Econometrica, Vol. 75, No. 5 (September, 2007), 1243–1284

ESTIMATION AND CONFIDENCE REGIONS FOR PARAMETER SETS IN ECONOMETRIC MODELS[1]

BY VICTOR CHERNOZHUKOV, HAN HONG, AND ELIE TAMER

This paper develops a framework for performing estimation and inference in econometric models with partial identification, focusing particularly on models characterized by moment inequalities and equalities. Applications of this framework include the analysis of game-theoretic models, revealed preference restrictions, regressions with missing and corrupted data, auction models, structural quantile regressions, and asset pricing models.

Specifically, we provide estimators and confidence regions for the set of minimizers Θ_I of an econometric criterion function $Q(\theta)$. In applications, the criterion function embodies testable restrictions on economic models. A parameter value θ that describes an economic model satisfies these restrictions if $Q(\theta)$ attains its minimum at this value. Interest therefore focuses on the set of minimizers, called the identified set. We use the inversion of the sample analog, $Q_n(\theta)$, of the population criterion, $Q(\theta)$, to construct estimators and confidence regions for the identified set, and develop consistency, rates of convergence, and inference results for these estimators and regions. To derive these results, we develop methods for analyzing the asymptotic properties of sample criterion functions under set identification.

KEYWORDS: Set estimator, contour sets, moment inequalities, moment equalities, resampling, bootstrap.

1. INTRODUCTION

THIS PAPER PROVIDES estimators and confidence regions for minimizers of an extremum criterion function describing some testable restrictions on an economic model. A parameter value that describes an economic model satisfies the restrictions if the population criterion function attains its minimum at this value. When this criterion function is minimized uniquely at a particular parameter value, it is well known that one can obtain consistent estimators and confidence regions for this parameter value using a sample analog of this criterion function. This paper extends this criterion-based estimation and inference to models where the criterion function is minimized on a set of parameter values, called here the identified set, or identified region. (See Manski (2003).)

[1]This is a revision of a previous paper, "Parameter Set Inference in a Class of Econometric Models" (2002), which was circulated as a MIT working paper. We thank A. Abadie, D. Bertsimas, A. Belloni, G. Chamberlain, M. Cohen, A. Galichon, R. Guiteras, J. Hahn, C. Hansen, M. Henry, B. Honoré, G. Imbens, S. Izmalkov, Y. Kitamura, I. Makarov, C. Manski, A. Milkusheva, I. Molchanov, W. Newey, A. Rosen, O. Rytchkov, A. Simsek, and Neshe Yildiz for valuable comments that substantially improved the paper. We also thank four referees and Oliver Linton for valuable comments that also substantially improved the paper. We also thank seminar participants at many institutions where this paper was presented. Chernozhukov gratefully acknowledges research support from the Castle Krob Chair, National Science Foundation, and the Sloan Foundation. Hong gratefully acknowledges research support from the National Science Foundation. Tamer gratefully acknowledges research support from the National Science Foundation and the Sloan Foundation.

We focus the analysis on moment condition models defined by either moment inequalities or moment equalities.

This paper uses the inversion of the sample criterion functions as the building principle for the estimators and confidence regions. The resulting estimators and confidence regions are appropriate contour sets of the sample criterion functions. We develop consistency, rates of convergence, and inference results for these sets. Specifically, we show that in moment condition models an appropriate lower contour set of the sample criterion function converges in the Hausdorff metric to the identified set at essentially the $1/\sqrt{n}$ rate. In more general problems, the convergence occurs at other rates. We also develop a method for determining the appropriate level of the contour set so that it covers the identified set with a specified probability. For this purpose, we derive the asymptotics of several inferential statistics, the quantiles of which determine the appropriate level of the contour set.

Some of the most immediate applications of the estimation and inference methods developed in this paper include empirical game-theoretic models, empirical revealed preference analysis, econometric analysis with missing and mismeasured data, bounds analysis in auction models, structural quantile models and other simultaneous equation models without additivity, bounds analysis in asset pricing models, and inference on dominance regions in stochastic dominance analysis.[2] In most of these problems, the parameters in the economic models of interest must satisfy a collection of moment inequalities, and the resulting criterion functions are typically minimized on a set. Our paper develops estimators and confidence regions for these sets.

This paper advances the previous econometric literature on partial identification. The literature on partially identified models in econometrics has been largely initiated and popularized in an impressive body of work by Manski (see, in particular Manski (1990, 2003, 2007) and Manski and Tamer (2002)). The early ideas of set identification go back to Frisch (1934) and Marschak and Andrews (1944). Marschak and Andrews constructed the identified set as a collection of parameters representing different production functions that cannot be rejected by the data and that are consistent with the functional restrictions that the authors consider. Frisch (1934) constructed consistent interval bounds on parameters of structural regression equations that are subject to measurement error. Klepper and Leamer (1984) generalized the Frisch bounds to multivariate regression models with measurement errors and constructed consistent estimates. Gilstein and Leamer (1983) provided set consistent estimators

[2]For a detailed discussion of each problem, see, for example, Varian (1982), Hansen, Heaton, and Luttmer (1995), McFadden (2005), Blundell, Browning, and Crawford (2005), Manski and Tamer (2002), Manski (2003), Molinari (2004), Haile and Tamer (2003), Ciliberto and Tamer (2003), Chernozhukov and Hansen (2005), Chernozhukov, Hansen, and Jansson (2005), and Linton, Post, and Whang (2005).

in a class of nonlinear regression models where the identified set is an interval of parameters that are robust to the misspecification of the distribution of the error term. In a different development, Phillips (1989) suggested that multicollinearity of a general kind may be a cause for partial identification in a number of econometric models and provided a set of asymptotic results for Wald statistics under such conditions. Hansen, Heaton, and Luttmer (1995) proposed an estimator for the region of feasible means and variances of pricing kernels in asset pricing model and proved its consistency. Manski and Tamer (2002) developed a number of models with interval-censored data and proved several consistency results. In a previous version of this paper, Chernozhukov, Hong, and Tamer (2002) developed consistency and inference results and developed an empirical application of the linear moment inequality models. Our paper is also related to the literature on the weak identification problem, notably Dufour (1997) and Staiger and Stock (1997). However, our problem differs considerably, since the nature of the partial identification in our main applications cannot typically be approximated by the weak identification framework.

Our work also advances the statistical literature on partial identification. Hannan (1982) emphasized and studied the partial identification problem in several time series models. Redner (1981) and Hannan and Deistler (1988) showed that any sequence of maximum likelihood estimators has its limit points within the identified set Θ_I (see also Pötscher and Prucha (1997) and van der Vaart (1998)). Our analysis of consistency and rate of convergence for general extremum problems, and moment condition problems in particular, substantively goes beyond these earlier results. Veres (1987), Dacunha-Castelle and Gassiat (1999), Liu and Shao (2003), and Fukumizu (2003) investigated the behavior of the likelihood ratio test under loss of identifiability in correctly specified likelihood models, with a special focus on the mixture, autoregressive moving average and neural network models. These results do not readily extend to the extremum problems and moment condition models analyzed in this paper. There is also a connection to the literature on image processing dealing with support estimation of a density; see, for example, Korostelëv, Simar, and Tsybakov (1995) and Cuevas and Fraiman (1997). The structure of such problems differs considerably from the general extremum and moment condition problems analyzed in this paper.

The rest of the paper is organized as follows. Section 2 outlines the moment condition models and presents several examples that will serve to illustrate the analysis. Section 2 also informally outlines the main results of the paper. Section 3 develops consistency, rates of convergence, and inference results that apply generically. Section 4 analyzes the moment inequality and moment equality models in detail and verifies the conditions of Section 3. Section 5 briefly discusses pointwise regions and develops their properties as estimators of identified sets. Section 6 offers conclusions and discusses recent research on the problem of set estimation and inference. The Appendix collects proofs and a definition of notations used in the paper.

2. PROBLEM DEFINITION AND INFORMAL DISCUSSION OF THE MAIN RESULTS

Suppose that we have a collection of empirical restrictions on economic models. The restrictions are embodied in a nonnegative population criterion function $Q(\theta)$, where θ is a parameter describing the models. The parameter θ belongs to the parameter space Θ, which is a compact subset of the Euclidean space $\mathbb{R}^d$. We say that a model satisfies our testable empirical restrictions if $Q(\theta) = 0$, that is, if the population criterion function is minimized at the value of θ that indexes the model. The set of parameter values that satisfy our empirical restrictions, $\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\}$, will generally be multivalued. This paper develops estimators and confidence regions for the set Θ_I constructed using the contour sets of the sample criterion function $Q_n(\theta)$. Section 5 also discusses a related problem of constructing confidence regions for a particular point, θ^*, in Θ_I.

This section begins with a review of the main econometric models and economic examples that motivate the framework described above. An informal review of the methods and the results obtained in this paper follows.

2.1. *The General Framework of Moment Condition Models*

This paper is primarily concerned with applications to two main types of moment conditions: moment equalities and moment inequalities. In empirical analysis, the moment conditions represent testable restrictions on economic models. Economic models are described by the finite-dimensional parameter θ, and we are interested in the set of parameter values that satisfy these testable restrictions.

Moment-inequality restrictions take the form

$$E_P[m_i(\theta)] \leq 0. \tag{2.1}$$

The quantity $m_i(\theta) := m(\theta, w_i)$ is a vector of moment functions parameterized by θ and determined by a vector of real random variables w_i. The expectation E_P is computed with respect to the population probability law P of the data w_i. Thus, the set of parameter values θ that pass the restrictions in (2.1) is given by $\Theta_I = \{\theta \in \Theta : E_P[m_i(\theta)] \leq 0\}$.

It is interesting to comment on the structure of the set Θ_I in this model. When the moment functions are linear in parameters, the set Θ_I is given by an intersection of linear half-spaces; it could be a triangle, a trapezoid, or a polyhedron, as in Examples 1 and 2 introduced below. When the moment functions are nonlinear, the set Θ_I is a region bounded by nonlinear manifolds.

The set Θ_I can be characterized as the set of minimizers of the criterion function[3]

$$Q(\theta) := \left\| E_P[m_i(\theta)]' W^{1/2}(\theta) \right\|_+^2, \tag{2.2}$$

[3]Let $\|x\|_+ = \|(x)_+\|$ and $\|x\|_- = \|(x)_-\|$, where $(x)_+ := \max(x, 0)$ and $(x)_- := \max(-x, 0)$.

where $W(\theta)$ is a continuous and diagonal matrix with strictly positive diagonal elements for each $\theta \in \Theta$. Therefore, inference on Θ_I may be based on the empirical analog of Q,

$$(2.3) \qquad Q_n(\theta) := \left\| E_n[m_i(\theta)]' W_n^{1/2}(\theta) \right\|_+^2, \quad E_n[m_i(\theta)] := \frac{1}{n}\sum_{t=1}^{n} m_t(\theta),$$

where the matrix $W_n(\theta) = W_n(\theta, w_1, \dots, w_n)$ is an estimate that converges to $W(\theta)$ uniformly in $\theta \in \Theta$. In applications, $W_n(\theta)$ can be taken to be an identity matrix or chosen to weigh the individual empirical moments by estimates of the inverses of their individual variances.

Moment-equality restrictions are more traditional in empirical analysis and take the form

$$(2.4) \qquad E_P[m_i(\theta)] = 0, \quad \text{that is,} \quad \Theta_I = \left\{\theta \in \Theta : E_P[m_i(\theta)] = 0\right\}.$$

When the moment functions are linear in parameters, the identified set Θ_I is either a singleton or a set of points defined by an intersection of a hyperplane with the parameter space Θ. When the moment functions are nonlinear in parameters, the set Θ_I is typically a manifold, which also includes the case of isolated points (a zero-dimensional manifold).

The set Θ_I can be characterized by the set of minimizers of the generalized method of moments function

$$(2.5) \qquad Q(\theta) := \left\| E_P[m_i(\theta)]' W^{1/2}(\theta) \right\|^2,$$

where $W(\theta)$ is a continuous and positive-definite matrix for each $\theta \in \Theta$. The inference on Θ_I can be based on the conventional sample analog

$$(2.6) \qquad Q_n(\theta) := \left\| E_n[m_i(\theta)]' W_n^{1/2}(\theta) \right\|^2,$$

where $W_n(\theta)$ is an estimate that converges uniformly to $W(\theta)$. In applications, $W_n(\theta)$ is typically an estimate of the inverse of the asymptotic covariance matrix of the empirical moment functions.

In many situations, we can use the normalized objective function for estimation and inference:

$$Q_n'(\theta) = Q_n(\theta) - \inf_{\theta' \in \Theta} Q_n(\theta').$$

This normalization is useful in cases where Q_n does not attain the population value zero in finite samples. In such cases, using the modified objective function often yields improvements in power, as is well known in the point-identified cases (e.g., Mikusheva (2006)).

2.2. *Motivating Examples*

There are several interesting examples for the moment condition models described above, where the identified set Θ_I is not a single point, but rather a collection of points.

EXAMPLE 1—Interval Data: The first example is motivated by missing data problems, where Y is an unobserved real random variable bracketed below by Y_1 and above by Y_2, both of which are observed real random variables. The parameter of interest $\theta = E_P[Y]$ is known to satisfy the restriction

$$E_P[Y_1] \leq \theta \leq E_P[Y_2].$$

Hence the identified set is an interval, $\Theta_I = \{\theta : E_P[Y_1] \leq \theta \leq E_P[Y_2]\}$. This example falls in the moment-inequality framework with moment function

$$m_i(\theta) = (Y_{1i} - \theta, \theta - Y_{2i})'.$$

Therefore, Θ_I can be characterized as the set of minimizers of $Q(\theta) = \|E_P[m_i(\theta)]\|_+^2 = (E_P[Y_{1i}] - \theta)_+^2 + (E_P[Y_{2i}] - \theta)_-^2$, which has sample analog $Q_n(\theta) = (E_n[Y_{1i}] - \theta)_+^2 + (E_n[Y_{2i}] - \theta)_-^2$.

EXAMPLE 2—Interval Outcomes in Regression Models: A regression generalization of the previous example is immediate. Suppose a regressor vector X_i is available and the conditional mean of unobserved Y_i is modeled using the linear function $X_i'\theta$. The parameters of this function can be bounded using the inequalities $E_P[Y_{1i}|X_i] \leq X_i'\theta \leq E_P[Y_{2i}|X_i]$. These conditional restrictions imply the inequalities

$$E_P[Y_{1i}Z_i] \leq \theta' E_P[X_i Z_i] \leq E_P[Y_{2i}Z_i],$$

where Z_i is a vector of positive transformations of X_i, for example, $Z_i = \{1(X_i \in \mathcal{X}_j), j = 1, \ldots, J\}'$ for a suitable collection of sets $\mathcal{X}_j$. These inequalities define the identified set Θ_I, which is therefore given by an intersection of linear half-spaces in $\mathbb{R}^d$. This example also falls in the moment-inequality framework, with moment function given by

$$m_i(\theta) = ((Y_{1i} - \theta' X_i)Z_i', -(Y_{2i} - \theta' X_i)Z_i')'.$$

In auction analysis, the bracketing of the latent response Y, which represents the bidder's valuation, by functions of observed bids, Y_1 and Y_2, is very natural and occurs in a variety of settings (Haile and Tamer (2003)). Analogous situations occur in income surveys, where only income brackets are available instead of actual income (Manski and Tamer (2002)). Chernozhukov, Hong, and Tamer (2002) analyzed this linear moment inequality setup in detail.

EXAMPLE 3—Optimal Choice of Economic Agents and Game Interactions: Another application of (2.1) is the analysis of the optimal choice behavior of firms and economic agents. Suppose that a firm can make one of two choices: $D_i = 0$ or $D_i = 1$. Suppose that the profit of the firm from making the choice D_i is given by $\pi(W_i, D_i, \theta) + U_i$, where U_i is a disturbance such that $E_P[U_i|X_i] = 0$, X_i represents the information available to the firm making the decision, and W_i are various determinants of the firm's profit, some of which may be included in X_i. For example, W_i may include actions of other firms that affect the firm's profit. From a revealed preference principle, the fact that the firm chooses D_i necessarily implies that

$$E_P[\pi(W_i, D_i, \theta)|X_i] \geq E_P[\pi(W_i, 1 - D_i, \theta)|X_i]. \tag{2.7}$$

Therefore, we can take the moment condition in (2.1) to be

$$m_i(\theta) = (\pi(W_i, 1 - D_i, \theta) - \pi(W_i, D_i, \theta))Z_i, \tag{2.8}$$

where Z_i is the set of positive instrumental variables defined as in the previous example.

This simple example highlights the structure of empirically testable restrictions arising from the optimizing behavior of firms and economic agents. These testable restrictions can be seen as moment-inequality conditions. Note that this simple example also allows for game-theoretic interactions among economic agents. The moment-inequality conditions of the kind above are ubiquitous in structural problems (Ciliberto and Tamer (2003), Bajari, Benkard, and Levin (2006), Ryan (2005)). Related ideas also appear in the area of stochastic revealed preference analysis (Varian (1984), McFadden (2005), Blundell, Browning, and Crawford (2005)).

