Econometrica, Vol. 91, No. 2 (March, 2023), 709–736

CONSTRAINED CONDITIONAL MOMENT RESTRICTION MODELS

VICTOR CHERNOZHUKOV
Department of Economics, M.I.T.

WHITNEY K. NEWEY
Department of Economics, M.I.T.

ANDRES SANTOS
Department of Economics, U.C.L.A.

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.

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

1. INTRODUCTION

SHAPE RESTRICTIONS HAVE PLAYED A CENTRAL ROLE IN ECONOMICS as both testable implications of classical theory and sufficient conditions for obtaining informative counterfactual predictions. 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)). 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 (2016)). The empirical analysis of shape restrictions, however, goes well beyond this important application with recent examples including, among others, studies into the monotonicity of the state price density (Jackwerth (2000)) and the existence of complementarities in demand (Gentzkow (2007)).

Shape restrictions are often equivalent to inequality restrictions on parameters of interest and on certain unknown functions. For example, Slutsky negative semidefiniteness and monotonicity require that certain functions satisfy inequality restrictions. Inference with

Victor Chernozhukov: vchern@mit.edu
Whitney K. Newey: wnewey@mit.edu
Andres Santos: andres@econ.ucla.edu
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. Whitney K. Newey's research supported by NSF Grant 1757140. Andres Santos' research supported by NSF Grant SES-1426882.

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. 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, under the null hypothesis, equal the nominal level of the test in many applications. Our results apply to a class of conditional moment restriction models that encompasses parametric (Hansen (1982)), semiparametric (Ai and Chen (2003)), and nonparametric (Newey and Powell (2003)) instrumental variable (IV) models, as well as 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 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, under the null hypothesis, 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. Indeed, these methods have been used by Torgovitsky (2019) to construct informative confidence intervals for 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/nonparametric 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 are 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, that is, 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 setup 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. 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. The latter two examples are discussed in detail in the Supplemental Material (Chernozhukov, Newey, and Santos (2023)).

Our paper contributes to an extensive literature studying semiparametric and nonparametric models under partial identification. Freyberger and Horowitz (2015), for instance, developed inference methods for shape restricted partially identified discrete IV models; their approach, however, is based on limiting distributions that are discontinuous in the true parameters leading to nonuniform inference. 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. In that context, our results are complementary to those of Bugni, Canay, and Shi (2017) and Kaido, Molinari, and Stoye (2019), who provided 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 Santos (2012), Tao (2014), and Chen, Tamer, and Torgovitsky (2011) who studied inference on functionals of potentially partially identified structural functions, but did not allow for shape constraints as we do.

Following the original version of this paper, Zhu (2019) and Fang and Seo (2019) 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 developed uniform inference for functionals under shape restrictions while imposing point identification. Our paper is of course part of a large literature on shape restrictions. 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; see, for example, Armstrong (2015) and references therein. 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.

The remainder of the paper is organized as follows. In Section 2, we show how to implement our tests in a linear IV 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 heterogeneous demand model of Hausman and Newey (2016). The Supplemental Material includes further applications of our general results and additional background on AM spaces. All mathematical derivations may be found in the working paper Chernozhukov, Newey, and Santos (2022).

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).

2.1. *Linear Instrumental Variables*

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

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

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

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

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 Hansen (1982). To this end, we define the criterion

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

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

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

for example, we may consider a test that rejects for large values of $I_n(R) - I_n(\Theta)$. In what follows, we also let $\hat{\theta}_n$ and $\hat{\theta}_n^{\mathrm{u}}$ denote the minimizers of Q_n over $\Theta \cap R$ and Θ, respectively.

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

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

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

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

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 jth row of G and g_j the jth 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

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

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

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

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

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, that is, 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 1 (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"

[1]We follow previous work (e.g., Hansen (1996)) in considering Gaussian $\{\omega_i\}_{i=1}^n$ because it simplifies the proofs of our main results. We expect our analysis extends to other distributions of $\{\omega_i\}_{i=1}^n$—for example, for ω_i following an exponential distribution, which results in a version of the Bayesian bootstrap.

analogue of $\hat{\theta}_n^{\mathrm{u}}$ and some $\gamma_n \to 0$, we recommend setting r_n to satisfy

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

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 $G\theta_0 \le g$ on θ_0 (e.g., sign or monotonicity restrictions on $w \mapsto w'\theta_0$). For example, for $\theta^{(k)}$ the kth coordinate of $\theta \in \mathbf{R}^{d_\theta}$, we may set $R_\lambda = \{\theta \in \mathbf{R}^{d_\theta} : \theta^{(k)} = \lambda, G\theta \le g\}$ and obtain a confidence region for $\theta_0^{(k)}$ 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, that is, 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) documented that the impact of childbearing on labor force participation depends on observable characteristics. In particular, their two stage least 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

[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.

