a_0 + h_1)\|_o \|h_2\|_{\mathbf{A}} \leq K_0 \|h_1\|_{\mathbf{A}} \|h_2\|_{\mathbf{A}} \leq \frac{\|h_2\|_{\mathbf{A}}}{4K_0} \end{aligned} \tag{S.277}$$

by Condition (iii) and $\|h_1\|_{\mathbf{A}} \leq \delta_0 \leq (4K_0^2)^{-1}$. Similarly, for any $h_2 \in M_0$ we have

$$\|F(a_0 + h_1 + h_2) - F(a_0 + h_1) - \nabla F(a_0 + h_1)[h_2]\|_{\mathbf{C}} \leq K_0 \|h_2\|_{\mathbf{A}}^2 \leq \frac{\|h_2\|_{\mathbf{A}}}{4K_0} \tag{S.278}$$

due to $a_0 + h_1 \in \mathcal{B}_{\epsilon_0}(a_0)$ and Condition (ii). Moreover, since $F(a_0) = 0$ by hypothesis, Conditions (ii) and (iv), $\|h_1\|_{\mathbf{A}} \leq \delta_0^2$, and $\delta_0 \leq (4K_0^2)^{-1}$ yield that

$$\|F(a_0+h_1)\|_{\mathbf{C}} = \|F(a_0+h_1) - F(a_0)\|_{\mathbf{C}} \leq K_0 \|h_1\|_{\mathbf{A}}^2 + \|\nabla F(a_0)\|_o \|h_1\|_{\mathbf{A}} \leq \frac{\delta_0}{2K_0}. \tag{S.279}$$

Hence, by (S.274) and (S.277)-(S.279) we obtain for any $h_2 \in M_0$ and h_1 with $\|h_1\|_{\mathbf{A}} \leq \delta_0^2$

$$\|g(h_1, h_2)\|_{\mathbf{C}} \leq \frac{\|h_2\|_{\mathbf{A}}}{2K_0} + \frac{\delta_0}{2K_0} \leq \frac{\delta_0}{K_0}. \tag{S.280}$$

Thus, since $\|\nabla F(a_0)^{-1}\|_o \leq K_0$ by Condition (v), result (S.280) implies $T_{h_1} : M_0 \to M_0$, and (S.276) yields $\|T_{h_1}(h_2) - T_{h_1}(\tilde{h}_2)\|_{\mathbf{A}} \leq 2^{-1}\|h_2 - \tilde{h}_2\|_{\mathbf{A}}$ 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 (S.274) and (S.275).

Next, we note that since $h_2^\star(h_1)$ is a fixed point of T_{h_1}, we can conclude that

$$\|h_2^\star(h_1)\|_{\mathbf{A}} = \|T_{h_1}(h_2^\star(h_1))\|_{\mathbf{A}} \leq \|T_{h_1}(h_2^\star(h_1)) - T_{h_1}(0)\|_{\mathbf{A}} + \|T_{h_1}(0)\|_{\mathbf{A}}. \tag{S.281}$$

Thus, since $\|T_{h_1}(h_2^\star(h_1)) - T_{h_1}(0)\|_{\mathbf{A}} \leq 2^{-1}\|h_2^\star(h_1)\|_{\mathbf{A}}$ by (S.276) and $\|\nabla F(a_0)^{-1}\|_o \leq K_0$, it follows from result (S.281) and $T_{h_1}(0) \equiv -\nabla F(a_0)^{-1}F(a_0 + h_1)$ that

$$\begin{aligned}\frac{1}{2}\|h_2^\star(h_1)\|_{\mathbf{A}} &\leq \|T_{h_1}(0)\|_{\mathbf{A}} \leq K_0\|F(a_0 + h_1)\|_{\mathbf{C}} \\ &\leq K_0\{K_0\|h_1\|_{\mathbf{A}}^2 + \|\nabla F(a_0)\|_o\|h_1\|_{\mathbf{A}}\} \leq 2K_0^2\|h_1\|_{\mathbf{A}},\end{aligned} \tag{S.282}$$

where in the second inequality we employed $\|\nabla F(a_0)^{-1}\|_o \leq K_0$, in the third inequality we used (S.279), and in the final inequality we exploited $\|h_1\|_{\mathbf{A}} \leq 1$. While the estimate in (S.282) applies for generic $\mathbf{A}_1$, we note that if in addition $\mathbf{A}_1 = \mathcal{N}(\nabla F(a_0))$, then

$$\frac{1}{2}\|h_2^\star(h_1)\|_{\mathbf{A}} \leq \|T_{h_1}(0)\|_{\mathbf{A}} \leq K_0\|F(a_0 + h_1)\|_{\mathbf{C}} \leq K_0^2\|h_1\|_{\mathbf{A}}^2 ,$$

due to $F(a_0) = 0$ and $\nabla F(a_0)[h_1] = 0$, and thus the final claim of the lemma follows. ■

S.6 Coupling via Koltchinskii (1994)

In this section we develop uniform coupling results for empirical processes that help verify Assumption 3.3(i) in specific applications. Our analysis is based on the Hungarian construction of Massart (1989) and Koltchinskii (1994), and we state the results in a notation that abstracts from the rest of the paper due to their potential independent interest. Thus, in what follows we consider $V \in \mathbf{R}^d$ to be a generic random variable distributed according to $P \in \mathbf{P}$, denote its support under P by $\Omega(P) \subset \mathbf{R}^d$, and let λ denote the Lebesgue measure on $\mathbf{R}^d$. For any function f we further set

$$\mathbb{G}_n(f) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n(f(V_i) - E_P[f(V)]).$$

The rates obtained through a Hungarian construction crucially depend on the ability of the functions indexing 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. For a measure P and a function $f : \mathbf{R}^d \to \mathbf{R}$, the integral modulus of continuity of f is the function $\varpi(f,\cdot,P):\mathbf{R}_+\to\mathbf{R}_+$ defined for every $h\in\mathbf{R}_+$ as

$$\varpi(f,h,P)\equiv\sup_{\|s\|\le h}\Big(\int_{\Omega(P)}(f(v+s)-f(v))^2 1\{v+s\in\Omega(P)\}dP(v)\Big)^{\frac{1}{2}}. \tag{S.283}$$

Intuitively, the integral modulus of continuity quantifies the "smoothness" of a function f by examining the difference between f and its own translation. For example, it is straightforward to verify that $\varpi(f,h,P)\lesssim h$ whenever f is Lipschitz. In contrast indicator functions such as $f(v)=1\{v\le t\}$ typically satisfy $\varpi(f,h,P)\lesssim h^{1/2}$.

We impose the following assumptions to establish the uniform coupling results.

Assumption S.6.1. *(i) For all $P\in\mathbf{P}$, $P\ll\lambda$ and $\Omega(P)\subset\mathbf{R}^d$ 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 S.6.2. *(i) For each $P\in\mathbf{P}$ there is a continuously differentiable bijection $T_P:[0,1]^d\to\Omega(P)$; (ii) The Jacobian JT_P and its determinant $|JT_P|$ satisfy $\inf_{P\in\mathbf{P}}\inf_{v\in[0,1]^d}|JT_P(v)|>0$ and $\sup_{P\in\mathbf{P}}\sup_{v\in[0,1]^d}\|JT_P(v)\|_o<\infty$.*

Assumption S.6.3. *The classes $\mathcal{F}_n$ satisfy: (i) $\sup_{P\in\mathbf{P}}\sup_{f\in\mathcal{F}_n}\varpi(f,h,P)\le\varphi_n(h)$ for some $\varphi_n:\mathbf{R}_+\to\mathbf{R}_+$ satisfying $\varphi_n(Ch)\le C^\kappa\varphi_n(h)$ for all n, $C\ge 1$, and some $\kappa>0$; and (ii) $\sup_{f\in\mathcal{F}_n}\|f\|_\infty\le K_n$ for some $K_n>0$.*

In Assumption S.6.1 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 S.6.2 requires that the support of V under each P be "smooth" in the sense that it be a differentiable transformation of the unit square. Together, Assumptions S.6.1 and S.6.2 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 S.6.1. As a result, the approximation error by the Haar basis implied by each partition can be controlled uniformly by the integral modulus of continuity; see Lemma S.6.2. Together with Assumption S.6.3, which imposes conditions on the integral modulus of continuity of $\mathcal{F}_n$ uniformly in P, we can obtain a uniform coupling result through the analysis in Koltchinskii (1994). We note that the homogeneity condition on φ_n in Assumption S.6.3(i) is not necessary, but we impose it to simplify the bound.

The next theorem establishes a coupling for the empirical process $\mathbb{G}_n$.

Theorem S.6.1. *Let Assumptions S.6.1-S.6.3 hold, $\{V_i\}_{i=1}^n$ be i.i.d. with $V_i\sim P\in\mathbf{P}$ and for any $\delta_n\downarrow 0$ let $N_n\equiv\sup_{P\in\mathbf{P}}N_{[]}(\delta_n,\mathcal{F}_n,\|\cdot\|_{P,2})$, $J_n\equiv\sup_{P\in\mathbf{P}}J_{[]}(\delta_n,\mathcal{F}_n,\|\cdot\|_{P,2})$,*

$$S_n\equiv\Big(\sum_{i=0}^{\lceil\log_2 n\rceil}2^i\varphi_n^2(2^{-\frac{i}{d}})\Big)^{\frac{1}{2}}. \tag{S.284}$$

If $N_n \uparrow \infty$, there is a Gaussian $\mathbb{G}_P$ (possible depending on n) so that uniformly in $P \in \mathbf{P}$

$$\|\mathbb{G}_n - \mathbb{G}_P\|_{\mathcal{F}_n} = O_P(\frac{K_n \log(nN_n)}{\sqrt{n}} + \frac{K_n\sqrt{\log(nN_n)\log(n)}S_n}{\sqrt{n}} + J_n(1+\frac{J_nK_n}{\delta_n^2\sqrt{n}})). \quad \text{(S.285)}$$

Theorem S.6.1 is a mild modification of the results in Koltchinskii (1994). The proof of Theorem S.6.1 relies on a coupling of the empirical process on a sequence of grids of cardinality N_n, and employs the equicontinuity of $\mathbb{G}_n$ and $\mathbb{G}_P$ to obtain a coupling on the entire class $\mathcal{F}_n$. The conclusion of Theorem S.6.1 applies to any choice of grid accuracy δ_n. In order to obtain the best rate, δ_n must be chosen to balance the terms in (S.285) and thus depends on the metric entropy of $\mathcal{F}_n$ through the terms N_n and J_n.

Below, we include the proof of Theorem S.6.1 and auxiliary results.

Proof of Theorem S.6.1: Let $\{\Delta_i(P)\}$ be the partitions of $\Omega(P)$ in Lemma S.6.1 and $\mathcal{B}_{P,i}$ the σ-algebra generated by $\Delta_i(P)$. By Lemma S.6.2 and Assumption S.6.3,

$$\sup_{P\in\mathbf{P}}\sup_{f\in\mathcal{F}_n}(\sum_{i=0}^{\lceil\log_2 n\rceil} 2^i E_P[(f(V)-E_P[f(V)|\mathcal{B}_{P,i}])^2])^{\frac{1}{2}}$$
$$\leq C_1(\sum_{i=0}^{\lceil\log_2 n\rceil} 2^i\varphi_n^2(2^{-\frac{i}{d}}))^{\frac{1}{2}} \equiv C_1S_n \quad \text{(S.286)}$$

for some constant $C_1 > 0$ and for S_n as defined in (S.284). 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 $\|\cdot\|_{P,2}$. Since $N(\epsilon,\mathcal{F}_n,\|\cdot\|_{P,2}) \leq N_{[]}(\epsilon,\mathcal{F}_n,\|\cdot\|_{P,2})$, it follows from the definition of N_n that we may choose $\mathcal{F}_{P,n,\delta_n}$ so that

$$\sup_{P\in\mathbf{P}} \operatorname{card}(\mathcal{F}_{P,n,\delta_n}) \leq \sup_{P\in\mathbf{P}} N_{[]}(\delta_n,\mathcal{F}_n,\|\cdot\|_{P,2}) \equiv N_n. \quad \text{(S.287)}$$

By Theorem 3.5 in Koltchinskii (1994), (S.286) and (S.287), it follows that for each $n \geq 1$ there exists an isonormal process $\mathbb{G}_P$, such that for all $\eta_1 > 0$, $\eta_2 > 0$

$$\sup_{P\in\mathbf{P}} P(\frac{\sqrt{n}}{K_n}\|\mathbb{G}_n - \mathbb{G}_P\|_{\mathcal{F}_{P,n,\delta_n}} \geq \eta_1 + \sqrt{\eta_1}\sqrt{\eta_2}(C_1S_n+1))$$
$$\lesssim N_n\exp\{-C_2\eta_1\} + n\exp\{-C_2\eta_2\}, \quad \text{(S.288)}$$

for some $C_2 > 0$. Since $N_n \uparrow \infty$, (S.288) 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

$$\sup_{P\in\mathbf{P}} P(\|\mathbb{G}_n - \mathbb{G}_P\|_{\mathcal{F}_{P,n,\delta_n}} \geq C_4K_n \times \frac{\log(nN_n) + \sqrt{\log(N_n)\log(n)}S_n}{\sqrt{n}}) < \varepsilon. \quad \text{(S.289)}$$

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 $\sup_{P\in\mathbf{P}}\sup_{f\in\mathcal{F}_n}\|f-\Gamma_{n,P}(f)\|_{P,2} \leq \delta_n$. Let $D(\epsilon,\mathcal{F}_n,\|\cdot\|_{P,2})$ denote the ϵ-packing number for $\mathcal{F}_n$ under $\|\cdot\|_{P,2}$, and note $D(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2}) \leq N_{[\,]}(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2})$. Therefore, by Corollary 2.2.8 in van der Vaart and Wellner (1996) we can conclude that

$$\sup_{P\in\mathbf{P}} E_P[\|\mathbb{G}_P - \mathbb{G}_P \circ \Gamma_{n,P}\|_{\mathcal{F}_n}] \lesssim \sup_{P\in\mathbf{P}} \int_0^{\delta_n} \sqrt{\log D(\epsilon, \mathcal{F}_n, \|\cdot\|_{P,2})} d\epsilon \leq \sup_{P\in\mathbf{P}} J_{[\,]}(\delta_n, \mathcal{F}_n, \|\cdot\|_{P,2}) \equiv J_n. \quad \text{(S.290)}$$

Similarly, employing Lemma 3.4.2 in van der Vaart and Wellner (1996) yields that

$$\sup_{P\in\mathbf{P}} E_P[\|\mathbb{G}_n - \mathbb{G}_n \circ \Gamma_{n,P}\|_{\mathcal{F}_n}] \lesssim \sup_{P\in\mathbf{P}} J_{[\,]}(\delta_n, \mathcal{F}_n, \|\cdot\|_{P,2})(1 + \sup_{P\in\mathbf{P}} \frac{J_{[\,]}(\delta_n, \mathcal{F}_n, \|\cdot\|_{P,2})K_n}{\delta_n^2\sqrt{n}}) \equiv J_n(1 + \frac{J_n K_n}{\delta_n^2\sqrt{n}}). \quad \text{(S.291)}$$

Therefore, combining (S.289), (S.290), and (S.291) together with the decomposition

$$\|\mathbb{G}_n - \mathbb{G}_P\|_{\mathcal{F}_n} \leq \|\mathbb{G}_n - \mathbb{G}_P\|_{\mathcal{F}_{P,n,\delta_n}} + \|\mathbb{G}_n - \mathbb{G}_n \circ \Gamma_{n,P}\|_{\mathcal{F}_n} + \|\mathbb{G}_P - \mathbb{G}_P \circ \Gamma_{n,P}\|_{\mathcal{F}_n},$$

establishes the claim of the theorem by Markov's inequality. ■

Lemma S.6.1. *Let $\mathcal{B}_P$ denote the completion of the Borel $\sigma-$algebra on $\Omega(P)$ with respect to P. If Assumptions S.6.1 and S.6.2 hold, then for each $P \in \mathbf{P}$ there exists a sequence $\{\Delta_i(P)\}$ of partitions of $(\Omega(P), \mathcal{B}_P, P)$ 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* $i \geq 0$.

(iii) $P(\Delta_{i+1,2k}(P)) = P(\Delta_{i+1,2k+1}(P)) = 2^{-i-1}$ *for* $k = 0, \ldots 2^i - 1$, $i \geq 0$.

(iv) $\sup_{P\in\mathbf{P}} \max_{0\leq k\leq 2^i-1} \sup_{v,v'\in\Delta_{i,k}(P)} \|v - v'\|_2 = O(2^{-\frac{i}{d}})$.

(v) $\mathcal{B}_P$ *equals the completion with respect to* P *of the* σ*-algebra generated by* $\bigcup_i \Delta_i(P)$.

Proof: Let $\mathcal{A}$ denote the Borel σ-algebra on $[0,1]^d$, and for any $A \in \mathcal{A}$ define

$$Q_P(A) \equiv P(T_P(A)), \quad \text{(S.292)}$$

where $T_P(A) \in \mathcal{B}_P$ due to T_P^{-1} being measurable. Moreover, $Q_P([0,1]^d) = 1$ 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 (S.292) is a probability measure on $([0,1]^d, \mathcal{A})$. In addition, for λ the Lebesgue measure, we obtain from Theorem 3.7.1 in Bogachev (2007) that

$$Q_P(A) = P(T_P(A)) = \int_{T_P(A)} \frac{dP}{d\lambda}(v) d\lambda(v) = \int_A \frac{dP}{d\lambda}(T_P(a))|JT_P(a)|d\lambda(a), \quad \text{(S.293)}$$

where $|JT_P(a)|$ denotes the Jacobian of T_P at any point $a \in [0,1]^d$. 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

$a \in [0,1]^d$. Next, let $a = (a_1, \ldots, a_d)' \in [0,1]^d$ and define for any $t \in [0,1]$

$$G_{l,P}(t|A) \equiv \frac{Q_P(a \in A : a_l \leq t)}{Q_P(A)}, \tag{S.294}$$

for any $A \in \mathcal{A}$ with $Q_P(A) > 0$ and $1 \leq l \leq d$. Also let $\mathrm{m}(i) \equiv i - \lfloor \frac{i-1}{d} \rfloor \times d$ (i.e. $\mathrm{m}(i)$ equals i modulo d), set $\tilde{\Delta}_{0,0}(P) = [0,1]^d$, and inductively define the partitions of $[0,1]^d$

$$\begin{aligned}
\tilde{\Delta}_{i+1,2k}(P) &\equiv \{a \in \tilde{\Delta}_{i,k}(P) : G_{\mathrm{m}(i+1),P}(a_{\mathrm{m}(i+1)}|\tilde{\Delta}_{i,k}(P)) \leq \frac{1}{2}\} \\
\tilde{\Delta}_{i+1,2k+1}(P) &\equiv \tilde{\Delta}_{i,k}(P) \setminus \tilde{\Delta}_{i+1,2k}(P)
\end{aligned} \tag{S.295}$$

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 $[0,1]^d$ – i.e. it is of the general form

$$\mathrm{cl}(\tilde{\Delta}_{i,k}(P)) = \prod_{j=1}^{d} [l_{i,k,j}(P), u_{i,k,j}(P)].$$

Moreover, since g_P is positive on $[0,1]^d$ by Assumptions S.6.1(ii) and S.6.2(ii), it follows that for any $i \geq 0$, $0 \leq k \leq 2^i - 1$ and $1 \leq j \leq d$, we have

$$\begin{aligned}
l_{i+1,2k,j}(P) &= l_{i,k,j}(P) \\
u_{i+1,2k,j}(P) &= \begin{cases} u_{i,k,j}(P) \text{ if } j \neq \mathrm{m}(i+1) \\ \text{solves } G_{\mathrm{m}(i+1),P}(u_{i+1,2k,j}(P)|\tilde{\Delta}_{i,k}(P)) = \frac{1}{2} \text{ if } j = \mathrm{m}(i+1) \end{cases}
\end{aligned} \tag{S.296}$$

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

$$u_{i+1,2k+1,j}(P) = u_{i,k,j}(P) \qquad l_{i+1,2k+1,j}(P) = \begin{cases} l_{i,k,j}(P) \text{ if } j \neq \mathrm{m}(i+1) \\ u_{i+1,2k,j}(P) \text{ if } j = \mathrm{m}(i+1) \end{cases} \tag{S.297}$$

Since $Q_P(\mathrm{cl}(\tilde{\Delta}_{i+1,2k}(P))) = Q_P(\tilde{\Delta}_{i+1,2k}(P))$ by $Q_P \ll \lambda$, (S.294) and (S.296) yield

$$\begin{aligned}
Q_P(\tilde{\Delta}_{i+1,2k}(P)) &= Q_P(a \in \tilde{\Delta}_{i,k}(P) : a_{\mathrm{m}(i+1)} \leq u_{i+1,2k,\mathrm{m}(i+1)}(P)) \\
&= G_{\mathrm{m}(i+1),P}(u_{i+1,2k,\mathrm{m}(i+1)}(P)|\tilde{\Delta}_{i,k}(P))Q_P(\tilde{\Delta}_{i,k}(P)) \\
&= \frac{1}{2} Q_P(\tilde{\Delta}_{i,k}(P)).
\end{aligned}$$

Therefore, since $\tilde{\Delta}_{i,k}(P) = \tilde{\Delta}_{i+1,2k}(P) \cup \tilde{\Delta}_{i+1,2k+1}(P)$, it follows $Q_P(\tilde{\Delta}_{i+1,2k+1}(P)) = \frac{1}{2} Q_P(\tilde{\Delta}_{i,k}(P))$ for $0 \leq k \leq 2^i - 1$ as well. In particular, $Q_P(\tilde{\Delta}_{0,0}(P)) = 1$ implies that

$$Q_P(\tilde{\Delta}_{i,k}(P)) = \frac{1}{2^i} \tag{S.298}$$

for any integers $i \geq 1$ and $0 \leq k \leq 2^i - 1$. Moreover, we note that result (S.293) and