EXAMPLE 4—Structural Equations: Consider instrumental variable estimation of the returns to schooling. Potential income Y is related to education E through a flexible quadratic functional form $Y = \theta_0 + \theta_1 E + \theta_2 E^2 + \epsilon = X'\theta + \epsilon$, where $\theta = (\theta_0, \theta_1, \theta_2)$ and $X = (1, E, E^2)'$. Although parsimonious, this simple quadratic model is not point-identified in the presence of the standard single quarter-of-birth instrument I (Angrist and Krueger (1992)). We have only one instrument I for two endogenous variables E and E^2. In the absence of point identification, all parameter values θ consistent with the instrumental orthogonality restriction $E_P[(Y - \theta' X)Z] = 0$, where $Z = (1, I)'$, are of interest in economic analysis. Phillips (1989) developed a number of related important examples. Similar partial identification problems also arise in nonlinear moment and instrumental variables problems (Demidenko (2000), Chernozhukov and Hansen (2005)). In Chernozhukov and Hansen (2005), the parameter values θ of the structural quantile functions for the returns to schooling must satisfy the restrictions $E_P[(\tau - 1(Y \leq X'\theta))Z] = 0$, where $\tau \in (0, 1)$ is the quantile of interest. This is an example of a nonlinear instrumental variable model, where, in the absence of point identification, the identification region is generally given by a nonlinear manifold. Chernozhukov and Hansen (2004) and Chernozhukov, Hansen, and Jansson (2005) presented empirical applications of a returns-to-schooling model and a structural demand model with partial identification.

2.3. *Overview of Methods and Results*

The objective of this paper is to construct sets that are consistent estimates of Θ_I, that converge to Θ_I at the fastest rates, and that have the confidence regions property defined below. The sets constructed take the form of a contour set of level c, denoted as $C_n(c)$, of the sample criterion function $a_n Q_n$:

$$C_n(c) := \{\theta \in \Theta : a_n Q_n(\theta) \leq c\}$$

for some normalizing sequence a_n. In Examples 1–4, the normalizing sequence $a_n = n$, and we use it in this section for simplicity. This sequence is selected so that the inferential statistic

$$\mathcal{C}_n := \sup_{\theta \in \Theta_I} a_n Q_n(\theta) \tag{2.9}$$

is stochastically bounded and has a nondegenerate large sample distribution.

The level $c = \widehat{c}$, possibly data dependent, should be selected efficiently, meaning that the estimator $C_n(\widehat{c})$ should converge to Θ_I at the fastest possible rate. A general approach is to select $c = \widehat{c}$ to be as small as possible, but no smaller than $\mathcal{C}_n$, with probability approaching 1. A feasible way to achieve this is to select the level $\widehat{c}$ to grow very slowly; for concreteness, we select $\widehat{c} \propto \ln n$. In problems with a degeneracy property, defined below, a more special approach is possible: We can feasibly select $\widehat{c}$ not to grow at all, namely set $\widehat{c} \geq q_n := \inf_{\theta \in \Theta} a_n Q_n(\theta)$ such that $\widehat{c} = O_p(1)$. Indeed, in Example 1, setting $\widehat{c} = 0$ gives us $C_n(0) = [E_n[Y_1], E_n[Y_2]]$, which clearly is consistent for the region $[E_P[Y_1], E_P[Y_2]]$ at the $1/\sqrt{n}$ rate.

The formal analysis of the rates of convergence and consistency makes use of the Hausdorff distance between sets, defined as

$$d_H(A, B) := \max\Big[\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A)\Big],$$

where $d(b, A) := \inf_{a \in A} \|b - a\|$ and $d_H(A, B) := \infty$ if either A or B is empty. This metric is a natural generalization of the Euclidean distance. Our general consistency result $d_H(C_n(\widehat{c}), \Theta_I) \to_p 0$ follows from the uniform convergence of the sample function Q_n to the limit continuous function Q over the compact parameter space Θ, with the $1/a_n$ rate of convergence over the identified set Θ_I.

The rates of convergence follow from the existence of polynomial minorants on $Q_n(\theta)$ over suitable neighborhoods of Θ_I, as defined formally in Section 3. The existence of quadratic minorants on Q_n occurring in Examples 1–4, as verified in Section 4, implies that

$$d_H(C_n(\widehat{c}), \Theta_I) = O_p(\sqrt{\max(\widehat{c}, 1)/n}).$$

With the choice of $\widehat{c}$ described above, the rate of convergence is nearly $1/\sqrt{n}$, and, in many moment-inequality problems, exactly $1/\sqrt{n}$. We can attain the exact $1/\sqrt{n}$ rate under the *degeneracy property*—there are subsets of Θ that can approximate Θ_I at the $1/\sqrt{n}$ rate and the criterion function $a_n Q_n$ vanishes or is equal to its minimum value on these subsets. Section 3 discusses this property further, and Section 4 shows that this property typically holds in Examples 1–3. On the other hand, this property does not appear to hold generally in moment-equality problems.

We also want $C_n(c)$ to be a confidence region for Θ_I, namely

$$\lim_{n\to\infty} P(\Theta_I \subseteq C_n(c)) = \alpha$$

for a specified confidence level $\alpha \in (0, 1)$.[4] The event $\{\Theta_I \subseteq C_n(c)\}$ is equivalent to the event $\{\mathcal{C}_n \leq c\}$, where $\mathcal{C}_n$ is the inferential statistic defined in (2.9). Therefore, we need to select $c = \widehat{c}$ such that $\widehat{c}$ is a consistent estimate of the α-quantile of $\mathcal{C}_n$.

The estimate $\widehat{c}$ can be obtained from the limit distributions of (2.9) presented in Section 4 or by the generic subsampling method presented in Section 3.5. For instance, in Example 1, if $\sqrt{n}((E_n[Y_1] - E_P[Y_1]), (E_n[Y_2] - E_P[Y_2])) \to_d (W_1, W_2) = N(0, \Omega)$, then $\mathcal{C}_n \to_d \mathcal{C} = \max[(W_1)_+^2, (W_2)_-^2]$. In this case, we can easily estimate quantiles of $\mathcal{C}$ by simulation. For cases with less tractable limits $\mathcal{C}$, we can estimate the quantiles by the generic subsampling method developed in Section 3.5. The method subsamples the statistic (2.9) after replacing the identified set Θ_I by an efficient estimator $C_n(\widehat{c})$.

When the degeneracy property does not hold, the confidence region may be an inconsistent estimator of Θ_I. Inconsistency means that the confidence region may be very far away from the identified set Θ_I in the Hausdorff distance. Thus, to ensure consistency, we need to take an expanded region $C_n(\widehat{c} + \kappa_n)$, which is a $\sqrt{\ln n/n}$ consistent estimator of Θ_I in regular cases. This expanded region is also a confidence region with an asymptotic coverage of 1.

There are two difficulties arising in the asymptotic analysis of general moment problems. The first difficulty is nonequicontinuity of the empirical process $\theta \mapsto a_n Q_n(\theta)$ in moment-inequality problems. We address this

[4]We discuss robustness of estimation and coverage results under contiguous perturbations of P in the online supplement (Chernozhukov, Hong, and Tamer (2007)).

difficulty by adapting the notions of epiconvergence and stochastic equi-semicontinuity (Knight (1999)) to the set-identified case. The second difficulty is the parameter on the boundary problem, arising when the identified set Θ_I does not lie in the interior of Θ. For instance, in Example 4, the identified set Θ_I is an intersection of a hyperplane with Θ, necessarily containing common points with the boundary of Θ. We address this difficulty through a generalization of the Chernoff regularity to the set-identified case. In the point-identified case, the Chernoff regularity requires convergence of the local parameter space to a cone (Chernoff (1954), Andrews (2001)). Our generalized condition requires a convergence of the graph of the local parameter space to a limit graph.

3. GENERAL ESTIMATION AND INFERENCE IN LARGE SAMPLES

In this section we formally define the estimators and confidence regions, and develop the basic results on the consistency, rates of convergence, and coverage properties of these regions. We develop general conditions that parallel those used in extremum estimation in point-identified cases (Amemiya (1985), Newey and McFadden (1994), van der Vaart (1998)). In the subsequent Section 4, we illustrate and verify these conditions for the moment condition models.

We use the following notation in the sequel: The ϵ-expansion of Θ_I in Θ is defined as $\Theta_I^\epsilon := \{\theta \in \Theta : d(\theta, \Theta_I) \leq \epsilon\}$, and the ϵ-contraction of the set Θ_I in Θ is defined as $\Theta_I^{-\epsilon} := \{\theta \in \Theta_I : d(\theta, \Theta \setminus \Theta_I) \geq \epsilon\}$, where $\epsilon \geq 0$. Unless an ambiguity arises, $\sup_A f$ is used to denote $\sup_{a \in A} f(a)$. The notions of stochastic convergence are defined as in van der Vaart and Wellner (1996). For any two numbers a and b, $a \wedge b$ denotes $\min(a, b)$ and $a \vee b$ denotes $\max(a, b)$. For convenience, the Appendix collects other definitions and notation.

3.1. *Consistency and Rates of Convergence in the General Case*

We assume the following condition holds.

CONDITION C.1—Consistency: (a) *The parameter space Θ is a nonempty compact subset of $\mathbb{R}^d$.* (b) *There is a lower semicontinuous population criterion function $Q : \Theta \mapsto \mathbb{R}_+$ such that $\inf_\Theta Q = 0$. Let $\Theta_I := \arg\inf_\Theta Q$ be the set of its minimizers, called the identified set.* (c) *There is a sample criterion function $Q_n(\theta) = Q_n(\theta, w_1, \ldots, w_n)$ that takes values in $\mathbb{R}_+$ and is jointly measurable in the parameter $\theta \in \Theta$ and the data $w_1, \ldots, w_n$ defined on a complete probability space $(\Omega, \mathcal{F}, P)$.* (d) *The sample criterion function is uniformly no smaller than the population function in large samples, that is, $\sup_\Theta (Q - Q_n)_+ = O_p(1/b_n)$ for a sequence of constants $b_n \to \infty$.* (e) *The sample criterion converges to the limit criterion function over the identified set Θ_I at the $1/a_n$ rate, that is, $\sup_{\Theta_I} Q_n = O_p(1/a_n)$ for a sequence of normalizing constants $a_n \to \infty$.*

Condition C.1 assumes one-sided uniform convergence for the criterion function Q_n to the limit criterion function Q. This condition includes the usual uniform convergence, $\sup_\Theta |Q_n - Q| = o_p(1)$, as a special case. Section 4 verifies the uniform convergence condition for moment condition models with $b_n = \sqrt{n}$. Condition C.1 also identifies Θ_I as the minimizer of the limit criterion function Q. The normalization to nonnegativity, $Q \geq 0$ and $Q_n \geq 0$, is not restrictive.[5] Condition C.1(c) assumes measurability of the criterion function with respect to the product of the sigma-field $\mathcal{F}$ and the Borel sigma-field of Θ, and thus guarantees measurability of $\sup_{\Theta_I} Q_n$ and related statistics (e.g., van der Vaart and Wellner (1996, p. 47).[6] Condition C.1 also defines the principal quantity

$$\mathcal{C}_n = \sup_{\theta \in \Theta_I} a_n Q_n(\theta), \tag{3.1}$$

which plays the crucial role in the analysis of consistency, rates of convergence, and inference. Section 4 verifies that in moment condition models $a_n = n$.

In order to study convergence rates, we impose the following condition, a replacement for the weaker Condition C.1(d).

CONDITION C.2—Existence of a Polynomial Minorant: *There exist positive constants* (δ, κ, γ) *such that for any* $\varepsilon \in (0, 1)$ *there are* $(\kappa_\varepsilon, n_\varepsilon)$ *such that for all* $n \geq n_\varepsilon$,

$$Q_n(\theta) \geq \kappa \cdot [d(\theta, \Theta_I) \wedge \delta]^\gamma$$

uniformly on $\{\theta \in \Theta : d(\theta, \Theta_I) \geq (\kappa_\varepsilon / a_n)^{1/\gamma}\}$, *with probability at least* $1 - \varepsilon$.

Condition C.2 states that Q_n can be stochastically bounded below, over a neighborhood of the identified set Θ_I, by a polynomial in the distance from Θ_I. Condition C.2 parallels the conditions used to derive the rate of convergence of estimators in the point-identified case.

The contour sets of $a_n Q_n$ form the class of estimators considered here. The contour set of $a_n Q_n$ of level c is defined as

$$C_n(c) := \{\theta \in \Theta : a_n Q_n(\theta) \leq c\}, \tag{3.2}$$

where $c \geq 0$. The estimates and confidence regions will generally take the form $\widehat{\Theta}_I := C_n(\widehat{c})$ for some possibly data-dependent choice of $\widehat{c}$.

[5]Indeed, for any $\tilde{f} : \Theta \to \mathbb{R}$, $f(\theta) = \tilde{f}(\theta) - \inf_{\theta' \in \Theta} \tilde{f}(\theta') \geq 0$, provided the last term is finite.

[6]These conditions, imposed due to a referee's request, are not strictly necessary, since we can allow for stochastic convergence in the sense of Hoffmann–Jorgensen.

THEOREM 3.1—Coverage, Consistency, and Rates of Convergence of $C_n(\widehat{c})$: *Let* $\widehat{\Theta}_I = C_n(\widehat{c})$, *where* $\widehat{c} \geq \mathcal{C}_n$ *with probability approaching* 1 *but* $\widehat{c}/a_n \to_p 0$. *Suppose that* $\Theta_I \neq \Theta$. *Then* (1) *Condition* C.1 *implies that* $\Theta_I \subseteq \widehat{\Theta}_I$ *with probability approaching* 1 *and* $d_H(\widehat{\Theta}_I, \Theta_I) = o_p(1)$, *and* (2) *Conditions* C.1 *and* C.2 *imply that* $d_H(\widehat{\Theta}_I, \Theta_I) = O_p((1 \vee \widehat{c})/a_n)^{1/\gamma}$. *Suppose that* $\Theta_I = \Theta$, *then* (3) *Condition* C.1 *implies that* $d_H(\widehat{\Theta}_I, \Theta_I) = 0$ *with probability approaching* 1.

Parts (1) and (2) of Theorem 3.1 address the case of partial identification, $\Theta_I \neq \Theta$. Our interest lies primarily in this case. The consistency results (1) and (2), stated in terms of the Hausdorff metric, generalize those obtained for point-identified cases (see, e.g., van der Vaart (1998)). Both the consistency and rate results are new for the problem studied in this paper.[7] Part (3) of Theorem 3.1 also addresses the case of the complete nonidentification, that is, $\Theta_I = \Theta$. In this case, the estimator converges to Θ_I in the Hausdorff metric faster than any rate. This case is not of prime interest and is stated for completeness.

Theorem 3.1 also suggests that the level $c = \widehat{c}$ should be selected *efficiently*, that is, the estimator $C_n(\widehat{c})$ should converge to Θ_I at the fastest possible rate. A general approach is to select $\widehat{c}$ to be as small as possible subject to the constraint $\widehat{c} \geq \mathcal{C}_n$ holding with probability approaching 1. Setting $\widehat{c} = O_p(1)$ subject to this constraint gives the efficient rate $a_n^{-1/\gamma}$, but this choice is not feasible in practice. A feasible method for implementing nearly efficient estimation is to let $\widehat{c}$ grow very slowly with n, for example, $\widehat{c} \propto \ln n$.

In moment condition models, we verify in Section 4 that Conditions C.1 and C.2 hold with $a_n = n$ and $\gamma = 2$. Thus, it follows by Theorem 3.1 that with the choice of $\widehat{c}$ given above, the convergence rate is essentially $1/\sqrt{n}$, namely $\sqrt{\log n/n}$. We will discuss the choice of $\widehat{c}$ further when we consider inference.

3.2. *Consistency and Rates of Convergence with Degeneracy*

In many moment-inequality problems, the exact rate of convergence $a_n^{-1/\gamma}$ can be attained *without* setting $\widehat{c} \geq \mathcal{C}_n$. The main reason is that in moment-inequality problems, the criterion function Q_n can be degenerate over subsets of Θ that can approximate Θ_I. The discussion of Example 1 below provides the simplest instance where this is possible. Consistency and rate results then follow, because $C_n(\widehat{c})$ includes these subsets and, by the previous Condition C.2, cannot be much larger than these subsets. We state the degeneracy property as follows.

[7]The consistency result differs from an earlier result by Manski and Tamer (2002) that derives consistency of the set $\{\theta \in \Theta : Q_n(\theta) \leq c/b_n\}$, where $b_n = \sqrt{n}$ in regular cases. We show consistency of smaller sets $\{\theta \in \Theta : Q_n(\theta) \leq c/a_n\}$, where $a_n = n$ in regular cases.

CONDITION C.3—Degeneracy: (a) *There is a sequence of subsets* Θ_n *of* Θ, *which could be data dependent, such that* Q_n *vanishes on these subsets, that is,* $Q_n(\theta) - \inf_{\theta\in\Theta} Q_n(\theta) = 0$ *for each* $\theta \in \Theta_n$, *for each* n, *and these sets can approximate the identified set arbitrarily well in the Hausdorff distance, that is,* $d_H(\Theta_n, \Theta_I) \leq \epsilon_n$ *for some* $\epsilon_n = o_p(1)$. (b) *The sequence* ϵ_n *is of stochastic order* $a_n^{-1/\gamma}$.

Condition C.3 together with Conditions C.1 and C.2 allow us to claim results on consistency and convergence rates of the contour set $C_n(\widehat{c})$. Condition C.3 alone is generally *not* sufficient for consistency, since it does not guarantee that the contour sets are not much larger than the identified set Θ_I. Section 4 verifies this condition in our main applications. It is useful to note that Condition C.3 allows us to cover the classical point-identified case as a special case. In such a case, Conditions C.1 and C.2 in fact imply Condition C.3(a) and (b), respectively, provided Q_n is lower semicontinuous.

In moment-inequality models, the following situation can often arise: Let $\Theta_I^{-\epsilon}$ denote an ϵ-contraction of the identified set Θ_I in Θ, as previously defined. Under conditions specified in Section 4, all finite-sample moment inequalities become satisfied on a contraction $\Theta_n = \Theta_I^{-\epsilon_n}$ with $\epsilon_n = O_p(1/\sqrt{n})$, and the criterion function vanishes on this contraction. Moreover, if Θ_I is regular enough, then ϵ_n-contractions can approximate Θ_I in the Hausdorff metric, $d_H(\Theta_I, \Theta_I^{-\epsilon_n}) = O(\epsilon_n)$. Condition C.3 covers this case, but allows for other types of degeneracy. Of course, Condition C.3 trivially covers the point-identified case, as noted earlier.