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

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

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-value 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. 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). Imposing monotonicity 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. The third panel of Figure 1 shows that additionally imposing LATE$(\cdot)$ be nonpositive 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

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

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

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

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

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

for Υ_F a known possibly nonlinear function—for example, $\Upsilon_F(\theta) = F\theta - f$ 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

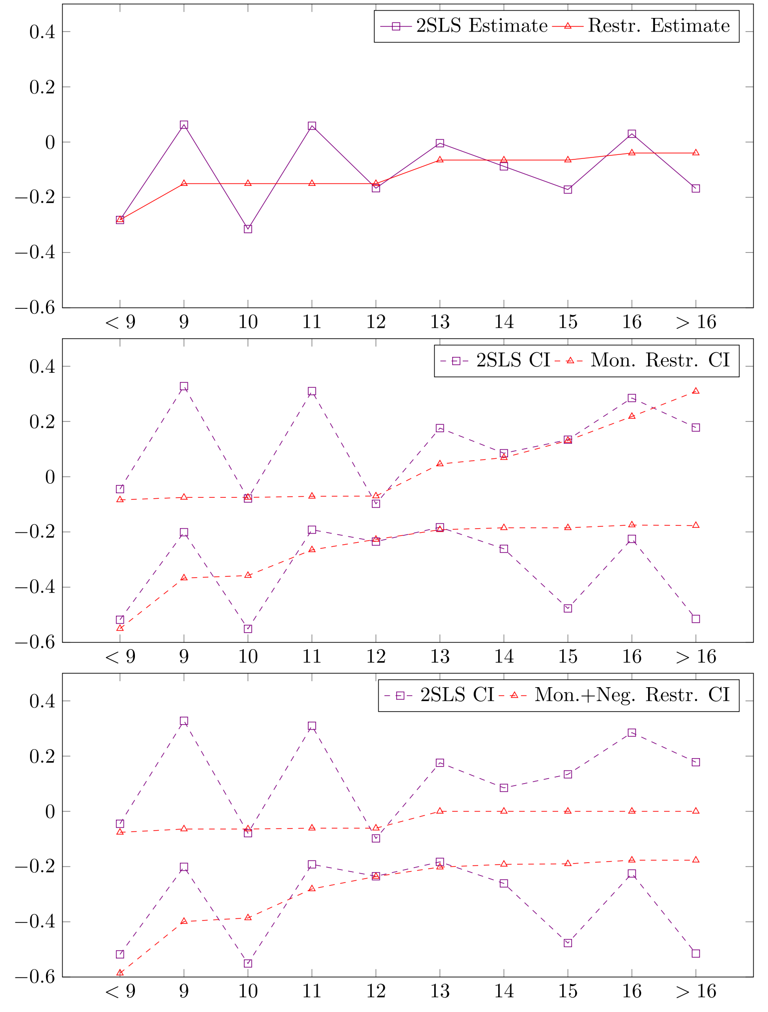

FIGURE 1.—First panel: Unconstrained and shape restricted LATE estimates (imposing monotonicity or monotonicity and negativity yield the same estimates). Second and third panels: 95% Confidence intervals for LATE at different education levels.

$I_n(\Theta) = 0$ (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

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

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

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

(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 $\hat{\theta}^{\mathrm{u}}_n$ may be necessary; see Section 2.2.1 for an example. Second, our bootstrap approximation employs an estimator $\hat{\Theta}^{\mathrm{r}}_n$ for $\Theta_0 \cap R$. To this end, we set

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

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

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

Thus, to obtain a level α test, we reject the null hypothesis whenever $I_n(R)$ exceeds the $1 - \alpha$ quantile of $\hat{U}^b_n(R)$ 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. However, setting $\tau_n = 0$ can result in lower power in applications for which the corresponding $\hat{\Theta}^{\mathrm{r}}_n$ is not (Hausdorff) consistent for $\Theta_0 \cap R$—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

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

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 $\hat{\Theta}^{\mathrm{r}}_n$ 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 Υ_F, 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

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

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. 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 sets of 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 (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 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 let $\tau_n = 0$ and set r_n according to (7) with $\gamma_n = 0.05$ and where we used the distribution of estimators of identified parameters for their partially identified counterparts.[3] 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 the estimates by 0.011. Nonetheless, the unrestricted confidence intervals are large as the estimates for the identified set are large—a result driven by the low proportion of compliers (5% on average across S). 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