Assumptions S.6.1(ii) and S.6.2(ii) together imply that the density g_P of Q_P satisfies

$$0 < \inf_{P\in\mathbf{P}} \inf_{a\in[0,1]^d} g_P(a) < \sup_{P\in\mathbf{P}} \sup_{a\in[0,1]^d} g_P(a) < \infty, \tag{S.299}$$

and therefore $Q_P(A) \asymp \lambda(A)$ uniformly in $A \in \mathcal{A}$ and $P \in \mathbf{P}$. Hence, since by (S.296) $u_{i+1,2k,j}(P) = u_{i,k,j}(P)$ and $l_{i+1,2k,j}(P) = l_{i,k,j}(P)$ for all $j \neq \mathrm{m}(i+1)$, we obtain

$$\begin{aligned}\frac{(u_{i+1,2k,\mathrm{m}(i+1)}(P) - l_{i+1,2k,\mathrm{m}(i+1)}(P))}{(u_{i,k,\mathrm{m}(i+1)}(P) - l_{i,k,\mathrm{m}(i+1)}(P))} &= \frac{\prod_{j=1}^d (u_{i+1,2k,j}(P) - l_{i+1,2k,j}(P))}{\prod_{j=1}^d (u_{i,k,j}(P) - l_{i,k,j}(P))} \\ &= \frac{\lambda(\tilde{\Delta}_{i+1,2k}(P))}{\lambda(\tilde{\Delta}_{i,k}(P))} \asymp \frac{Q_P(\tilde{\Delta}_{i+1,2k}(P))}{Q_P(\tilde{\Delta}_{i,k}(P))} = \frac{1}{2} \end{aligned} \tag{S.300}$$

uniformly in $P \in \mathbf{P}$, $i \geq 0$, and $0 \leq k \leq 2^i - 1$ by results (S.298) and (S.299). Moreover, by identical arguments but using (S.297) instead of (S.296) we conclude

$$\frac{(u_{i+1,2k+1,\mathrm{m}(i+1)}(P) - l_{i+1,2k+1,\mathrm{m}(i+1)}(P))}{(u_{i,k,\mathrm{m}(i+1)}(P) - l_{i,k,\mathrm{m}(i+1)}(P))} \asymp \frac{1}{2} \tag{S.301}$$

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 $1 \leq j \leq d$ we obtain from $\mathrm{m}(i) = i - \lfloor \frac{i-1}{d} \rfloor \times d$, results (S.300) and (S.301), and proceeding inductively that

$$(u_{i,k,j}(P) - l_{i,k,j}(P)) \asymp 2^{-\frac{i}{d}}, \tag{S.302}$$

uniformly in $P \in \mathbf{P}$, $i \geq 0$, $0 \leq k \leq 2^i - 1$, and $1 \leq j \leq d$. Thus, result (S.302) yields

$$\begin{aligned} &\sup_{P\in\mathbf{P}} \max_{0\leq k\leq 2^i-1} \sup_{a,a'\in\tilde{\Delta}_{i,k}(P)} \|a - a'\| \\ &\qquad \leq \sup_{P\in\mathbf{P}} \max_{0\leq k\leq 2^i-1} \max_{1\leq j\leq d} \sqrt{d} \times (u_{i,j,k}(P) - l_{i,j,k}(P)) = O(2^{-\frac{i}{d}}). \end{aligned} \tag{S.303}$$

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 $[0,1]^d$. To this end, set

$$\Delta_{i,k}(P) \equiv T_P(\tilde{\Delta}_{i,k}(P))$$

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 (S.295). In addition, $\{\Delta_i(P)\}$ satisfies condition (iii) since by definition (S.292) and result (S.298) we have

$$P(\Delta_{i,k}(P)) = P(T_P(\tilde{\Delta}_{i,k}(P))) = Q_P(\tilde{\Delta}_{i,k}(P)) = 2^{-i},$$

for all $0 \leq k \leq 2^i - 1$. Moreover, by Assumption S.6.2(ii), $\sup_{P\in\mathbf{P}} \sup_{a\in[0,1]^d} \|JT_P(a)\|_o < \infty$, and hence by the mean value theorem we can conclude that

$$\sup_{P\in\mathbf{P}} \max_{0\leq k\leq 2^i-1} \sup_{v,v'\in\Delta_{i,k}(P)} \|v - v'\| = \sup_{P\in\mathbf{P}} \max_{0\leq k\leq 2^i-1} \sup_{a,a'\in\tilde{\Delta}_{i,k}(P)} \|T_P(a) - T_P(a')\|$$

$$\lesssim \sup_{P\in\mathbf{P}} \max_{0\leq k\leq 2^i-1} \sup_{a,a'\in\tilde{\Delta}_{i,k}(P)} \|a - a'\| = O(2^{-\frac{i}{d}}),$$

by result (S.303), 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

$$\mathcal{A} = \sigma(\bigcup_{i\geq 0} \tilde{\Delta}_i(P)), \tag{S.304}$$

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

$$D_i(P) \equiv \bigcup_{k:\tilde{\Delta}_{i,k}(P)\cap A\neq\emptyset} \tilde{\Delta}_{i,k}(P).$$

Notice that since $\{\tilde{\Delta}_i(P)\}$ is a partition of $[0,1]^d$, $A \subseteq D_i(P)$ 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 (S.303) 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 (S.304) follows. To conclude, we then note that

$$\sigma(\bigcup_{i\geq 0} \Delta_i(P)) = \sigma(\bigcup_{i\geq 0} T_P(\tilde{\Delta}_i(P))) = T_P(\sigma(\bigcup_{i\geq 0} \tilde{\Delta}_i(P))) = T_P(\mathcal{A}), \tag{S.305}$$

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 (S.305) implies $\{\Delta_i(P)\}$ satisfies condition (v) establishing the lemma. ■

Lemma S.6.2. *Let $\{\Delta_i(P)\}$ be as in Lemma S.6.1, and $\mathcal{B}_{P,i}$ denote the σ-algebra generated by $\Delta_i(P)$. If Assumptions S.6.1 and S.6.2 hold, then there are $K_0 > 0$, $K_1 \geq 1$ such that for all $P \in \mathbf{P}$ and any f satisfying $f \in L^2_P$ for all $P \in \mathbf{P}$:*

$$E_P[(f(V) - E_P[f(V)|\mathcal{B}_{P,i}])^2] \leq K_0 \times \varpi^2(f, K_1 \times 2^{-\frac{i}{d}}, P).$$

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^2_P by

$$E_P[f(V)|\mathcal{B}_{P,i}] = 2^i \sum_{k=0}^{2^i-1} 1\{V \in \Delta_{i,k}(P)\} \int_{\Delta_{i,k}(P)} f(v)dP(v).$$

Hence, employing that $P(\Delta_{i,k}(P)) = 2^{-i}$ 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 $(f(v) - f(\tilde{v}))1\{v \in \Omega(P)\}1\{\tilde{v} \in \Delta_{i,k}(P)\}$ we can conclude that

$$\begin{aligned}
&E_P[(f(V) - E_P[f(V)|\mathcal{B}_{P,i}])^2] \\
&\quad = \sum_{k=0}^{2^i-1} \int_{\Delta_{i,k}(P)} (f(v) - 2^i \int_{\Delta_{i,k}(P)} f(\tilde{v})dP(\tilde{v}))^2 dP(v) \\
&\quad = \sum_{k=0}^{2^i-1} 2^{2i} \int_{\Delta_{i,k}(P)} (\int_{\Delta_{i,k}(P)} (f(v) - f(\tilde{v}))1\{v \in \Omega(P)\}dP(\tilde{v}))^2 dP(v) \\
&\quad \leq \sum_{k=0}^{2^i-1} 2^{2i} P(\Delta_{i,k}(P)) \int_{\Delta_{i,k}(P)} \int_{\Delta_{i,k}(P)} (f(v) - f(\tilde{v}))^2 1\{v \in \Omega(P)\}dP(\tilde{v})dP(v) \\
&\quad = \sum_{k=0}^{2^i-1} 2^i \int_{\Delta_{i,k}(P)} \int_{\Delta_{i,k}(P)} (f(v) - f(\tilde{v}))^2 1\{v \in \Omega(P)\}dP(\tilde{v})dP(v).
\end{aligned}$$

Let $D_i \equiv \sup_{P \in \mathbf{P}} \max_{0 \leq k \leq 2^i-1} \mathrm{diam}\{\Delta_{i,k}(P)\}$, where $\mathrm{diam}\{\Delta_{i,k}(P)\}$ is the diameter of $\Delta_{i,k}(P)$. Further note that by Lemma S.6.1(iv), $D_i = O(2^{-\frac{i}{d}})$ and hence we have $\lambda(\{s \in \mathbf{R}^d : \|s\| \leq D_i\}) \leq M_1 2^{-i}$ 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 S.6.1(ii), and doing the change of variables $s = v - \tilde{v}$ we then obtain for some constant $M_0 > 0$ that

$$\begin{aligned}
&E_P[(f(V) - E_P[f(V)|\mathcal{B}_{P,i}])^2] \\
&\quad \leq M_0 \sum_{k=0}^{2^i-1} 2^i \int_{\Delta_{i,k}(P)} \int_{\|s\| \leq D_i} (f(\tilde{v} + s) - f(\tilde{v}))^2 1\{s + \tilde{v} \in \Omega(P)\}d\lambda(s)dP(\tilde{v}) \\
&\quad \leq M_0 M_1 \sup_{\|s\| \leq D_i} \sum_{k=0}^{2^i-1} \int_{\Delta_{i,k}(P)} (f(\tilde{v} + s) - f(\tilde{v}))^2 1\{\tilde{v} + s \in \Omega(P)\}dP(\tilde{v}). \qquad \text{(S.306)}
\end{aligned}$$

Hence, since $\{\Delta_{i,k}(P) : k = 0 \ldots 2^i - 1\}$ is a partition of $\Omega(P)$, $\varpi(f, h, P)$ is decreasing in h, and $D_i \leq K_1 2^{-\frac{i}{d}}$ for some $K_1 \geq 1$ by Lemma S.6.1(iv), we obtain

$$E_P[(f(V) - E_P[f(V)|\mathcal{B}_{P,i}])^2] \leq M_0 M_1 \times \varpi^2(f, K_1 \times 2^{-\frac{i}{d}}, P) \qquad \text{(S.307)}$$

by (S.306). Setting $K_0 \equiv M_0 \times M_1$ in (S.307) then establishes the lemma. ■

S.7 Uniform Bootstrap Coupling

We next provide uniform coupling results for the multiplier bootstrap that allow us to verify Assumption 3.11 in a variety of problems. The results in this appendix may be of independent interest, as they extend the validity of the multiplier bootstrap to suitable non-Donsker classes $\mathcal{F}_n$. For this reason, as in Section S.6, we state the results in a notation that abstracts from the rest of the paper. Hence, here $V \in \mathbf{R}^d$ should be interpreted as a generic random variable whose distribution is given by $P \in \mathbf{P}$. For $\{\omega_i\}_{i=1}^n$ i.i.d. standard normal random variables independent of $\{V_i\}_{i=1}^n$ we also set

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

Our coupling results rely on a series approximation to the elements of $\mathcal{F}_n$. To this end, we will assume that for each $P \in \mathbf{P}$ there is a basis $\{f_{d,n,P}\}_{d=1}^{d_n}$, with d_n possibly diverging to infinity, that provides a suitable approximation to every $f \in \mathcal{F}_n$. Formally, for $f_{n,P}^{d_n}(v) \equiv (f_{1,n,P}(v), \ldots, f_{d_n,n,P}(v))'$, we impose the following:

Assumption S.7.1. *For each $P \in \mathbf{P}$ there is an array of functions $\{f_{d,n,P}\}_{d=1}^{d_n} \subset L_P^2$ such that: (i) The eigenvalues of $E_P[f_{n,P}^{d_n}(V) f_{n,P}^{d_n}(V)']$ are bounded by $1 \leq C_n$ uniformly in $P \in \mathbf{P}$; (ii) $\sup_{P\in\mathbf{P}} \max_{1\leq d\leq d_n} \|f_{d,n,P}\|_\infty \leq K_n$ with $1 \leq K_n$ finite.*

Assumption S.7.2. *For every $f \in \mathcal{F}_n$ and $P \in \mathbf{P}$ there is a $\beta_{n,P}(f) \in \mathbf{R}^{d_n}$ such that: (i) The class $\mathcal{G}_{n,P} \equiv \{(f - \int f dP) - f_{n,P}^{d_n\prime}\beta_{n,P}(f) : f \in \mathcal{F}_n\}$ has envelope $G_{n,P}$ which satisfies $\|g\|_{P,2} \leq \delta_n \|G_{n,P}\|_{P,2}$ for all $P \in \mathbf{P}$, $g \in \mathcal{G}_{n,P}$, and some $\delta_n > 0$ with*

$$J_{1n} \equiv \sup_{P\in\mathbf{P}} \{J_{[]}(\delta_n\|G_{n,P}\|_{P,2}, \mathcal{G}_{n,P}, \|\cdot\|_{P,2}) + \sqrt{n} E_P[G_{n,P}(V) \exp\{-\frac{n\delta_n^2 \|G_{n,P}\|_{P,2}^2}{G_{n,P}^2(V)\eta_{n,P}}\}]\}$$

finite and $\eta_{n,P} \equiv 1 + \log N_{[]}(\delta_n\|G_{n,P}\|_{P,2}, \mathcal{G}_{n,P}, \|\cdot\|_{P,2})$; (ii) The set $\mathcal{B}_n \equiv \{\beta_{n,P}(f) : f \in \mathcal{F}_n, P \in \mathbf{P}\} \cup \{0\}$ satisfies $J_{2n} \equiv \int_0^\infty \sqrt{\log(N(\epsilon, \mathcal{B}_n, \|\cdot\|_2))} d\epsilon < \infty$.

Assumption S.7.1 imposes our regularity conditions on the approximating functions $\{f_{d,n,P}\}_{d=1}^{d_n}$. We emphasize that the functions $\{f_{d,n,P}\}_{d=1}^{d_n}$ need not be known: They are only employed in the theoretical construction of the coupling. In certain applications, such as when $\mathcal{F}_n$ is finite dimensional, a basis $\{f_{d,n,P}\}_{d=1}^{d_n}$ may be naturally available. The approximating requirements on $\{f_{d,n,P}\}_{d=1}^{d_n}$ are imposed in Assumption S.7.2. In particular, Assumption S.7.2(i) requires that the remainder of the approximation of $\mathcal{F}_n$ by $\{f_{d,n,P}\}_{d=1}^{d_n}$ not be "too large." Intuitively, Assumption S.7.2(i) controls the "bias" in a series approximation of $\mathcal{F}_n$ by linear combinations of $\{f_{d,n,P}\}_{d=1}^{d_n}$. Assumption S.7.2(ii) in turn controls the "variance" of the series approximation by demanding that the class of approximating functions have a finite entropy.

We next show Assumptions S.7.1 and S.7.2 suffice for coupling $\hat{\mathbb{G}}_n$.

Theorem S.7.1. *Let Assumptions S.7.1, S.7.2 hold, $\{(\omega_i, V_i)\}_{i=1}^n$ be i.i.d. with $V_i \sim P \in \mathbf{P}$, $\omega_i \sim N(0,1)$, ω_i and V_i independent, and $d_n \log(1+d_n)K_n^2 C_n = o(n)$. Then: (i) There is a linear Gaussian $\mathbb{G}_P^\star$ (possibly depending on n) independent of $\{V_i\}_{i=1}^n$ with*

$$\|\hat{\mathbb{G}}_n - \mathbb{G}_P^\star\|_{\mathcal{F}_n} = O_P(J_{2n}\{\frac{K_n^2 C_n d_n \log(1+d_n)}{n}\}^{1/4} + J_{1n})$$

uniformly in $P \in \mathbf{P}$ with $E[\mathbb{G}_P^\star(f)] = 0$ and $E[(\mathbb{G}_P^\star(f)\mathbb{G}_P^\star(g)] = \mathrm{Cov}_P\{f(V), g(V)\}$ for any $f, g \in \mathcal{F}_n$. (ii) If in addition $\sup_{P\in\mathbf{P}} \|(\mathrm{Var}_P\{f_{n,P}^{d_n}(V)\})^{-1}\|_{o,2} \leq \xi_n < \infty$ and $\xi_n\sqrt{d_n \log(1+d_n)C_n}K_n/\sqrt{n} = o(1)$, then uniformly in $P \in \mathbf{P}$

$$\|\hat{\mathbb{G}}_n - \mathbb{G}_P^\star\|_{\mathcal{F}_n} = O_P(\frac{J_{2n}K_n\sqrt{\xi_n C_n d_n \log(1+d_n)}}{\sqrt{n}} + J_{1n}).$$

Theorem S.7.1(i) derives a rate of convergence for the coupled process, while Theorem S.7.1(ii) improves on the rate under the additional requirement that $\mathrm{Var}_P\{f_{n,P}^{d_n}(V)\}$ be bounded away from singularity. The rates of both Theorems S.7.1(i) and S.7.1(ii) depend on the selected sequence d_n, which should be chosen optimally. Heuristically, the proof of Theorem S.7.1 proceeds in two steps. First, we construct a multivariate normal random variable $\mathbb{G}_P^\star(f_{n,P}^{d_n}) \in \mathbf{R}^{d_n}$ that is coupled with $\hat{\mathbb{G}}_n(f_{n,P}^{d_n}) \in \mathbf{R}^{d_n}$, and then employ the linearity of $\hat{\mathbb{G}}_n$ to obtain a suitable coupling on the subspace $\mathbb{S}_{n,P} \equiv \overline{\mathrm{span}}\{f_{n,P}^{d_n}\}$. Second, we employ Assumption S.7.2(i) to show that a successful coupling on $\mathbb{S}_{n,P}$ leads to the desired construction since $\mathcal{F}_n$ is well approximated by $\{f_{d,n,P}\}_{d=1}^{d_n}$.

Below, we include the proof of Theorem S.7.1 and auxiliary results.

Proof of Theorem S.7.1: We first couple $\hat{\mathbb{G}}_n$ on a finite dimensional subspace and then show that such a result suffices for coupling $\hat{\mathbb{G}}_n$ and $\mathbb{G}_P^\star$ on $\mathcal{F}_n$. To this end, let $\mathbb{S}_{n,P} \equiv \overline{\mathrm{span}}\{f_{n,P}^{d_n}\}$ and note that Assumption S.7.2(ii) and Lemma S.7.1 imply that there exists a linear Gaussian process $\mathbb{G}_P^{(1)}$ on $\mathbb{S}_{n,P}$ and a sequence $R_n = o(1)$ such that

$$\sup_{\beta\in\mathcal{B}_n} |\hat{\mathbb{G}}_n(f_{n,P}^{d_n\prime}\beta) - \mathbb{G}_P^{(1)}(f_{n,P}^{d_n\prime}\beta)| = O_P(J_{2n}R_n) \tag{S.308}$$

uniformly in $P \in \mathbf{P}$, $E[\mathbb{G}_P^{(1)}(f_{n,P}^{d_n\prime}\beta)] = 0$, and also $E[(\mathbb{G}_P^{(1)}(f_{n,P}^{d_n\prime}\beta_1))(\mathbb{G}_P^{(1)}(f_{n,P}^{d_n\prime}\beta_2))] = \mathrm{Cov}_P(f_{n,P}^{d_n}(V)'\beta_1, f_{n,P}^{d_n}(V)'\beta_2\}$. To establish part (i) of the theorem we will set $R_n = (d_n \log(1+d_n)C_n K_n^2/n)^{1/4}$ and employ Lemma S.7.1(i), while to establish part (ii) we will set $R_n = (\xi_n d_n \log(1+d_n)C_n K_n^2/n)^{1/2}$ and employ Lemma S.7.1(ii) instead.