The following theorem establishes consistency and rates of convergence under degeneracy.

THEOREM 3.2—Consistency and Rates of Convergence of $C_n(\widehat{c})$ with Degenerate Subsets: *Let* $\widehat{\Theta}_I$ *denote* $C_n(\widehat{c}')$, *where* $\widehat{c}' \geq q_n =: \inf_{\theta\in\Theta} a_n Q_n(\theta)$ *with probability* 1 *and* $\widehat{c}' = O_p(1)$. *Then* (1) *Conditions* C.1 *and* C.3(a) *imply consistency of the estimator, that is,* $d_H(\widehat{\Theta}_I, \Theta_I) = o_p(1)$, *and* (2) *Conditions* C.2 *and* C.3 *imply the convergence rate* $d_H(\widehat{\Theta}_I, \Theta_I) = O_p(a_n^{-1/\gamma})$. *Moreover, if* $\Theta_I = \Theta$ *and* $\sup_{\Theta_I} a_n Q_n = q_n$ *with probability approaching* 1, *then* (3) $d_H(\widehat{\Theta}_I, \Theta_I) = 0$ *with probability approaching* 1.

Parts (1) and (2) state that the rate $1/a_n^{1/\gamma}$ is achieved exactly under Condition C.3. In particular, the smallest nonempty contour set, the argmin, is consistent and converges to Θ_I at the rate $1/a_n^{1/\gamma}$. Section 4 shows that in many moment inequality examples, Condition C.3(b) and (c) hold with $\gamma = 2$ and $a_n = n$, yielding the rate of convergence $1/\sqrt{n}$. Part (3) addresses the less typical case of the complete nonidentification, $\Theta_I = \Theta$, where the estimator converges to Θ_I in the Hausdorff metric faster than any rate.

EXAMPLE 1—Continued: Recall that $Q_n(\theta) = (E_n[Y_1] - \theta)_+^2 + (E_n[Y_2] - \theta)_-^2$ and $Q(\theta) = (E_P[Y_1] - \theta)_+^2 + (E_P[Y_2] - \theta)_-^2$. Suppose that $(\sqrt{n}(E_n[Y_1] - E_P[Y_1]), \sqrt{n}(E_n[Y_2] - E_P[Y_2]))' \to_d (W_1, W_2)' \sim N(0, \Omega)$. Then $\sup_\Theta |Q_n - Q| = O_p(1/\sqrt{n})$ while $\sup_{\Theta_I} |Q_n - Q| = O_p(1/n)$, so that $b_n = \sqrt{n}$ and $a_n = n$. By Theorem 3.1, $C_n(\ln n)$ consistently estimates $\Theta_I = [E_P[Y_1], E_P[Y_2]]$. Further, we see that Condition C.2 holds with $\gamma = 2$. Hence by Theorem 3.1, the set $C_n(\ln n)$ is consistent at the $\sqrt{\ln n/n}$ rate. Note, however, that the set $C_n(0) = [E_n[Y_1], E_n[Y_2]]$ is consistent at the $1/\sqrt{n}$ rate. Notice that this is precisely the example with the degeneracy property Condition C.3: $Q_n(\theta) = 0$ on $\theta \in \Theta_n = [E_n[Y_1], E_n[Y_2]]$ and $d_H(\Theta_n, \Theta_I) = O_p(1/\sqrt{n})$. Therefore, in this example, the set $C_n(c)$ with $c \geq 0$ is consistent at the rate of $1/\sqrt{n}$.

EXAMPLE 4—Continued: Finally, we emphasize that we cannot attain the sharp rate of convergence $a_n^{-1/\gamma}$ by setting $c = 0$ or constant in all cases. Recall Example 4, but for simplicity of the argument, consider the case where the linear instrumental variable model is point-identified, the number of strong instruments is k, and the number of endogenous variables is $d < k$. Then the minimized criterion function $\inf_\Theta nQ_n$, known as the J-statistic, is approximately distributed as χ^2_{k-d}, the chi-squared variable with $k - d$ degrees of freedom. Then $C_n(c)$ is empty with probability $P(\chi^2_{k-d} > c) > 0$ in large samples. Therefore, $d_H(C_n(c), \Theta_I) > 1$ with this probability in large samples, implying that $C_n(c)$ is inconsistent.

3.3. *Confidence Regions*

The question that arises next is how to choose $\widehat{c}$ to guarantee that $C_n(\widehat{c})$ has a confidence region property. Observe that the inferential properties of sets $C_n(c)$ are determined by the statistic $\mathcal{C}_n = \sup_{\theta \in \Theta_I} a_n Q_n(\theta)$. Indeed, the event $\{\mathcal{C}_n \leq c\}$ is equivalent to the event $\{\Theta_I \subseteq C_n(c)\}$. Therefore, if we know quantiles of $\mathcal{C}_n$ or good upper bounds on them, we can conduct finite-sample inference. For instance, we can obtain the upper bounds on quantiles using finite-sample simulation methods (e.g., Dufour (2006)) or using the maximal inequalities for empirical processes (e.g., Vapnik (1998)). In this paper, we focus on obtaining asymptotic estimates of quantiles of $\mathcal{C}_n$, using either a generic subsampling method (developed in Section 3.5), or the asymptotic approximation to the distribution for $\mathcal{C}_n$ in the moment condition problems (developed in Section 4).

The following basic condition is required to hold.

CONDITION C.4—Convergence of $\mathcal{C}_n$: *$P\{\mathcal{C}_n \leq c\} \to P\{\mathcal{C} \leq c\}$ for each $c \in [0, \infty)$, where the distribution function of $\mathcal{C}$ is nondegenerate and continuous on $[0, \infty)$.*

Section 4 verifies Condition C.4 for moment condition models. For instance, in Example 1 we have that $\mathcal{C}_n$ converges in distribution to $\mathcal{C} =$

$\max((W_1)_+^2, (W_2)_-^2)$ in $\mathbb{R}_+$, where $(W_1, W_2) \sim N(0, \Omega)$. Quantiles of this distribution can be easily estimated using methods discussed in Section 4.

LEMMA 3.1—Basic Large Sample Inference: *Suppose that Condition* C.4 *holds. Then for any* $\widehat{c} \to_p c(\alpha) := \inf\{c \geq 0 : P\{\mathcal{C} \leq c\} \geq \alpha\}$ *for* $\alpha \in (0, 1)$, *such that* $\widehat{c} \geq 0$ *with probability* 1, *we have that as* $n \to \infty$, $P\{\Theta_I \subseteq C_n(\widehat{c})\} = P\{\mathcal{C}_n \leq \widehat{c}\} = P\{\mathcal{C} \leq c(\alpha)\} + o(1) = \alpha + o(1)$ *if* $c(\alpha) > 0$ *and* $P\{\Theta_I \subseteq C_n(\widehat{c})\} = P\{\mathcal{C}_n \leq \widehat{c}\} \geq P\{\mathcal{C} = 0\} + o(1) \geq \alpha + o(1)$ *if* $c(\alpha) = 0$.

It is worth repeating our earlier comment on the consistency and inconsistency of the confidence regions. Under the degeneracy Condition C.3, the confidence region is a consistent estimator of the identified set Θ_I. When the degeneracy condition does not hold, the confidence region may be an inconsistent estimator of Θ_I. Inconsistency means that the confidence region may be very far away from the identified set Θ_I in the Hausdorff distance. To ensure consistency, we need to expand the confidence region. Taking $C_n(\widehat{c} + \kappa_n)$ with $\kappa_n \propto \log n$ gives the $(\log n/a_n)^{1/\gamma}$ consistent estimator of Θ_I. This expanded region is a conservative confidence region that has an asymptotic coverage probability equal to 1.

3.4. *Generic Estimation of the Critical Value by Subsampling*

This section develops a generic subsampling method for consistent estimation of the critical value. The method estimates the quantiles of $\mathcal{C}_n$ using many data subsamples of size b. The following condition facilitates the construction.

CONDITION C.5—Approximability of $\mathcal{C}_n$: *Let* Θ_n *be any sequence of subsets of* Θ *such that* $d_H(\Theta_n, \Theta_I) = o_p(1/a_n^{1/\gamma})$ *and define* $\mathcal{C}_n' = \sup_{\theta \in \Theta_n} a_n Q_n(\theta)$. *Then for any* $c \geq 0$, *we have that* $P[\mathcal{C}_n' \leq c] = P[\mathcal{C} \leq c] + o(1)$.

Section 4 verifies Condition C.5 for models of Section 2. In the subsampling algorithm defined below, Condition C.5 allows us to replace the identified set Θ_I by an estimator $C_n(\widehat{c})$ that has an efficient or near-efficient rate of convergence.

Consider the following generic subsampling algorithm. At a preliminary stage, for cases when data $\{W_t\}$ are an independent and identically distributed sequence, consider all subsets of size $b \ll n$.[8] Denote the number of subsets by B_n. (In applications, since the number of such subsets is large, it suffices to consider a smaller number, B_n, of randomly chosen subsets of size b

[8]For cases when $\{W_t\}$ is a stationary strongly mixing time series, construct $B_n = n - b + 1$ subsets of size b of the form $\{W_j, \ldots, W_{j+b-1}\}$. We do not explicitly discuss this case, but the proof of Theorem 3.3 goes through by using an upper bound on covariance given in the proof of Theorem 3.2.1 in Politis, Romano, and Wolf (1999).

such that $B_n \to \infty$ as $n \to \infty$.) The algorithm has four steps: (1) Initialize a starting value $\widehat{c}_0 \geq q_n := \inf_{\theta\in\Theta} a_n Q_n(\theta)$, which can be data dependent, such that $\widehat{c}_0 = O_p(1)$. Set $\kappa_n \propto \ln n$. If Condition C.3 is known to hold, we can set $\widehat{c}_0 = q_n$ and $\kappa_n = 0$. Also set $l = 1$. (2) Compute $\widehat{c}_l$ as the α-quantile of the sample $\{\widehat{C}_{j,b} := \sup_{\theta\in C_n(\widehat{c}_{l-1}\vee q_n+\kappa_n)} a_b Q_{j,b}(\theta), j = 1,\dots,B_n\}$, where $Q_{j,b}$ denotes the criterion function evaluated using the jth subsample of size b. (3) This step is optional. Repeat Step 2 for $l = 2,\dots,L$. (4) Let $\widehat{c} = \widehat{c}_L$ be the final value produced by the algorithm. Report $C_n(\widehat{c})$ as a confidence region. Report $C_n(\widehat{c} \vee q_n + \kappa_n)$ as a consistent estimator and a confidence region.

THEOREM 3.3—General Validity of Subsampling: *Suppose* (a) *data* $\{w_1,\dots, w_n\}$ *are independent and identically distributed*, (b) $b \to \infty$, $b/n \to 0$ *at polynomial rates as* $n \to \infty$, *and* (c) $a_n \to \infty$ *at least at a polynomial rate in* n. *Suppose that Conditions* C.1, C.2, C.4, *and* C.5 *hold. Let* $\alpha \in (0,1)$ *denote the desired coverage level. Then* (1) $\widehat{c} \to_p c(\alpha) := \inf\{c : P\{\mathcal{C} \leq c\} \geq \alpha\}$, (2) $\lim_n P\{\Theta_I \subseteq C_n(\widehat{c})\} = \alpha$ *if* $c(\alpha) > 0$, *and* $\liminf_n P\{\Theta_I \subseteq C_n(\widehat{c})\} \geq \alpha$ *if* $c(\alpha) = 0$.

Therefore, any finite iteration of the algorithm produces consistent estimates of $c(\alpha)$. The iterations are asymptotically equivalent and thus, for the purposes of asymptotics, form a single step procedure. The resulting regions $C_n(\widehat{c})$ cover Θ_I with P-probability α in large samples. Further, in order to get confidence regions that also consistently estimate Θ_I, we should expand them, namely take $C_n(\widehat{c} \vee q_n + \kappa_n)$ for κ_n defined above.

REMARK 3.1—Maintaining the Level: It follows from the proof of Theorem 3.3 that Condition C.4 alone suffices for $C_n(\widehat{c})$ constructed with $\kappa_n \propto \log n$ to cover Θ_I with P-probability at least α.

REMARK 3.2—Iterations: It is easy to see that the sequence of critical values generated by the algorithm, c_l, is weakly decreasing in $l \geq 1$. Therefore, the algorithm can be terminated when the change between the subsequent critical values is small.

REMARK 3.3—Choice of κ_n: Note that under Condition C.3, we can set $\kappa_n = 0$. When Condition C.3 does not hold, we still can set $\kappa_n = 0$ in the third step without affecting the coverage properties of the region (see Romano and Shaikh (2006)).

REMARK 3.4—Constants in Moment Problems: We discuss implementation in detail in a companion work. A recommendation for moment problems is to use a small number of iterations $L \in \{1,2\}$, set c_0 equal to the α-quantile of a χ^2_J variable, and set $c_0 + \kappa_n$ equal to the $\alpha + \beta_n$ quantile of a χ^2_J variable, where $\alpha + \beta_n \to 1$. This recommendation is based on simulation experiments using a linear moment-inequality model.

3.5. *Asymptotics of $\mathcal{C}_n$ and Related Inferential Statistics*

This section develops methods for obtaining the limits of $\mathcal{C}_n$ and related inferential statistics that determine the probabilities of false coverage. With the exception of simple examples, the difficulties include the lack of equicontinuous behavior of the underlying empirical process and, in some cases, the parameter on the boundary problem. This section outlines a framework for obtaining these limits by relying on the concepts of stochastic equi-semicontinuity and generalizations of the Chernoff (1954) regularity.

Define $\epsilon_n = \delta/a_n^{1/\gamma}$ for $\delta \geq 0$. Consider the statistic $\mathcal{C}_n(\delta) := \sup_{\Theta_I^{\epsilon_n}} a_n Q_n$, where $\Theta_I^{\epsilon_n}$ is ϵ_n-expansion of Θ_I. Notice that $\mathcal{C}_n$ is a special case of this statistic, with $\mathcal{C}_n = \mathcal{C}_n(0)$. Suppose that for each $\delta \geq 0$,

$$\mathcal{C}_n(\delta) \to_d \mathcal{C}(\delta). \tag{3.3}$$

Relation (3.3) implies that the probability of the confidence region for Θ_I covering the false local region $\Theta_I^{\epsilon_n}$ satisfies

$$P\{\Theta_I^{\epsilon_n} \subseteq C_n(c(\alpha))\} = P\{\mathcal{C}_n(\delta) \leq c(\alpha)\} \to P\{\mathcal{C}(\delta) \leq c(\alpha)\}, \tag{3.4}$$

provided $c(\alpha)$ is the continuity point of the distribution of $\mathcal{C}(\delta)$. Then the asymptotic probability of false coverage satisfies $P\{\mathcal{C}(\delta) \leq c(\alpha)\} \leq P\{\mathcal{C} \leq c(\alpha)\}$, with strict inequality holding whenever the distribution function of $\mathcal{C}(\delta)$ differs from the distribution function of $\mathcal{C}$ at $c(\alpha)$. From a testing perspective, we can view $\Theta_I^{\epsilon_n} = \Theta_I + B_\delta/a_n^{1/\gamma}$ as a local alternative to Θ_I, so that statements about false coverage are statements about local power.

The asymptotic distribution of $\mathcal{C}_n(\delta)$ is determined by the empirical process

$$\ell_n(\theta, \lambda) := a_n Q_n(\theta + \lambda/a_n^{1/\gamma}), \quad (\theta, \lambda) \in V_n^\delta, \tag{3.5}$$

where

$$V_n^\delta := \{(\theta, \lambda) : \theta \in \Theta_I, \lambda \in V_n^\delta(\theta)\}, \quad V_n^\delta(\theta) := a_n^{1/\gamma}(\Theta - \theta) \cap B_\delta, \tag{3.6}$$

where B_δ denotes the closed ball in $\mathbb{R}^d$ of radius δ centered at the origin, and $a_n^{1/\gamma}(\Theta - \theta)$ denotes the parameter space translated by θ and multiplied by the scaling rate $a_n^{1/\gamma}$. The parameter λ represents the local deviation from θ and ranges over the local parameter space $V_n^\delta(\theta)$, and the parameter θ ranges over the identified set Θ_I. Together θ and λ range over V_n^δ, the graph of the correspondence $\theta \rightrightarrows V_n^\delta(\theta)$ defined over the domain Θ_I.

The inferential statistics in (3.3) are suprema of the empirical process (3.5) over the graph V_n^δ:

$$\mathcal{C}_n(\delta) = \sup_{(\theta,\lambda)\in V_n^\delta} \ell_n(\theta, \lambda).$$

The limit properties of $\mathcal{C}_n(\delta)$ will therefore depend on the limit properties of the graph V_n^δ. Let V_∞^δ denote the graph of some other correspondence $\theta \rightrightarrows V_\infty^\delta(\theta)$, with a domain Θ_I. We require V_n^δ to "converge" to V_∞^δ in a statistical sense. Moreover, we make a technical assumption that the criterion function is defined over a neighborhood of the parameter space.