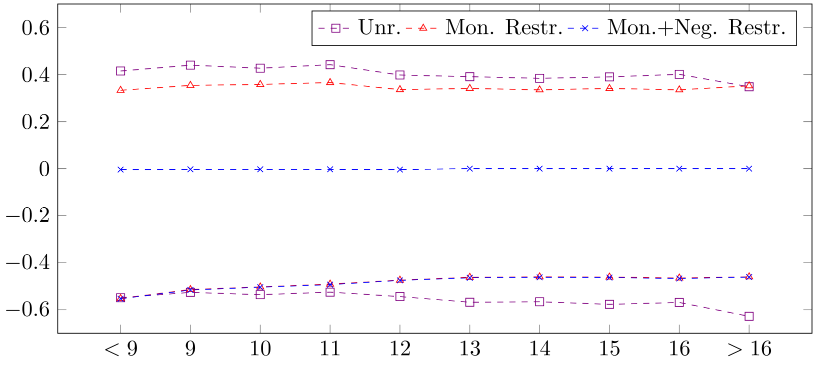

FIGURE 2.—95% Confidence intervals for ATE at different education levels. "Unr." uses bounds implied by $Y_d \in \{0, 1\}$; "Mon. Restr." adds that average treatment effects be increasing in education for all types; "Mon. + Neg. Restr." also requires they be negative.

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

TABLE I

POINT ESTIMATES AND 95% CONFIDENCE INTERVALS FOR THE AVERAGE TREATMENT EFFECT AT DIFFERENT GROUPS DEFINED BY SCHOOLING LEVELS UNDER DIFFERENT SHAPE RESTRICTIONS.

	Estimate		
Subgroup	Unrestricted	Mon. Restr.	Mon. + Neg Restr.
HS Drop	[−0.520, 0.426]	[−0.489, 0.346]	[−0.489, −0.008]
Coll. Drop	[−0.561, 0.380]	[−0.447, 0.325]	[−0.447, −0.004]
Coll. Grad	[−0.579, 0.375]	[−0.446, 0.328]	[−0.446, −0.002]
All	[−0.545, 0.395]	[−0.467, 0.328]	[−0.467, −0.008]

	95% Confidence Interval		
Subgroup	Unrestricted	Mon. Restr.	Mon. + Neg Restr.
HS Drop	[−0.526, 0.432]	[−0.500, 0.356]	[−0.501, −0.008]
Coll. Drop	[−0.566, 0.385]	[−0.460, 0.337]	[−0.462, 0.000]
Coll. Grad	[−0.586, 0.382]	[−0.462, 0.339]	[−0.464, 0.000]
All	[−0.547, 0.398]	[−0.477, 0.338]	[−0.478, −0.003]

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 I 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, 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

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

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 $Z \equiv (Z_1, \dots, Z_{\mathcal{J}})$ 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

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

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

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

that is, 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 $\mathbf{1}_{\mathbf{G}}$.*

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 Material. 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 $\mathbf{1}_{\mathbf{G}} \in \mathbf{G}$ is an element such that for any $\theta \in \mathbf{G}$, there exists a scalar $\lambda > 0$ satisfying $|\theta| \leq \lambda \mathbf{1}_{\mathbf{G}}$ (e.g., when $\mathbf{G} = \mathbf{R}^d$ we may set $\mathbf{1}_{\mathbf{G}} \equiv (1, \dots, 1)' \in \mathbf{R}^d$). These properties of an AM space will prove instrumental in our analysis. In particular, the order unit $\mathbf{1}_{\mathbf{G}}$ will provide a crucial link between the partial order "$\leq$" and 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)—for example, it may be employed to implement generalizations of the J or incremental J tests. In this section, we first introduce $I_n(R)$, obtain an approximation to its distribution, and devise a bootstrap procedure for estimating its quantiles.