Next note that since $\hat{\mathbb{G}}_n(f-c) = \hat{\mathbb{G}}_n(f)$ for any $c \in \mathbf{R}$ and $f \in L_P^2$, we may assume without loss of generality that $E_P[f(V)] = 0$ for all $f \in \mathcal{F}_n$. For any closed linear subspace $\mathbb{A}_{n,P} \subseteq L_P^2$ let $\mathrm{Proj}\{f|\mathbb{A}_{n,P}\}$ denote the $\|\cdot\|_{P,2}$ projection of f onto $\mathbb{A}_{n,P}$ and $\mathbb{A}_{n,P}^\perp \equiv \{f \in L_P^2 : f = g - \mathrm{Proj}\{g|\mathbb{A}_{n,P}\}$ for some $g \in L_P^2\}$. Assuming the underlying probability space is suitably enlarged to carry a linear isonormal Gaussian process $\mathbb{G}_P^{(2)}$ on $\{\mathrm{Proj}\{f|\mathbb{S}_{n,P}^{\perp}\} : f \in \mathcal{F}_n \cup \mathcal{G}_{n,P}\}$ independent of $\mathbb{G}_P^{(1)}$ and $\{V_i\}_{i=1}^n$, we then set

$$\mathbb{G}_P^{\star}(f) \equiv \mathbb{G}_P^{(1)}(\mathrm{Proj}\{f|\mathbb{S}_{n,P}\}) + \mathbb{G}_P^{(2)}(\mathrm{Proj}\{f|\mathbb{S}_{n,P}^{\perp}\}),$$

which is linear in f by linearity of $f \mapsto \mathrm{Proj}\{f|\mathbb{S}_{n,P}\}$, $\mathbb{G}_P^{(1)}$, and $\mathbb{G}_P^{(2)}$, and satisfies $E[\mathbb{G}_P^{\star}(f)] = 0$ and $E[\mathbb{G}_P^{\star}(f)\mathbb{G}_P^{\star}(g)] = \mathrm{Cov}_P\{f(V), g(V)\}$. Moreover, since $\mathbb{G}_P^{\star}$ is sub-Gaussian with respect to $\|\cdot\|_{P,2}$, it follows from Corollary 2.2.8 in van der Vaart and Wellner (1996), $N(\delta_n\|G_{n,P}\|_{P,2}, \mathcal{G}_{n,P}, \|\cdot\|_{P,2}) = 1$ due to $\|g\|_{P,2} \leq \delta_n\|G_{n,P}\|_{P,2}$ for all $g \in \mathcal{G}_{n,P}$ and $P \in \mathbf{P}$, bracketing numbers being larger than covering numbers, Jensen's inequality, and the definition of J_{1n} in Assumption S.7.2(i) that

$$\begin{aligned} E_P[\sup_{g\in\mathcal{G}_{n,P}} |\mathbb{G}_P^{\star}(g)|] &\lesssim \delta_n\|G_{n,P}\|_{P,2} + \int_0^{\infty} \sqrt{\log(N(\epsilon, \mathcal{G}_{n,P}, \|\cdot\|_{P,2}))}d\epsilon \\ &\leq \delta_n\|G_{n,P}\|_{P,2} + \int_0^{\delta_n\|G_{n,P}\|_{P,2}} \sqrt{1 + \log(N_{[\,]}(\epsilon, \mathcal{G}_{n,P}, \|\cdot\|_{P,2}))}d\epsilon \lesssim J_{1n}. \end{aligned} \quad \text{(S.309)}$$

To obtain an analogous bound for $\hat{\mathbb{G}}_n$, note $\sup_{g\in\mathcal{G}_{n,P}} \|g\|_{P,2} \leq \delta_n\|G_{n,P}\|_{P,2}$ by Assumption S.7.2(i) and $|E_P[g(V)]| \leq \|g\|_{P,2}$ by Jensen's inequality imply that

$$\begin{aligned} \sup_{g\in\mathcal{G}_{n,P}} |\hat{\mathbb{G}}_n(g)| &\leq \sup_{g\in\mathcal{G}_{n,P}} |\frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_i g(V_i)| \\ &+ |\frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_i| \times \{\sup_{g\in\mathcal{G}_{n,P}} |\frac{1}{n}\sum_{i=1}^n g(V_i) - E_P[g(V)]| + \delta_n\|G_{n,P}\|_{P,2}\}. \end{aligned} \quad \text{(S.310)}$$

Next, define the class $\tilde{\mathcal{G}}_{n,P} \equiv \{(\omega, v) \mapsto \omega g(v) : g \in \mathcal{G}_{n,P}\}$, and with some abuse of notation let P index the joint distribution of (V, ω). Further note that if $\{[g_{i,l,P}, g_{i,u,P}]\}_i$ is a bracket for $\mathcal{G}_{n,P}$, then the functions $\{[\tilde{g}_{i,l,P}, \tilde{g}_{i,u,P}]\}$ given by

$$\begin{aligned} \tilde{g}_{i,l,P}(\omega, v) &\equiv \max\{\omega, 0\}g_{i,l,P}(v) + \min\{\omega, 0\}g_{i,u,P}(v) \\ \tilde{g}_{i,u,P}(\omega, v) &\equiv \min\{\omega, 0\}g_{i,l,P}(v) + \max\{\omega, 0\}g_{i,u,P}(v) \end{aligned}$$

form a bracket for $\tilde{\mathcal{G}}_{n,P}$. Moreover, since $E[\omega^2] = 1$ and ω and V are independent, it follows that $\|\tilde{g}_{i,u,P} - \tilde{g}_{i,l,P}\|_{P,2} = \|g_{i,u,P} - g_{i,l,P}\|_{P,2}$. Setting $\tilde{G}_{n,P}(\omega, v) \equiv |\omega|G_{n,P}(v)$, then note that $\tilde{G}_{n,P}$ is an envelope for $\tilde{\mathcal{G}}_{n,P}$, which satisfies $\|\tilde{G}_{n,P}\|_{P,2} = \|G_{n,P}\|_{P,2}$. For $\eta_{n,P} \equiv 1 + \log N_{[\,]}(\delta_n\|G_{n,P}\|_{P,2}, \mathcal{G}_{n,P}, \|\cdot\|_{P,2})$ (as in Assumption S.7.2(i)) we then obtain by Theorem 2.14.2 in van der Vaart and Wellner (1996) that

$$E_P[\sup_{g\in\mathcal{G}_{n,P}}|\frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_i g(V_i)|] \lesssim J_{[\,]}(\delta_n\|G_{n,P}\|_{P,2}, \mathcal{G}_{n,P}, \|\cdot\|_{P,2})$$
$$+ \sqrt{n}E_P[|\omega|G_{n,P}(V)1\{|\omega|\frac{G_{n,P}(V)}{\|G_{n,P}\|_{P,2}} > \frac{\sqrt{n}\delta_n}{\sqrt{\eta_{n,P}}}\}]. \quad \text{(S.311)}$$

Moreover, since ω follows a standard normal distribution, we have $E[|\omega|1\{|\omega|>a\}] \lesssim \exp\{-a^2/2\}$ for any $a \geq 0$. Therefore, the independence of ω and V implies

$$E_P[|\omega|G_{n,P}(V)1\{|\omega|\frac{G_{n,P}(V)}{\|G_{n,P}\|_{P,2}} > \frac{\sqrt{n}\delta_n}{\sqrt{\eta_{n,P}}}\}] \lesssim E_P[G_{n,P}(V)\exp\{-\frac{n\delta_n^2\|G_{n,P}\|_{P,2}^2}{2G_{n,P}^2(V)\eta_{n,P}}\}]$$

which together with result (S.311) and the definition of J_{1n} in Assumption S.7.2(i) yields

$$E_P[\sup_{g\in\mathcal{G}_{n,P}}|\frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_i g(V_i)|] \lesssim J_{1n}. \quad \text{(S.312)}$$

Moreover, by Lemmas 2.3.1 and 2.9.1 in van der Vaart and Wellner (1996) we have

$$E_P[\sup_{g\in\mathcal{G}_{n,P}}|\frac{1}{\sqrt{n}}\sum_{i=1}^n g(V_i) - E_P[g(V)]|] + \delta_n\|G_{n,P}\|_{P,2}$$
$$\lesssim E_P[\sup_{g\in\mathcal{G}_{n,P}}|\frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_i g(V_i)|] + \delta_n\|G_{n,P}\|_{P,2} \lesssim J_{1n}, \quad \text{(S.313)}$$

where the final inequality follows from (S.312) and the definition of J_{1n}. Thus, (S.310), (S.312), and (S.313) together with Markov's inequality imply that uniformly in $P \in \mathbf{P}$

$$\|\hat{\mathbb{G}}_n\|_{\mathcal{G}_{n,P}} = O_P(J_{1n}). \quad \text{(S.314)}$$

Next, we use the linearity of the processes $f \mapsto \hat{\mathbb{G}}_n(f)$ and $f \mapsto \mathbb{G}_P^\star(f)$ to obtain that

$$\|\hat{\mathbb{G}}_n - \mathbb{G}_P^\star\|_{\mathcal{F}_n} \leq \sup_{f\in\mathcal{F}_n}|\hat{\mathbb{G}}_n(f_{n,P}^{d_n\prime}\beta_{n,P}(f)) - \mathbb{G}_P^\star(f_{n,P}^{d_n\prime}(\beta_{n,P}(f)))| + \|\hat{\mathbb{G}}_n - \mathbb{G}_P^\star\|_{\mathcal{G}_{n,P}}$$
$$\leq \sup_{\beta\in\mathcal{B}_n}|\hat{\mathbb{G}}_n(f_{n,P}^{d_n\prime}\beta) - \mathbb{G}_P^\star(f_{n,P}^{d_n\prime}\beta)| + O_P(J_{1n}) = O_P(J_{2n}R_n + J_{1n}),$$

where the second inequality holds uniformly in $P \in \mathbf{P}$ by (S.309) and Markov's inequality, result (S.314), and set inclusion, while the equality holds uniformly in $P \in \mathbf{P}$ by result (S.308). The first claim of the theorem then follows by using Lemma S.7.1(i) to set $R_n = (d_n \log(1+d_n)C_nK_n^2/n)^{1/4}$ in (S.308), and the second part of the theorem follows from using Lemma S.7.1(ii) to set $R_n = (\xi_n d_n \log(1+d_n)C_nK_n^2/n)^{1/2}$ instead. ■

Lemma S.7.1. *Let $\{(\omega_i, V_i)\}_{i=1}^n$ be i.i.d. with $V_i \sim P \in \mathbf{P}$, $\omega_i \sim N(0,1)$, and ω_i and V_i independent. Suppose Assumption S.7.1 holds, $d_n\log(1+d_n)K_n^2C_n = o(n)$, and*

$\mathcal{B}_n \subset \mathbf{R}^{d_n}$ satisfies $0 \in \mathcal{B}_n$ and $J_{2n} \equiv \int_0^\infty \sqrt{\log(N(\epsilon, \mathcal{B}_n, \|\cdot\|_2))}d\epsilon < \infty$. Then: (i) There is a linear Gaussian process $\mathbb{G}_P^\star$ on $\mathbb{S}_{n,P} \equiv \overline{\text{span}}\{f_{n,P}^{d_n}\}$ independent of $\{V_i\}_{i=1}^n$ with

$$\sup_{\beta \in \mathcal{B}_n} |\hat{\mathbb{G}}_n(f_{n,P}^{d_n\prime}\beta) - \mathbb{G}_P^\star(f_{n,P}^{d_n\prime}\beta)| = O_P(J_{2n}\{\frac{d_n \log(1+d_n)C_n K_n^2}{n}\}^{1/4})$$

uniformly in $P \in \mathbf{P}$ and satisfying $E[\mathbb{G}_P^\star(f_{n,P}^{d_n\prime}\beta)] = 0$ and $E[\mathbb{G}_P^\star(f_{n,P}^{d_n\prime}\beta_1)\mathbb{G}_P^\star(f_{n,P}^{d_n\prime}\beta_2)] = \text{Cov}_P\{f_{n,P}^{d_n}(V)'\beta_1, f_{n,P}^{d_n}(V)'\beta_2\}$. (ii) If in addition $\sup_{P\in\mathbf{P}} \|\text{Var}_P^{-1}\{f_{n,P}^{d_n}(V)\}\|_{o,2} \leq \xi_n < \infty$ and $\xi_n\sqrt{d_n \log(1+d_n)C_n}K_n/\sqrt{n} = o(1)$, then uniformly in $P \in \mathbf{P}$

$$\sup_{\beta \in \mathcal{B}_n} |\hat{\mathbb{G}}_n(f_{n,P}^{d_n\prime}\beta) - \mathbb{G}_P^\star(f_{n,P}^{d_n\prime}\beta)| = O_P(\frac{J_{2n}\sqrt{\xi_n d_n \log(1+d_n)C_n}K_n}{\sqrt{n}}).$$

PROOF: First note that $\hat{\mathbb{G}}_n(f - c) = \hat{\mathbb{G}}_n(f)$ for any $c \in \mathbf{R}$ and $f \in L_P^2$. We may therefore assume without loss of generality that $E_P[f_{n,P}^{d_n}(V)] = 0$, and for every $P \in \mathbf{P}$ we let $\Sigma_n(P) \equiv \text{Var}_P\{f_{n,P}^{d_n}(V)\} = E_P[f_{n,P}^{d_n}(V)f_{n,P}^{d_n}(V)']$ and define

$$\hat{\Sigma}_n(P) \equiv \frac{1}{n}\sum_{i=1}^n (f_{n,P}^{d_n}(V_i) - \frac{1}{n}\sum_{j=1}^n f_{n,P}^{d_n}(V_j))(f_{n,P}^{d_n}(V_i) - \frac{1}{n}\sum_{j=1}^n f_{n,P}^{d_n}(V_j))'.$$

For a sequence R_n with $R_n = o(1)$, and any constant $M > 0$ and $P \in \mathbf{P}$ define the event

$$A_{n,P}(M) \equiv \{\|\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P)\|_{o,2} \leq MR_n\}. \tag{S.315}$$

Further note that by Lemma S.7.2 it follows we may select $R_n = o(1)$ such that we have

$$\liminf_{M\uparrow\infty} \liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P(\{V_i\}_{i=1}^n \in A_{n,P}(M)) = 1. \tag{S.316}$$

In particular, to establish part (i) we will set $R_n = (d_n \log(1+d_n)C_n K_n^2/n)^{1/4}$ and employ Lemma S.7.2(i), while to establish part (ii) we will set $R_n = (\xi_n d_n \log(1+d_n)C_n K_n^2/n)^{1/2}$ and employ Lemma S.7.2(ii) instead.

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

$$\mathbb{Z}_{n,P} \equiv \sum_{d:\hat{\lambda}_d\neq 0} \hat{\nu}_d \frac{(\hat{\nu}_d'\hat{\mathbb{G}}_n(f_{n,P}^{d_n}))}{\hat{\lambda}_d^{1/2}} + \sum_{d:\hat{\lambda}_d=0} \hat{\nu}_d(\hat{\nu}_d'\mathcal{N}_{d_n}). \tag{S.317}$$

Then note that since $\hat{\mathbb{G}}_n(f_{n,P}^{d_n}) \sim N(0, \hat{\Sigma}_n(P))$ 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,

$$E[\mathbb{Z}_{n,P}\mathbb{Z}_{n,P}'|\{V_i\}_{i=1}^n] = \sum_{d=1}^{d_n} \hat{\nu}_d \hat{\nu}_d' = I_{d_n}$$

by direct calculation for I_{d_n} the $d_n \times d_n$ identity matrix. 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 $\hat{\mathbb{G}}_n(f_{n,P}^{d_n}) \sim N(0, \hat{\Sigma}_n(P))$ conditional on $\{V_i\}_{i=1}^n$, it follows that $\hat{\mathbb{G}}_n(f_{n,P}^{d_n})$ 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 $\{\hat{\nu}_d : \hat{\lambda}_d \neq 0\}_{d=1}^{d_n}$ spans the range of $\hat{\Sigma}_n(P)$, we conclude from (S.317) that for any $\beta \in \mathbf{R}^{d_n}$

$$\beta' \hat{\Sigma}_n^{1/2}(P)\mathbb{Z}_{n,P} = \beta' \sum_{d:\hat{\lambda}_d \neq 0} \hat{\nu}_d(\hat{\nu}_d' \hat{\mathbb{G}}_n(f_{n,P}^{d_n})) = \hat{\mathbb{G}}_n(\beta' f_{n,P}^{d_n}).$$

Analogously, we define for any $\beta \in \mathbf{R}^{d_n}$ the linear Gaussian process $\mathbb{G}_P^\star$ on $\mathbb{S}_{n,P}$ by

$$\mathbb{G}_P^\star(\beta' f_{n,P}^{d_n}) \equiv \beta' \Sigma_n^{1/2}(P)\mathbb{Z}_{n,P},$$

which by construction is independent of $\{V_i\}_{i=1}^n$ and satisfies $E[\mathbb{G}_P^\star(f_{n,P}^{d_n\prime}\beta)] = 0$ and $E[\mathbb{G}_P^\star(f_{n,P}^{d_n\prime}\beta_1)\mathbb{G}_P^\star(f_{n,P}^{d_n\prime}\beta_2)] = \mathrm{Cov}_P\{f_{n,P}^{d_n}(V)'\beta_1, f_{n,P}^{d_n}(V)'\beta_2\}$. Setting

$$\bar{\mathbb{G}}_P(\beta) \equiv (\beta'(\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P))\mathbb{Z}_{n,P})1\{A_{n,P}(M)\}, \tag{S.318}$$

where $1\{A_{n,P}(M)\}$ denotes the indicator for the event $\{V_i\}_{i=1}^n \in A_{n,P}(M)$, then note

$$\sup_{\beta \in \mathcal{B}_n} |\hat{\mathbb{G}}_n(f_{n,P}^{d_n\prime}\beta) - \mathbb{G}_P^\star(f_{n,P}^{d_n\prime}\beta)|1\{A_{n,P}(M)\} = \sup_{\beta \in \mathcal{B}_n} |\bar{\mathbb{G}}_P(\beta)|. \tag{S.319}$$

Moreover, we note that conditional on $\{V_i\}_{i=1}^n$, $\bar{\mathbb{G}}_P$ is sub-Gaussian under the semi-metric $\rho_n(\tilde{\beta}, \beta) \equiv \|(\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P))(\tilde{\beta} - \beta)\|_2$. Since $\|\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P)\|_{o,2} \leq MR_n$ whenever $1\{A_{n,P}(M)\} = 1$ we obtain, whenever $\{V_i\}_{i=1}^n \in A_{n,P}(M)$, that

$$\begin{aligned}\int_0^\infty \sqrt{\log(N(\epsilon, \mathcal{B}_n, \rho_n))}d\epsilon &\leq \int_0^\infty \sqrt{\log(N(\epsilon/MR_n, \mathcal{B}_n, \|\cdot\|_2))}d\epsilon \\ &= MR_n \int_0^\infty \sqrt{\log(N(u, \mathcal{B}_n, \|\cdot\|_2))}du, \end{aligned} \tag{S.320}$$

where the equality follows from the change of variables $\epsilon = MR_n u$. Therefore, since $0 \in \mathcal{B}_n$, Corollary 2.2.8 in van der Vaart and Wellner (1996) and (S.320) imply

$$E[\sup_{\beta \in \mathcal{B}_n} |\bar{\mathbb{G}}_P(\beta)||\{V_i\}_{i=1}^n] \lesssim MR_n \int_0^\infty \sqrt{\log(N(u, \mathcal{B}_n, \|\cdot\|_2))}du \equiv MR_n J_{2n}. \tag{S.321}$$

Next, note (S.318), (S.319), and (S.321) together with Markov's inequality imply that

$$P(\sup_{\beta\in\mathcal{B}_n} |\hat{\mathbb{G}}_n(f_{n,P}^{d_n\prime}\beta) - \mathbb{G}_P^\star(f_{n,P}^{d_n\prime}\beta)| > M^2 R_n J_{2n};\ A_{n,P}(M)) \\ \leq P(\sup_{\beta\in\mathcal{B}_n} |\bar{\mathbb{G}}_P(\beta)| > M^2 R_n J_{2n}) \lesssim \frac{1}{M} \quad \text{(S.322)}$$

for all $P \in \mathbf{P}$. Therefore, combining results (S.316) and (S.322), we can finally conclude

$$\limsup_{M\uparrow\infty} \limsup_{n\to\infty} \sup_{P\in\mathbf{P}} P(\sup_{\beta\in\mathcal{B}_n} |\hat{\mathbb{G}}_n(f_{n,P}^{d_n\prime}\beta) - \mathbb{G}_P^\star(f_{n,P}^{d_n\prime}\beta)| > M^2 R_n J_{2n}) \\ \lesssim \limsup_{M\uparrow\infty} \limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \{\frac{1}{M} + P(\{V_i\}_{i=1}^n \notin A_{n,P}(M))\} = 0.$$

The first claim of the lemma then follows by employing Lemma S.7.2(i) to set $R_n = (d_n \log(1+d_n) C_n K_n^2/n)^{1/4}$ in (S.315), while the second claim follows by employing Lemma S.7.2(ii) to set $R_n = (\xi_n d_n \log(1+d_n) C_n K_n^2/n)^{1/2}$. ■

Lemma S.7.2. *Let $\{V_i\}_{i=1}^n$ be i.i.d. with $V \sim P \in \mathbf{P}$, suppose Assumption S.7.1 holds, define $\Sigma_n(P) \equiv \mathrm{Var}_P\{f_{n,P}^{d_n}(V)\}$ and its sample analogue $\hat{\Sigma}_n(P)$ to equal*

$$\hat{\Sigma}_n(P) \equiv \frac{1}{n}\sum_{i=1}^n (f_{n,P}^{d_n}(V_i) - \frac{1}{n}\sum_{j=1}^n f_{n,P}^{d_n}(V_j))(f_{n,P}^{d_n}(V_i) - \frac{1}{n}\sum_{j=1}^n f_{n,P}^{d_n}(V_j))',$$

and assume $d_n \log(1+d_n) K_n^2 C_n = o(n)$. (i) Then, it follows that uniformly in $P \in \mathbf{P}$

$$\|\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P)\|_{o,2} = O_P(\{\frac{d_n \log(1+d_n) C_n K_n^2}{n}\}^{1/4}).$$

(ii) If in addition $\sup_{P\in\mathbf{P}} \|\Sigma_n^{-1}(P)\|_{o,2} \leq \xi_n < \infty$ and $\xi_n \sqrt{d_n \log(1+d_n) C_n} K_n/\sqrt{n} = o(1)$, then we can also conclude uniformly in $P \in \mathbf{P}$ that

$$\|\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P)\|_{o,2} = O_P(\frac{\sqrt{\xi_n d_n \log(1+d_n) C_n} K_n}{\sqrt{n}}).$$

Proof: First note that we may without loss of generality assume that $E_P[f_{n,P}^{d_n}(V)] = 0$. Next note that Assumption S.7.1(ii) implies that for all $P \in \mathbf{P}$ we must have

$$\|\frac{1}{n}\{f_{n,P}^{d_n}(V_i) f_{n,P}^{d_n}(V_i)' - E_P[f_{n,P}^{d_n}(V) f_{n,P}^{d_n}(V)']\}\|_{o,2} \leq \frac{2 d_n K_n^2}{n} \quad \text{(S.323)}$$

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 K_n^2. Similarly, employing $\|f_{n,P}^{d_n}(V_i) f_{n,P}^{d_n}(V_i)'\|_{o,2} \leq d_n K_n^2$ almost surely we obtain

$$\|\frac{1}{n} E_P[\{f_{n,P}^{d_n}(V) f_{n,P}^{d_n}(V)' - E_P[f_{n,P}^{d_n}(V) f_{n,P}^{d_n}(V)']\}^2]\|_{o,2} \leq \frac{2 d_n K_n^2 C_n}{n}. \quad \text{(S.324)}$$

Thus, employing results (S.323) and (S.324), together with $d_n \log(1+d_n)K_n^2 C_n = o(n)$, we obtain by Theorem 6.1(ii) in Tropp (2012) that for all $P \in \mathbf{P}$

$$P(\|\frac{1}{n}\sum_{i=1}^n f_{n,P}^{d_n}(V_i)f_{n,P}^{d_n}(V_i)' - E_P[f_{n,P}^{d_n}(V)f_{n,P}^{d_n}(V)']\|_{o,2} > \frac{M\sqrt{d_n \log(1+d_n)C_n}K_n}{\sqrt{n}})$$
$$\leq d_n \exp\{-\frac{M^2(d_n \log(1+d_n)K_n^2 C_n)}{2n}\frac{n}{MBd_nK_n^2C_n}\} \quad \text{(S.325)}$$

for some $B < \infty$. Hence, we can conclude from (S.325) that uniformly in $P \in \mathbf{P}$

$$\|\frac{1}{n}\sum_{i=1}^n f_{n,P}^{d_n}(V_i)f_{n,P}^{d_n}(V_i)' - E_P[f_{n,P}^{d_n}(V)f_{n,P}^{d_n}(V)']\|_{o,2} = O_P(\frac{\sqrt{d_n \log(1+d_n)C_n}K_n}{\sqrt{n}}). \quad \text{(S.326)}$$

Recalling that we had without loss of generality set $E_P[f_{n,P}^{d_n}(V)] = 0$, next note that $E_P[f_{d,n,P}^2(V)] \leq \|E_P[f_{n,P}^{d_n}(V)f_{n,P}^{d_n}(V)']\|_o \leq C_n$, Markov's inequality, and Lemmas 2.2.9 and 2.2.10 in van der Vaart and Wellner (1996) imply, uniformly in $P \in \mathbf{P}$, that

$$\|\frac{1}{n}\sum_{i=1}^n f_{n,P}^{d_n}(V_i)\|_2 \leq \sqrt{d_n}\max_{1\leq d\leq d_n}|\frac{1}{n}\sum_{i=1}^n f_{d,n,P}(V_i)|$$
$$= O_P(\frac{K_n \log(1+d_n)\sqrt{d_n}}{n} + \frac{\sqrt{C_n d_n \log(1+d_n)}}{\sqrt{n}}). \quad \text{(S.327)}$$

Therefore, since for any $a, b \in \mathbf{R}^{d_n}$ we have $\|ab'\|_{o,2} \leq \|a\|_2\|b\|_2$, results (S.326) and (S.327) together with the triangle inequality yield, uniformly in $P \in \mathbf{P}$, that

$$\|\hat{\Sigma}_n(P) - \Sigma_n(P)\|_{o,2} = O_P(\frac{\sqrt{d_n \log(1+d_n)C_n}K_n}{\sqrt{n}}). \quad \text{(S.328)}$$

Finally, since $\hat{\Sigma}_n(P) \geq 0$ and $\Sigma_n(P) \geq 0$, Theorem X.1.1 in Bhatia (1997) implies that

$$\|\hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P)\|_{o,2} \leq \|\hat{\Sigma}_n(P) - \Sigma_n(P)\|_{o,2}^{1/2} \quad \text{(S.329)}$$

almost surely, and hence the first claim the lemma follows from (S.328) and (S.329).