CONDITION S.1—Generalized Chernoff Regularity: (A) *Q_n is defined on a neighborhood Θ' of Θ in $\mathbb{R}^d$, and is jointly measurable in $\theta \in \Theta'$ and data $w_1, \dots, w_n$ defined on a complete probability space $(\Omega, \mathcal{F}, P)$.* (B) *For any $\varepsilon > 0$ and $\delta \geq 0$ there exists n_ε such that for all $n \geq n_\varepsilon$, $P\{|\sup_{V_n^\delta} \ell_n - \sup_{V_\infty^\delta} \ell_n| \geq \varepsilon\} \leq \varepsilon$.*

The limit graph V_∞^δ introduced by this condition need not be unique. However, we can treat all the limit graphs that satisfy this condition as an equivalence class. In the case that $\delta = 0$, arising in the analysis of $\mathcal{C}_n = \mathcal{C}_n(0)$, we can take $V_n^0 = V_\infty^0 = \Theta_I \times \{0\}$ in Condition S.1(B). In many cases, Condition S.1(B) holds with

$$V_\infty^\delta = \Theta_I \times B_\delta. \tag{3.7}$$

We call this case the *parameter in the interior* case. The simplest sufficient condition for (3.7) is as follows: There exists $\eta > 0$ such that $B_\eta(\theta) \subset \Theta$ for each $\theta \in \Theta_I$, where $B_\eta(\theta)$ is a closed ball in $\mathbb{R}^d$ of radius η centered at θ. Then we have that $V_n^\delta = \Theta_I \times B_\delta$ for all sufficiently large n. This is reasonable in applications when Θ is a rectangle or a convex body in $\mathbb{R}^d$ and Θ_I is in the interior of Θ. On the other hand, we call any case that does not admit relation (3.7) *the parameter on the boundary* case. In this case, the limit graph V_∞^δ will have a form that depends on the structure of Θ. Our definitions extend definitions given by Chernoff (1954) and Andrews (1999) for point-identified cases.

An important example of the parameter on the boundary is when Θ is its own boundary in $\mathbb{R}^d$, defined by linear or nonlinear equalities. Lemmas 4.1 and 4.2 in Section 4 derive V_∞^δ for the moment condition models of Section 2 for such cases.

We next consider the following condition.

CONDITION S.2—Weak Sup Convergence: *For any finite set $\Delta \subset [0, \infty)$ with cardinality $|\Delta|$, $(\sup_{V_\infty^\delta} \ell_n, \delta \in \Delta) \to_d (\sup_{V_\infty^\delta} \ell_\infty, \delta \in \Delta)$ in $\mathbb{R}^{|\Delta|}$, where $(\theta, \lambda) \mapsto \ell_\infty(\theta, \lambda)$ is a nonnegative stochastic process.*

We call the process ℓ_∞ the sup limit of ℓ_n. This condition adapts the concept of weak epiconvergence in Knight (1999) and Molchanov (2005) to our framework. The sup convergence is more general than the uniform convergence to the finite-dimensional limit. In particular, the latter fails in the moment-inequality model, while the former does not. The following condition is helpful in verifying sup convergence.

CONDITION S.3—Weak Finite-Dimensional Convergence and Approximability: (A) *Finite-Dimensional Convergence*: *For any* $\delta > 0$ *and any finite subset* M *of* V_∞^δ, $(\ell_n(\theta, \lambda), (\theta, \lambda) \in M) \to_d (\ell_\infty(\theta, \lambda), (\theta, \lambda) \in M)$ *in* $\mathbb{R}^{|M|}$, *where* $(\theta, \lambda) \mapsto \ell_\infty(\theta, \lambda)$ *is a nonnegative stochastic process*. (B) *Finite-Dimensional Approximability*: *For any* $\varepsilon > 0$ *and* $\delta \geq 0$, *there is a finite subset* $M(\varepsilon)$ *of* V_∞^δ *such that for all* $n \in [n_\varepsilon, \infty]$, $P\{\sup_{V_\infty^\delta} \ell_n - \max_{M(\varepsilon)} \ell_n \geq \varepsilon\} \leq \varepsilon$.

These conditions establish when the finite-dimensional limit and the sup limit coincide. Otherwise, the two types of limits may not agree. The finite-dimensional approximability is motivated by and extends Knight's (1999) concept of stochastic equi-semicontinuity to set-identified models.

LEMMA 3.2: (1) *Conditions* S.1 *and* S.2 *imply* (3.3) *with* $\mathcal{C}(\delta) = \sup_{V_\infty^\delta} \ell_\infty$; *in particular*, $\mathcal{C} = \mathcal{C}(0) = \sup_{V_\infty^0} \ell_\infty(\theta, \lambda)$. (2) *Condition* S.3 *implies Condition* S.2.

EXAMPLE 1—Continued: In Example 1, $Q_n(\theta) = (E_n[Y_1] - \theta)_+^2 + (E_n[Y_2] - \theta)_-^2$. Then $\ell_n(\theta, \lambda) = n(E_n[Y_1] - \theta - \lambda/\sqrt{n})_+^2 + n(E_n[Y_2] - \theta - \lambda/\sqrt{n})_-^2$. Suppose that $(\sqrt{n}(E_n[Y_1] - E_P[Y_1]), \sqrt{n}(E_n[Y_2] - E_P[Y_2]))' \to_d (W_1, W_2)' = N(0, \Omega)$. Then the finite-dimensional limit of $\ell_n(\theta, \lambda)$ is $\ell_\infty(\theta, \lambda) = (W_1 - \lambda)_+^2 1(\theta = E_P[Y_1]) + (W_2 - \lambda)_-^2 1(\theta = E_P[Y_2])$. The limit is not continuous in θ at $\theta = E_P[Y_1]$ and at $\theta = E_P[Y_2]$. Suppose that $\Theta_I = [E_P[Y_1], E_P[Y_2]]$ is in the interior of Θ. Then Condition S.1 holds, as we can take $V_n^\delta = V_\infty^\delta = \Theta_I \times B_\delta$ for all sufficiently large n. It is easy to verify the finite-dimensional approximability Condition S.3(B). Therefore, Condition S.3 holds. Consequently, the sup limit of $\ell_n(\theta, \lambda)$ equals the finite-dimensional limit $\ell_\infty(\theta, \lambda)$. It follows by Lemma 3.2 that $\mathcal{C}_n(\delta) \to_d \mathcal{C}(\delta) = \sup_{(\theta,\lambda)\in\Theta_I\times B_\delta} \ell_\infty(\theta, \lambda)$; in particular, $\mathcal{C}_n \to_d \mathcal{C} = \sup_{\theta\in\Theta_I} \ell_\infty(\theta, 0) = \max((W_1)_+^2, (W_2)_-^2)$.

4. ANALYSIS OF MOMENT CONDITION MODELS

This section verifies the main conditions of Section 3 in moment condition models. Section 4.1 begins the analysis with moment equalities. Section 4.2 provides the analysis for moment inequalities.

4.1. *Moment Equalities*

Recall the moment-equality setup in Section 2, where the identification region takes the form $\Theta_I = \{\theta \in \Theta : E_P[m_i(\theta)] = 0\}$. We first assume that the following partial identification condition holds: There exist positive constants C and δ such that for all $\theta \in \Theta$,

$$\|E_P[m_i(\theta)]\| \geq C \cdot (d(\theta, \Theta_I) \wedge \delta). \tag{4.1}$$

This condition states that once θ is bounded away from Θ_I, the moment equations are bounded below by a number proportional to the distance to the identified set.

In the point-identified case, the full rank and continuity of the Jacobian $\nabla_\theta E_P[m_i(\theta)]$ near Θ_I ordinarily imply (4.1). In the set-identified case, the Jacobian may be degenerate, which requires a more involved condition (4.1). For example, in the linear IV model of Example 4 we have that $E_P[m_i(\theta)] = E_P[ZX'](\theta - \theta^*)$, where θ^* is the closest point to θ in Θ_I. Provided that $\|\theta^* - \theta\| > 0$, the vector $(\theta - \theta^*)$ is orthogonal to the hyperplane $\{v : E_P[ZX']v = 0\}$. Hence if the rank of $E_P[ZX']$ is nonzero, then $\|E_P[ZX'](\theta - \theta^*)\| \geq C \cdot \|\theta - \theta^*\|$, where C is the square root of the minimal nonzero eigenvalue of $E_P[XZ']E_P[ZX']$.

We also require moment functions to have a Donsker property, which is broadly applicable and easily verifiable (van der Vaart (1998)): (1) In the metric space $L^\infty(\Theta')$,

$$\mathbb{G}_n[m_i(\theta)] := \sqrt{n}\big(E_n[m_i(\theta)] - E_P[m_i(\theta)]\big) \quad \Rightarrow \quad \Delta(\theta), \tag{4.2}$$

where $\Delta(\theta)$ is a mean zero Gaussian process with almost sure continuous paths, $\mathrm{Var}_P[\Delta(\theta)] > 0$ for each $\theta \in \Theta'$, and Θ' is a neighborhood of Θ in $\mathbb{R}^d$. (2) The probability space $(\Omega, \mathcal{F}, P)$ is rich enough to support the representation $\mathbb{G}_n[m_i(\theta)] =_d \Delta(\theta) + o_p(1)$ in $L^\infty(\Theta')$. The symbol $=_d$ means equality in law. The representation exists without loss of generality due to the Skorohod–Dudley–Wichura construction (Dudley (1985), van der Vaart and Wellner (1996)).

We summarize these assumptions as well as others as follows.

CONDITION M.1: *Suppose the following conditions hold for the moment-equality model of Section* 2: (a) *The parameter space Θ is a nonempty compact subset of $\mathbb{R}^d$, and the real-valued criterion function $Q_n(\theta)$ is defined on a neighborhood Θ' of Θ in $\mathbb{R}^d$, and is jointly measurable in $\theta \in \Theta'$ and the data $w_1, \dots, w_n$ defined on a complete probability space $(\Omega, \mathcal{F}, P)$.* (b) *The parameter space Θ is such that the graph of the local parameter space V_n^δ converges to some set V_∞^δ in the Hausdorff metric for each $\delta \geq 0$, where V_∞^δ is nondecreasing in $\delta \geq 0$.* (c) *The collection $\{m_i(\theta), \theta \in \Theta'\}$ satisfies the P-Donsker condition stated above.* (d) *The moment function $E_P[m_i(\theta)]$ satisfies the partial identification condition* (4.1) *and has a continuous Jacobian $G(\theta) = \nabla_\theta E_P[m_i(\theta)]$ for each $\theta \in \Theta'$.* (e) *The weighting matrix is such that $W_n(\theta) = W(\theta) + o_p(1)$ uniformly in $\theta \in \Theta'$, where $W(\theta)$ is positive definite and continuous for all $\theta \in \Theta'$.*

We use these assumptions to verify the main conditions of Section 3. Condition M.2(b) is a Chernoff regularity condition needed in Section 3.6 and in the theorem below. In parameters in the interior cases, which appear to be the most relevant in practice, this condition holds with $V_\infty^\delta = \Theta_I \times B_\delta$. Lemma 4.1 below provides simple sufficient conditions for Condition M.1(b).

THEOREM 4.1—Moment Equalities: (1) (a) *Suppose Condition* M.1(a) *and* (c)–(e) *holds. Then Conditions* C.1, C.2, C.4, *and* C.5 *hold with* $\gamma = 2$, $a_n = n$, $b_n = \sqrt{n}$, *and*

$$\mathcal{C} = \sup_{\theta \in \Theta_I} \|\Delta(\theta)' W^{1/2}(\theta)\|^2. \tag{4.3}$$

(b) *Suppose Condition* M.1 *holds. Then Conditions* S.1–S.3 *hold. The sup limit of* $\ell_n(\theta, \lambda) := nQ_n(\theta + \lambda/\sqrt{n})$ *equals* $\ell_\infty(\theta, \lambda) = \|(\Delta(\theta) + G(\theta)\lambda)' W^{1/2}(\theta)\|^2$.
(2) *Suppose that we use* $Q_n'(\theta) = Q_n(\theta) - \inf_{\theta' \in \Theta} Q_n(\theta')$ *for estimation and inference, and that Condition* M.1 *holds. Then* (a) *Conditions* C.1, C.2, C.4, *and* C.5 *hold with* $\gamma = 2$, $a_n = n$, $b_n = \sqrt{n}$, *and where, for* $V_\infty^\infty := \lim_{\delta \uparrow \infty} V_\infty^\delta$,

$$\mathcal{C} = \sup_{\theta \in \Theta_I} \|\Delta(\theta)' W^{1/2}(\theta)\|^2 - \inf_{(\theta, \lambda) \in V_\infty^\infty} \left\|(\Delta(\theta) + G(\theta)\lambda)' W^{1/2}(\theta)\right\|^2. \tag{4.4}$$

(c) *Conditions* S.1–S.3 *also hold. The sup limit of* $\ell_n'(\theta, \lambda) := nQ_n'(\theta + \lambda/\sqrt{n})$ *equals* $\ell_\infty'(\theta, \lambda) = \ell_\infty(\theta, \lambda) - \inf_{(\theta', \lambda') \in V_\infty^\infty} \ell_\infty(\theta', \lambda')$.

REMARK 4.1: In the parameter in the interior case, $V_\infty^\delta = \Theta_I \times B_\delta$ and $V_\infty^\infty = \Theta_I \times \mathbb{R}^d$.

Theorem 4.1 verifies conditions of Section 3 for moment equalities. Therefore, the most basic corollary of Theorem 4.1 is the following: Let $\widehat{c}$ be any consistent estimate of $c(\alpha)$, the α-quantile of the limit statistic $\mathcal{C}$. Then the region $C_n(\widehat{c})$ covers the identified set Θ_I with asymptotic probability α. Since Condition C.3 may not hold, this region may be an inconsistent estimator of Θ_I. To ensure consistency, we can take the expanded region $C_n(\widehat{c} + \kappa_n)$, which is a $\sqrt{\ln n/n}$ consistent estimator of Θ_I. This expanded region is also a confidence region with an asymptotic coverage of 1. Theorem 4.1 also obtains the sup limit ℓ_∞ of the empirical process ℓ_n, which is needed to describe the power of the test that underlies the confidence region (see Section 3.6).

Finally, we can estimate the quantiles of $\mathcal{C}$ in (4.3) either by the generic subsampling method of Section 3.5 or by the simulation method defined below.

REMARK 4.2—Quantiles of (4.3) by Simulation and Bootstrap: If the data are independent and identically distributed, we can estimate the quantiles of $\mathcal{C}$ by simulating the distribution of the variable $\mathcal{C}_n^* := \sup_{\theta \in \widehat{\Theta}_I} \mathcal{C}_n^*(\theta)$, where $\mathcal{C}_n^*(\theta) := \|\Delta_n^*(\theta)' W_n^{1/2}(\theta)\|^2$, $\Delta_n^*(\theta) := n^{-1/2} \sum_{i=1}^n [m_i(\theta) z_i^*]$, $\widehat{\Theta}_I$ is a consistent estimator of Θ_I, and $(z_i^*, i \leq n)$ is an n-vector of independent and identically distributed $N(0, 1)$ variables. Note that $\Delta_n^*(\theta)$ is a zero-mean Gaussian process in $L^\infty(\Theta)$ with covariance function $E_n[m_i(\theta) m_i(\bar{\theta})']$. Then $E_n[m_i(\theta) m_i(\bar{\theta})'] = E_P[m_i(\theta) m_i(\bar{\theta})'] + o_p(1)$ uniformly in $(\theta, \bar{\theta}) \in \Theta \times \Theta$. Thus, in the weak convergence metric, the distance between the laws of $\Delta_n^*(\theta)$ and $\Delta(\theta)$ converges in probability to zero. Finally, the distance between the laws of $\mathcal{C}_n^*$

and $\mathcal{C}$ converges in probability to zero. This follows because $\Delta_n^*(\theta)$ is stochastically equicontinuous, so that preliminary estimation of Θ_I and $W(\theta)$ has no impact on the asymptotic distribution of the statistic $\mathcal{C}_n^*$. Similarly, we can simulate $\Delta_n^*(\theta)$ using the bootstrap. For this purpose, we take $\Delta_n^*(\theta) := n^{-1/2}\sum_{i=1}^n [m_i(\theta)z_i^*]$, where $z_i^* = k_i^* - 1$ for each i and $(k_i^*, i \leq n)$ is an n-vector following the multinomial distribution with success probabilities $1/n$, defined over n trials.

REMARK 4.3—Quantiles of (4.4) by Simulation and Bootstrap: We can estimate the quantiles of $\mathcal{C}$ in (4.4) by simulating the distribution of the variable $\mathcal{C}_n^* := \sup_{\theta\in\widehat{\Theta}_I} \mathcal{C}_n^*(\theta)$, where $\mathcal{C}_n^*(\theta) := \|\Delta_n^*(\theta)' W_n^{1/2}(\theta)\|^2 - \inf_{\theta+\lambda/\sqrt{n}\in\Theta_n} \|(\Delta^*(\theta) + \widehat{G}(\theta) \times\lambda)' W_n^{1/2}(\theta)\|^2$, $\Delta_n^*(\theta)$ and $\widehat{\Theta}_I$ are obtained as in Remark 4.2, $\widehat{G}(\theta)$ is a uniformly consistent estimate of $\nabla_\theta E_P[m_i(\theta)]$, and $\Theta_n = C_n(\kappa_n)$ with $\kappa_n \propto \log n$.

The lemma below calculates the limit graphs of local parameter spaces for several examples.