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 $q_\jmath^{k_{n,\jmath}}(z_\jmath) \equiv (q_{1,\jmath}(z_\jmath), \dots, q_{k_{n,\jmath},\jmath}(z_\jmath))'$. Recalling that $Z \equiv (Z_1, \dots, Z_{\mathcal{J}})$, we further set $k_n \equiv \sum_{\jmath=1}^{\mathcal{J}} k_{n,\jmath}$, $q^{k_n}(z) \equiv (q_1^{k_{n,1}}(z_1)', \dots, q_{\mathcal{J}}^{k_{n,\mathcal{J}}}(z_{\mathcal{J}})')'$, $\rho(x, \theta) \equiv (\rho_1(x, \theta), \dots, \rho_{\mathcal{J}}(x, \theta))'$, and let

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

that is, 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

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

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 $p \geq 2$—that is, $\|a\|_p \equiv (\sum_{i=1}^{d} |a^{(i)}|^p)^{1/p}$ for any $a \equiv (a^{(1)},\dots,a^{(d)})' \in \mathbf{R}^d$. Letting $\Theta_n \cap R$ be a finite dimensional subset of $\Theta \cap R$ that grows dense in $\Theta \cap R$, 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.

Intuitively, $\sqrt{n}Q_n$ 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 test of Hansen (1982). Alternatively, we may build a test by considering the recentered test statistic $I_n(R) - I_n(\Theta)$, which is similar in spirit to the incremental J test. 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 next obtain a strong approximation to $I_n(R)$. To this end, we first 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 $\|f\|_{P,2} \equiv (E_P[f^2(V)])^{1/2}$ 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 impose requirements on $\Theta_n \cap R$ and the transformations q^{k_n}.

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* $E_P[\|\rho(X,\theta_1)-\rho(X,\theta_2)\|_2^2]\le K_\rho^2\|\theta_1-\theta_2\|_{\mathbf{E}}^{2\kappa_\rho}$ *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}—for example, 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 (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 Supplemental Material, we verify Assumption 3.3(i) by relying on existing results or a novel extension of Koltchinskii's coupling. 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, Hong, and Tamer (2007)), 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

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

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

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

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

(*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* 1 *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_{0n}^{r} after establishing the consistency of $\hat\theta_n$ through standard arguments. 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_{0n}^{r}. 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, Chen, and Kristensen (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_{0n}^{r} that shrinks at a rate

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

where recall B_n and J_n were introduced in Assumption 3.2. Under assumptions on the (Hausdorff) distance between Θ_{0n}^{r} 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_{0n}^{r} (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 (2015)). Specifically, for each $1 \leq \jmath \leq \mathcal{J}$ and $\theta \in \Theta$, we set

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

that is, $m_{P,\jmath}$ 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 $\nabla m_{P,\jmath}(\theta) : \mathbf{B} \to L_P^2$, and constants $\epsilon > 0$ and $K_m, M < \infty$ such that for all $P \in \mathbf{P}$, $h \in \mathbf{B}_n$, $1 \leq \jmath \leq \mathcal{J}$, and elements $\theta_1, \theta_2 \in \{\theta \in \Theta_n \cap R : \overrightarrow{d}_H(\theta, \Theta_{0n}^{\mathrm{r}}, \|\cdot\|_{\mathbf{E}}) \leq \epsilon\}$, 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*) $\sup_{P\in\mathbf{P}_0} \sup_{\theta\in\Theta_{0n}^{\mathrm{r}}} \sqrt{n}\|E_P[\rho(X,\theta) * q^{k_n}(Z)]\|_{\Sigma_P,p} = o(a_n)$.

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 reparameterization. Formally, we denote the local deviations from $\theta \in \Theta_n \cap R$ by

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

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

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

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

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

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*

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

(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

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

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

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

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 $\hat{\Theta}^{\mathrm{r}}_n$ for Θ^{r}_{0n} (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

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

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

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

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

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

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

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

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

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, Hong, and Tamer (2007), Andrews and Shi (2013)). Specifically, for a sequence r_n and any $\theta \in \Theta_n \cap R$, we define

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

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

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

that is, 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 $\|h/\sqrt{n}\|_{\mathbf{B}} \leq \ell_n$.

Before establishing the asymptotic validity of the proposed bootstrap procedure, we require some additional notation. For any set $A \subseteq \mathbf{B}_n$, we let $(A)^\epsilon \equiv \{\theta \in \mathbf{B}_n : \inf_{a \in A} \|a - \theta\|_{\mathbf{B}} \leq \epsilon\}$. 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 what follows, 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 an $h \in \mathbf{B}_n \cap \mathcal{N}(\nabla\Upsilon_F(\theta))$ 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.[4] Thus, while Assumptions 3.9(i)–(iii) impose conditions analogous to those required of Υ_G, Assumption 3.9(iv) additionally demands that $\nabla\Upsilon_F(\theta)$ possess a norm

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

bounded right inverse on $(\Theta_{0n}^{\mathrm{r}})^{\epsilon}$—the existence of a right inverse is equivalent to a classical rank condition.[5] 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* $M < \infty$, $\|h\|_{\mathbf{E}} \le M\|h\|_{\mathbf{B}}$ *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* 1 *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* 1 *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. More generally, we note that our analysis remains valid if the multiplier bootstrap is replaced with any other re-sampling 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 is nonlinear. The main conditions on ℓ_n, employed in constructing $\hat{V}_n(\theta, R|\ell_n)$, 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 $\hat{\Theta}_n^{\mathrm{r}}$. Assumption 3.13(iii) is always satisfied if $r_n = +\infty$,

[5]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 $\nabla\Upsilon_F(\theta) : \mathbf{B}_n \to \mathbf{F}_n$.