For the second claim, let $\underline{\text{eig}}\{A\}$ denote the smallest eigenvalue of any Hermitian matrix A. Since $\|\Sigma_n^{-1}(P)\|_{o,2} = 1/\underline{\text{eig}}\{\Sigma_n(P)\}$, $\sup_{P\in\mathbf{P}}\|\Sigma_n^{-1}(P)\|_{o,2} \leq \xi_n$, result (S.328), Corollary III.2.6 in Bhatia (1997), and $\xi_n\sqrt{d_n \log(1+d_n)C_n}K_n/\sqrt{n} = o(1)$ imply

$$\liminf_{n\to\infty}\inf_{P\in\mathbf{P}} P(\underline{\text{eig}}\{\hat{\Sigma}_n(P)\} > \frac{1}{2\xi_n})$$
$$\geq \liminf_{n\to\infty}\inf_{P\in\mathbf{P}} P(\underline{\text{eig}}\{\Sigma_n(P)\} > \frac{1}{2\xi_n} + \|\hat{\Sigma}_n(P) - \Sigma_n(P)\|_{o,2}) = 1.$$

Therefore, Applying Theorem X.3.8 in Bhatia (1997) we can then conclude that

$$\limsup_{M\uparrow\infty}\limsup_{n\to\infty}\sup_{P\in\mathbf{P}} P(\|\hat{\Sigma}_n^{1/2}(P)-\Sigma_n^{1/2}(P)\|_{o,2} > M\frac{\sqrt{\xi_n d_n\log(1+d_n)C_n}K_n}{\sqrt{n}})$$
$$\leq \limsup_{M\uparrow\infty}\limsup_{n\to\infty}\sup_{P\in\mathbf{P}} P(\|\hat{\Sigma}_n(P)-\Sigma_n(P)\|_{o,2} > M\frac{\sqrt{d_n\log(1+d_n)C_n}K_n}{\sqrt{n}}) = 0,$$

where the final equality follows from result (S.328). ■

Lemma S.7.3. *For any positive random variable U with $E[U^2]<\infty$ and finite constant $A>0$ it follows that $E[U\exp\{-A/U^2\}]\leq E[U]\exp\{-A/E[U^2]\}+E[U^2]/\sqrt{2A}$.*

Proof: First note $u\mapsto u\exp\{-A/u^2\}$ is convex on $u\in(0,\sqrt{2A}]$. Therefore Jensen's inequality, $u\mapsto u\exp\{-A^2/u^2\}$ being increasing in $u\in(0,\infty)$, $E[1\{0<U<\sqrt{2A}\}U]\leq E[U]$ due to U being positive a.s., and $\exp\{-A/U^2\}\leq 1$ due to $A>0$, imply

$$E[U\exp\{-\frac{A}{U^2}\}] = E[1\{0<U\leq\sqrt{2A}\}U\exp\{-\frac{A}{U^2}\}]+E[1\{U>\sqrt{2A}\}U\exp\{-\frac{A}{U^2}\}]$$
$$\leq E[U]\exp\{-\frac{A}{E[U^2]}\}+E[1\{U>\sqrt{2A}\}U].$$

The claim of the lemma therefore follows from $E[1\{U>\sqrt{2A}\}U]\leq E[U^2]/\sqrt{2A}$ by the Cauchy Schwarz inequality and Markov's inequality. ■

References

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis – A Hitchhiker's Guide*. Springer-Verlag, Berlin.

Bhatia, R. (1997). *Matrix Analysis*. Springer, New York.

Bogachev, V. I. (1998). *Gaussian Measures*. American Mathematical Society, Providence.

Bogachev, V. I. (2007). *Measure Theory: Volume I*. Springer-Verlag, Berlin.

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive approximation*, vol. 303. Springer Science & Business Media.

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.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. Wiley, New York.

Massart, P. (1989). Strong approximation for multivariate empirical and related processes, via kmt constructions. *The Annals of Probability*, **17** 266–291.

Pollard, D. (2002). *A user's guide to measure theoretic probability*, vol. 8. Cambridge 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.

WALKUP, D. W. and WETS, R. J.-B. (1969). A lipschitzian characterization of convex polyhedra. *Proceedings of the American Mathematical Society* 167–173.

ZEIDLER, E. (1985). *Nonlinear Functional Analysis and its Applications I*. Springer-Verlag, Berlin.

ZHAI, A. (2018). A high-dimensional clt in w_2 distance with near optimal convergence rate. *Probability Theory and Related Fields*, **170** 821–845.

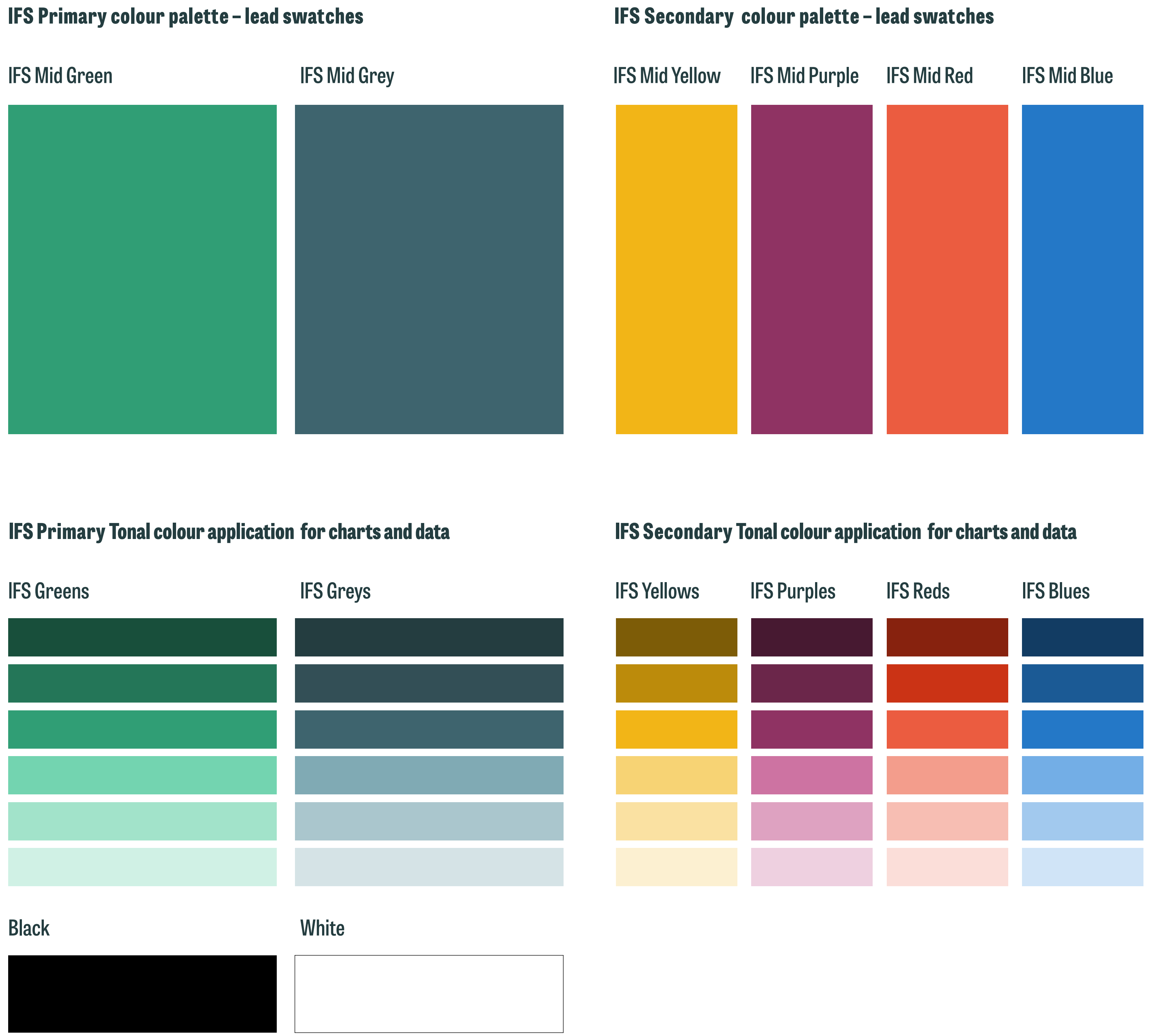
Colour overview

The IFS colour palette has been designed with impact and colour accessibility in mind. It provides a harmonious and flexible colour palette to support the recognisable IFS Green. For the IFS primary palette the lead mid green is anchored with a cool mid grey, allowing the visual tone to shift to suit the specific project. This is accompanied by a vibrant and flexible secondary colour palette.

The 'Mid' hues are designed for use over larger areas of block colour, such as publication covers. 'Hue' is the colour property we refer to when we say, for example, that something looks red, yellow, green etc.

This is supported by a set of tonal colours for charts and other data visualizations. 'Tone' is produced either by lightening or darkening a colour.

The palette has been carefully selected to give the most consistency when applied to MS office applications. These colours can also be applied to Beamer manually. See pages 20–22 for more details on colour values.



Primary colour values

A wide selection of swatches comprise IFS's palette. The Primary palette comprises of greens and greys, all with light and dark (tonal) options.

The 'Mid' colours are intended to lead, adding interest, supported by the other swatches to add tonal differentiation to infographics and charts where required.

Always use the colour values as stated in these guidelines, or the colour swatch files provided within the internal templates. All templates have been created to be sympathetic to these colour sets.

IFS Dark Green 2

C	87	R	24	#184F3B
M	41	G	79	
Y	75	B	59	
K	45			

IFS Dark Green 1

C	77	R	36	#247658
M	12	G	118	
Y	65	B	88	
K	1			

IFS Mid Green

C	77	R	48	#309E75
M	12	G	158	
Y	65	B	117	
K	1			

IFS Light Green 1

C	55	R	115	#73D4B0
M	0	G	212	
Y	41	B	176	
K	0			

IFS Light Green 2

C	40	R	162	#A2E3CA
M	0	G	227	
Y	29	B	202	
K	0			

IFS Light Green 3

C	22	R	208	#D0F1E5
M	0	G	241	
Y	15	B	229	
K	0			

IFS Dark Grey 2

C	82	R	36	#243D40
M	54	G	61	
Y	54	B	64	
K	57			

IFS Dark Grey 1

C	78	R	51	#334F56
M	50	G	79	
Y	48	B	86	
K	42			

IFS Mid Grey

C	74	R	64	#40646D
M	43	G	100	
Y	42	B	109	
K	29			

IFS Light Grey 1

C	54	R	128	#80AAB4
M	21	G	170	
Y	26	B	180	
K	3			

IFS Light Grey 2

C	38	R	170	#AAC6CD
M	13	G	198	
Y	18	B	205	
K	0			

IFS Light Grey 3

C	20	R	213	#D5E3E6
M	6	G	227	
Y	10	B	230	
K	0			

Secondary colour values

A wider selection of swatches comprise IFS's secondary palette. These colours are intended to add vibrancy to outputs and tonal differentiation to infographics and other similar outputs, where required.

The secondary palette comprises four colour sets – yellows, purples, reds and blues, all with light and dark (tonal) options. When using the secondary colour palette in large areas of colour, such as report covers, lead with the 'Mid' option, as illustrated on page 19.

For the application within charts and infographics follow the advice on page 30 and 31.

Always use the colour values as stated in these guidelines, or the colour swatch files provided within the internal templates, all templates have been created to be sympathetic to these colour sets.

IFS Dark Red 2

C	29	R	135	#87220E
M	95	G	34	
Y	100	B	14	
K	35			

IFS Dark Red 1

C	13	R	203	#CB3315
M	90	G	51	
Y	100	B	21	
K	35			

IFS Mid Red

C	0	R	235	#EB5C40
M	75	G	92	
Y	74	B	64	
K	0			

IFS Light Red 1

C	0	R	243	#F39D8C
M	49	G	157	
Y	40	B	140	
K	0			

IFS Light Red 2

C	1	R	247	#F7BEB3
M	33	G	190	
Y	26	B	179	
K	0			

IFS Light Red 3

C	0	R	251	#FBDED9
M	18	G	222	
Y	13	B	217	
K	0			

IFS Dark Blue 2

C	100	R	18	#123C63
M	76	G	60	
Y	35	B	99	
K	25			

IFS Dark Blue 1

C	92	R	27	#1B5A95
M	63	G	90	
Y	15	B	149	
K	2			

IFS Mid Blue

C	82	R	36	#2478C7
M	47	G	120	
Y	0	B	199	
K	0			

IFS Light Blue 1

C	57	R	115	#73AEE6
M	21	G	174	
Y	0	B	230	
K	0			

IFS Light Blue 2

C	40	R	162	#A2C9EE
M	12	G	201	
Y	0	B	238	
K	0			

IFS Light Blue 3

C	22	R	208	#D0E4F7
M	5	G	228	
Y	0	B	247	
K	0			

Secondary colour values

A wider selection of swatches comprise IFS's secondary palette. These colours are intended to add vibrancy to outputs and tonal differentiation to infographics and other similar outputs, where required.

The secondary palette comprises four colour sets – yellows, purples, reds and blues, all with light and dark (tonal) options. When using the secondary colour palette in large areas of colour, such as report covers, lead with the 'Mid' option, as illustrated on page 19.

For the application within charts and infographics follow the advice on page 30 and 31.

Always use the colour values as stated in these guidelines, or the colour swatch files provided within the internal templates, all templates have been created to be sympathetic to these colour sets.

	C	M	Y	K	R	G	B	
IFS Dark Yellow 2	37	51	100	29	125	92	07	#7D5C07
IFS Dark Yellow 1	23	42	100	7	188	139	11	#BC8B0B
IFS Mid Yellow	5	31	93	0	242	181	23	#F2B517
IFS Light Yellow 1	4	17	63	0	247	211	116	#F7D374
IFS Light Yellow 2	3	12	44	0	250	225	162	#FAE1A2
IFS Light Yellow 3	2	6	23	0	252	240	209	#FCF0D1
IFS Dark Purple 2	57	92	42	63	71	25	49	#471931
IFS Dark Purple 1	49	91	35	40	107	38	74	#6B264A
IFS Mid Purple	41	88	28	19	143	51	99	#8F3363
IFS Light Purple 1	20	66	9	0	205	115	162	#CD73A2
IFS Light Purple 2	12	45	7	0	222	162	193	#DEA2C1
IFS Light Purple 3	6	24	4	0	238	208	224	#EED0E0

Chart colour overview

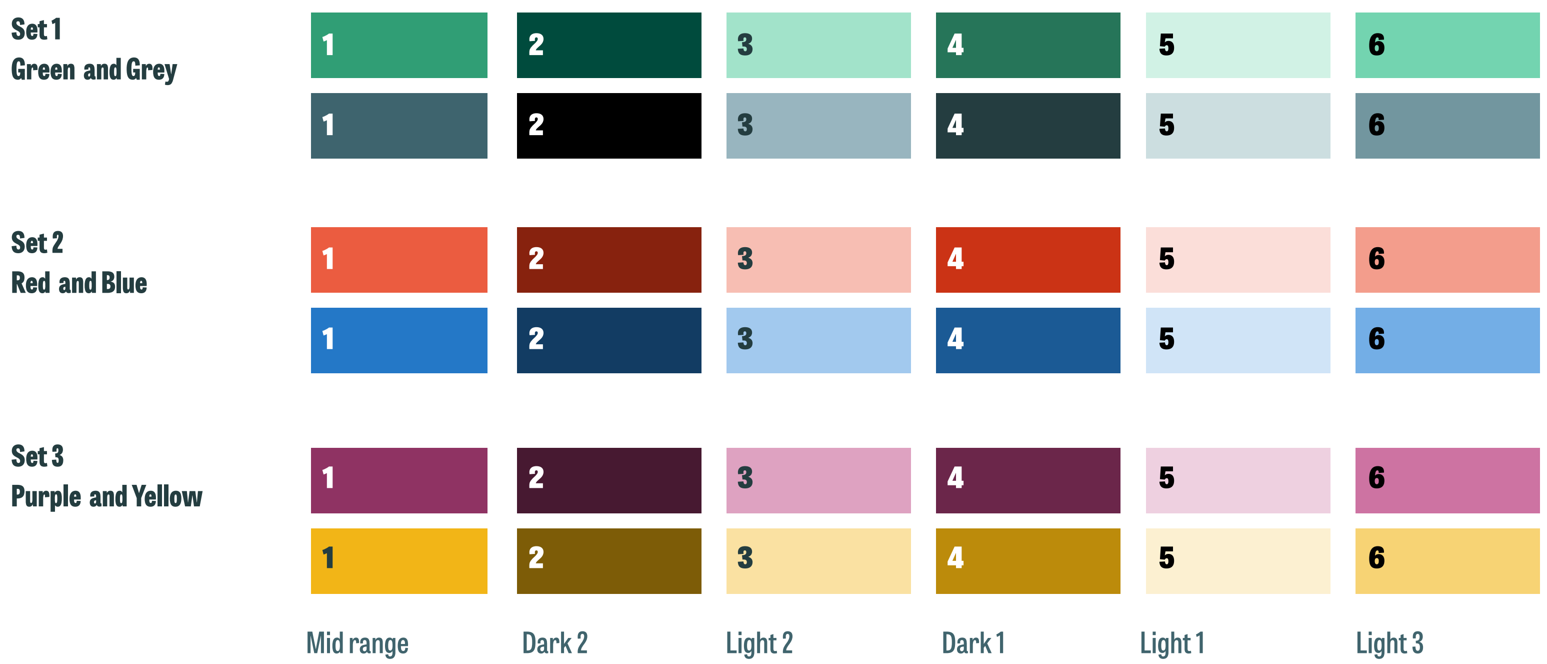
Designing consistent and legible charts, maps, and infographics is vital to ensure your information can be understood by the widest audience, colour application is a crucial component of this.

The IFS colour palette has been created to give optimal contrast when using data. Certain sets of colours from the primary and secondary palette can be combined when designing more complex graphics.

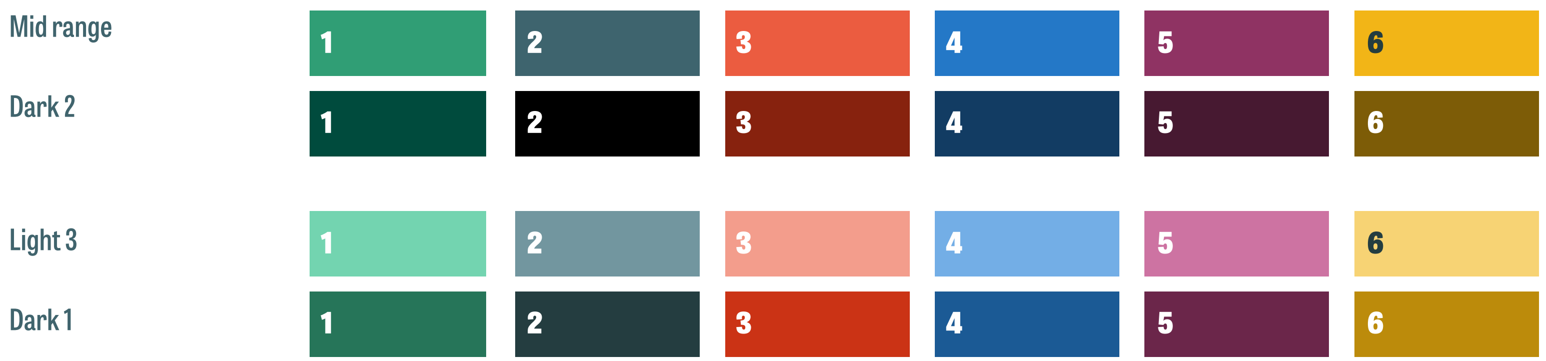
You can either start with a single tonal colour set or run through each hue to create brighter charts. The colours as indicated are the recommended sequence to provide optimal contrast within data sets. Always use the colour values as specified on pages 20–22. For more information on tone and hue, see page 19.

The secondary colour pairs indicated here are best used together for legibility, and have been selected for their suitability when diverging sets are needed.

Tonal colour application and suggested sets



Hue colour application and suggested order



Institute for Fiscal Studies

Trustees' Report

Year ended 31 December 2019

Contents

Company information

Company registered office

7 Ridgmount Street

London

WC1E 7AE

Company registered number

0954616 (Incorporated in England and Wales)

Registered charity

258815

Company bankers

National Westminster Bank plc

City of London Office

1 Princes Street

London EC2R 8BP

Auditor

BDO LLP

55 Baker Street

London W1U 7EU

Introduction from the Chair of Trustees

I am pleased to present the Trustees' report of the activities of IFS in 2019, which marked the fiftieth anniversary of the founding of the Institute in 1969. To illustrate the contribution that IFS has made over the years to the use of economic analysis to inform policy, we held a series of celebratory events and published a suite of papers in our journal *Fiscal Studies*, written by IFS researchers, past and present. Covering topics such as taxation, the working-age benefit system, education, and pensions policy, the events brought together panels of experts from IFS and elsewhere, along with a range of stakeholders and members of the public.