LEMMA 4.1: *Sufficient conditions for the parameter space Θ to satisfy Condition* M.1(b) *include either of the following cases. In each case, the limit graph is of the form $V_\infty^\delta = \{(\theta, \lambda) \in \Theta_I \times B_\delta : \lambda \in V_\infty^\delta(\theta)\}$ up to a closure, where $V_\infty(\theta)$ depends on the case.* (1) *Convex Case: Θ is a compact and convex set. Then $V_\infty^\delta(\theta) := \{\lambda \in B_\delta : \lambda \in \sqrt{n'}(\Theta - \theta)$ for some $n' \geq 1\}$.* (2) *Nonlinear Restrictions: For each $\theta \in \Theta_I$, there exists $\eta > 0$ such that $\Theta \cap B_\eta(\theta) = \{\theta' \in B_\eta(\theta) : g(\theta') = 0\}$ for some function $g : \mathbb{R}^d \to \mathbb{R}^{d_g}$, $d_g \leq c < \infty$, that has a continuous Jacobian $\nabla_\theta g(\cdot)$ of full row rank on $B_\eta(\theta)$. Then $V_\infty^\delta(\theta) := \{\lambda \in B_\delta(\theta) : \nabla_\theta g(\theta)\lambda = 0\}$.* (3) *Interior Case: Any condition that implies convergence of V_n^δ in the Hausdorff distance to the graph $\Theta_I \times B_\delta$, including* (a) *for some $\eta > 0$, $B_\eta(\theta) \subseteq \Theta$ for each $\theta \in \Theta_I$ and* (b) *a closure of a graph V_∞^δ found in the convex case equals $\Theta \times B_\delta$.*

Lemma 4.1 obtains several conditions for the generalized Chernoff regularity. Case (1) covers convex parameter spaces and case (2) covers parameter spaces generated by nonlinear restrictions. In both cases, the values of the limit graph are given by (truncated) tangent cones of Θ at points θ. Case (3) is the parameter in the interior case, the primary case. Lemma 4.1 extends the results of Chernoff (1954), Geyer (1994), and Andrews (1997) to the set-identified case.

Let us illustrate case (3) with the partially identified linear instrumental variable model of Example 4. Suppose that the parameter space Θ is a rectangle or other convex body in $\mathbb{R}^d$, and that the intersection of Θ_I and the interior of Θ is nonempty. The closure of a limit graph calculated in Lemma 4.1 (case (1)) equals $\Theta_I \times B_\delta$, which is therefore a limit graph in this case.

4.2. *Moment Inequalities*

Recall the setup of the moment-inequality model in Section 2. We have the identified set $\Theta_I = \{\theta \in \Theta : \|E_P[m_i(\theta)]\|_+ = 0\}$. We first assume that the following partial identification condition holds: There exist positive constants C and δ such that for all $\theta \in \Theta$,

$$\|E_P[m_i(\theta)]\|_+ \geq C \cdot (d(\theta, \Theta_I) \wedge \delta). \tag{4.5}$$

This condition states that once θ is bounded away from Θ_I, the moment equations are bounded below by a number proportional to the distance from the identified set.

Moreover, in many cases, the following condition applies: There exist positive constants (C, M, δ) such that

$$\max_j E_P[m_{ij}(\theta)] \leq -C \cdot (\epsilon \wedge \delta) \quad \text{for all } \theta \in \Theta_I^{-\epsilon}, \tag{4.6}$$
$$d_H(\Theta_I^{-\epsilon}, \Theta_I) \leq M\epsilon \quad \text{for all } \epsilon \in [0, \delta],$$

where $(m_{ij}(\theta), j = 1, \ldots, J)$ are components of the vector $m_i(\theta)$ and $\Theta_I^{-\epsilon} = \{\theta \in \Theta_I : d(\theta, \Theta \setminus \Theta_I) \geq \epsilon\}$. Without loss of generality, the constants (C, δ) can be taken to be the same as in (4.5). Equation (4.6) states that moment equations are strictly negative for all θ in the contractions of Θ_I and that these contractions $\Theta_I^{-\epsilon}$ can approximate Θ_I. Condition (4.6) need not hold generally, but it appears to hold in many of the empirical examples listed in Section 2. It is one condition that implies the degeneracy Condition C.3. Point identification is another condition that implies Condition C.3.

In order to state the regularity conditions, define $\Theta_{\mathcal{J}} := \{\theta \in \Theta_I : E_P[m_{ij}(\theta)] = 0\ \forall j \in \mathcal{J}, E_P[m_{ij}(\theta)] < 0\ \forall j \in \mathcal{J}^c\}$, where $\mathcal{J}$ is any subset of $\{1, \ldots, J\}$ such that $\Theta_{\mathcal{J}}$ is nonempty and $\mathcal{J}^c$ is the complement of $\mathcal{J}$ relative to $\{1, \ldots, J\}$.

CONDITION M.2: *Suppose the following conditions hold for the moment inequality model of Section* 2: (a) *The parameter space Θ is a nonempty compact subset of $\mathbb{R}^d$, and the criterion function $Q_n(\theta)$ is defined on a neighborhood Θ' of Θ in $\mathbb{R}^d$, and is jointly measurable in $\theta \in \Theta'$ and the data $w_1, \ldots, w_n$ defined on a complete probability space $(\Omega, \mathcal{F}, P)$.* (b) *The parameter space Θ is such that for each $\delta \geq 0$ and each $\mathcal{J}$, every restriction of the graph of the local parameter space of the form $V_{n,\mathcal{J}}^{\delta} := V_n^{\delta}|\theta \in \Theta_{\mathcal{J}}$ converges to a corresponding restriction $V_{\infty,\mathcal{J}}^{\delta} = V_{\infty}^{\delta}|\theta \in \Theta_{\mathcal{J}}$ of the limit graph $V_{\infty}^{\delta} = \bigcup_{\mathcal{J}} V_{\infty,\mathcal{J}}^{\delta}$ in the Hausdorff metric, where $V_{\infty,\mathcal{J}}^{\delta}$ is nondecreasing in $\delta \geq 0$.* (c) *The collection $\{m_i(\theta), \theta \in \Theta'\}$ satisfies the P-Donsker condition stated in Section* 4.1. (d) *The moment function $E_P[m_i(\theta)]$ satisfies the partial identification condition* (4.5) *and has a continuous Jacobian $G(\theta) = \nabla_\theta E_P[m_i(\theta)]$ for each $\theta \in \Theta'$.* (e) *The weighting matrix is such that $W_n(\theta) = W(\theta) + o_p(1)$ uniformly in $\theta \in \Theta'$, where $W(\theta)$ is a diagonal matrix with positive diagonal elements and is continuous for all $\theta \in \Theta'$.* (f) *Condi-*

tion (4.6) *holds*.

We use these assumptions to verify the main conditions of Section 3. Condition M.2(b) is an assumption of Chernoff type needed in Section 3.6 and below. In parameters in the interior cases, which appear to be the most relevant for practice, this condition holds with $V_\infty^\delta = \Theta_I \times B_\delta$. Lemma 4.2 below provides other simple sufficient conditions. Finally, we state Condition M.2(f) as a sufficient condition for the degeneracy Condition C.3. The point identification is another sufficient condition for degeneracy.

THEOREM 4.2—Moment Inequalities: *Suppose Condition* M.2(a) *and* (c)–(e) *holds*. (1)(a) *Then Conditions* C.1, C.2, C.4, *and* C.5 *hold with* $\gamma = 2$, $a_n = n$, $b_n = \sqrt{n}$, *and*

$$\mathcal{C} = \sup_{\theta \in \Theta_I} \left\| (\Delta(\theta) + \xi(\theta))' W^{1/2}(\theta) \right\|_+^2, \tag{4.7}$$

where $\xi(\theta) := (\xi_j(\theta), j \le J)$ *with* $\xi_j(\theta) = -\infty$ *if* $E_P[m_{ij}(\theta)] < 0$ *and* $\xi_j(\theta) = 0$ *if* $E_P[m_{ij}(\theta)] = 0$. *If Condition* M.2(f) *is also true, then Condition* C.3 *also holds*. (b) *Suppose that Condition* M.2(b) *also holds. Then Conditions* S.1–S.3 *hold. The sup-limit of* $\ell_n(\theta, \lambda) := nQ_n(\theta + \lambda/\sqrt{n})$ *equals* $\ell_\infty(\theta, \lambda) = \|(\Delta(\theta) + G(\theta)\lambda + \xi(\theta))' W^{1/2}(\theta)\|_+^2$. (2) *Suppose that we use the criterion* $Q_n'(\theta) = Q_n(\theta) - \inf_{\theta' \in \Theta} Q_n(\theta')$ *for estimation and inference and that Condition* M.2(a)–(e) *holds. Then* (a) *Conditions* C.1, C.2, C.4, *and* C.5 *hold with* $\gamma = 2$, $a_n = n$, $b_n = \sqrt{n}$, *and*

$$\begin{aligned} \mathcal{C} = {} & \sup_{\theta \in \Theta_I} \left\| (\Delta(\theta) + \xi(\theta))' W^{1/2}(\theta) \right\|_+^2 \\ & - \inf_{(\theta, \lambda) \in V_\infty^\infty} \left\| (\Delta(\theta) + G(\theta)\lambda + \xi(\theta))' W^{1/2}(\theta) \right\|_+^2, \end{aligned} \tag{4.8}$$

where the second term equals zero if Condition M.2(f) *holds, and* $V_\infty^\infty := \lim_{\delta \uparrow \infty} V_\infty^\delta$. (b) *Conditions* S.1–S.3 *hold. The sup limit of* $\ell_n'(\theta, \lambda) := nQ_n'(\theta + \lambda/\sqrt{n})$ *equals* $\ell_\infty'(\theta, \lambda) = \ell_\infty(\theta, \lambda) - \inf_{(\theta', \lambda') \in V_\infty^\infty} \ell_\infty(\theta', \lambda')$, *where the second term equals zero if Condition* M.2(f) *holds*.

REMARK 4.4: In the parameter in the interior case, $V_\infty^\delta = \Theta_I \times B_\delta$ and $V_\infty^\infty = \Theta_I \times \mathbb{R}^d$.

Theorem 4.2 verifies conditions of Section 3 for moment inequalities. Therefore, the most basic corollary of Theorem 4.2 is the following: Let $\widehat{c}$ be any consistent estimate of $c(\alpha)$, the α-quantile of variable $\mathcal{C}$. Then the region $C_n(\widehat{c})$ covers the identified set Θ_I with asymptotic probability α. Moreover, this region is a consistent estimator of the identified set in the Hausdorff distance and has convergence rate $1/\sqrt{n}$. Theorem 4.2 also obtains the sup limit ℓ_∞ of the empirical process ℓ_n, which is needed for power analysis in Section 3.6. We can estimate the quantiles of $\mathcal{C}$ in (4.7) either by the generic subsampling method of Section 3.5 or by the simulation method defined below.

REMARK 4.5—Quantiles of (4.7) by Simulation and Bootstrap: If the data are independent and identically distributed, we can estimate the quantiles of $\mathcal{C}$ by simulating the distribution of the variable $\mathcal{C}_n^* := \sup_{\theta\in\widehat{\Theta}_I} \mathcal{C}_n^*(\theta)$, where $\mathcal{C}_n^*(\theta) := \|(\Delta_n^*(\theta)+\widehat{\xi}(\theta))'W_n^{1/2}(\theta)\|_+^2$, $\widehat{\xi}(\theta) := (\widehat{\xi}_j(\theta), j=1,\dots,J)'$ with $\widehat{\xi}_j(\theta) := -\infty$ if $E_n[m_{ij}(\theta)] \le -c_j\sqrt{\log n/n}$, and $\widehat{\xi}_j(\theta) := 0$ if $E_n[m_{ij}(\theta)] > -c_j\sqrt{\log n/n}$, for some positive constants $c_j > 0$. Again $\widehat{\Theta}_I$ is a consistant estimator of Θ_I and we simulate $\Delta_n^*(\theta)$ according to Remark 4.2 and hold other quantities fixed.

REMARK 4.6—Quantiles of (4.8) by Simulation and Bootstrap: If the data are independent and identically distributed, we can estimate the quantiles of $\mathcal{C}$ by simulating the distribution of the variable $\mathcal{C}_n^* := \sup_{\theta\in\widehat{\Theta}_I} \mathcal{C}_n^*(\theta)$, where $\mathcal{C}_n^*(\theta) := \|(\Delta_n^*(\theta)+\widehat{\xi}(\theta))'W_n^{1/2}(\theta)\|_+^2 - \inf_{\theta'=\theta+\lambda/\sqrt{n}\in\Theta_n} \|(\Delta_n^*(\theta)+\widehat{G}(\theta)\lambda+\widehat{\xi}(\theta'))'W_n^{1/2}(\theta)\|_+^2$, the estimate $\widehat{G}(\theta)$ converges uniformly to $\nabla_\theta E_P[m_i(\theta)]$, and $\Theta_n = C_n(\kappa_n)$ with $\kappa_n \propto \log n$. The quantity $\Delta_n^*(\theta)$ is simulated according to Remark 4.2. Other quantities are held fixed.

The form of Θ plays an important role in determining the limit form of the local parameter spaces and of the statistic $\mathcal{C}_n(\delta)$, whose behavior determines the probability of false coverage.

LEMMA 4.2: *Sufficient conditions for the parameter space Θ to satisfy Condition* M.2(b) *include any of the conditions stated in Lemma* 4.1. *In particular, the limit graph is of the form $V_{\infty,\mathcal{J}}^\delta = \{(\theta,\lambda)\in\Theta_{\mathcal{J}}\times B_\delta : \lambda\in V_\infty^\delta(\theta)\}$ up to a closure, with the forms of $V_\infty^\delta(\theta)$ specified as in Lemma* 4.1.

Comments that are similar to those stated after Lemma 4.1 apply.

5. POINTWISE APPROACH

Suppose we are interested in a particular parameter value θ^* inside Θ_I. The inference on a point value θ^* is well motivated where there is a sense in which θ^* is the true parameter value. This is the case when a model can correctly represent a true complex real-world phenomenon. The purpose of this section is to briefly review pointwise confidence regions, to derive their properties as estimators of the identified set, and to obtain the limit distribution of the relevant inferential statistics in moment condition models. The previous analysis is quite helpful in obtaining these results.

Let us impose the following condition, which is a pointwise analog of Condition C.4.

CONDITION C.6: *Suppose there exists a sequence of positive numbers $a_n\to\infty$ such that, for $\mathcal{C}_n(\theta) := a_nQ_n(\theta)$, $P\{\mathcal{C}_n(\theta)\le c\}\to P\{\mathcal{C}(\theta)\le c\}$ for each $c\ge 0$*

and each $\theta \in \Theta_I$, where $\mathcal{C}(\theta)$ is a real random variable that has a continuous distribution function on $[0, \infty)$ *and α-quantile denoted as $c(\alpha, \theta)$. Moreover, for at least one $\theta \in \Theta_I$, $\mathcal{C}(\theta) > 0$ with positive probability.*

A classical approach to constructing confidence intervals for a parameter is the inversion of tests of simple hypotheses (Lehmann (1997, p. 90)): For each $\theta \in \Theta$, test whether $Q(\theta) = 0$ by checking whether $a_n Q_n(\theta)$ is less than an estimate of its α-quantile, calculated under the assumption that the hypothesis is right. Then collect all $\theta \in \Theta$ that pass the test to form a confidence region. One implementation of a confidence region of this type can be written as

$$C_n(\bar{c}(\cdot)) = \{\theta \in \Theta : a_n Q_n(\theta) \le \bar{c}(\theta)\}, \tag{5.1}$$

where $\bar{c}(\theta) := \widehat{c}(\theta) \wedge \widehat{c}$, such that $\widehat{c}(\theta) = c(\alpha, \theta) + o_p(1) \ge 0$ for each $\theta \in \Theta_I$, and $\widehat{c}$ is an estimate of an upper bound on $\sup_{\theta \in \Theta_I} c(\theta)$ such that $\widehat{c} = O_p(1)$. This region may be seen as a generalized contour set, but need not be an ordinary contour set of the type considered in Section 3. If the latter is desired, typically for practical reasons, then it suffices to take $C_n(\widehat{c}) = \{\theta \in \Theta : a_n Q_n(\theta) \le \widehat{c}\}$. Asymptotic pointwise validity of these regions follows from the fact that the probability that θ^* is in either region is no smaller than

$$\begin{aligned} P\big\{a_n Q_n(\theta^*) &\le [c(\alpha, \theta^*) + o_p(1)] \vee 0\big\} \\ &= P\{\mathcal{C}(\theta^*) \le c(\alpha, \theta^*)\} + o(1) \ge \alpha + o(1), \end{aligned} \tag{5.2}$$

where the equation follows under the weak convergence and continuity assumptions stated above. In econometrics, the use of pointwise testing and derived confidence regions (5.1) for partially identified models goes back to Anderson and Rubin (1949) and appears in a variety of cases. For example, Dufour (1997), Staiger and Stock (1997), Hu (2002), Kleibergen (2002), Chernozhukov and Hansen (2004), Imbens and Manski (2004), Chernozhukov, Hansen, and Jansson (2005), Rosen (2006), Galichon and Henry (2006b), Romano and Shaikh (2006), and Andrews and Guggenberger (2006) all employed regions of this or similar type.

The first part of the following theorem reviews the basic inferential properties of these regions, while the second part establishes their estimation properties.

THEOREM 5.1: *Suppose that Condition* C.6 *holds. Let there be an estimator $\widehat{c}(\theta) \ge 0$ such that $\widehat{c}(\theta) = c(\alpha, \theta) + o_p(1)$ for each $\theta \in \Theta_I$, let there be an estimator $\widehat{c} \ge 0$ such that $\widehat{c} \ge \sup_{\theta \in \Theta_I} c(\alpha, \theta) + o_p(1)$ and $\widehat{c} = O_p(1)$, and let $\bar{c}(\theta) := \widehat{c}(\theta) \wedge \widehat{c}$.* (1) *For any $\theta^* \in \Theta_I$, $\liminf_{n\to\infty} P\{\theta^* \in C_n(\widehat{c})\} \ge \liminf_{n\to\infty} P\{\theta^* \in C_n(\bar{c}(\cdot))\} \ge \alpha$.* (2) *Consider the estimators $\widehat{\Theta}_I := C_n(\bar{c}(\cdot) \vee q_n + \kappa_n)$ and $\widetilde{\Theta}_I := C_n(\widehat{c} \vee q_n + \kappa_n)$, where $\kappa_n = 0$ when Condition* C.3 *is known to hold and $\kappa_n \propto \log n$ otherwise. Then $d_H(\widehat{\Theta}_I, \Theta_I) = O_p((1/a_n)^{1/\gamma})$ if Condition* C.3 *is known* *to hold and* $d_H(\widehat{\Theta}_I, \Theta_I) = O_p((\kappa_n/a_n)^{1/\gamma})$ *otherwise. The same results also hold for* $\widetilde{\Theta}_I$.