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, as discussed in Section 2.2, such a choice can lead to a loss of power in certain partially identified models.

Our next result provides a coupling result for our bootstrap statistic. In its statement, $U_P^\star(R|\ell_n)$ is defined identically to $U_P(R|\ell_n)$ but with $\mathbb{W}_P^\star$ 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 $P \in \mathbf{P}_0$,*

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

Theorem 3.2 shows that with probability tending to 1 uniformly on $P \in \mathbf{P}_0$, our bootstrap statistic is bounded from below by a random variable that is independent of the data. Crucially, 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 if 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 the following to hold: (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 n; and (iii) in partially identified settings, τ_n must tend to zero sufficiently slowly so that $\hat{\Theta}_n^{\mathrm{r}}$ is Hausdorff consistent for Θ_{0n}^{r}.

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 τth quantile given $\{V_i\}_{i=1}^n$. For example, we have that

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

3.2.1. *Tests Based on $I_n(R)$*

We first examine a test that employs $I_n(R)$ as a test statistic. As has been shown in the literature, uniform consistent estimation of approximating distributions is not sufficient for characterizing the asymptotic size of a test. 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 are $\eta \geq 0$ 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. 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 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. For instance, when $U_P(R|\tilde{\ell}_n)$ is a convex function of $\mathbb{W}_P$, as in Section 2.1.1, the distribution of $U_P(R|\tilde{\ell}_n)$ can readily be shown to be continuous on $(0, +\infty)$.

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*

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

In Algorithm 1 below, we describe how to compute the p-value of 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 whenever the conditions of Theorem 3.1(ii) hold. The consistency of the test against any $P \in \mathbf{P} \setminus \mathbf{P}_0$ for which $\max_{\jmath} \|E_P[\rho_{\jmath}(X, \theta)|Z_{\jmath}]\|_{P,2}$ is bounded away from zero (in $\theta \in \Theta \cap R$) is also straightforward to establish. Finally, we 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.

Algorithm 1 Computing the p-value of the test based on $I_n(R)$.

Require: $\Theta_n, \Upsilon_F, \Upsilon_G, \{\rho(X_i, \theta) * q^{k_n}(Z_i)\}_{i=1}^n, \hat{\Sigma}_n, r_n, \tau_n, \ell_n$

▷ Compute the test statistic

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

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

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

▷ Prepare variables for bootstrap problem

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

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

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

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

▷ 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

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

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

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

end for

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

This modification to the critical value was originally proposed in a different context by Andrews and Shi (2013), who suggested setting $\delta = 10^{-6}$.

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

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

The power functions of the corresponding tests are not necessarily ranked. It can therefore 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. The asymptotic validity of this test follows from Theorems 3.1 and 3.2 under a 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—for example, $\mathcal{F}_n^{\mathrm{u}}$ 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 $U_P(\Theta|\ell_n^{\mathrm{u}})$, we define

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

that is, $\hat{\Theta}_n^{\mathrm{u}}$ 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

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

that is, 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*

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

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. 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

As an example, we next illustrate how to conduct inference in the heterogeneous demand model of Hausman and Newey (2016); for alternative models of demand under conditional moment restrictions, see Chen and Christensen (2018) and references therein. Specifically, for $Y \in [0, 1]$ the expenditure share on a commodity, $W \in \mathbf{W}$ a vector of prices, income, and covariates, and η unobserved individual heterogeneity, suppose

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

where g is a known function of (W, η). As in Hausman and Newey (2016), we note that the unobserved heterogeneity η can potentially be infinite dimensional.