Over the year, IFS published research findings on a wide range of topics, including its Green Budget and annual reviews of trends in living standards, poverty and inequality and of education spending, as well as a new review of local government spending. 2019 also saw the launch of a large and ambitious project, which, over the next five years, will examine inequality in its many forms and make policy recommendations. The Deaton Review of Inequalities is led by Nobel Laureate, Sir Angus Deaton, and has brought together leading academics both in economics and in other disciplines; the review will also hear evidence from policymakers and others with an interest, as well as focus groups drawn from the general public.

The academic excellence of the Institute's research and researchers has once again been recognised. Research Director, Professor Imran Rasul, received the Yrjö Jahnsson Award in Economics, an award for a European economist under 45 years old who has made a particularly significant contribution in theoretical and applied economic research. Research Director, Professor Rachel Griffith, became President of the Royal Economic Society, the Society's first female President in over 35 years. IFS's ESRC Centre Director, Professor Sir Richard Blundell, was elected by US National Academy of Sciences, as one of only six European 'foreign member associates'.

Of course, the year was also marked by significant political events, not least the general election in December. IFS researchers worked to inform the election debate, as well as discussion around key fiscal events, with background papers and briefings.

This report highlights these, along with a small selection of the research and activities that took place over the year.

Careful scrutiny of the finances of IFS is an important part of the Trustees' work; as ever, this has been helped by clear and timely presentation of the facts to the committee by IFS officials. Whilst we, in common with other organisations that seek funding for academic research, currently face challenges in raising the finances to cover our ambitious programme of work, I am reassured that our financial position is healthy. In 2020, IFS's ESRC Centre – which has now attained 'Institute status' – will receive a further five years of Research Council funding. This will greatly contribute to future stability. The Institute has also been successful in gaining 'impact acceleration' funding from the ESRC to broaden and deepen the impact of its research, which will be used to invest in digital expansion and public engagement. We have continued with this programme during 2020, for example producing a range of new digital materials and running a series of public lectures at venues across the UK.

At the time of writing IFS, along with much of the country, is in lockdown. This has not stopped the Institute continuing to produce up-to-date, careful analysis of the extraordinary events as we navigate a health and economic crisis. This is a tribute to the commitment and determination of the staff at IFS.

I would like to thank my fellow Trustees for giving their time and expertise so generously throughout the year.

David Miles

Chair of Trustees

Institute for Fiscal Studies

Objectives and activities

The objects of the Institute

The objects of IFS are the advancement of education, for the benefit of the public, by promotion on a non-political basis of the study and discussion of, and the exchange and dissemination of information and knowledge concerning, the economic and social effects and influences of:

- existing taxes;
- proposed changes in fiscal systems; and
- other aspects of public policy,

in each case whether in the United Kingdom (UK) or elsewhere in the world.

So as to advance these objectives, it is IFS's policy to retain the right to publish its reports openly in order to inform public debate and policymaking. The Members of the Executive Committee confirm that they have complied with the duty in Section 17 of the Charities Act 2011 and have taken due regard of the Charity Commission's general guidance on public benefit. Examples of how the Institute has aimed to meet its public benefit are given in the review of 2019, where the Institute's achievements are reported.

Strategic framework

IFS operates within a strategic framework agreed by the Executive Committee; the committee meets every year to discuss strategy with IFS staff, discuss issues and difficulties, and agree on objectives. These discussions cover maintaining excellence in research, preserving independence and impartiality in policy analysis, engaging with a wide range of stakeholders, financial viability and good management, good governance, and supporting Institute members.

How has the Institute tried to further these aims?

During the year, the Institute has carried out a wide range of research and has publicised the resulting findings as widely as possible through publications and conference participation, on its own website and in the media. We believe that success lies in the scientific quality of our research and the efficacy with which our findings have informed the public debate. The following pages outline how this has been done.

Review of 2019

In 2019, IFS continued to undertake rigorous research to inform public understanding of crucial policy issues. IFS research spans a broad spectrum of topics and is presented to and discussed with audiences from academics at international conferences to UK policymakers to undergraduate students.

Academic excellence

In recognition of the contribution made by IFS research and researchers to the advancement of economic understanding, a number of staff received awards and honours for their work.

- Research Director, Professor Imran Rasul, received the Yrjö Jahnsson Award in Economics (joint with Oriana Bandiera). The award is given to a European economist under 45 years old who has made a particularly significant contribution in theoretical and applied economic research. Imran was also elected as Fellow of the British Academy.

- Research Director, Professor Rachel Griffith, became President of the Royal Economic Society. She is the Society's first female President in over 35 years and becomes only the second woman to hold the post in the Society's 129-year history.

- IFS's ESRC Centre Director, Professor Sir Richard Blundell, was elected by US National Academy of Sciences, as one of only six European 'foreign member associates'. He is the only economist foreign member associate this year, and there is only one other in the UK in total. Members are elected in recognition of their distinguished and continuing achievements in original research.

- As an illustration of how research rooted in academic expertise can inform our understanding of society, IFS won the Royal Statistical Society's 'Statistic of the Year' award. The winning statistic was 58%: the proportion of those in relative poverty who live in a working household. The judging panel chose this figure as it highlights both the growth of in-work poverty and the need to rise to fresh welfare challenges. The last 20 years have seen a major shift in Britain, from poverty being largely seen as a problem of unemployment to an issue that is now seen to afflict working households too. The number was taken from the IFS report, 'Living standards, poverty and inequality in the UK: 2019'.

Key new research grants

IFS research is funded through research grants, from the UK Research Councils and elsewhere (see financial review on page 21 for details). During the year, IFS was notified of the outcome of 54 research proposals, of which 29 were approved for funding (54% success rate). Given that the length of the decision process varies somewhat across funders, the number of applications evaluated was broadly comparable to 2018, but there was a more noticeable drop in the success rate (in 2018: 62 evaluated, 46 approved, success rate of 74%). A total of 99 funded research projects were active in 2019, which is around the same number as in 2018 (102).

Deaton Review of Inequalities

A major new project, one of the largest IFS has ever undertaken, was launched in May 2019, to look at inequalities which are at the forefront of today's public and policy debates. They have been linked to some of the most important political events and have sparked worldwide protest movements. There could hardly be a more pressing time to understand how inequalities arise, which ones matter, why they matter and how they should be addressed.

We see inequalities all the time, whether at the school gates, the hospital, when travelling round the country – or even a single town – or when turning on the news. But at any moment we typically encounter, or hear about, one specific type of inequality, a specific alleged cause of it or a specific proposed solution. Inequalities are too pervasive and too complex for us to stop at that. We need to step back and ask: how are different kinds of inequality related, which matter most, what are the big forces that combine to create them and what is the right mix of policies to tackle them?

The IFS Deaton Review, led by Sir Angus Deaton, aims to rise to that challenge. In the most ambitious study of its kind yet attempted, with funding from the Nuffield Foundation, we will aim to understand inequality not just of income, but of health, wealth, political participation, and opportunity; and not just between rich and poor but by gender, ethnicity, geography, age and education. We will cover the full breadth of the population – not just what is happening at the very top and very bottom. We will examine what concerns people about inequality, what aspects of it are perceived to be fair and unfair, and how those concerns relate to the actual levels of inequality and the processes by which they are created. We will examine the big forces that drive inequalities – from technological change, globalisation, labour markets and corporate behaviour to family structures and education systems.

The project involves a larger number of IFS staff, as well as researchers elsewhere in the UK and overseas. Over the course of the next four years, the Review will draw on the leading minds across the social sciences to assemble the evidence on the causes and consequences of different forms of inequalities, and the ways that they can best be reduced or mitigated.

A multinational, multi-disciplinary panel of experts are leading the review. The panel comprises:

- Angus Deaton, Princeton
- Orazio Attanasio, IFS and Yale
- James Banks, IFS and Manchester
- Lisa Berkman, Harvard
- Tim Besley, London School of Economics
- Richard Blundell, IFS and University College London
- Paul Johnson, IFS
- Robert Joyce, IFS
- Kathleen Kiernan, York
- Pinelopi Koujianou Goldberg, Yale and World Bank
- Lucinda Platt, London School of Economics
- Imran Rasul, IFS and University College London
- Debra Satz, Stanford
- Jean Tirole, Toulouse School of Economics

More information can be found at: www.ifs.org.uk/inequality/

Workers in health and social care

The National Institute for Health Research (NIHR) has launched a new set of Research Units to undertake research to inform decision-making by government and arms-length bodies. King's College London is hosting the Health and Social Care Workforce Research Unit (HSCWRU), in partnership with the IFS and Imperial College London. Of every 100 people working in England today 13 of them have jobs in health and social care. Nearly £2 out of every £3 spent on the NHS goes on paying its staff. The HSCWRU aims to help government by providing the answers to the workforce questions that affect both the quality and cost of health and social care services. The Unit is tackling a set of research questions agreed by government. In particular, IFS research focuses on analysing recruitment and retention.

Doctors' labour supply

There is mounting evidence of wide variation across regions and providers in healthcare costs, treatments provided and patient outcomes. Variation exists even among clinicians working in the same hospitals and treating similar patients. This has led to recent efforts to understand better the effect of individual clinicians and different ways of organising care on patient outcomes and medical productivity. But there is still limited research on this topic relating to healthcare in the UK: the aim of this project, funded by the ESRC, is to bring about a step change in the understanding of the determinants of variation in patient outcomes arising from the organisation of medical professionals.

Graduate earnings

Newly available Longitudinal Education Outcomes (LEO) data shows how much UK graduates of different courses at different universities are earning either one, three or five years since graduating. They do this by linking up tax, benefits, and student loans data. IFS researchers have been using these data, in work funded by the ESRC, to increase understanding of the individual and social returns to higher education degrees over the entire lifecycle. The work will also advance the academic literature on modelling earnings dynamics. The research estimates the impact of undergraduate and postgraduate degrees in specific subjects and from specific institutions on the earnings and employment patterns of graduates over their lifetimes.

Sanitation and use of toilets in India

With funding from the ESRC, researchers are looking at investments in sanitation in India. The research will improve our understanding of the causes of low levels of investment in, and use of, preventative health care in low-income settings. In particular, it will examine how the bargaining process within households affects the use of community toilets, and consider what role the incentives provided to different household members affect the use of sanitation.

New businesses

The ESRC has awarded a New Investigator grant to early-career IFS researcher, Kate Smith. She is leading a comprehensive study of how specific features of the tax and broader policy environment affect the decision to start a business. This includes how to organise that business (e.g. the decision about whether to incorporate or not) and the ongoing decisions, such as on investment, that affect subsequent growth and survival. The research uses and develops state-of-the-art techniques in combination with novel panel data from UK administrative tax records.

School effectiveness

School effectiveness is most commonly assessed through 'value-added' scores, which measure school outcomes relative to predicted outcomes based on school inputs (e.g. pupil prior attainment). A major concern with these measures is that they might be biased by factors that are not allowed for in this prediction. The aim of this project, which has been funded by the ESRC, is to estimate accurately school effectiveness and use these estimates to assess the extent of the bias in commonly used value-added measures.

Pensions Consortium

IFS researchers have been pursuing an ongoing programme of work into pensions and saving. Funding was secured for two years from a consortium of funders across the pensions industry, as well as the ESRC. The last decade has seen substantial reforms affecting how people save for retirement. The programme of research is investigating changes in expectations, attitudes and behaviour that affect individuals' preparedness for retirement. Four main questions are included in the programme, which will help shape our understanding of how policy can support people in preparing for retirement:

- How have attitudes to saving and retirement changed in an evolving pensions landscape?
- Who is opting out after being automatically enrolled into a workplace pension?
- How are the self-employed saving for their retirement?
- How do people make choices between housing and pension saving?

Understanding Society: impact of tax and benefit changes

IFS researchers are using data from the Understanding Society survey – the largest longitudinal household panel study of its kind, which provides vital evidence on life changes and stability. In collaboration with researchers from the University of Essex and building on existing research, this work uses Understanding Society data in conjunction with IFS's tax and benefit microsimulation model, TAXBEN. The project is developing the infrastructure of the simulation model so that in future it can investigate the impacts of tax and benefit policies on different people over long periods of their life, rather than simply in a single snapshot.

To coincide with the Budget and other fiscal events, IFS produces uniquely high-profile analysis of the distributional and incentive effects of tax and benefit policy. This analysis has hitherto been restricted to the snapshot effects by the lack of suitable longitudinal data. The analysis, which is used by government to inform policy decisions, will in future be more insightful and illuminating. Current work is also looking in particular at the gender pay gap and producing a labour market model, to assess the role of gender differences in commuting patterns and job-skill mismatch in driving the gender pay gap over the lifecycle.

Labour market specialisation and low-skilled workers

The project, with funding from the Alan Turing Institute, examines the widespread concern that changes in the labour market may have curtailed the opportunities for workers to secure stable jobs offering decent prospects of career progression, and that the effects of these changes may have been especially severe for low-paid low-skilled workers. Central to this discussion are the impacts of technological advances which, while leading to higher aggregate income, might also affect inequality in labour market opportunities. Researchers are looking at the role of firm specialisation, including outsourcing, in shaping the opportunities of workers with different skills. The research employs data science techniques in conjunction with novel large-scale administrative data. These data will allow us to understand the changing structure of our labour market in unprecedented detail.

Communication and stakeholders

IFS won the 2019 Prospect award for best UK economic and financial affairs think tank. The award was given in particular for 'highly authoritative number-crunching on education and, following the 70th anniversary of the creation of the NHS, also health'.

In 2019, IFS received UKRI funding for five years specifically to enhance the impact of our research in the form of a renewed Impact Acceleration Account. This has been and will continue to be, used to develop our relationships with key stakeholders – business, central government, and local and devolved governments – and to improve the resources available to the public to aid their understanding of economic issues.

Conferences and lectures

2019 was the fiftieth anniversary of the foundation of IFS, and a number events took place to mark this birthday. In addition, with a general election and the ongoing debate about Brexit, our events brought together policymakers, academics and other experts to discuss significant issues of critical importance to the country. Some highlights are listed below.

- The 'IFS at 50' series of four events were attended by a total of 891 people, including over 200 delegates from national and local government and 185 from the private sector.
- More than 200 IFS staff, alumni and high profile invitees came to the Institute's fiftieth birthday celebration in May, including the then Secretary of State for Business, Energy and Industrial Strategy, Greg Clark MP.
- IFS held two general election events: a briefing event to launch manifesto analysis had 111 delegates, including 33 members of the press and 44 representatives from political parties and the civil service (including delegates from the head office of the Conservative, Labour and Liberal Democrat parties); and a panel event, including speakers from the Institute for Government and the UKRI, 'Separating fact from fiction' attracted 314 attendees from a broad range of backgrounds.
- We launched the IFS Deaton Review to a high-profile audience of 137, including six lords (including a former Chancellor), two MPs (including a former Leader of the Opposition), the adviser to the Prime Minister on equalities, the Head of Analysis at the Government Equalities Office, the Chief Economic Advisor at HMT, the Chief Economist at DfE, senior professors from a range of universities, the Chief Executive of the Behavioural Insights Team, the Chief Executive of the British Academy, the Global Head of Research at Citigroup, and the General Secretary of the TUC.
- Fifteen IFS researchers presented papers at the Royal Economic Society annual conference, held at the University of Sussex. In addition, researchers presented at a range of international conferences including the annual conferences of the American Economic Association in Atlanta and of the European Economic Association in Manchester.
- The IFS annual lecture with Penny Goldberg (World Bank and Yale) was attended by 297 people; this made it the best-attended annual lecture since at least 2012.

- To cope with increasing demand, the IFS Green Budget launch was spread out over three events: a press briefing, a corporate member briefing and a public briefing, which was attended by over 250 people.
- Researchers gave two public talks in partnership with the University of Manchester this year: Jack Britton on 'Is it fair to charge £9,250 for university tuition fees?' and George Stoye on 'Who should pay for health and social care?' The first was to an audience of over 200 people, primarily university students but also containing some senior school students and members of the public.
- We held a series of four online-only events for the first time, as part of the ESRC Festival of Social Science. All four have been watched by between 200 and 300 people each. Funding from the ESRC was used to purchase high-quality filming equipment and will allow us to hold similar in-house events for free in the future.
- We held three joint debates with the Chartered Institute of Taxation, covering: 'The powers of HMRC and the responsibility of citizens in today's world', 'Taxing commercial property – time to tweak business rates or replace with a land value tax?'; and 'The digital services tax'. Together these debates attracted around 300 people from business and government.

Research findings and reports

A key strength of IFS is that its analysis of policy and its contributions to the public debate are grounded in rigorous empirical research. **Significant peer-reviewed journal publications**, in leading academic and field journals, produced by IFS researchers and associates included:

- Oriana Bandiera, Myra Mohnen, Imran Rasul, Martina Viarengo, 'Nation-building through compulsory schooling during the age of mass migration', *Economic Journal,* January 2019, 10.1111/ecoj.12624
- Raquel Bernal, Orazio Attanasio, Ximena Peña, Marcos Vera-Hernández , 'The effects of the transition from home-based childcare to childcare centers on children's health and development in Colombia', *Early Childhood Research Quarterly*, April 2019, 10.1016/j.ecresq.2018.08.005
- Mike Brewer, James Browne, Carl Emmerson, Andrew Hood, Robert Joyce, 'The curious incidence of rent subsidies: evidence of heterogeneity from administrative data', *Journal of Urban Economics*, November 2019, 10.1016/j.jue.2019.103198
- Jack Britton, Lorraine Dearden, Neil Shephard, Anna Vignoles, 'Is improving access to university enough? Socio-economic gaps in the earnings of English graduates', *Oxford Bulletin of Economics & Statistics*, April 2019, 10.1111/obes.12261
- Jonathan Cribb, 'Intergenerational differences in income and wealth: evidence from Britain', *Fiscal Studies*, October 2019, 10.1111/1475-5890.12202
- Rachel Griffith, Martin O'Connell, Kate Smith, Tax design in the alcohol market, *Journal of Public Economics*, April 2019, 10.1016.j.jpubeco.2018.12.005
- Bo Hou, James Nazroo, James Banks, Alan Marshall, 'Are cities good for health? A study of the impacts of planned urbanization in China', *International Journal of Epidemiology*, August 2019, 10.1093/ije/dyz031

Researchers published a range of reports relating to a broad spectrum of important policy areas.

The **IFS annual report on living standards, poverty and inequality** examines how living standards – most commonly measured by households' incomes – have changed for different groups in the UK, and the consequences that these changes have for income inequality and for measures of deprivation and poverty. In the 2019 report, we focussed in particular on those people who are poorest in society, with two of the three main chapters focusing on poverty.

In September, researchers launched the second **IFS annual report for 2019 on education spending** in England. Education spending is the second-largest element of public service spending in the UK behind health, representing about £91 billion in 2018–19 in today's prices or about 4.2% of national income. The level of UK education spending has risen significantly in real terms over time, growing particularly fast from the late 1990s through to the late 2000s, before falling in real terms from 2010 onwards. Whilst important, such overall trends in total education spending tell us little about what has happened to the different areas of education spending. The report provided measures of spending per student in the early years, schools, further education and higher education back to the early 1990s. These series of day-to-day spending per pupil allowed researchers to understand how policy decisions have affected the resources available to students in different stages of education over the long run, and to inform both policymakers and those seeking to influence policymakers about the extent to which funding is currently below previous peak levels in different areas. The report also analysed the effects of the 2019 Spending Round and the longer-term spending options for policymakers.

Researchers looking at **local government** published a number of reports during the year. These examined aspects of local government funding, devolved taxation and support schemes for localised council tax. In addition, we published the first of what will be an annual report on the state of local government spending in the UK.

Publications were also launched around **key political and fiscal events**.

To inform debate surrounding the Conservative leadership campaign, researchers published two briefing notes: 'Boris Johnson's tax policies: what would they cost and who would benefit?' and 'Jeremy Hunt's tax and spending policies: what would they cost and who would benefit?'.

The IFS Green Budget 2019 was published ahead of the Chancellor's Budget. When a government plans to pass a law, it often publishes a green paper. This is an opportunity to share its thinking and provoke

discussion. The Finance Bill is a law Parliament passes to renew taxes, propose new taxes and maintain the administration of the tax system. It enacts proposals announced in the Budget, which the Chancellor writes in secret. There's no green paper. This means important decisions about taxes, spending and public policy are made without consultation. So our annual Green Budget analyses the issues and challenges facing the Chancellor as he prepares for the Budget. The areas covered in the 2019 Green Budget by IFS researchers, and partners at Citi and the Institute for Government, were: the global outlook; recent trends to the UK economy; the UK economic outlook under different Brexit scenarios; the state of the public finances; fiscal targets and policy; the 2019 spending round; barriers to delivering new domestic policies; options for cutting direct personal taxes and supporting low earners; and a road map for motoring taxation.

Following the Budget statement itself, IFS researchers helped to explain its implications by answering questions from journalists from all the national papers, as well as conducting interviews on the BBC, ITV and other major broadcasters. In addition, as ever, research was disseminated via local radio and newspapers and through a range of online media outlets. Analysis was presented on the day following the Budget to journalists and key civil servants, to explain the implications for the public finances, businesses and households. Similar comment and analysis were carried out earlier in the year in response to the government's spending review.

Clearly a major feature of the 2019 political landscape was the snap election towards the end of the year. IFS created a special election website to share comment and new research with journalists and the public. We initially set out the six big economic challenges that needed addressing, which did not relate directly to Brexit. Although it was crucial to consider whether or not the UK was set to leave the European Union, and if so on what terms, other areas also needed to be explored and considered as part of the election debate.