Notice that these properties closely parallel those of the regions considered in Section 3. The following theorem derives the limit distributions of the inferential statistics $\mathcal{C}_n(\theta) = a_n Q_n(\theta)$ in the moment condition models. We need the quantiles of these statistics in the construction of confidence regions.

THEOREM 5.2—Limits of $\mathcal{C}_n(\theta)$ in Moment Condition Models: (1) *Suppose Condition* M.1 *holds for the moment-equality model. Then Condition* C.6 *holds with*

$$\mathcal{C}(\theta) := \|\Delta(\theta)' W^{1/2}(\theta)\|^2 \tag{5.3}$$

$$:= \|\Delta(\theta)' W^{1/2}(\theta)\|^2 - \inf_{(\theta',\lambda)\in V_\infty^\infty} \left\|(\Delta(\theta') + G(\theta')\lambda)' W^{1/2}(\theta')\right\|^2 \tag{5.4}$$

for the cases when we use $Q_n(\theta)$ *and* $Q_n'(\theta) = Q_n(\theta) - \inf_{\theta'\in\Theta} Q_n(\theta')$, *respectively.* (2) *Suppose Condition* M.2 *holds for the moment-inequality model. Then Condition* C.6 *holds with*

$$\mathcal{C}(\theta) := \left\|(\Delta(\theta) + \xi(\theta))' W^{1/2}(\theta)\right\|_+^2 \tag{5.5}$$

$$\begin{aligned} &:= \left\|(\Delta(\theta) + \xi(\theta))' W^{1/2}(\theta)\right\|_+^2 \\ &\quad - \inf_{(\theta',\lambda)\in V_\infty^\infty} \left\|(\Delta(\theta') + G(\theta')\lambda + \xi(\theta))' W^{1/2}(\theta')\right\|_+^2 \end{aligned} \tag{5.6}$$

for the cases when we use $Q_n(\theta)$ *and* $Q_n'(\theta) = Q_n(\theta) - \inf_{\theta'\in\Theta} Q_n(\theta')$, *respectively. For the results* (5.3) *and* (5.5), *the Donsker condition on moment functions can be replaced by a weaker condition that* $n^{-1/2}(\sum_{t=1}^n (m_t(\theta) - E_P[m_t(\theta)])) \to_d \Delta(\theta) = N(0, E_P[\Delta(\theta)\Delta(\theta)'])$.

REMARK 5.1—Quantiles of $\mathcal{C}(\theta)$ by Simulation and Bootstrap: In the case of independent and identically distributed data, the quantiles of $\mathcal{C}(\theta)$, specified in (5.3), (5.4), (5.5), and (5.6), can be estimated by simulation of the variable $\mathcal{C}_n^*(\theta)$ specified, respectively, in Remarks 4.2, 4.3, 4.5, and 4.6. This procedure uses data to select moment inequalities in the calculation of the critical value. Soares (2006) showed that a similar procedure has a uniform asymptotic validity and is not conservative. The critical values $\widehat{c}(\theta)$ constructed in this way are naturally truncated above by some $\widehat{c} = O_p(1)$, uniformly in $\theta \in \Theta$. Therefore, we can set $\bar{c}(\theta) = \widehat{c}(\theta)$.

REMARK 5.2—Quantiles of $\mathcal{C}(\theta)$ by Subsampling and Bootstrap: The estimate $\widehat{c}(\theta)$ can also be obtained by subsampling, namely by taking the α-quantile of $\{a_b Q_{j,b}(\theta), j = 1, \ldots, B_n\}$. Our construction uses explicit truncation of such values, $\bar{c}(\cdot) = \widehat{c}(\cdot) \wedge \widehat{c}$, where $\widehat{c} = O_p(1)$. The reason for truncation is that if $\theta \notin \Theta_I$, we have that $\widehat{c}(\theta) \to_p \infty$, typically at the rate a_b, while $\widehat{c}(\theta) \wedge \widehat{c}$ remains bounded by $\widehat{c}$. The bounding value $\widehat{c}$ can be constructed by taking $\sup_{\theta \in \widehat{\Theta}_I} \widehat{c}(\theta)$ for $\widehat{\Theta}_I = C_n(\kappa_n)$ with $\kappa_n \propto \log n$, using the regionwise critical value of Section 3, or calculating such value analytically. The idea of truncation builds on the point made in Chernozhukov and Fernandez-Val (2005), who demonstrated that bounding the subsampling critical values leads to substantial finite-sample power improvements.[9] It should be noted that the pointwise validity of canonical subsampling regions $C_n(\widehat{c}(\cdot))$ has already been shown by Politis, Romano, and Wolf (1999, Section 2.6) for an arbitrary problem. However, for the reasons just stated, we prefer the region $C_n(\bar{c}(\cdot))$ constructed above.

6. CONCLUSION AND DISCUSSION

In this paper, we consider criterion functions that are minimized on a nonempty set of parameters. The object of interest, Θ_I, is the set that minimizes the given criterion function. We develop practical methods for performing inference on Θ_I. Under general conditions, we show that properly defined contour sets of a sample analog of the criterion function are Hausdorff-consistent estimators of Θ_I. Moreover, we characterize the rate of set convergence as a function of the topology of the set and of the behavior of the criterion function. We highlight the importance of a particular statistic, $\mathcal{C}_n = \sup_{\Theta_I} a_n Q_n(\theta)$, which allows us to compute a sequence of contour sets with a confidence property. In the limit, this sequence covers the identified set with a specified probability α. In particular, the α-quantile of the above statistic defines the level of the contour set that guarantees this coverage. We provide a generic subsampling procedure that consistently estimates the α-quantile of $\mathcal{C}_n$. This subsampling procedure is general and can be applied in many cases. However, in moment condition models, we can also estimate the α-quantile of $\mathcal{C}_n$ by simulating or bootstrapping the asymptotic distribution of $\mathcal{C}_n$. Although the paper is focused on obtaining methods to estimate and perform inference on identified sets, we show how our procedures can be modified to produce the Anderson–Rubin style regions that cover points within the identified set with a specified probability.

There are many interesting research directions for further work, as highlighted in the recent papers on estimation and inference in partially identified models. Beresteanu and Molinari (2006) developed an approach based on random set theory (Molchanov (1990, 2005)) to perform estimation and inference on identified sets. In particular, they employed the Wald statistic, which measures the Hausdorff distance between the identified set and a set-valued estimator, and developed large sample and bootstrap inference procedures for this statistic. The Wald statistic-based approach complements the

[9] They also attribute the improvement to a gain in the speed of rejection of nonlocal alternatives, that is, Bahadurian efficiency.

criterion-based approach of this paper in an important way. Galichon and Henry (2006b) developed a criterion-based inference in an incomplete econometric model using the one-sided Kolmogorov–Smirnov test. They proposed implementing the test by simulation, bootstrap, or subsampling. Their paper also shows a valuable connection to the Monge–Kantarovich optimal transportation theory (e.g., Villani (2003)). Rosen (2006) provided alternative formulations of criterion functions and derived analytical critical values. He also established a useful link to the literature on constrained statistical inference (e.g., Silvapulle and Sen (2005)). Rytchkov (2006) used finite-sample simulation methods for performing inference in a set-identified Kalman filtering model. Finite-sample methods can be very useful in a variety of other set-identified models. Bugni (2006) revisited our analysis of criterion functions and showed that a properly constructed bootstrap procedure provides valid inferences in a subclass of models that covers interval data. He also derived second order properties of the bootstrap procedure. Santos (2006) proposed methods for performing inference in the nonparametric simultaneous equations model, where point identification seems difficult to attain. Analogous issues also arise in nonadditive simultaneous equation models (Chernozhukov and Hansen (2004), Chernozhukov, Hansen, and Jansson (2005)). Finally, Romano and Shaikh (2006) provided further useful subsampling procedures for criterion-based confidence regions.

Recently, Andrews, Berry, and Jia (2004), Galichon and Henry (2006a), and Pakes, Porter, Ishii, and Ho (2006) investigated the inference problem using projection methods. The idea is to construct a region for a point-identified high-dimensional nuisance parameter and then further project the region to obtain a confidence region for partially identified functionals of this parameter. Galichon and Henry's (2006a) approach proceeds by creating a confidence region for a multivariate quantile function and then performing the projection by a perfect matching algorithm.

As emphasized by Imbens and Manski (2004), robustness, or uniform coverage, is an important issue in set-identified models. Andrews and Guggenberger (2006) developed global uniformity results for subsampling, developed hybrid methods for moment inequality and other problems, and also covered dependent data. Romano and Shaikh (2006) analyzed global uniformity of subsampling procedures in several moment inequality examples. Chernozhukov, Hong, and Tamer (2007) suggested examining robustness to local (contiguous) deviations from a fixed data generating process. A useful feature of the local approach is that the conditions for robustness are independent of the way consistent critical values are generated (be it by bootstrap or simulation), making it simple to check robustness of various confidence regions to small perturbations of the data generating process. Soares (2006) examined uniformity of analytical and simulation-based inference in moment inequality models.

Dept. of Economics and Operations Research Center, MIT, 50 Memorial Drive, Cambridge, MA 02139, U.S.A.; vchern@mit.edu, *Dept. of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305-6072, U.S.A.; han.hong@duke.edu,*

and

Dept. of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, U.S.A.; tamer@northwestern.edu.

Manuscript received April, 2004; final revision received March, 2007.

APPENDIX A

A.1. *Notation*

We use standard notation for empirical processes: $E_n[f_t] := \frac{1}{n}\sum_{t=1}^n f(w_t)$ and $\mathbb{G}_n[f_t] := \frac{1}{\sqrt{n}}\sum_{t=1}^n (f(w_t) - E_P[f(w_t)])$. We use notions of stochastic convergence as in van der Vaart and Wellner (1996). For example, wp $\to 1$ means "with the probability approaching 1," where the inner probability is used for nonmeasurable events. The notation $=_d$ means equality in law: given two elements X and Y mapping Ω to a metric space $\mathbb{D}$, we write $X =_d Y$ if $E_P^*[f(X)] = E_P^*[f(Y)]$ for every bounded $f : \mathbb{D} \mapsto \mathbb{R}$. The symbol E_P^* denotes outer expectation with respect to P. Further, we let $\|x\|_+ = \|\max(x, 0)\|$ and $\|x\|_- = \|\max(-x, 0)\|$, where the operation max is performed on vectors elementwise. B_δ denotes a closed ball of diameter δ centered at the origin. In many instances, we use the notation $\sup_A f$ to mean $\sup_{a\in A} f(a)$ if it causes no ambiguity. The Hausdorff distance between sets A and B is defined as $d_H(A, B) := \max[\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)]$, where $d(b, A) := \inf_{a\in A}\|b - a\|$ and $d_H(A, B) := \infty$ if either A or B is empty. The ϵ-expansion of Θ_I is defined as $\Theta_I^\epsilon := \{\theta \in \Theta : d(\theta, \Theta_I) \le \epsilon\}$ and the ϵ-contraction of Θ_I is defined as $\Theta_I^{-\epsilon} := \{\theta \in \Theta_I : d(\theta, \Theta \setminus \Theta_I) \ge \epsilon\}$, where $\epsilon \ge 0$.

A.2. *Proof of Theorem 3.1*

PROOF OF PART (1):

STEP (a): With probability approaching 1, $\widehat{c} \ge C_n = \sup_{\Theta_I} a_n Q_n$, which implies $\Theta_I \subseteq \widehat{\Theta}_I$, which implies $\sup_{\theta\in\Theta_I} d(\theta, \widehat{\Theta}_I) = 0$.

STEP (b): For any $\epsilon > 0$, $\inf_{\Theta\setminus\Theta_I^\epsilon} Q_n \ge_{(\mathrm{i})} \inf_{\Theta\setminus\Theta_I^\epsilon} Q + o_p(1) \ge_{(\mathrm{ii})} \delta(\epsilon) + o_p(1)$ for some $\delta(\epsilon) > 0$, where (i) follows from the one-sided uniform convergence by Condition C.1(d) and (ii) follows from Q being minimized on Θ_I by Condition C.1(b). Similarly, $\sup_{\widehat{\Theta}_I} Q \le_{(\mathrm{i})} \sup_{\widehat{\Theta}_I} Q_n + o_p(1) \le_{(\mathrm{ii})} \widehat{c}/a_n + o_p(1) =_{(\mathrm{iii})} o_p(1)$, where (i) holds by Condition C.1(d), (ii) holds by construction of $\widehat{\Theta}_I$, and (iii) holds by $\widehat{c}/a_n \to_p 0$ occurring by construction. Hence $\sup_{\widehat{\Theta}_I} Q < \delta(\epsilon) = \inf_{\Theta\setminus\Theta_I^\epsilon} Q$, where $\delta(\epsilon) > 0$ wp $\to 1$. Hence $\widehat{\Theta}_I \cap (\Theta \setminus \Theta_I^\epsilon)$ is empty wp $\to 1$, which implies $\widehat{\Theta}_I \subseteq \Theta_I^\epsilon$. Thus, $\sup_{\theta\in\widehat{\Theta}_I} d(\theta, \Theta_I) \le \epsilon$.

Combining Steps (a) and (b), we conclude that $d_H(\widehat{\Theta}_I, \Theta_I) \le \epsilon$ wp $\to$ 1. Since $\epsilon > 0$ is arbitrary, the result is proven. *Q.E.D.*

PROOF OF PART (2): We have that $C_n \le \widehat{c}$ and $\widehat{c}/a_n \to_p 0$ by construction. For any $\varepsilon > 0$, let the positive constants $(\kappa, \delta, \gamma, n_\varepsilon, \kappa_\varepsilon)$ be as specified in Condition C.2. Let $\bar{c} := (\kappa \cdot \kappa_\varepsilon) \vee \widehat{c}$. There exists $n'_\varepsilon \ge n_\varepsilon$ such that for all $n > n'_\varepsilon$ with probability of at least $1 - \varepsilon$, the following is true: $\epsilon_n := [\bar{c}/(a_n\kappa)]^{1/\gamma} \le \delta$ and $\epsilon_n \ge (\kappa_\varepsilon/a_n)^{1/\gamma}$. Therefore, by Condition C.2, $\inf_{\Theta \setminus \Theta_I^{\epsilon_n}} a_n Q_n \ge \kappa \cdot a_n \cdot (\epsilon_n \wedge \delta)^\gamma \ge \kappa a_n \cdot \epsilon_n^\gamma = \bar{c}$. By construction of $\widehat{\Theta}_I$, $\sup_{\widehat{\Theta}_I} a_n Q_n \le \widehat{c} \le \bar{c}$. Therefore, $\widehat{\Theta}_I \subseteq \Theta_I^{\epsilon_n}$. Combining with Step (a) of the Proof of Part (1), we conclude that $d_H(\widehat{\Theta}_I, \Theta_I) \le \epsilon_n$. Therefore, $d_H(\widehat{\Theta}_I, \Theta_I) = O_p([\bar{c}/a_n]^{1/\gamma})$. *Q.E.D.*

PROOF OF PART (3): This part is immediate.

A.3. *Proof of Lemma 3.1*

The proof is elementary and is therefore omitted.

A.4. *Proof of Theorem 3.2*

PROOF OF PART (1): Fix any $\epsilon > 0$. Let (ϵ_n, Θ_n) be the sequence specified in Condition C.3. Then wp $\to 1$, $\Theta_n \subseteq C_n(\widehat{c}') \subseteq C_n(\widehat{c}) \subseteq \Theta_I^\epsilon$. This follows by using that $q_n \le \widehat{c}'$, choosing $\widehat{c} \ge \widehat{c}'$ as a sequence to satisfy conditions in Part (1) of Theorem 3.1, and applying Part (1) of Theorem 3.1. Since $d_H(\Theta_n, \Theta_I) \le \epsilon$ wp $\to$ 1 by Condition C.3(a) and $d_H(\Theta_I^\epsilon, \Theta_I) \le \epsilon$ by definition of Θ_I^ϵ, it follows that $d_H(C_n(\widehat{c}'), \Theta_I) \le \epsilon$ wp $\to$ 1. *Q.E.D.*

PROOF OF PART (2): Let (ϵ_n, Θ_n) be the sequence specified in Condition C.3. We have that $\Theta_n \subseteq C_n(q_n) \subseteq C_n(\widehat{c}') \subseteq C_n(\widehat{c})$ wp $\to$ 1, where $\widehat{c} = O_p(1)$ is a sequence that satisfies the conditions of Part (2) of Theorem 3.1. By Condition C.3, $d_H(\Theta_n, \Theta_I) \le \epsilon_n = O_p(a_n^{-1/\gamma})$. By Part (2) of Theorem 3.1, $d_H(C_n(\widehat{c}), \Theta_I) = O_p(a_n^{-1/\gamma})$. We conclude that $d_H(C_n(\widehat{c}'), \Theta_I) = O_p(a_n^{-1/\gamma})$. *Q.E.D.*

PROOF OF PART (3): This part is immediate.