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

$$P(Y \le c|W) = P\big(g(W, \eta) \le c|W\big) = \int 1\big\{g(W, \eta) \le c\big\}\mu_0(d\eta), \tag{24}$$

where μ_0 denotes the unknown distribution of η. Result (24) 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) studied functionals

$$\int \Psi(g, \eta)\mu_0(d\eta), \tag{25}$$

which is the average across individuals. By evaluating the set of values of (25) which can be generated by a distribution μ_0 satisfying (24) at a grid $\{c_j\}_{j=1}^{\mathcal{J}}$, Hausman and Newey (2016) provided 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 (25) 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 (25). To this end, we let $F_P(c_j|W) \equiv P(Y \le c_j|W)$ for a given grid $\{c_j\}_{j=1}^{\mathcal{J}}$. 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 $\|\cdot\|_{\mathrm{TV}}$ denote the total variation norm. Assuming $F_P(c_j|\cdot) \in C_B(\mathbf{W})$ for $C_B(\mathbf{W})$ the space of continuous and bounded functions on $\mathbf{W}$, we set $\mathbf{B} = (\bigotimes_{j=1}^{\mathcal{J}} C_B(\mathbf{W})) \times \mathcal{M}$, for any $(\{F(c_j|\cdot)\}_{j=1}^{\mathcal{J}}, \mu) = \theta \in \mathbf{B}$ let $\|\theta\|_{\mathbf{B}} = \sum_{j=1}^{\mathcal{J}} \|F(c_j|\cdot)\|_\infty + \|\mu\|_{\mathrm{TV}}$, and set

$$\Theta = \Big\{\big(\{F(c_j|\cdot)\}_{j=1}^{\mathcal{J}}, \mu\big) = \theta \in \mathbf{B} : \max_{1\le j\le \mathcal{J}} \big\|F(c_j|\cdot)\big\|_\infty \le 2\Big\}, \tag{26}$$

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 $1 \le \jmath \le \mathcal{J}$, we then define

$$\rho_\jmath(X, \theta) = 1\{Y \le c_\jmath\} - F(c_\jmath|W), \tag{27}$$

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 (25) by setting

$$R = \bigg\{ \big(\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu\big) : \mu(\Omega) = 1, \mu(B) \ge 0 \text{ for all } B \in \mathcal{B}, \int \Psi(g, \eta)\mu(d\eta) = \lambda,$$
$$\text{and } F(c_\jmath|w_l) = \int 1\{g(w_l, \eta) \le c_\jmath\}\mu(d\eta) \text{ for all } 1 \le \jmath \le \mathcal{J}, 1 \le l \le \mathcal{L} \bigg\}. \tag{28}$$

Thus, the null hypothesis that $\Theta_0 \cap R$ is nonempty corresponds to requiring that there exist a distribution μ for η satisfying the restrictions in (24) at the points $(c_\jmath, w_l)$ and yielding a value for the functional in (25) 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

$$\Upsilon_G(\theta)(B) = -\mu(B). \tag{29}$$

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

$$\begin{aligned} \Upsilon_F^{(e)}(\theta) &= \bigg\{ F(c_\jmath|w_l) - \int 1\{g(w_l, \eta) \le c_\jmath\}\mu(d\eta) \bigg\}_{1 \le \jmath \le \mathcal{J}, 1 \le l \le \mathcal{L}}, \\ \Upsilon_F^{(\mu)}(\theta) &= \mu(\Omega) - 1, \\ \Upsilon_F^{(s)}(\theta) &= \int \Psi(g, \eta)\mu(d\eta) - \lambda. \end{aligned} \tag{30}$$

Given these definitions, we may then map R (as introduced in (28)) into the framework of Section 3 by noting that $R = \{\theta \in \mathbf{B} : \Upsilon_F(\theta) = 0 \text{ and } \Upsilon_G(\theta) \le 0\}$.

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—that is, 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

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

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 $p^{j_n}(w) = (p_1(w), \dots, p_{j_n}(w))'$, we set as our sieve

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

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 $q^{k_n}(w) = (q_1(w), \dots, q_{k_n}(w))'$, and for any $(\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu) = \theta$ define

$$Q_n(\theta) = \left\{ \sum_{\jmath=1}^{\mathcal{J}} \left\| \frac{1}{n} \sum_{i=1}^{n} \big(1\{Y_i \le c_\jmath\} - F(c_\jmath|W_i) \big) q^{k_n}(W_i) \right\|_{\hat{\Sigma}_{\jmath,n},2}^2 \right\}^{1/2}. \tag{31}$$

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) $\max_{1 \le k \le k_n} \| q_k \|_\infty \lesssim \sqrt{k_n}$; (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 $1 \le \jmath \le \mathcal{J}$: (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 Assumptions 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 $k_n \sqrt{j_n} \times \log^2(n)(\ell_n \vee \ell_n^{\mathrm{u}}) = o(1)$ and $k_n \sqrt{j_n} \log(n)/\sqrt{n} = o(\ell_n \wedge \ell_n^{\mathrm{u}})$, then uniformly in $P \in \mathbf{P}_0$*