Over the course of the election campaign, researchers published a series of briefing notes on topics relating to recent government policy and the parties' proposals, including:

- Levels and incidence of taxation, public spending and austerity; we looked in particular at the levels of taxation in other countries, with a view to putting some of Labour's proposed tax rises in context
- Early education and childcare spending, and higher education spending and reforms
- Public sector pay, employment, in-work poverty and the minimum wage
- Benefit changes, distributional impact, and increases in the state pension age
- Health care spending and provision
- Labour's nationalisation policy
- Distributional impact of personal tax and benefit reforms, 2010 to 2019
- Effect of taxes and benefits on UK inequality

During the year, a number of reports were published outlining the effectiveness of programmes that IFS researchers had studied, aiming to **improve outcomes for children or young people**. These included: the health effects of Sure Start in the UK; sustainable total sanitation in Nigeria; a home-visiting programme for disadvantaged young children in the UK; and a scheme promoting adolescent engagement, knowledge and health in Rajasthan, India.

Capacity building

IFS contributes to the UK social science environment by training excellent economists – both our own researchers and those working elsewhere. IFS researchers who move on typically take up positions in academia, or in the civil service or the media where they will put into practice the research and communication skills they have learned at the Institute. During 2019, in-house training for research staff included media training, writing and presentation skills, Stata and other analytical skills, while there was training for support staff in social media, design, membership management and other communication skills.

Six new graduate economists were taken on in 2019 (2018: two), as well as two postdoctoral researchers (2018: three). In addition to research staff, the capacity to support research and its dissemination was increased by adding a Head of Digital, whose role is to enhance our digital offering, in particular our website, and help bring IFS research findings to a wide audience.

The Institute also runs a summer internship programme, and in 2019 eight students (2018: eight) were employed for six-week placements, working with research teams on projects that gave them a taste of the type of work undertaken by new research economists.

In order to encourage diversity and openness in our recruitment process, we took steps during the year to make our recruitment materials more accessible, to provide information to demystify the recruitment and interview process, and to advertise our vacancies more widely to reach a wider group. We worked on ensuring that the language and imagery used in our recruitment and other materials reflect our policy to embrace diversity. IFS researchers and communications staff are involved in the Royal Economic Society initiative, #DiscoverEconomics, which aims to attract more women, minority students and students from state schools and colleges to study the subject at university. IFS has also been working with a range of think-tanks and social policy research organisations to run recruitment events aimed at minority and potentially disadvantaged groups.

Each year, IFS holds a day of talks on issues in public economics of interest to undergraduates in economics and related disciplines. The aim is to focus on the policy implications of research carried out at

the Institute. The day also includes a session with IFS researchers talking about their careers in order to promote both IFS recruitment opportunities and working as an economist in public policy more generally. . Over 200 students signed up to attend the 2020 lectures in London. Support from the ESRC allowed us to film the lectures and make them available online. As part of the ESRC Festival of Social Science, we also held a series of live-streamed talks for students, which have been watched over 1,000 times in total.

The Centre for Microdata Methods and Practice (Cemmap) at IFS provides training courses and masterclasses for policymakers, practitioners, academics and students. During 2019, five training courses (2018: four) were held, as well as two masterclasses (2018: three) and 24 seminars (2018: 27).

During the year, staff served on a number of boards and committees contributing to better policymaking and understanding of public policy. These included: Orazio Attanasio as President of the European Economic Association, and on the Council of the Royal Economic Society; Carl Emmerson on the Social Security Advisory Committee and the advisory panel of the Office for Budget Responsibility; Paul Johnson on the Committee on Climate Change and the Banking Standards Board; Robert Joyce on the Social Metrics Commission; Helen Miller as chair of the Royal Economic Society's Communications Committee; and David Phillips on the Welsh Government's Tax Advisory Group, and part of the Scottish Parliament's external expert panel.

2019 in numbers

Top five journal* articles past decade (2010 – 2019)	**59**
Top field journal◊ articles past decade	**150**

Front pages 2019	**129**
Press interviews 2019	**Today: 19 LBC: 19 BBC TV news: 14**

Academic and policy publications and events	2019	2018
Journal articles	42	44
Top five*	5	3
Top field journals◊	10	8
Working papers	64	68
IFS reports	43	39
Observations	36	29
Newspaper articles and blogs	61	73
IFS events	40	31
Event attendance	3,900	2,399
Hansard mentions	165	188

* *American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics, Review of Economic Studies*

◊ *Journal of Health Economics, Journal of Labor Economics, Journal of Human Resources, Review of Economic Dynamics, Journal of Public Economics, Journal of Econometrics, RAND Journal of Economics, The Review of Economics and Statistics, Journal of Economic Literature, The Economic Journal, Journal of the European Economic Association, European Economic Review, Journal of Monetary Economics, Quantitative Economics*

Public engagement	2019	2018
Press releases	32	30
Broadcast mentions	8,492	8,475
Print mentions	3,272	3,135
Front pages	129	82
Online mentions	20,479	17,068
Interviews given	180	160
Website visitors	710,570	503,057
Twitter impressions	738,000	663,000

Election highlights	2019	2017	2015
Broadcast mentions	3,494	2,196	2,300
Hard copy mentions	1,004	808	1,000
Front pages	58	30	75
Internet mentions	10,407	6,119	4,500
Election microsite visits	420,000	130,000	98,000
Election briefing notes	13		
Election Observations	12		
Election manifesto analysis views on YouTube	5,000		

Priorities for 2020 and beyond

Governance

In response to the coronavirus (COVID-19) crisis, IFS has taken a number of steps to ensure that work can continue, whilst protecting the health and safety of staff and partners. During lockdown, all staff have been able to work from home, with secure remote access to our internal network. Measures have been put in place to ensure that all staff are in regular contact with their managers and teams; staff meetings and seminars also take place frequently online.

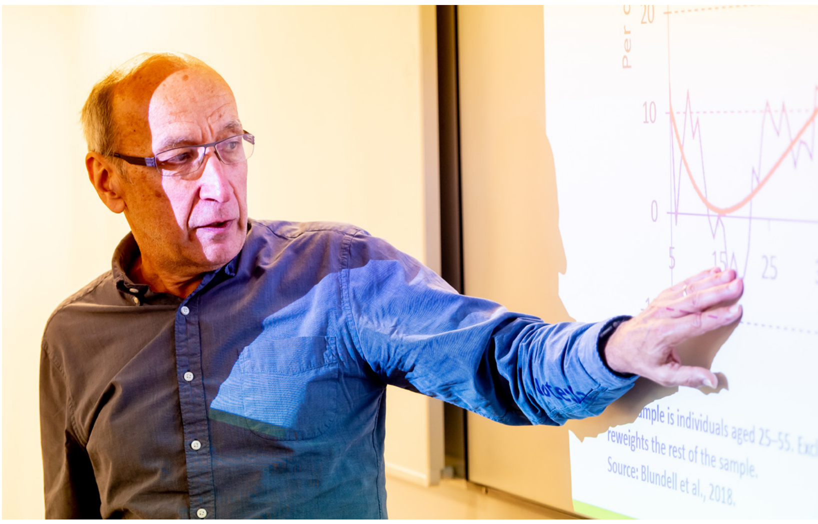
Some issues have arisen relating to access to sensitive data sets, used for a number of research projects. In consultation with data owners, we have found solutions to allow access to these in most cases. For a small number of projects, where data collection is underway, different arrangements are being implemented and projects rearranged to accommodate the current situation. At the time of writing, there are only two projects that have been delayed beyond their expected deadlines.

In line with our mission to inform the public debate and support policymakers, IFS researchers are responding to the crisis with comment and analysis relating to the economic consequences of the virus and the lockdown. As set out below, we are in the process of applying for further funding for a range of research to help inform policymaking decisions at this difficult time.

Academic excellence

The ESRC Centre for the Microeconomic Analysis of Public Policy (CPP) at IFS receives funding for five-year periods. It covers a broad research and dissemination programme and makes possible a flexible response to both scientific and policy developments. The CPP has been accorded Institute status – one of just two in the country. The ESRC states that this is 'to recognise its global centres of excellence with official ESRC Research Institute status. The move acknowledges those centres which have demonstrated sustained strategic value to the Council, as well as to the broader social science research landscape, with long-term, five-year funding.' Funding for five years from 2020 has been confirmed at a similar level to the current grant.

The current Centre is funded until Autumn 2020; after this, the new status should mean that this funding stream will be more reliable and will have the potential to cover a broader research programme. Over the past years of the Centre, research has spanned multiple subject areas, and has been unified by a desire to develop a rigorous empirical foundation for improving public policy in a changing economic and social environment. Our future agenda will continue this focus and also address new challenges. We will exploit new data, including administrative data linkages, in the UK, the US, European countries and developing countries. We will interact with researchers worldwide, exploiting our unique research environment for capacity building in empirical policy research.



The core objective of the ESRC Institute at IFS is to inform and improve the quality of public debates around economic policy in the UK and internationally. We do this by conducting world-class research, acting as a national resource by collaborating with a wide range of researchers in the UK and abroad, engaging with policymakers and practitioners, and building capacity through training new generations of researchers. We are strongly committed to bringing the high-quality and rigorous insights from our research, and the research of others, to bear on issues of current public interest through many forms of media and communication.

Our research agenda is ambitious and will yield policy-relevant academic research that we expect to make important scientific advances and result in papers published in the most prestigious peer-reviewed journals. This agenda is driven by our core areas of expertise, covers a broad spectrum of interrelated topics and is designed to address major challenges the UK and other economies face in ensuring the resilience of households, firms and the broader economy. It will continue to evolve in response to the changing policy landscape.

In terms of the research programme, our agenda is organised around five interconnected themes: inequalities and living standards; tax and benefit reform; human capital and productivity; the challenges of an ageing population; and demands on public expenditure and public services. We will carry out research in the areas of public finances and public spending, education and skills, health and social care, employment, pay and welfare, firm taxation and productivity, the tax and benefit system, consumer behaviour and indirect taxation, pensions and saving, devolution, the regions and local government, and policies and interventions in developing countries and in the UK.

Key new research grants

We expect to carry out research across the full range of our areas of expertise in 2020. The following specific projects are already funded and due to begin during the year.

Research relating to the consequences of the Coronavirus

In response to the urgent need for research to inform public policy in the face of the unprecedented situation brought about by the spread of the Coronavirus, we will divert resources to this area. With funding from the ESRC, UKRI, the Nuffield Foundation and the Standard Life Foundation, we will look at issues including:

- Food purchasing behaviour during the crisis
- Analysis of the impact of COVID-19 on other types of hospital care
- The impact of COVID-19 on personal finances
- Research to support fiscal policy decisions during the crisis
- The effect of the pandemic on families' time investments and child development
- The effects on gender and family inequalities
- The impacts of school and early years centre closures
- Sanitation and COVID-19 in developing countries

Deaton Review of Inequalities

As outlined above, work will continue on this project, in terms of the research programme, evidence gathering and communication with policymakers and the public.

Extending working lives

This project is a partnership and programme of work with the Centre for Ageing Better (CfAB) that will significantly expand the evidence base around paid work among those approaching later life. This is a crucial area of interest given increasing longevity at older ages. Extending working lives is a key government objective, and fulfilling work has proven potential to improve individuals' financial security, health and well-being into and through retirement.

Through this research we will be addressing the following important research questions.

- How is the nature of paid work at older ages evolving over time, in terms of the characteristics of employment and the rate of employment churn, and how does this vary across different types of individuals?
- How prevalent are different pathways into retirement – including via reduced hours, a 'bridge job' or a spell in self-employment? How is this changing over time, and how do pathways vary depending on individuals' characteristics and the nature of their work?

- What is the effect of the increase in the state pension age to 66 on the labour market activity of men and women at older ages?
- What is the effect of the increase in the state pension age to 66 on household incomes and living standards?
- Are emerging changes in patterns of paid work at older ages consistent with projections for future labour market activity produced by official forecasters?

Taxing sugary drinks

Eating too much sugar increases the risk of many health conditions, including diabetes, obesity, and dental caries. In the UK, people consume, on average, over 50% more added sugar than is recommended; adolescents and young adults consume more than double the recommended levels. Soft drinks are an important contributor to dietary sugar, contributing roughly 20% of the sugar consumption of adolescents and young adults, and 10% at other ages. Concern about the health costs generated by people eating too much sugar motivated the introduction of the UK's tax on soft drinks in 2018. With funding from the ESRC, our research aims to understand how effective the tax was at reducing sugar consumption, both in the whole population and among groups about whom policymakers are especially concerned, such as people with very high-sugar diets and young people. We also aim to understand how the tax reduced the amount of sugar that people buy. This will contribute to the evidence on the effectiveness of taxes on soft drinks and inform how their design could be improved.

Support for young people in education

The ESRC will fund a project to investigate the labour market impact of the Education Maintenance Allowance (EMA), a financial scheme that paid young people across the UK up to £3,600 for staying in education or training beyond age 16. While there is evidence that this programme increased post-16 education participation, it is not known whether this additional education has translated into better labour market outcomes later on. The research will use the LEO data set, a large administrative dataset containing complete linked education and tax records for anyone who has taken GCSEs in England since 2002.

Inequality across the generations

Children of economically successful parents tend to be economically successful themselves, having relatively high levels of income, education and wealth. By some estimates, the intergenerational association of economic status in the UK is the strongest in the developed world and has grown stronger in recent decades. Given that equality of opportunity is a common policy goal, disparities in economic outcomes determined by parental background are in urgent need of attention. Tackling such disparities requires an understanding of how they arise and which policies can have an impact. The ESRC will fund an ambitious programme of research that will further our understanding of the mechanisms whereby parents transmit their economic status to their children, and how this process is influenced by various government policies.

National Living Wage

In a project funded by the Low Pay Commission, researchers will provide a comprehensive assessment of the impact of the National Living Wage. It incorporates three strands: impacts on wages, hours, employment and earnings for those aged 25 and over; impacts on wages, hours, employment and earnings for those aged 24 and under; and impacts on family and household incomes

Family support through economic changes

There is growing recognition that sudden changes in workers' economic environment, through for instance changes in global trade patterns or advances in technology and automation, can have large and persistent effects on the labour market. However, much research on these questions has focused on the effect of these shocks on individual workers, ignoring the role of the family as a potential source of support for individuals. This research, funded by a grant from the British Academy, will shed new light on the role families play in helping workers adapt to economic shocks. The work will be directly relevant for policymakers designing measures to assist households affected by dislocations associated with globalisation and other labour market shocks.

Communication and stakeholders

As an institute our overarching aim is to conduct wide-ranging, high-quality microeconomic research to help inform evidence-based policymaking and improve the quality of public scrutiny and debate at local and national levels. Strengthening and extending our knowledge exchange and impact strategies and encouraging learning, development and innovation are therefore key to our success.

We have three strategies to help us achieve these key goals:

- Develop stronger engagement, relationships and impact with three key stakeholder groups: business, central government, and local and devolved governments. This in turn will lead to improved understanding, engagement and knowledge exchange. This will inform our own research programmes, help us to build coalitions of funders, and impact on these actors' understanding and policymaking.
- Improve public understanding of our research, economic principles and public policy. This is a huge task to set ourselves as a small organisation but it is an increasingly vital role for research organisations wanting to have the ultimate effect of improving policy. To provide information directly to the public, we have secured funding from Friends Provident to build a website, TaxLab, which will hold accessible materials, including videos, graphics and summaries of research, on the subject of tax. The aim is to explain the workings of the tax system and policy choices in a way that is interesting and relevant to members of the public.
- Train and develop research and support staff at all career stages. The ultimate objectives are to ensure the sustainability of our capacity to carry out excellent research, maximise its impact capacity and to ensure that we build on our past successes in creating new generations of researchers who can go on to influential positions in academia and public policy, where they can have long-term positive impact on policy and public understanding.

The situation with the Coronavirus will temporarily change the way research can be communicated. In particular, it will not be possible to hold events

to be attended in person. But we are investigating ways to launch work digitally, either with recorded presentations of work, other enhanced digital content, or live events conducted virtually.

Capacity building

Our aim to train and develop research and support staff at all career stages has the ultimate objectives stated above.

To this end, five new and recent graduates will start work at IFS in Autumn 2020. They will be trained in research and communication skills, working alongside more experienced researchers and Research Fellows and Associates, who are leaders in their fields from universities in the UK and overseas.

We plan to take on a further three Postdoctoral Fellows from September 2020 on two-year contracts, as well as an additional one-year placement for a post-doctoral researcher at a UK institution, with funding from the ESRC to increase the skills and policy understanding of early-career researchers.

The Institute will also host a number of graduate students, who will work on PhDs under the supervision of senior staff and work alongside researchers whose research interests they share. The specific expertise of these individuals will feed into related research programmes and will enrich the knowledge of colleagues through frequent seminars and interchange of views. The researchers themselves will also benefit from the stimulating intellectual environment at IFS and they are likely to go on to research or teaching posts in the future, where they will be able to apply what they have learned.

Under the auspices of the Centre for Microdata Methods and Practice (cemmap), we will continue to run training courses, masterclasses and workshops until the end of the academic year 2020–21.

Over the summer, we plan to host seven economics students in paid internships, although this will depend on lockdown conditions related to COVID-19. They will work on projects with IFS researchers to give them a flavour of what policy-relevant research is like. We also plan to host work experience students in collaboration with the Higher Education Access Network, as part of our commitment to diversity. Throughout our recruitment process, we will continue to look for ways to encourage diverse applicants to apply and to recruit staff from a range of backgrounds.

During 2020 we will launch a website, Communicating Economics, containing resources targeted at students, which will help communicate economic ideas to a range of audiences.

Strategic report

Financial review

The results for the year ended 31 December 2019 are presented in the statement of financial activities on page 30. The level of activity was very similar from 2018 to 2019. Total income was £9,272,321 (2018: £8,870,307) and total expenditure was £9,093,975 (2018: £8,630,409).

The statement of financial activities shows an overall surplus for the year ended 31 December 2019 of £178,346 (2018: £239,898), representing a surplus on charitable activities of £21,051 (2018: £117,907).

The Institute attempts to raise its research funds from a range of organisations so that it is not dependent upon a single source of funding. Although 40% of the income recognised in 2019 was provided by the Economic and Social Research Council, this funding covered a wide range of projects through over 20 different grants.

The investment policy of the Executive Committee has been to invest cash reserves in interest-bearing accounts and not to risk any of the principal. At the end of the year, £1,239,733 was held in a COIF Charities Deposit Fund (2018: £1,232,717) and £3,120,380 (2018: £3,197,027) was held in cash. The CAF Bond held with Principality Building Society (£515,151 at 31 December 2018) matured during the year and was held in cash or invested in short term deposits.

Reserves policy

The reserves policy is twofold: one, to hold funds for working capital purposes and as a contingency, should sufficient new funding not emerge or should existing contracts be cancelled; and two, to reflect the net book value of fixed assets.

As at 31 December 2019, the Institute's total reserves were £2,984,770 (2018: £2,806,424), comprising the unrestricted General Fund of £2,853,977 (2018: £2,642,079) and the unrestricted Fixed Asset Fund of £130,793 (2018: £164,345).

The General Fund reflects the Institute's net current

	2019 Unrestricted	2019 Restricted	2019 Total	2018 Unrestricted	2018 Restricted	2018 Total
Cash and cash equivalents	3,022,054	1,338,059	**4,360,113**	2,850,062	2,094,833	**4,944,895**
Less net grants received in advance	96,075	(1,208,948)	**(1,112,873)**	(88,457)	(2,174,807)	**(2,263,264)**
Cash holdings (excluding net project grants received in advance)	3,118,129	129,111	**3,247,240**	2,761,605	(79,974)	**2,681,631**
Other working capital	(264,152)	(129,111)	**(393,263)**	(119,526)	79,974	**(39,552)**
General Fund	2,853,977	-	**2,853,977**	2,642,079	-	**2,642,079**
No. of months of forecast expenditure (excluding direct project costs)	4.8 months			4.8 months		
Target level for the General Fund: (6 months' forecast expenditure, excluding direct project costs)	£3.5m			£3.3m		

assets and is considered to be the amount of reserves that could be easily converted to cash, should the need arise. The target is for the General Fund to be maintained at a level to cover up to six months' expenditure (excluding direct project costs). The Trustees wish to continue to raise modest surpluses so that the General Fund meets this target.

The Fixed Asset Fund was established in 2010 such that this fund would be equivalent in value to the net book value of the Institute's fixed asset. The value of IFS fixed assets was lower at year-end than at the beginning of the year and so the fund has been decreased accordingly with a transfer to the IFS General Fund. The Reserves Policy is subject to active review in the light of prevailing circumstances.

Principal risks and uncertainties

The Executive Committee has overall responsibility for ensuring that the Institute has appropriate systems of control, both financial and operational. These systems are designed to provide reasonable, but not absolute, assurance against material misstatement or loss.

During the year, the Executive Committee continued to review the major financial and operational risks facing the Institute. It continues to monitor, on an annual basis, the implementation of any changes necessary to ensure that, as far as is reasonable, controls are in place to protect the Institute, its members, its staff, the general public and other stakeholders.

The primary risks relate to financial issues and in particular to the reliance on the ESRC for a large proportion of the Institute's research funds. However, this funding represents a mix of long-term and short-term funding, which reduces the immediate risk. Additionally, a significant proportion of our staffing costs relates to staff from UK universities whose funding is explicitly aligned with ESRC funding, meaning that these costs can be reduced or terminated in line with the funding stream. The Institute continues to seek to diversify its funding sources in order to spread the risk.

Another key risk is in relation to our people and the risk of losing key staff. We attach a high priority to supporting our staff in developing their skills, whether through further study or by giving them opportunities to become involved with all aspects of research and communication throughout their careers. New Research Economists are provided with mentors and are given the opportunity to take on managerial responsibility as and when they are ready. Staff representatives, elected by peers, include in their remit the discussion of staffing issues with senior management. Regular reviews of selection procedures and conditions of service take place, together with periodic monitoring of salaries offered elsewhere. Staffing requirements are planned as far in advance as possible, and good relationships are maintained with top universities and institutions, both in the UK and overseas.

IFS is a leading academic institute, and it is imperative to maintain the quality of our research. Quality assurance procedures are in place that require the involvement of senior staff for all projects. Staff adhere to the IFS code of good practice in research, Social Research Association (SRA) ethical guidelines, and rulings of the UCL Research Ethics Committee. Any interactions with research participants are governed by this code and by established ethics principles and obligations. There is regular discussion of ongoing research at senior management meetings and, in addition, the Advisory Boards for the ESRC Centres have oversight of the Centres' research programmes.