A.5. *Proof of Theorem 3.3*

PROOF OF PART (1): It suffices to prove the result for $\widehat{c}_1$ produced with a single iteration. The proof for any subsequent iteration is identical to this proof, since the starting value of the algorithm can be data dependent. Both steps of the proof are special to our problem.

STEP 1: By Theorem 3.1 or 3.2 wp $\to 1$, we have that $d_H(C_n(\widehat{c}_0+\kappa_n),\Theta_I)\leq \epsilon_n \propto (\ln^2 n/a_n)^{1/\gamma}$. Fix a particular subsample j and the corresponding objective function $Q_{j,b}$ for that subsample. Define $\overline{\mathcal{C}}_{j,b} := \sup_{\theta\in\Theta_I^{\epsilon_n}} a_b Q_{j,b}(\theta)$ and $\underline{\mathcal{C}}_{j,b} := \inf_{K\in\mathcal{K}_n}\sup_{\theta\in K} a_b Q_{j,b}(\theta)$, where $\mathcal{K}_n$ are subsets $K\subseteq\Theta$ such that $d_H(K,\Theta_I)\leq \epsilon_n$. For the latter optimization problem there is a subset $\Theta_{j,b}$ such that $\underline{\mathcal{C}}_{j,b} = \sup_{\theta\in\Theta_{j,b}} a_b Q_{j,b}(\theta)$. Therefore, wp $\to 1$, $\underline{\mathcal{C}}_{j,b}\leq \widehat{\mathcal{C}}_{j,b} := \sup_{C_n(\widehat{c})} a_b Q_{j,b} \leq \overline{\mathcal{C}}_{j,b}$ for all $j=1,\ldots,B_n$, where index j denotes that the statistic was computed using the jth subsample; the total number of subsamples is B_n. Define $\widehat{G}(x) := B_n^{-1}\sum_{j=1}^{B_n} 1\{\widehat{\mathcal{C}}_{j,b}\leq x\}$. Hence wp $\to 1$,

$$\underline{G}(x) := B_n^{-1}\sum_{j=1}^{B_n} 1\{\overline{\mathcal{C}}_{j,b,n}\leq x\} \leq \widehat{G}(x) \leq \overline{G}(x) := B_n^{-1}\sum_{j=1}^{B_n} 1\{\underline{\mathcal{C}}_{j,b}\leq x\}. \tag{A.1}$$

By Step 2 below, $\underline{G}(x)\to_p P\{\mathcal{C}\leq x\}$ and $\overline{G}(x)\to_p P\{\mathcal{C}\leq x\}$, for each $x\geq 0$. This proves that $\widehat{G}(x)\to_p P\{\mathcal{C}\leq x\}$ for each $x\geq 0$. Convergence of the distribution function at continuity points implies convergence of the quantile function at continuity points. By Condition C.4, $c(\alpha) := \inf\{x : P\{\mathcal{C}\leq x\}\geq\alpha\}$ is continuous in $\alpha\in(0,1)$. Therefore, $\widehat{c} := \inf\{x : \widehat{G}(x)\geq\alpha\}\to_p c(\alpha)$ for each $\alpha\in(0,1)$.

STEP 2: Write $\underline{G}(x) \overset{(1)}{=} E_P[\underline{G}(x)] + o_p(1) \overset{(2)}{=} P\{\overline{\mathcal{C}}_{j,b}\leq x\} + o_p(1) \overset{(3)}{=} P\{\mathcal{C}\leq x\} + o_p(1)$ at each $x\geq 0$. Equality (1) follows by $\mathrm{Var}_P(B_n^{-1}\sum_{j=1}^{B_n} 1\{\overline{\mathcal{C}}_{j,b}\leq x\}) = o(1)$, which follows from $B_n\to\infty$ and the Hoeffding inequality for bounded U-statistics for independent and identically distributed data. Equality (2) is by nonreplacement sampling. Equality (3) follows by Condition C.5 and by $\epsilon_n\propto(\ln^2 n/a_n)^{1/\gamma} = o(1/a_b^{1/\gamma})$ due to restrictions on the subsample size b and the rate a_n stated in conditions (b) and (c) of this theorem. Likewise, we conclude that $\overline{G}(x)\to_p P\{\mathcal{C}\leq x\}$.

Q.E.D.

PROOF OF PART (2): The result follows from Lemma 3.1.

A.6. *Proof of Lemma 3.2*

PROOF OF PART (1): Conditions S.1 and S.2 immediately imply (3.3).

PROOF OF PART (2): This part shows that Condition S.3 implies Condition S.2. Note that for any $\delta\geq 0$ and $\varepsilon>0$ there exists a finite set $M(\varepsilon)\subset V_\infty^\delta$ such that $\limsup_{n\to\infty} P\{\sup_{V_\infty^\delta}\ell_n\leq r\} \leq_{(\mathrm{i})} \limsup_{n\to\infty} P\{\max_{M(\varepsilon)}\ell_n\leq r\} \leq_{(\mathrm{ii})}$

$P\{\max_{M(\varepsilon)} \ell_\infty \leq r\} \leq_{(\text{iii})} P\{\sup_{V_\infty^\delta} \ell_\infty \leq r + \varepsilon\} + \varepsilon$, where inequality (i) follows from $M(\varepsilon) \subset V_\infty^\delta$, (ii) from the finite-dimensional convergence Condition S.3(A), and (iii) from the finite-dimensional approximability Condition S.3(B) applied for $n = \infty$. Since ε is arbitrary, $\limsup_{n\to\infty} P\{\sup_{V_\infty^\delta} \ell_n \leq r\} \leq P\{\sup_{V_\infty^\delta} \ell_\infty \leq r\}$. Further, for any $\delta \geq 0$ and $\varepsilon > 0$ there exists a finite set $M(\varepsilon) \subset V_\infty^\delta$ such that $\liminf_{n\to\infty} P\{\sup_{V_\infty^\delta} \ell_n < r\} \geq_{(\text{i})} \liminf_{n\to\infty} P\{\max_{M(\varepsilon)} \ell_n < r - \varepsilon\} - \varepsilon \geq_{(\text{ii})} P\{\max_{M(\varepsilon)} \ell_\infty < r - \varepsilon\} - \varepsilon \geq_{(\text{iii})} P\{\sup_{V_\infty^\delta} \ell_\infty < r - \varepsilon\} - \varepsilon$, where inequality (i) follows from the finite-dimensional approximability Condition S.3(B), (ii) follows from finite-dimensional convergence Condition S.3(A), and (iii) follows from $M(\varepsilon) \subset V_\infty^\delta$. Since ε is arbitrary, $\liminf_{n\to\infty} P\{\sup_{V_\infty^\delta} \ell_n < r\} \geq P\{\sup_{V_\infty^\delta} \ell_\infty < r\}$. We conclude by the Portmanteau lemma that $\sup_{V_n^\delta} \ell_n \to_d \sup_{V_\infty^\delta} \ell_\infty$. The joint convergence of $(\sup_{V_\infty^\delta} \ell_n, \delta \in \Delta)$ for the finite set Δ in Condition S.2 follows similarly. *Q.E.D.*

A.7. *Proof of Theorem 4.1*

PROOF OF PART (1): The proof is organized in the following steps. Step 1 verifies Conditions C.1 and C.2. Step 2 gives an auxiliary basic approximation for ℓ_n. Using Step 2, Step 3 verifies Condition C.4, Step 4 verifies Condition C.5, and Step 5 verifies Conditions S.1–S.3.

STEP 1—Conditions C.1 and C.2: Uniform Convergence and Quadratic Minorants: Condition C.1 is immediate from Condition M.1(a) and (c)–(e). In particular, uniform convergence and the rates of convergence $a_n = n$ and $b_n = \sqrt{n}$ in Condition C.1 follow from $\{m_i(\theta), \theta \in \Theta\}$ being P-Donsker and having $E_P[m_i(\theta)] = 0$ on Θ_I. To verify Condition C.2, observe that wp $\to 1$, uniformly in $\theta \in \Theta$,

$$
\begin{aligned}
nQ_n(\theta) &= \left\| \left(\mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)]\right)' W_n^{1/2}(\theta) \right\|^2 \qquad \text{(by definition)} \\
&\geq \zeta \cdot \left\| \mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)] \right\|^2 \\
&\qquad \text{(by } \inf_{\theta\in\Theta} \text{mineig } W_n(\theta) \geq \zeta > 0, \text{wp} \to 1, \text{by Condition M.1(e))} \\
&\geq \zeta \cdot \left| \sqrt{n} \left\| E_P[m_i(\theta)] \right\| - \left\| \mathbb{G}_n[m_i(\theta)] \right\| \right|^2 \\
&\qquad \text{(by inequality } \|x + y\| \geq |\|y\| - \|x\||) \\
&\geq \zeta \cdot \left| C \cdot \sqrt{n}(d(\theta, \Theta_I) \wedge \delta) - O_p(1) \right|^2 \\
&\qquad \text{(by } \sup_{\theta\in\Theta} \|\mathbb{G}_n[m_i(\theta)]\| = O_p(1) \text{ and Condition M.1(d))},
\end{aligned}
$$

where $\sup_{\theta\in\Theta} \|\mathbb{G}_n[m_i(\theta)]\| = O_p(1)$ is by the P-Donsker property. Therefore, for any $\varepsilon > 0$ we can choose $(\kappa_\varepsilon, n_\varepsilon)$ large enough so that for all $n \geq n_\varepsilon$ with probability at least $1-\varepsilon$,

$$nQ_n(\theta) \geq \frac{1}{2}\cdot\zeta\cdot C^2\cdot n\cdot[d(\theta,\Theta_I)\wedge\delta]^2$$
$$\text{uniformly on} \quad \{\theta\in\Theta : d(\theta,\Theta_I)\geq(\kappa_\varepsilon/n)^{1/2}\}.$$

This verifies Condition C.2.

STEP 2—An Auxiliary Expansion: The claim of this step is that

$$\ell_n(\theta,\lambda) =_d \underbrace{\left\|(\Delta(\theta)+G(\theta)\lambda)'W^{1/2}(\theta)\right\|^2}_{\ell_\infty(\theta,\lambda)} + o_p(1) \quad \text{in} \quad L^\infty(\Theta_I\times K).$$

For future use, let us also note that $\ell_\infty(\theta,\lambda)$ is stochastically equicontinuous in $L^\infty(\Theta_I\times K)$, because the map $(\theta,\lambda)\mapsto(\Delta(\theta),G(\theta)\lambda,W(\theta))$ is stochastically equicontinuous in $L^\infty(\Theta_I\times K)$. To show the claim, write $\ell_n(\theta,\lambda)=\|\sqrt{n}E_n[m_i(\theta+\lambda/\sqrt{n})]'W_n^{1/2}(\theta+\lambda/\sqrt{n})\|^2=\|(\mathbb{G}_n[m_i(\theta+\lambda/\sqrt{n})]+\sqrt{n}E_P[m_i(\theta+\lambda/\sqrt{n})])'W_n^{1/2}(\theta+\lambda/\sqrt{n})\|^2$. For any nonempty compact subset K of $\mathbb{R}^d$, we have uniformly in $(\theta,\lambda)\in\Theta\times K$: (1) $\mathbb{G}_n[m_i(\theta+\lambda/\sqrt{n})]=_d\Delta(\theta+\lambda/\sqrt{n})+o_p(1)=_d\Delta(\theta)+o_p(1)$ by P-Donskerness and the stochastic equicontinuity of $\Delta(\theta)$, (2) $W_n(\theta+\lambda/\sqrt{n})=W(\theta)+o_p(1)$ by Condition M.1(e); (3) $\sqrt{n}E_P[m_i(\theta+\lambda/\sqrt{n})]=G(\theta)\lambda+o(1)$ by Condition M.1(d) and by $E_P[m_i(\theta)]=0$ for all $\theta\in\Theta_I$. The claim follows.

STEP 3—Condition C.4: Convergence of $\mathcal{C}_n$: By Step 2, $\mathcal{C}_n=_d\sup_{\theta\in\Theta_I}\|\Delta(\theta)'\times W^{1/2}(\theta)\|^2+o_p(1)\equiv\mathcal{C}+o_p(1)$, where $\mathcal{C}>0$ almost surely and has a continuous distribution function by Theorem 11.1 of Davydov, Lifshits, and Smorodina (1998). This verifies Condition C.4.

STEP 4—Condition C.5: Approximability of $\mathcal{C}_n$: By expansions in Step 2, for any sequence of sets Θ_n such that $d_H(\Theta_n,\Theta_I)=o_p(1/\sqrt{n})$, we have that $\mathcal{C}_n'=\sup_{\theta\in\Theta_n}nQ_n(\theta)=_d\sup_{\theta\in\Theta_n}\|\Delta(\theta)'W^{1/2}(\theta)\|+o_p(1)=_d\sup_{\theta\in\Theta_I}\|\Delta(\theta)'\times W^{1/2}(\theta)\|+o_p(1)$, with the last equality holding by the stochastic equicontinuity of $\theta\mapsto\Delta(\theta)'W^{1/2}(\theta)$. This verifies Condition C.5.

STEP 5—Conditions S.1–S.3: Limits of Related Statistics: This step shows that if Condition M.1(b) holds in addition to Condition M.1(a) and (c)–(e), then Conditions S.1 and S.3 hold. Condition S.3 implies Condition S.2 by Lemma 3.2. Condition M.1(b) states that $d_H(V_n^\delta,V_\infty^\delta)=o(1)$. Then $|\sup_{V_n^\delta}\ell_n-\sup_{V_\infty^\delta}\ell_n|=|\sup_{V_n^\delta}\ell_\infty-\sup_{V_\infty^\delta}\ell_\infty+o_p(1)|=o_p(1)$ by Step 2 and stochastic equicontinuity of $\ell_\infty(\theta,\lambda)$. This and Condition M.1(a) imply Condition S.1, the generalized Chernoff regularity. Step 2 above verified Condition S.3(A), the finite-dimensional convergence. Step 2 and stochastic equicontinuity of $\ell_\infty(\theta,\lambda)$ imply Condition S.3(B), the finite-dimensional approximability. *Q.E.D.*

PROOF OF PART (2): The proof is similar to the proof of Part (1). In particular, we have that $nQ_n'(\theta) = nQ_n(\theta) - n\inf_{\theta'\in\Theta} Q_n(\theta')$, where asymptotic approximations for the first term are identical to the proof of Part (1). The second term $\inf_{\theta'\in\Theta} nQ_n(\theta')$ can be arbitrarily well approximated by $\inf_{(\theta,\lambda)\in V_n^\delta} nQ_n(\theta+\lambda/\sqrt{n})$ for a sufficiently large δ. Then as in Part (1) it follows that $\inf_{(\theta,\lambda)\in V_n^\delta} nQ_n(\theta+\lambda/\sqrt{n}) =_d \inf_{(\theta,\lambda)\in V_\infty^\delta} \ell_\infty(\theta,\lambda)+o_p(1)$. We can set δ arbitrarily large to conclude that $\inf_{\theta'\in\Theta} nQ_n(\theta') = \inf_{(\theta,\lambda)\in V_\infty^\infty} \ell_\infty(\theta,\lambda)+o_p(1)$. The limit $\inf_{(\theta,\lambda)\in V_\infty^\infty} \ell_\infty(\theta,\lambda)$ exists and is tight due to monotone convergence: as $\delta \uparrow \infty$, $V_\infty^\delta \uparrow V_\infty^\infty$, and $\inf_{(\theta,\lambda)\in V_\infty^\delta} \ell_\infty(\theta,\lambda) \downarrow \inf_{(\theta,\lambda)\in V_\infty^\infty} \ell_\infty(\theta,\lambda) \geq 0$ almost surely. Further details are omitted for brevity. *Q.E.D.*

A.8. *Proof of Lemma 4.1*

PROOF OF PART (1): Define $V_\infty^\delta(\theta) = \{\lambda \in B_\delta : \sqrt{n'}(\Theta-\theta) \text{ for some } n'\}$. Define $V_n^\delta = \{(\theta,\lambda) : \theta\in\Theta_I, \lambda\in B_\delta, \lambda\in\sqrt{n}(\Theta-\theta)\}$. Note that $V_n^\delta \subseteq V_\infty^\delta$ and $V_n^\delta \uparrow V_\infty^\delta$ monotonically in the set-theoretic sense by convexity of Θ. This implies convergence in the Hausdorff distance because V_n^δ and V_∞^δ are subsets of the compact set $\Theta_I \times B_\delta$. *Q.E.D.*

PROOF OF PART (2):

STEP (a): For each $\theta\in\Theta_I$, we have that $d_H(V_n^\delta(\theta), V_\infty^\delta(\theta)) = o(1)$. This follows by standard arguments (e.g., Lemma 2 in Andrews (1997)). Therefore, for any (θ,λ) in $V_\infty^\delta(\theta)$, there is a sequence $(\theta_n,\lambda_n)\in V_n^\delta$ converging to it.

STEP (b): The claim of this step is that every sequence in $(\theta_n,\lambda_n)\in V_n^\delta \subseteq \Theta_I\times B_\delta$ has its limit points in V_∞^δ. To see this, take any convergent subsequence $(\theta_{n(k)},\lambda_{n(k)}) \to (\theta^*,\lambda^*) \in \Theta_I\times B_\delta$. For every n, we have that $\sqrt{n}g(\theta_{n(k)} + \lambda_{n(k)}/\sqrt{n}) = 0$. Further, $0 = \lim_n \sqrt{n}g(\theta_{n(k)} + \lambda_{n(k)}/\sqrt{n}) = \nabla_\theta g(\theta^*)\lambda^*$ by the assumptions on g. Therefore, $(\theta^*,\lambda^*)\in V_\infty^\delta$.

Combining Steps (a) and (b), conclude that $d_H(V_n^\delta, V_\infty^\delta) = o(1)$. *Q.E.D.*

PROOF OF PART (3): This part holds trivially.