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

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

$$G_n(\theta) = \big\{\big(\big\{p^{j_n\prime}\beta_{\jmath,h}\big\}_{\jmath=1}^{\mathcal{J}}, \mu_h\big) : \mu_h(B) \geq \sqrt{n}\min\big\{r_n - \mu(B), 0\big\} \text{ for all } B \in \mathcal{B}\big\} \tag{32}$$

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 (32) reduce to $\alpha_{sh} \geq \sqrt{n}\min\{r_n - \alpha_s, 0\}$ 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

$$\hat{V}_n(\theta, R| + \infty) = \big\{\big(\big\{p^{j_n\prime}\beta_{\jmath,h}\big\}_{\jmath=1}^{\mathcal{J}}, \mu_h\big) : h \in G_n(\theta), \Upsilon_F(h) = 0\big\}. \tag{33}$$

For each $\theta \in \Theta_n$, we denote the bootstrap process for the jth conditional moment by

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

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

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

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

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

$$\Gamma_n(\theta) \equiv \big\{\tilde{\mu} \in \mathcal{M}_n : \tilde{\theta} = \big(\big\{F(c_\jmath|\cdot)\big\}_{\jmath=1}^{\mathcal{J}}, \tilde{\mu}\big) \text{ satisfies } \Upsilon_F(\tilde{\theta}) = 0, \Upsilon_G(\tilde{\theta}) \leq 0\big\} \tag{34}$$

for any $\theta = (\{F(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu)$—that is, $\Gamma_n(\theta)$ 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 $\|\cdot\|_{\mathrm{TV}}$-Hausdorff distance between $\Gamma_n(\theta_1)$ and $\Gamma_n(\theta_2)$, which we denote by $d_H(\Gamma_n(\theta_1), \Gamma_n(\theta_2), \|\cdot\|_{\mathrm{TV}})$.

LEMMA 4.1: *If the probability measures $\{\delta_s\}_{s=1}^{s_n}$ are orthogonal, then, for every n, there is a $\zeta_n < \infty$ satisfying $d_H(\Gamma_n(\theta_1), \Gamma_n(\theta_2), \|\cdot\|_{\mathrm{TV}}) \leq \zeta_n \sum_{\jmath=1}^{\mathcal{J}} \|F_1(c_\jmath|\cdot) - F_2(c_\jmath|\cdot)\|_\infty$ 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*) $k_n^4 j_n^5 \log^5(n)/n = o(1)$; (*iv*) *$\Pi_n\theta = (\{F_n(c_\jmath|\cdot)\}_{\jmath=1}^{\mathcal{J}}, \mu_n)$ satisfies $\|F_n(c_\jmath|\cdot) -$*

$F_P(c_J|\cdot)\|_\infty = o(1)$ *uniformly in* $\theta \in \Theta_0 \cap R$ *and* $P \in \mathbf{P}_0$; (*v*) $k_n\sqrt{j_n}\log^2(n)\tau_n = o(1)$, *and* $\zeta_n(k_n j_n \log(n)/\sqrt{n} + \sqrt{j_n}\tau_n) = o(r_n)$.

The boundedness of $\Psi(g, \cdot)$ on Ω ensures $\Upsilon_F^{(s)}$ (as in (30)) 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 $F_P(c_J|\cdot)$. 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* $k_n\sqrt{j_n}\log(n)/\sqrt{n} = o(\ell_n \wedge \ell_n^{\mathrm{u}})$ *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 $I_n(R) - I_n(\Theta)$.

REFERENCES

AI, CHUNRONG, AND XIAOHONG CHEN (2003): "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1844. [710]

ANDREWS, DONALD W. K., AND XIAOXIA SHI (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666. [726,730]

ANGRIST, JOSHUA D., AND WILLIAM N. EVANS (1998): "Children and Their Parents' Labor Supply: Evidence From Exogenous Variation in Family Size," *The American Economic Review*, 88, 450–477. [710,712,714]

ARMSTRONG, TIMOTHY (2015): "Adaptive Testing on a Regression Function at a Point," *The Annals of Statistics*, 43, 2086–2101. [711]

BLUNDELL, RICHARD, XIAOHONG CHEN, AND DENNIS KRISTENSEN (2007): "Semi-Nonparametric iv Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613–1669. [722]

BUGNI, FEDERICO A., IVAN A. CANAY, AND XIAOXIA SHI (2017): "Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models," *Quantitative Economics*, 8, 1–38. [711]