Like all organisations, IFS has been affected by the COVID-19 pandemic and the resultant economic effects and uncertainty. As an organisation, our primary concern is the safety and well-being of our employees and their families, our research partners and suppliers. The future impact of the outbreak is uncertain and amongst other things will depend on actions taken to contain the coronavirus. The Executive Committee considers that IFS has adequate financial resources and is well placed to manage the risks associated with the COVID-19 pandemic. Financial projections, scenario testing and key risk identification have taken into consideration the current and expected economic climate, and its potential impact on IFS's sources of income and planned expenditure. There were no circumstances which arose from this scenario testing and the COVID-19 pandemic that resulted in an adjustment to the IFS financial statements as at 31 December 2019.

In March 2020, the UK government implemented significant measures to contain the spread of COVID-19, which had been declared as a pandemic by the World Health Organisation (WHO) during the same month. The measures implemented in the UK are similar to those implemented in many other countries around the world, and have significantly impacted many businesses, both operationally and financially.

For IFS, the key impacts that may arise or have arisen include:

- The cancellation of events hosted by IFS, in the interest of the health and safety of our stakeholders and in compliance with government guidelines to limit the risk of transmission of the virus
- The reduction to income that may arise from delays to our research activities or those of our research partners
- The operational impact to IFS arising from the risk of staff illness

As there remains uncertainty around the period over which governments' measures worldwide will remain in place, the precise impact cannot be determined. However, IFS has modelled the impact of these measures assuming that cash flows on a number of research projects may be delayed, in some cases for up to 12 months. This work concluded that IFS will have sufficient liquid resources (cash and investments that can be converted to cash) to continue to operate for at least 12 months from the date of approval of these financial statements.

This work also considered the possible implementation of risk mitigation measures that are available to IFS, including:

- Associated cost savings from the postponement/cancellation of work or events
- Deferral of projects that were planned to be executed in 2020
- Modifying our operations to conform to the current circumstances

The Executive Committee also considered other scenarios in which IFS may not receive the expected income and cash flows for 2020. Under all scenarios modelled, IFS would still have sufficient resources to be able to fulfil its existing commitments for the next 12 months. The Executive Committee remains of the view that there are no material uncertainties that call into doubt IFS's ability to continue. The financial statements have therefore been prepared on the basis that IFS is a going concern.

Governance and management

Constitution

The Institute for Fiscal Studies (IFS) was incorporated by guarantee on 21 May 1969. It is a private company limited by guarantee and has no share capital. It is a registered charity. The guarantee of each Company Law member ('Member') is limited to £1. The governing document is the Memorandum and Articles of Association of the Company and the members of the Executive Committee are the Directors of the Company and the Trustees.

Company Law members consist of the IFS Council members. At the end of November 2019, the number of guarantors was therefore 41 (30 at the end of November 2018). The Articles contain the provision that the IFS Council be expanded to no more than 50 persons and that when complete it shall consist of 45 members elected by Council and five members elected by the wider IFS membership.

Members of the Executive Committee

The Executive Committee, which is made up of the Trustees of the Institute, is established by the IFS Council: Trustees are elected by the Council from among themselves, and consist of at least seven and no more than twelve people, one of whom is the President of the Council. Trustees serve three-year terms, and will usually only serve a maximum of three terms. The Executive Committee met five times during the year. Committee membership during 2019 was:

- Jonathan Athow
- James Bell
- John F. Chown
- Margaret Cole
- David Gregson
- Caroline Mawhood
- Ian Menzies-Conacher (retired November 2019)
- David Miles (Chair)
- Gus O'Donnell (President, IFS Council)
- Michael Ridge
- Nicholas Timmins

As part of the organisation's governance review (see below), the Executive Committee set up two committees during 2019 to help improve scrutiny of the Institute's operations. These are a Nominations Committee and an Audit Committee. The remits of the committees are as follows.

Audit Committee

The Audit Committee's overall objective is to give advice to the Executive Committee on

- The overall processes for risk, control and governance
- Management assurances and appropriate actions from external audit and internal audit (if appropriate) findings, risk analysis and reporting undertaken
- The financial control framework and supporting compliance culture
- Accounting policies and material judgements, the accounts and the annual report and management's letter of representation to the external auditors
- Whistle-blowing arrangements for confidentially, raising and investigating concerns over possible improprieties in the conduct of IFS business
- Processes to protect against fraud and corruption
- The planned activity of internal audit (if appropriate) and external audit

Nominations Committee

The Nominations Committeee's roles are

- To develop and maintain rigorous and transparent procedures for appointments and re-appointments to the Council and the President, Trustees and its committees. To propose candidates for appointment to the Council and to the board of trustees.
- To formulate plans for succession and ensure

that there is a transparent and fair procedure for the appointment of the President, Chair of Trustees, honorary officers and members of the Council and Board of Trustees

- To review regularly the composition of the Board and its committees (including their diversity, balance of skills, knowledge and experience) and make recommendations to the Board with regard to any adjustments that are deemed necessary
- To review the results of the Board performance evaluation process that relate to the composition of the Board

Induction and training of Trustees

New Trustees receive training and induction following their appointment. Trustees are kept up-to-date with IFS research by a rolling programme of research presentations made at each meeting of the Executive Committee.

Remuneration policy

The salary of the Director is determined by the Executive Committee when renewing his contract and is normally adjusted each year for a cost-of-living adjustment, in line with salaries across the Institute. The pay of all other staff is reviewed by the Director and, where appropriate, other members of senior management annually and is also usually increased by a cost-of-living adjustment. From time to time, the salary scales of the Institute are benchmarked against comparable organisations. In 2019, the services of the Research Directors, Orazio Attanasio and Rachel Griffith, were provided by UCL and the University of Manchester respectively under contracts that reimburse the universities for an agreed percentage of the individual's salary, National Insurance and pension costs. Further details on these amounts are included in note 8 to the accounts.

Organisational structure of the Institute and the decision-making process

The overall management of IFS is carried out by the Director, Paul Johnson, who reports to the Trustees on a quarterly basis. The Director is part of the senior management team of the Institute, which also comprises the Deputy Director, Carl Emmerson, and the Research Directors, Professors Rachel Griffith, Fabien Postel-Vinay and Imran Rasul (the last twobecame Research Directors in 2020).

The Executive Committee delegates the operational responsibilities of the Institute via a 'Scheme of Delegation' to the Director of the Institute, who in turn delegates various duties to senior staff.

The Institute employed directly an average of 82 (2018: 83) full- and part-time staff based at its offices in London. Research staff are divided into sectors, and a small core of administrative and secretarial staff provide support facilities.

The Institute also employed indirectly 15 (2018: 16) senior academic staff based at UK universities on a part-time basis. In addition, a number of other academics from both UK and overseas institutions work with the staff as Research Fellows and Research Associates on an ad hoc collaborative basis.

Statement of policy on fundraising

Section 162a of the Charities Act 2011 requires us to make a statement regarding fundraising activities. We do not undertake widespread fundraising activities with members of the public, although we do accept donations or offers from partners to contribute to work that we undertake. The legislation defines fundraising as 'soliciting or otherwise procuring money or other property for charitable purposes'. Such amounts receivable are presented in our accounts as 'donations and legacies'. We do not use professional fundraisers or 'commercial participators' or any other third parties to solicit donations. We are therefore not subject to any regulatory scheme or relevant codes of practice, nor have we received any complaints in relation to fundraising activities.

Charity Governance Code

In July 2017, the new Charity Governance Code was published setting out recommended practice. The Executive Committee is supportive of the broad principles set out in the code and is keen to ensure that these are built into the governance of the organisation. To this end, during 2019 Trustees carried out a detailed review of its governance policies and procedures with reference to the code.

In consultation with IFS's senior staff, the committee considered each of the code's provisions to ascertain whether the organisation and Trustees already complied, were in the process of implementing changes to ensure compliance, or did not comply. The Committee concluded that the IFS’s governance was effective subject to improvements in selected areas and an action was put in place to address these areas. This process also involved considering how to ensure that the organisation is governed, and run on a day-to-day basis, in a way that is commensurate with the broad principles recommended for charity governance.

Trustees' responsibilities

The Trustees are responsible for preparing the Trustees' annual report and the financial statements in accordance with applicable law and United Kingdom Accounting Standards (United Kingdom Generally Accepted Accounting Practice, including Financial Reporting Standard 102 *The Financial Reporting Standard applicable in the UK and Republic of Ireland)*.

Company law requires the Trustees to prepare financial statements for each financial year. Under company law, the Trustees must not approve the financial statements unless they are satisfied that they give a true and fair view of the state of affairs of the charity and of the incoming resources and application of resources, including income and expenditure, of the charity for the year. In preparing those financial statements, the Trustees are required:

- To select suitable accounting policies and then apply them consistently
- To observe the methods and principles in the Charities SORP
- To make judgements and accounting estimates that are reasonable and prudent
- To prepare the financial statements on the going-concern basis unless it is inappropriate to presume that the charity will continue in business

The Trustees are responsible for keeping adequate accounting records that are sufficient to show and explain the charity's transactions, to disclose with reasonable accuracy at any time the financial position of the charity and to enable them to ensure that the financial statements comply with the requirements of the Companies Act 2006. They are also responsible for safeguarding the assets of the charity and hence for taking reasonable steps for the prevention and detection of fraud and other irregularities.

- So far as each of the Trustees at the time the report is approved are aware
- There is no relevant audit information of which the auditor is unaware
- They have taken all the steps they ought to have taken to make themselves aware of any relevant audit information and to establish that the auditor is aware of that information

The Trustees are responsible for the maintenance and integrity of the corporate and financial information included on the charity's website. Legislation in the UK governing the preparation and dissemination of the financial statements and other information included in annual reports may differ from legislation in other jurisdictions.

Approved and authorised for issue by the Executive Committee and signed on its behalf by

David K. Miles

David Miles, Chair of the Executive Committee

Company registered number: 0954616
Registered charity: 258815

Auditor's report

Indepedent Auditor's report to the members of the Institute for Fiscal Studies

Opinion

We have audited the financial statements of the Institute for Fiscal Studies for the year ended 31 December 2019 which comprise the statement of financial activities, the balance sheet, the statement of cash flows and notes to the financial statements, including a summary of significant accounting policies. The financial reporting framework that has been applied in their preparation is applicable law and United Kingdom Accounting Standards, including Financial Reporting Standard 102 *The Financial Reporting Standard applicable in the UK and Republic of Ireland* (United Kingdom Generally Accepted Accounting Practice).

In our opinion, the financial statements:

- give a true and fair view of the state of the Charitable Company's affairs as at 31 December 2019 and of its income and expenditure for the year then ended;
- have been properly prepared in accordance with United Kingdom Generally Accepted Accounting Practice; and
- have been prepared in accordance with the requirements of the Companies Act 2006.

Basis for opinion

We conducted our audit in accordance with International Standards on Auditing (UK) (ISAs (UK)) and applicable law. Our responsibilities under those standards are further described in the Auditor's responsibilities for the audit of the financial statements section of our report. We are independent of the Charitable Company in accordance with the ethical requirements relevant to our audit of the financial statements in the UK, including the FRC's Ethical Standard, and we have fulfilled our other ethical responsibilities in accordance with these requirements. We believe that the audit evidence we have obtained is sufficient and appropriate to provide a basis for our opinion.

Conclusions related to going concern

We have nothing to report in respect of the following matters in relation to which the ISAs (UK) require us to report to you where:

- the Trustees' use of the going concern basis of accounting in the preparation of the financial statements is not appropriate; or
- the Trustees have not disclosed in the financial statements any identified material uncertainties that may cast significant doubt about the Charitable Company's ability to continue to adopt the going concern basis of accounting for a period of at least twelve months from the date when the financial statements are authorised for issue.

Other information

The other information comprises the information included in the report, other than the financial statements and our auditor's report thereon. The Trustees are responsible for the other information.

Our opinion on the financial statements does not cover the other information and, except to the extent otherwise explicitly stated in our report, we do not express any form of assurance conclusion thereon.

In connection with our audit of the financial statements, our responsibility is to read the other information and, in doing so, consider whether the other information is materially inconsistent with the financial statements or our knowledge obtained in the audit or otherwise appears to be materially misstated. If we identify such material inconsistencies or apparent material misstatements, we are required to determine whether there is a material misstatement in the financial statements or a material misstatement of the other information. If, based on the work we have performed, we conclude that there is a material misstatement of this other information, we are required to report that fact.

We have nothing to report in this regard.

Opinions on other matters prescribed by the Companies Act 2006

In our opinion, based on the work undertaken in the course of the audit:

- the information given in the Trustees' Report, which includes the Directors' Report and the Strategic report prepared for the purposes of Company Law, for the financial year for which the financial statements are prepared is consistent with the financial statements; and
- the Strategic report and the Directors' Report, which are included in the Trustees' Report, have been prepared in accordance with applicable legal requirements.

Matters on which we are required to report by exception

In the light of the knowledge and understanding of the Charitable Company and its environment obtained in the course of the audit, we have not identified material misstatements in the Strategic report or the Trustee's report.

We have nothing to report in respect of the following matters in relation to which the Companies Act 2006 requires us to report to you if, in our opinion:

- adequate accounting records have not been kept by the Charitable Company, or returns adequate for our audit have not been received from branches not visited by us; or
- the Charitable Company financial statements are not in agreement with the accounting records and returns; or
- certain disclosures of Directors' remuneration specified by law are not made; or
- we have not received all the information and explanations we require for our audit; or
- the trustees were not entitled to prepare the financial statements in accordance with the small companies regime and take advantage of the small companies' exemptions in preparing the directors' report and from the requirement to prepare a strategic report.

Responsibilities of Trustees

As explained more fully in the Trustees' responsibilities statement, the Trustees (who are also the directors of the charitable company for the purposes of company law) are responsible for the preparation of the financial statements and for being satisfied that they give a true and fair view, and for such internal control as the Trustees determines is necessary to enable the preparation of financial statements that are free from material misstatement, whether due to fraud or error.

In preparing the financial statements, the Trustees are responsible for assessing the Charitable Company's ability to continue as a going concern, disclosing, as applicable, matters related to going concern and using the going concern-basis of accounting unless the Trustees either intend to liquidate the Charitable Company or to cease operations, or have no realistic alternative but to do so.

Auditor's responsibilities for the audit of the financial statements

We have been appointed as auditor under section 144 of the Charities Act 2011 and report in accordance with the Act and relevant regulations made or having effect thereunder.

Our objectives are to obtain reasonable assurance about whether the financial statements as a whole are free from material misstatement, whether due to fraud or error, and to issue an auditor's report that includes our opinion. Reasonable assurance is a high level of assurance, but is not a guarantee that an audit conducted in accordance with ISAs (UK) will always detect a material misstatement when it exists. Misstatements can arise from fraud or error and are considered material if, individually or in the aggregate, they could reasonably be expected to influence the economic decisions of users taken on the basis of these financial statements.

A further description of our responsibilities for the audit of the financial statements is located at the Financial Reporting Council's ("FRC's") website at: https://www.frc.org.uk/auditorsresponsibilities.
This description forms part of our auditor's report.

Use of our report

This report is made solely to the Charitable Company's members, as a body, in accordance with Chapter 3 of Part 16 of the Companies Act 2006. Our audit work has been undertaken so that we might state to the Charitable Company's members those matters we are required to state to them in an auditor's report and for no other purpose. To the fullest extent permitted by law, we do not accept or assume responsibility to anyone other than the Charitable Company and the Charitable Company's members as a body, for our audit work, for this report, or for the opinions we have formed.

Fiona Condron (Senior Statutory Auditor)

For and on behalf of BDO LLP, statutory auditor

London

Date: 7 July 2020

BDO LLP is a limited liability partnership registered in England and Wales (with registered number OC305127).

Financial Reports

Statement of financial activities

Year ended 31 December	Notes	2019 Unrestricted £	2019 Restricted £	2019 Total £	2018 Total £
Income from:					
Membership and donations	2	191,944	-	191,944	156,853
Charitable activities	3	1,067,366	7,989,045	9,056,411	8,695,250
Investment income	4	18,196	-	18,196	15,835
Other income		5,770	-	5,770	2,369
Total income		**1,283,276**	**7,989,045**	**9,272,321**	**8,870,307**
Expenditure on:					
Raising funds	6	58,615	-	58,615	53,066
Charitable activities	6	820,642	8,214,718	9,035,360	8,577,343
Total expenditure		**879,257**	**8,214,718**	**9,093,975**	**8,630,409**
Net income/(expenditure)		**404,019**	**(225,673)**	**178,346**	**239,898**
Transfers between funds	13	**(225,673)**	**225,673**	**-**	**-**
Net movement in funds		**178,346**	**-**	**178,346**	**239,898**
Reconciliation of funds:					
Total funds brought forward	13	**2,806,424**	**-**	**2,806,424**	**2,566,526**
Total funds carried forward	13	**2,984,770**	**-**	**2,984,770**	**2,806,424**

There were no other recognised gains or losses other than the net income for the year. All amounts relate to continuing operations.

Balance sheet

As at 31 December	Notes	2019 £	2018 £
Fixed assets			
Tangible assets	10	130,793	164,345
Total fixed assets		**130,793**	**164,345**
Current assets			
Debtors	11	2,081,457	1,420,074
Short-term deposits		1,239,733	1,747,868
Cash at bank and in hand		3,120,380	3,197,027
Total current assets		**6,441,570**	**6,364,969**
Liabilities:			
Creditors: amounts falling due within one year	12	**(3,587,593)**	**(3,722,890)**
Net current assets		**2,853,977**	**2,642,079**
Net assets		**2,984,770**	**2,806,424**
Total funds:			
Unrestricted funds			
- General Fund	13	2,853,977	2,642,079
- Fixed Asset Fund	13	130,793	164,345
		2,984,770	**2,806,424**
Restricted	13	-	-
Total		**2,984,770**	**2,806,424**

Approved and authorised for issue by the Executive Committee and signed on its behalf by

David K. Miles

....

David Miles, Chair of the Executive Committee

7 July 2020

Company registered number: 0954616

Registered charity: 258815

Statement of cash flows

Year ended 31 December	2019 £	2018 £
Reconciliation of net income to net cash flow from operating activities		
Net income for the reporting periods (as per the statement of financial activities)	178,346	239,898
Adjustments for:		
Depreciation charges	83,062	60,304
Interest on investments	(18,196)	(15,835)
(Increase)/decrease in debtors and accrued income	(661,383)	891,493
Increase/(decrease) in creditors and accrued expenses	441,805	(28,727)
(Decrease) in grants received in advance of expenditure	(577,102)	(515,572)
Net cash (used in)/generated from operating activities	**(553,468)**	**631,561**
Interest on investments	18,196	15,835
Purchase of tangible fixed assets	(49,510)	(91,094)
Cash flows from investing activities	**(31,314)**	**(75,259)**
Change in cash and cash equivalents in the reporting period	**(584,782)**	**556,302**
Cash and cash equivalents at the beginning of the reporting period	4,944,895	4,388,593
Cash and cash equivalents at the end of the reporting period	**4,360,113**	**4,944,895**

Analysis of cash and cash equivalents	2019 £	2018 £
Short-term deposits	1,239,733	1,747,868
Cash at bank and in hand	3,120,380	3,197,027
Total cash and cash equivalents	**4,360,113**	**4,944,895**

1 Accounting policies

The principal accounting policies adopted, judgements and key sources of estimation uncertainty in the preparation of the financial statements are as follows:

a) Basis of preparation

The financial statements have been prepared in accordance with Accounting and Reporting by Charities: Statement of Recommended Practice applicable to charities preparing their accounts in accordance with the Financial Reporting Standard applicable in the UK and Republic of Ireland (FRS 102) – (Charities SORP (FRS 102)), the Financial Reporting Standard applicable in the UK and Republic of Ireland (FRS 102) and the Companies Act 2006.

The Institute for Fiscal Studies meets the definition of a public benefit entity under FRS 102. Assets and liabilities are initially recognised at historical cost or transaction value unless otherwise stated in the relevant accounting policy note(s).

Going concern
In March 2020, the UK government implemented significant measures to contain the spread of COVID-19, which had been declared as a pandemic by the World Health Organisation (WHO) during the same month. The measures implemented in the UK are similar to those implemented in many other countries around the world, and have significantly impacted many businesses, both operationally and financially. For IFS, the key impacts that may arise or have arisen include:

- the cancellation of events hosted by IFS, in the interest of the health and safety of our stakeholders and in compliance with government guidelines to limit the risk of transmission of the virus;
- the reduction to income that may arise from delays to our research activities or those of our research partners;
- the operational impact to IFS arising from the risk of staff illness.

As there remains uncertainty around the period over which governments' measures worldwide will remain in place, the precise impact cannot be determined. However, IFS has modelled the impact of these measures assuming that cash flows on a number of research projects may be delayed, in some cases for up to 12 months. This work concluded that the entity will have sufficient liquid resources (cash and investments that can be converted to cash) to continue to operate for at least 12 months from the date of approval of these financial statements. This work also considered the possible implementation of risk mitigation measures that are available to IFS, including:

- associated cost savings from the postponement/cancellation of work or events;
- deferral of projects that were planned to be executed in 2020;
- modifying our operations to conform to the current circumstances.

There were no circumstances which arose from the COVID-19 pandemic that resulted in an adjustment to the IFS financial statements as at 31 December 2019. The Executive Committee has considered the impact of the measures taken in the UK and internationally in response to the COVID-19 pandemic and concluded that the going-concern assumption remains appropriate for the preparation of these financial statements.

b) Tangible fixed assets and depreciation

All tangible fixed assets costing more than £1,000 (excluding VAT) are capitalised and depreciated. Depreciation of fixed assets is calculated to write off the cost of each asset over the term of its estimated useful life.

The Executive Committee has determined that all costs relating to the refurbishment of the premises and any furniture be depreciated over five years and all other assets depreciated over three years. Assets are written off on a straight-line basis commencing from the quarter after the date of purchase. Where the length of any remaining lease is less than five years then any refurbishment costs are depreciated up to the end of the year in which the lease comes to an end.

c) Income – membership and donations

Membership income is deferred to the extent that it relates to services to be provided in future periods. Donations are credited to the statement of financial activities at the date of receipt.

d) Income – publications

Royalty income receivable from the publisher of the IFS-owned journal, *Fiscal Studies*, is recognised on an accruals basis and in accordance with the substance of the publishing agreement.

e) Income – research activities

Income from research activities is recognised when the Institute has entitlement to the funds, when it is probable that the income will be received and the amount can be measured reliably.

The Institute is usually entitled to research income in stages over the course of a project, subject to performance-related conditions requiring a particular level of service or output, often approximating to when related expenditure is incurred. In such cases, research income is credited to the statement of financial activities when it falls due to be receivable to the extent that it is matched by related expenditure.