CHEN, XIAOHONG, AND TIMOTHY M. CHRISTENSEN (2018): "Optimal sup-Norm Rates and Uniform Inference on Nonlinear Functionals of Nonparametric iv Regression," *Quantitative Economics*, 9, 39–84. [731]

CHEN, XIAOHONG, AND DEMIAN POUZO (2015): "Sieve Wald and qlr Inferences on Semi/Nonparametric Conditional Moment Models," *Econometrica*, 83, 1013–1079. [722,723]

CHEN, XIAOHONG, ELIE TAMER, AND ALEXANDER TORGOVITSKY (2011): "Sensitivity Analysis in Semiparametric Likelihood Models," Working paper, Yale University. [711]

CHERNOZHUKOV, VICTOR, HAN HONG, AND ELIE TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [722,726]

CHERNOZHUKOV, VICTOR, WHITNEY K. NEWEY, AND ANDRES SANTOS (2022): "Constrained Conditional Moment Restriction Models," arXiv preprint, arXiv:1509.06311. [712]

——— (2023): "Supplement to 'Constrained Conditional Moment Restriction Models'," *Econometrica Supplemental Material*, 91, https://doi.org/10.3982/ECTA13830. [711]

CHETVERIKOV, DENIS, AND DANIEL WILHELM (2017): "Nonparametric Instrumental Variable Estimation Under Monotonicity," *Econometrica*, 85, 1303–1320. [711]

FANG, ZHENG, AND JUWON SEO (2021): "A Projection Framework for Testing Shape Restrictions That Form Convex Cones," *Econometrica*, 89. 2439–2458. [711]
FREYBERGER, JOACHIM, AND JOEL L. HOROWITZ (2015): "Identification and Shape Restrictions in Nonparametric Instrumental Variables Estimation," *Journal of Econometrics*, 189, 41–53. [711]
FREYBERGER, JOACHIM, AND BRANDON REEVES (2018): "Inference Under Shape Restrictions," Available at SSRN 3011474. [711]
GENTZKOW, MATTHEW (2007): "Valuing New Goods in a Model With Complementarity: Online Newspapers," *American Economic Review*, 97, 713–744. [709]
HANSEN, BRUCE E. (1996): "Inference When a Nuisance Parameter Is not Identified Under the Null Hypothesis," *Econometrica*, 413–430. [713]
HANSEN, LARS P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 891–916. [710,712,721]
HAUSMAN, JERRY A., AND WHITNEY K. NEWEY (2016): "Individual Heterogeneity and Average Welfare," *Econometrica*, 84, 1225–1248. [709,712,731,732]
IMBENS, GUIDO W., AND CHARLES F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [717]
IMBENS, GUIDO W., AND JOSHUA D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. [714]
JACKWERTH, JENS C. (2000): "Recovering Risk Aversion From Option Prices and Realized Returns," *Review of Financial Studies*, 13, 433–451. [709]
KAIDO, HIROAKI, FRANCESCA MOLINARI, AND JÖRG STOYE (2019): "Confidence Intervals for Projections of Partially Identified Parameters," *Econometrica*, 87, 1397–1432. [711]
KLINE, PATRICK, AND CHRISTOPHER WALTERS (2021): "Reasonable Doubt: Experimental Detection of Job-Level Employment Discrimination," *Econometrica*, 89, 765–792. [710]
MATZKIN, ROSA L. (1994): "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics*, Vol. IV, ed. by Robert Engle and Dan McFadden. Elsevier. [709]
NEWEY, WHITNEY K., AND JAMES L. POWELL (2003): "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578. [710]
ROMANO, JOSEPH P., AND AZEEM M. SHAIKH (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–211. [717]
SANTOS, ANDRES (2012): "Inference in Nonparametric Instrumental Variables With Partial Identification," *Econometrica*, 80, 213–275. [711]
TAO, JING (2016): "On the Asymptotic Theory for Semiparametric GMM Models with Partial Identification," Working paper, University of Washington. [711]
TORGOVITSKY, ALEXANDER (2019): "Nonparametric Inference on State Dependence in Unemployment," *Econometrica*, 87, 1475–1505. [710]
ZHU, YU (2020): "Inference in Non-Parametric/Semi-Parametric Moment Equality Models With Shape Restrictions," *Quantitative Economics*, 11, 609–636. [711]

Co-editor Guido Imbens handled this manuscript.

Manuscript received 24 September, 2015; final version accepted 2 August, 2022; available online 4 November, 2022.