Where donations or grants are received without performance-related conditions, entitlement usually arises on receipt and research income is credited to the statement of financial activities when it falls due to be received.

f) Interest receivable

Interest on funds held on deposit is included when receivable and the amount can be measured reliably.

g) Allocation of expenses

Direct and indirect expenses are included when incurred. The majority of expenses are directly attributable to specific activities. Indirect overhead costs (e.g. premises and administration) are allocated on a basis consistent with the use of the resource, usually on a per capita basis. Irrecoverable VAT is charged as a cost against the activity for which the expenditure was incurred.

h) Pension costs

The pension cost charge represents contributions payable by the Institute to employees' personal pension plans in respect of the year.

i) Operating leases

Leasing charges in respect of operating leases are charged to the statement of financial activities as they are incurred.

j) Current asset investments – short-term deposits

Current asset investments include cash on deposit and cash equivalents held for investment purposes rather than to meet short-term cash commitments as they fall due.

k) Foreign currency

The value of the balances in the Institute's Euro and US Dollar accounts at the end of the year was based on the exchange rate as at 31 December 2019. Transactions in foreign currencies are calculated at the exchange rate ruling at the date of the transaction and Institute-wide foreign exchange gains or losses made during the year are taken into account in arriving at the net income for the year.

l) Financial instruments

IFS only has financial assets and financial liabilities of a kind that qualify as basic financial instruments. Basic financial instruments are initially recognised at transaction value and subsequently measured at their settlement value.

m) Critical accounting estimates and areas of judgement

Preparation of the financial statements requires some judgements and estimates to be made. The items in the financial statements where judgements and estimates are made include:
• judging the progress of multi-year research projects;
• judging whether grants are restricted or unrestricted;
• estimating the useful economic life of tangible fixed assets; and
• estimates relating to the allocation of support costs across expenditure categories.

n) Funds

IFS maintains three internal funds, which include restricted and unrestricted funds:

Unrestricted – General Fund: This fund is derived from any unrestricted donations and grants received by IFS as well as from contracts for research which are unrestricted in nature. These are funds that can be used for any purpose within the charitable objects of IFS.

Unrestricted – Designated Fixed Asset Fund: This fund represents resources set aside to cover future capital expenditure. The value of this fund at the year-end represents the net book value of tangible and intangible fixed assets.

Restricted – research funds: These funds represent grants and donations received to cover project expenditure on research projects. The restrictions are imposed by the funder, usually with respect to the specific research project being undertaken. The nature of the portfolio of research grants and contracts is such that in most cases income and expenditure are closely matched.

Amounts are transferred from the General Fund to the Fixed Asset Fund to maintain the Fixed Asset Fund at an amount that represents the net book value of tangible and intangible fixed assets at the year-end. Amounts are transferred from the General Fund to Restricted research funds to cover any deficit arising on the restricted research grants completed during the year.

o) Prior year information

The basis on which grants are classified as either restricted or unrestricted was changed in 2019. In prior years, only funds held on specific trusts under charity law were classified as restricted funds and IFS had no such restricted funds. Under the new policy, grants are classified as restricted where the funder has identified restrictions in the grant agreement or award such that the funding can only be used to cover expenditure on the specific project that the grant relates to. At the year-end, restricted research funds represent the net assets for projects that are ongoing at the year-end. In most cases, income and expenditure are matched on a grant and therefore the total net assets of the fund are nil. For comparative purposes, the statement of financial activities for the year ended 31 December 2018 has been presented in note 19 showing the split between restricted and unrestricted funds on the new basis.

2 Membership and donations

	2019	2018
	£	£
Corporate membership	143,849	143,235
Individual membership	17,944	13,317
	161,793	**156,552**
Other donations	30,151	301
	191,944	**156,853**

3 Income from charitable activities

IFS frequently collaborates with universities and other research organisations. The income classification below is based on the ultimate funder of the research.

	2019 Unrestricted	2019 Restricted	2019 Total	2018 Unrestricted	2018 Restricted	2018 Total
	£	£	£	£	£	£
ESRC	-	3,575,962	3,575,962	-	3,434,924	3,434,924
Charitable trusts and foundations	15,795	1,882,528	1,898,323	188,349	2,361,994	2,550,343
Government (or similar)	945,602	2,272,373	3,217,975	805,234	1,440,859	2,246,093
Other organisations	23,138	258,182	281,320	56,984	296,110	353,094
Event income	28,406	-	28,406	61,268	-	61,268
Publications	54,425	-	54,425	49,528	-	49,528
	1,067,366	**7,989,045**	**9,056,411**	**1,161,363**	**7,533,887**	**8,695,250**

IFS receives funds in the form of project grants, directly and indirectly, from the UK and other national governments, other governmental agencies and international governmental bodies. These funds are tied to specific research-related activities in the course of the standard charitable activities of IFS. IFS does not receive any funding in the form of general government grants or assistance. Therefore, it is not felt to be necessary, useful or practical to disclose further analysis within these accounts.

4 Investment income

All investment income arises from money held in interest-bearing deposits.

5 Analysis of expenditure

Total costs include payments to third parties that work together with IFS on particular projects. Where the Institute is the lead organisation, it receives funding from the grant-giving body for all participating organisations for onward transmission. Gross receipts are reflected in the Institute's revenues and, depending on the types of project undertaken, may vary significantly from year to year.

	Total charitable activities £	Raising funds £	Governance costs £	Support costs £	2019 Total £	2018 Total £
Research collaborations and subcontracts	1,401,519	-	-	-	**1,401,519**	679,257
Data costs and data collection costs	446,936	-	-	-	**446,936**	1,135,127
IFS travel, accommodation and subsistence	130,271	-	-	-	**130,271**	155,489
Visitor travel, accommodation and subsistence	56,452	-	-	-	**56,452**	125,376
Event, publication and dissemination costs	473,362	4,966	-	41,226	**519,554**	320,593
Other direct costs	142,001	-	-	-	**142,001**	76,051
Premises	-	-	-	571,126	**571,126**	602,633
IT and office costs	-	-	-	173,620	**173,620**	179,677
Other staff costs	-	-	-	132,068	**132,068**	151,601
Insurance and professional fees	-	-	28,048	65,265	**93,313**	85,854
Other	-	-	196	69,828	**70,024**	33,451
Total costs (excluding staff costs)	**2,650,541**	**4,966**	**28,244**	**1,053,133**	**3,736,884**	3,545,109
Staff costs (universities)	714,465	-	-	-	**714,465**	753,490
Research Fellows and Research Associates	249,213	-	-	-	**249,213**	242,550
	963,678	**-**	**-**	**-**	**963,678**	996,040
IFS staff costs (research)	3,499,844	-	-	-	**3,499,844**	3,283,672
IFS staff costs (events and dissemination)	-	25,090	-	311,584	**336,674**	264,096
IFS staff costs (research services)	-	16,807	-	151,259	**168,066**	153,237
IFS staff costs (central)	-	-	17,773	371,056	**388,829**	388,255
	3,499,844	**41,897**	**17,773**	**833,899**	**4,393,413**	4,089,260
Total staff costs (including Fellows and Associates)	**4,463,522**	**41,897**	**17,773**	**833,899**	**5,357,091**	5,085,300
Total expenditure	**7,114,063**	**46,863**	**46,017**	**1,887,032**	**9,093,975**	8,630,409
Allocation of support costs (including governance)	1,921,297	11,752	(46,017)	(1,887,032)	-	-
Total expenditure	**9,035,360**	**58,615**	-	-	**9,093,975**	8,630,409

Analysis of expenditure 2018

	Total charitable activities £	Raising funds £	Governance costs £	Support costs £	2018 Total £
Research collaborations and subcontracts	679,257	-	-	-	**679,257**
Data costs and data collection costs	1,135,127	-	-	-	**1,135,127**
IFS travel, accommodation and subsistence	155,489	-	-	-	**155,489**
Visitor travel, accommodation and subsistence	125,376	-	-	-	**125,376**
Event, publication and dissemination costs	282,422	3,210	-	34,961	**320,593**
Other direct costs	76,051	-	-	-	**76,051**
Premises	-	-	-	602,633	**602,633**
IT and office costs	-	-	-	179,677	**179,677**
Other staff costs	-	-	-	151,601	**151,601**
Insurance and professional fees	-	-	25,211	60,643	**85,854**
Other	-	-	3,357	30,094	**33,451**
Total costs (excluding staff costs)	**2,453,722**	**3,210**	**28,568**	**1,059,609**	**3,545,109**
Staff costs (universities)	753,490	-	-	-	**753,490**
Research Fellows and Research Associates	242,550	-	-	-	**242,550**
	996,040	**-**	**-**	**-**	**996,040**
IFS staff costs (research)	3,283,672	-	-	-	**3,283,672**
IFS staff costs (events and dissemination)	-	23,296	-	240,800	**264,096**
IFS staff costs (research services)	-	15,324	-	137,913	**153,237**
IFS staff costs (central)	-	-	17,065	371,190	**388,255**
	3,283,672	**38,620**	**17,065**	**749,903**	**4,089,260**
Total staff costs (including Fellows and Associates)	**4,279,712**	**38,620**	**17,065**	**749,903**	**5,058,300**
Total expenditure	**6,733,434**	**41,830**	**45,633**	**1,809,512**	**8,630,409**
Allocation of support costs (including governance)	1,843,909	11,236	(45,633)	(1,809,512)	-
Total expenditure	**8,577,343**	**53,066**	-	-	**8,630,409**

6 Total expenditure

	Unrestricted	Restricted	2019 Total	2018 Total
	£	£	£	£
Cost of raising funds				
Direct costs (membership programme)	4,966	-	4,966	3,210
Staff costs (direct)	41,897	-	41,897	38,620
Support and governance costs (allocation)	11,752	-	11,752	11,236
	58,615	-	**58,615**	**53,066**
Charitable activities				
Project costs	309,856	2,340,685	2,650,541	2,453,722
Staff costs (total)	425,215	4,889,979	5,315,194	5,046,680
Support and governance costs (allocation)	85,571	984,054	1,069,625	1,076,941
	820,642	**8,214,718**	**9,035,360**	**8,577,343**
Total expenditure	**879,257**	**8,214,718**	**9,093,975**	**8,630,409**

2018	**Unrestricted**	**Restricted**	**2018 Total**
Cost of raising funds			
Direct costs (membership programme)	3,210	-	3,210
Staff costs (direct)	38,620	-	38,620
Support and governance costs (allocation)	11,236	-	11,236
	53,066	**-**	**53,066**
Charitable activities			
Project costs	253,811	2,199,911	2,453,722
Staff costs (total)	555,135	4,491,545	5,046,680
Support and governance costs (allocation)	118,462	958,479	1,076,941
	927,408	**7,649,935**	**8,577,343**
Total expenditure	**980,474**	**7,649,935**	**8,630,409**

IFS initially identifies the costs of its support functions. It then identifies those costs that relate to governance. The remaining support costs together with the governance costs are apportioned between charitable activities and the cost of raising funds.

The cost of raising funds includes costs related to the IFS membership programme and costs related to activities focused on seeking funding. This includes some direct costs and direct staff time, as well as an allocation of support costs. Support costs are allocated on the basis of staff time.

Governance costs include the costs of external audit. Other governance costs relate primarily to costs associated with the AGM and annual lecture and dinner and also include travel and accommodation expenses for one Council member. No expenses were claimed by the Trustees during the year (2018: £38).

7 Net income

Net income is stated after charging:

	2019	2018
	£	£
Depreciation	83,062	60,304
Auditors' remuneration		
- Audit fees	17,800	17,000
Operating lease rentals – property	375,000	375,000

Audit fees are stated net of VAT and disbursements.

8 Analysis of staff costs and key management personnel

	2019	2018
	£	£
Wages and salaries	3,791,893	3,543,506
Social security costs	373,556	354,364
Pension costs	227,964	191,390
	4,393,413	**4,089,260**
Comprising:		
Researchers	3,499,844	3,283,672
Support staff	893,569	805,588
IFS payroll staff	**4,393,413**	**4,089,260**
Staff costs (universities)	714,465	753,490
Research Fellow and Research Associate payments	249,213	242,550
	5,357,091	**5,085,300**

Staff costs (universities): IFS has agreements in place with several universities/institutions for the provision of an agreed proportion of the working time (typically 10–50%) of, during 2019, on average 15 (2018: 16) named, highly skilled individuals to carry out specific research duties at IFS in their areas of academic excellence. In 2019, £92,500 (2018: £97,500) of the amount for Research Fellows and Research Associates related to these individuals.

During 2019, the Institute's senior management team comprised: the Director, Paul Johnson, the Deputy Director, Carl Emmerson, and the Research Directors, Professor Rachel Griffith and, until 1 September 2019, Professor Orazio Attanasio. In 2019, the total compensation for these key management personnel, including amounts due to universities under contractual arrangements for the provision of an agreed amount of the Research Directors' time, was £473,982 (2.6 FTE) (2018: £512,573 (2.9 FTE))

The numbers of employees whose emoluments (excluding pension contributions) were in excess of £60,000 are shown in the ranges below. In addition, pension contributions were paid by the Institute on behalf of these employees. The total sum of these contributions was £102,187 (for 20 employees) (2018: £98,822 for 18 employees).

	2019 Number	2018 Number
£60,001–£70,000	8	7
£70,001–£80,000	5	8
£80,001–£90,000	5	1
£90,001–£100,000	1	1
£180,001–£190,000	-	1
£210,001–£220,000	1	-
	20	**18**

9 Staff numbers

	2019 FTE	Average number	2018 FTE	Average number
Research staff				
Permanent contracts	36.2	40.4	34.7	38.6
Fixed-term contracts	16.2	20.5	18.3	26.2
Variable-hour contracts	1.6	3.7	2.2	2.4
	54.0	64.6	55.2	67.2
Central staff				
Events, publications, dissemination	6.3	6.8	6.5	6.9
Finance, HR, IT, central support	7.3	7.6	6.3	6.6
Research services	3.0	3.0	2.3	2.3
	16.6	17.4	15.1	15.8
Total	**71**	**82**	**70**	**83**
Full-time		55		53
Part-time		27		30

10 Tangible fixed assets

	Fixtures and improvements to short leasehold premises	Office equipment	Total
	£	**£**	**£**
Cost			
At 1 January 2019	769,385	440,397	1,209,782
Additions	19,620	29,890	49,510
Disposals and assets no longer in use	(13,611)	(60,716)	(74,327)
At 31 December 2019	**775,394**	**409,571**	**1,184,965**
Depreciation			
At 1 January 2019	745,670	299,767	1,045,437
Charge for the year (1)	10,441	72,621	83,062
Disposals and assets no longer in use	(13,611)	(60,716)	(74,327)
At 31 December 2019	**742,500**	**311,672**	**1,054,172**
Net book value			
As at 31 December 2019	**32,894**	**97,899**	**130,793**
As at 31 December 2018	**23,715**	**140,630**	**164,345**

(1) The depreciation charge for the year included £29,760 (2018: £10,003) of depreciation on assets used on specific projects and reimbursed under the grant as direct project costs.

All fixed assets are held for use on a continuing basis for the purpose of charitable activities.

11 Debtors

	Unrestricted	Restricted	2019	2018
	£	£	£	£
Accrued income	276,892	1,047,486	1,324,378	751,089
Trade debtors	88,692	333,206	421,898	371,801
Other debtors	18,978	-	18,978	22,557
Prepayments	316,203	-	316,203	274,627
	700,765	**1,380,692**	**2,081,457**	**1,420,074**

12 Creditors

	Unrestricted	Restricted	2019	2018
	£	£	£	£
Amounts falling due within one year				
Trade payables	233,641	172,550	406,191	107,003
Taxation and social security	115,270	-	115,270	104,550
VAT	44,679	-	44,679	34,607
Accruals	294,435	289,767	584,202	462,377
	688,025	**462,317**	**1,150,342**	**708,537**
Deferred income				
Balance at 1 January	**256,501**	**2,757,852**	**3,014,353**	**3,529,925**
Amount released to income	(144,092)	(2,469,310)	(2,613,402)	(3,070,580)
Amount deferred in the year	68,408	1,967,892	2,036,300	2,555,008
Balance at 31 December	**180,817**	**2,256,434**	**2,437,251**	**3,014,353**
Total creditors: amounts falling due within one year	**868,842**	**2,718,751**	**3,587,593**	**3,722,890**

As at 31 December 2019, total deferred income was £2,437,251 (2018: £3,014,353). This includes amounts received on multi-year projects, where the timing of the related expenditure may be more than 12 months from the balance sheet date. A proportion of this deferred income will therefore not be released to income until 2021 or 2022.

13 Analysis of movement in funds

	At 1 Jan 2019	Income	Expenditure	Transfers	At 31 Dec 2019
	£	£	£	£	£
Unrestricted funds					
General Fund	2,642,079	1,283,276	(879,257)	(192,121)	2,853,977
Fixed Asset Fund	164,345	-	-	(33,552)	130,793
	2,806,424	**1,283,276**	**(879,257)**	**(225,673)**	**2,984,770**
Restricted funds					
Research funds	-	7,989,045	(8,214,718)	225,673	-
Total funds	**2,806,424**	**9,272,321**	**(9,093,975)**	**-**	**2,984,770**

2018	At 1 Jan 2018	Income	Expenditure	Transfers	At 31 Dec 2018
	£	£	£	£	£
Unrestricted funds					
General Fund	2,432,971	1,336,420	(980,474)	(146,838)	2,642,079
Fixed Asset Fund	133,555	-	-	30,790	164,345
	2,566,526	**1,336,420**	**(980,474)**	**(116,048)**	**2,806,424**
Restricted funds					
Research funds	-	7,533,887	(7,649,935)	116,048	-
Total funds	**2,566,526**	**8,870,307**	**(8,630,409)**	**-**	**2,806,424**

Amounts have been transferred from the General Fund to the Fixed Asset Fund to maintain the Fixed Asset Fund at an amount that represents the net book value of tangible and intangible fixed assets at the year-end.

Amounts have been transferred from the General Fund to Restricted research funds to cover the overall deficit arising on the restricted research grants that completed during the year.

Within restricted research funds are funds relating to projects where the agreement with the funder requests that the project funding is separately disclosed in the financial statements. During 2019, the income and expenditure on these grants was as set out below.

Project name	Funder	Start date	End date	2019 income and expenditure £	Accrued/ (Deferred) income as at 31 Dec 2019 £
The Centre for Tax Analysis in Developing Countries – Phase 2 (TAXDEV II)	DFID	1/11/2018	31/10/2022	982,089	315,520
Evaluation of Lively Minds educational play schemes in Ghana	Global Innovation Fund	1/2/2017	30/4/2019	143,071	-

2018

Project name	Funder	Start date	End date	2018 income and expenditure £	Accrued/ (Deferred) income as at 31 Dec 2018 £
Improving tax and benefit policy analysis and development in partner countries with the Institute for Fiscal Studies	DFID	1/2/2016	31/3/2018	201,462	-
The Centre for Tax Analysis in Developing Countries – Phase 2 (TAXDEV II)	DFID	1/11/2018	31/10/2022	72,198	72,198
Evaluation of Lively Minds educational play schemes in Ghana	Global Innovation Fund	1/2/2017	30/4/2019	135,834	(82,887)

14 Analysis of net assets between funds

	2019 Unrestricted £	2019 Restricted £	2019 Total £	2018 Unrestricted £	2018 Restricted £	2018 Total £
Tangible fixed assets	130,793	-	**130,793**	164,345	-	164,345
Cash at bank and in hand	3,022,054	1,338,059	**4,360,113**	2,850,062	2,094,833	4,944,895
Other net current liabilities	(168,077)	(1,338,059)	**(1,506,136)**	(207,983)	(2,094,833)	(2,302,816)
Net assets at 31 December	**2,984,770**	**-**	**2,984,770**	**2,806,424**	**-**	**2,806,424**

15 Operating lease commitments

The total of future minimum lease payments under non-cancellable operating leases is set out below for each of the following periods.

	2019	2018
	£	£
One year	178,767	375,000
Two to five years	-	178,767

16 Pension scheme

The total pension cost to IFS for contributions to employees' pension schemes under IFS's group personal pension plans with Scottish Widows was £206,842 (2018: £172,784). In addition, three members of staff (2018: three) participated in other personal pension schemes, of their own choice, to which the Institute contributed £21,122 (2018: £18,606).

	2019	2018
	£	£
Scottish Widows	206,842	172,784
Other	21,122	18,606
Total	**227,964**	**191,390**

17 Related party transactions

Lorraine Dearden, a member of Paul Johnson's close family, is paid as an IFS Research Fellow at the standard rate of £5,000 per annum (2018: £5,000). Her initial appointment pre-dates his term as Director and is reviewed annually by the Research Directors. In addition, IFS has an agreement with Lorraine Dearden's employer, the Institute of Education, for a buyout of 20% of her full employment costs for the year ended 31 December 2019. The buyout from the Institute of Education pre-dates Paul Johnson's appointment as Director and was agreed by his predecessor.

18 Post-balance sheet events

There were no circumstances which arose from the COVID-19 pandemic that resulted in an adjustment to the IFS financial statements as at 31 December 2019.

19 Comparative information: statement of financial activities for the year to 31 December 2018

2018	2018 Unrestricted £	2018 Restricted £	2018 Total £
Income from:			
Membership and donations	156,853	-	156,853
Charitable activities	1,161,363	7,533,887	8,695,250
Investment income	15,835	-	15,835
Other income	2,369	-	2,369
Total income	**1,336,420**	**7,533,887**	**8,870,307**
Expenditure on:			
Raising funds	53,066	-	53,066
Charitable activities	927,408	7,649,935	8,577,343
Total expenditure	**980,474**	**7,649,935**	**8,630,409**
Net income/(expenditure)	**355,946**	**(116,048)**	**239,898**
Transfers between funds	**(116,048)**	**116,048**	**-**
Net movement in funds	**239,898**	**-**	**239,898**
Reconciliation of funds:			
Total funds brought forward	**2,566,526**	**-**	**2,566,526**
Total funds carried forward	**2,806,424**	**-**	**2,806,424**

The Institute for Fiscal Studies
7 Ridgmount Street
London WC1E 7AE

Tel: +44 (0) 20 7291 4800
Fax: +44 (0) 20 7323 4780
Email: mailbox@ifs.org.uk