A.9. *Proof of Theorem 4.2*

PROOF OF PART (1): The proof is organized as follows: Step 1 verifies Conditions C.1, C.2, and C.3. Step 2 gives an auxiliary basic approximation for ℓ_n. Lemma A.1 gives another approximation. Using Step 2 and Lemma A.1, Step 3 verifies Condition C.4, Step 4 verifies Condition C.5, and Step 6 verifies Conditions S.1–S.3.

STEP 1—Verification of Conditions C.1, C.2, and C.3: Condition C.1 is immediate from Condition M.2(a) and (c)–(e). In particular, uniform convergence and the rates of convergence $a_n = n$ and $b_n = \sqrt{n}$ in Condition C.1 follow from $\{m_i(\theta), \theta \in \Theta\}$ being P-Donsker and $E_P[m_i(\theta)] \leq 0$ on $\theta \in \Theta_I$. To verify Condition C.2, observe that wp $\to 1$, uniformly in $\theta \in \Theta$,

$$\begin{aligned}
nQ_n(\theta) &\equiv \left\| \left(\mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)]\right)' W_n^{1/2}(\theta) \right\|_+^2 \\
&\geq \zeta \cdot \left\| \mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)] \right\|_+^2 \\
&= \zeta \cdot \left\| \sqrt{n}E_P[m_i(\theta)] \right\|_+^2 \\
&\quad \cdot \left(\left\| \mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)] \right\|_+^2 / \left\| \sqrt{n}E_P[m_i(\theta)] \right\|_+^2 \right),
\end{aligned} \tag{A.2}$$

where we have used $\inf_{\theta\in\Theta} \text{mineig } W_n(\theta) \geq \zeta > 0$ wp $\to 1$ by Condition M.2(e). By Condition M.2(d), we have that $\|\sqrt{n}E_P[m_i(\theta)]\|_+^2 \geq C \cdot n \cdot (d(\theta, \Theta_I) \wedge \delta)^2$ on Θ for some $C > 0$ and $\delta > 0$. Therefore, for any $\varepsilon > 0$, we can choose $(\kappa_\varepsilon, n_\epsilon)$ so that for all $n \geq n_\varepsilon$ with probability at least $1 - \varepsilon$,

$$\begin{aligned}
&nQ_n(\theta) \geq \frac{1}{2} \cdot \zeta \cdot C \cdot n \cdot (d(\theta, \Theta_I) \wedge \delta)^2 \\
&\quad \text{uniformly in} \quad \{\theta \in \Theta : d(\theta, \Theta_I) \geq (\kappa_\varepsilon/n)^{1/2}\}.
\end{aligned}$$

This follows by (A.2), by $\|y + x\|_+/\|x\|_+ \to 1$ as $\|x\|_+ \to \infty$ for any $y \in \mathbb{R}^J$, and by $\sup_{\theta\in\Theta} \|\mathbb{G}_n[m_i(\theta)]\| = O_p(1)$, holding by the P-Donsker property. This verifies Condition C.2.

To verify Condition C.3, observe that for some $C > 0$ and $\delta > 0$, wp $\to 1$, uniformly in $\theta \in \Theta_I$,

$$\begin{aligned}
nQ_n(\theta) &\leq \zeta' \cdot \left\| \mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)] \right\|_+^2 \\
&\leq \zeta' \cdot \sum_{j \leq J} \left| \mathbb{G}_n[m_{ij}(\theta)] + \sqrt{n}E_P[m_{ij}(\theta)] \right|_+^2 \\
&\leq \zeta' \cdot J \cdot \left| O_p(1) - \sqrt{n} \cdot C \cdot (d(\theta, \Theta \setminus \Theta_I) \wedge \delta) \right|_+^2
\end{aligned}$$

by Condition M.2(f) and by $\sup_{\theta\in\Theta} \text{maxeig } W_n(\theta) \leq \zeta' < \infty$ wp $\to 1$ by Condition M.2(e). Conclude that $Q_n(\theta) = 0$ on $\theta \in \Theta_I^{-\epsilon_n}$ with $\epsilon_n = O_p(1/\sqrt{n})$. Therefore, Condition C.3 holds under Condition M.2(f).

STEP 2—A Basic Approximation: The claim is that

$$\begin{aligned}
&\ell_n(\theta, \lambda) =_d \left\| \left(\Delta(\theta) + G(\theta)\lambda + \sqrt{n}E_P[m_i(\theta)]\right)' W^{1/2}(\theta) + o_p(1) \right\|_+^2, \\
&\quad \text{in} \quad L^\infty(\Theta_I \times B_\delta).
\end{aligned}$$

The proof is very similar to Step 2 in the proof of Theorem 4.1.

Steps 3, 4, and 5 also make use of the following result.

LEMMA A.1: *Define*

$$f_n(\theta,\lambda,x) := \left\|\left(\Delta(\theta)+G(\theta)\lambda+\sqrt{n}E_P[m_i(\theta)]\right)'W^{1/2}(\theta)+x\right\|_+^2, \tag{A.3}$$

$$g(\theta,\lambda,x) := \left\|(\Delta(\theta)+G(\theta)\lambda+\xi(\theta))'W^{1/2}(\theta)+x\right\|_+^2.$$

The approximation

$$\sup_{V_n^\delta}\ell_n(\theta,\lambda) =_d \sup_{V_n^\delta} f_n(\theta,\lambda,o_p(1)) = \sup_{V_\infty^\delta} g(\theta,\lambda,o_p(1)) \tag{A.4}$$

is true and $\sup_{V_\infty^\delta} g(\theta,\lambda,o_p(1)) = \max_{\mathcal{J}}\sup_{V_{\infty,\mathcal{J}}^\delta}\sum_{j\in\mathcal{J}}|(\Delta_j(\theta)+G_j(\theta)'\lambda)\times W_{jj}^{1/2}(\theta)+o_p(1)|_+^2$. *Here* $\Theta_{\mathcal{J}} := \{\theta\in\Theta_I : E_P[m_{ij}(\theta)]=0\ \forall j\in\mathcal{J}, E_P[m_{ij}(\theta)]<0\ \forall j\in\mathcal{J}^c\}$, $\mathcal{J}$ *denotes any subset of* $\{1,\dots,J\}$ *that gives nonempty* $\Theta_{\mathcal{J}}$, *and* $\xi_j(\theta) := 0$ *if* $E_P[m_{ij}(\theta)]=0$ *and* $\xi_j(\theta) := -\infty$ *if* $E_P[m_{ij}(\theta)]<0$.

When the subset $\mathcal{J}$ is empty, we interpret the summation over indices j in this subset $\mathcal{J}$ to return 0. The proof of this lemma follows below.

STEP 3—Condition C.4: Convergence of $\mathcal{C}_n$: Recall that $\mathcal{C}_n = \sup_{\theta\in\Theta_I} nQ_n(\theta)$. By Lemma A.1,

$$\begin{aligned}\mathcal{C}_n &=_d \max_{\mathcal{J}}\sup_{\theta\in\Theta_{\mathcal{J}}}\sum_{j\in\mathcal{J}}|\Delta_j(\theta)'W_{jj}^{1/2}(\theta)+o_p(1)|_+^2\\ &= \sup_{\theta\in\Theta_I}\left\|(\Delta(\theta)+\xi(\theta))'W^{1/2}(\theta)+o_p(1)\right\|_+^2.\end{aligned}$$

Hence $P[\mathcal{C}_n\le c]\to P[\mathcal{C}\le c]$ for each $c>0$, for $\mathcal{C}$ defined in the statement of the theorem. By Theorem 11.1 of Davydov, Lifshits, and Smorodina (1998), nondegeneracy of the covariance function of $\Delta(\theta)$ implies that $\mathcal{C}$ has a continuous distribution function on $[0,\infty)$ with a possible point mass at $c=0$. To show $P[\mathcal{C}_n=0]\to P[\mathcal{C}=0]$, note that the nondegeneracy implies $Y=\sup_{\mathcal{J}\neq\varnothing,\theta\in\Theta_{\mathcal{J}},j\in\mathcal{J}}[\Delta_j(\theta)W_{jj}^{1/2}(\theta)]$ has a continuous distribution function on $\mathbb{R}$. It follows that $P[\mathcal{C}_n\le 0]$ is bounded above (below) by $P[Y\le\epsilon_n]$ with some $\epsilon_n\downarrow 0$ ($\epsilon_n\uparrow 0$), and $P[Y\le\epsilon_n]\to P[Y\le 0]=P[\mathcal{C}\le 0]$. This verifies Condition C.4.

STEP 4—Condition C.5: Approximability of $\mathcal{C}_n$: Let Θ_n be any sequence of subsets of Θ such that $d_H(\Theta_n,\Theta_I)=o_p(1/\sqrt{n})$. Then $\mathcal{C}_n' =_d \sup_{\theta\in\Theta_n}\|(\Delta(\theta)+\sqrt{n}E_P[m_i(\theta)])'W^{1/2}(\theta)+o_p(1)\|_+^2 = \sup_{\theta\in\Theta_I}\|(\Delta(\theta)+\sqrt{n}E_P[m_i(\theta)])'W^{1/2}(\theta)+$

$o_p(1)\|_+^2$, where we used Step 2, the stochastic equicontinuity of $\theta \mapsto (\Delta(\theta), W^{1/2}(\theta))$, and that $\|\sqrt{n}(E_P[m_i(\theta)] - E_P[m_i(\theta')])\| = o(1)$ uniformly on $\{(\theta, \theta') \in \Theta : \|\theta' - \theta\| \leq o_p(1/\sqrt{n})\}$. Then it follows as in Step 3 that $P[\mathcal{C}_n' \leq c] \to P[\mathcal{C} \leq c]$ for each $c \geq 0$. This verifies Condition C.5.

STEP 5—Verification of Conditions S.1–S.3: Conditions S.1(A) follows from Condition M.2(a). Conditions S.1(B) and S.2 follow from Lemma A.1. Further, in Equation (A.4), for each $\mathcal{J}$, $\sup_{V_{\infty,\mathcal{J}}^\delta} \sum_{j\in\mathcal{J}} |(\Delta_j(\theta) + G_j(\theta)'\lambda)W_{jj}^{1/2}(\theta) + o_p(1)|_+^2$ admits a finite-dimensional approximation by the stochastic equicontinuity of $(\theta, \lambda) \mapsto (\Delta(\theta), G(\theta)\lambda, W(\theta))$ in $L^\infty(\Theta \times B_\delta)$, which verifies Condition S.3(B). By Step 2, the finite-dimensional limit of $\ell_n(\theta, \lambda)$ equals $\ell_\infty(\theta, \lambda) := \|(\Delta(\theta) + G(\theta)\lambda + \xi(\theta))'W^{1/2}(\theta)\|_+^2$, which verifies Condition S.3(A). *Q.E.D.*

PROOF OF PART (2): The proof is similar to the proof of Part (1) and it is therefore omitted.

A.10. *Proof of Lemma A.1*

The first equality in (A.4) is immediate by Step 2 of the proof of Theorem 4.2. The second equality in (A.4) is proven as follows.

STEP 1: The claim of this step is that wp $\to$ 1, for some $\epsilon_n \downarrow 0$, $\sup_{V_n^\delta} g(\theta, \lambda, -\epsilon_n) \leq \sup_{V_n^\delta} \ell_n(\theta, \lambda) \leq \sup_{V_n^\delta} f_n(\theta, \lambda, \epsilon_n)$. With probability approaching 1, for some $\epsilon_n \downarrow 0$, $g(\theta, \lambda, -\epsilon_n) \leq_{\text{(i)}} f_n(\theta, \lambda, -\epsilon_n) \leq_{\text{(ii)}} \ell_n(\theta, \lambda) \leq_{\text{(iii)}} f_n(\theta, \lambda, \epsilon_n)$. Here (i) follows by $\sqrt{n}E_P[m_i(\theta)] \geq \xi(\theta)$ for each $\theta \in \Theta_I$ and by monotonicity: $a \geq b$ implies $\|(\Delta(\theta) + G(\theta)\lambda + a)'W^{1/2}(\theta) + \epsilon_n\|_+^2 \geq \|(\Delta(\theta) + G(\theta)\lambda + b)'W^{1/2}(\theta) + \epsilon_n\|_+^2$, recalling that $W(\theta)$ is diagonal with positive diagonal entries, and (ii) and (iii) follow from Step 2 of the proof of Theorem 4.2. Therefore, the claim is true.

STEP 2: The claim of this step is that for any $\epsilon_n \uparrow 0$ or $\epsilon_n \downarrow 0$, $\sup_{V_n^\delta} g(\theta, \lambda, \epsilon_n) := \sup_{V_\infty^\delta} g(\theta, \lambda, o_p(1))$. To show this, write $\sup_{V_n^\delta} g(\theta, \lambda, \epsilon_n) = \max_{\mathcal{J}} \sup_{V_{n,\mathcal{J}}^\delta} g_{\mathcal{J}}(\theta, \lambda, \epsilon_n)$, where $g_{\mathcal{J}}(\theta, \lambda, \epsilon_n) = \sum_{j\in\mathcal{J}} |(\Delta_j(\theta) + G_j(\theta)'\lambda)\times W_{jj}^{1/2}(\theta) + \epsilon_n|_+^2$. By Condition M.2(b), $d_H(V_{n,\mathcal{J}}^\delta, V_{\infty,\mathcal{J}}^\delta) = o(1)$, which implies by the stochastic equicontinuity of $(\theta, \lambda) \mapsto (\Delta(\theta), G(\theta)\lambda, W(\theta))$ that $\sup_{V_{n,\mathcal{J}}^\delta} g_{\mathcal{J}}(\theta, \lambda, \epsilon_n) = \sup_{V_{\infty,\mathcal{J}}^\delta} g_{\mathcal{J}}(\theta, \lambda, o_p(1))$ for every $\mathcal{J}$. The claim follows.

STEP 3: The claim of this step is that for some $\epsilon_n' \downarrow 0$, $\sup_{V_n^\delta} f_n(\theta, \lambda, \epsilon_n) \leq \sup_{V_\infty^\delta} g(\theta, \lambda, \epsilon_n')$ wp $\to$ 1. To show the claim, observe that for any $\theta_n \in \Theta_I$ converging to $\theta \in \Theta_I$,

$$\text{(A.5)} \qquad \begin{aligned} \limsup_n \sqrt{n}E_P[m_{ij}(\theta_n)] &\leq \xi_j(\theta), \quad \text{if} \quad \xi_j(\theta) = 0, \\ &= \xi_j(\theta), \quad \text{if} \quad \xi_j(\theta) = -\infty. \end{aligned}$$

Let $\Omega_{n,\varepsilon}$ be the intersection of the event $\{\omega \in \Omega : \sup_{\theta \in \Theta_I} \|\Delta(\theta)\| \leq K_\varepsilon\}$ and the event that $\Delta(\theta)$ is continuous on Θ. For any $\varepsilon > 0$, there exists K_ε such that $P(\Omega_{n,\varepsilon}) \geq 1 - \varepsilon$ for all $n \geq n_\varepsilon$. Suppose that the claim of Step 3 does not hold. Then there must exist constants $\epsilon > 0$ and $\varepsilon > 0$ and a subsequence $(\omega_{n(k)}, \theta_{n(k)}, \lambda_{n(k)})$ with $\omega_{n(k)} \in \Omega_{n(k),\varepsilon}$, $(\theta_{n(k)}, \lambda_{n(k)}) \in V^\delta_{n(k)}$, such that

$$\lim_k \Big[f_{n(k)}(\theta_{n(k)}, \lambda_{n(k)}, \epsilon_{n(k)}) - \sup_{V^\delta_\infty} g(\theta, \lambda, \epsilon) \Big](\omega_{n(k)}) > 0. \tag{A.6}$$

Select a further subsequence such that $\theta_{n(k(l))} \to \theta^*$ and $\lambda_{n(k(l))} \to \lambda^*$, where (θ^*, λ^*) is in the closure of V^δ_∞ by $d_H(V^\delta_n, V^\delta_\infty) \leq \max_{\mathcal{J}} d_H(V^\delta_{n,\mathcal{J}}, V^\delta_{\infty,\mathcal{J}}) = o(1)$ and by $V^\delta_\infty \subseteq \Theta_I \times B_\delta$. By continuity of $\Delta(\theta)$, conclude that $\sup_{V^\delta_\infty} g(\theta, \lambda, \epsilon) = \sup_{\overline{V^\delta_\infty}} g(\theta, \lambda, \epsilon) \geq g(\theta^*, \lambda^*, \epsilon)$ for all $\omega \in \Omega_{n,\varepsilon}$, which together with (A.6) implies $\lim_l [f_{n(k(l))}(\theta_{n(k(l))}, \lambda_{n(k(l))}, \epsilon_{n(k(l))}) - g(\theta^*, \lambda^*, \epsilon)](\omega_{n(k(l))}) > 0$. Given the definition of f_n and g stated in (A.3), this inequality occurs only if $\limsup_l \sqrt{n(k(l))}\, E_P[m_{ij}(\theta_{n(k(l))})] > \xi_j(\theta^*)$ for some j. This gives a contradiction to (A.5). Therefore, the claim of Step 3 is correct.

Combining Steps 1, 2, and 3 implies the result of the lemma. *Q.E.D.*

A.11. *Proof of Lemma 4.2*

Apply the proof of Lemma 4.1 to $V^\delta_{n,\mathcal{J}}$ for each $\mathcal{J}$.

A.12. *Proof of Theorem 5.1*

PROOF OF PART (1): As has been noted, this is a consequence of Equation (5.2) in the main text.

PROOF OF PART (2): Since $C_n(q_n + \kappa_n) \subseteq \widehat{\Theta}_I \subseteq \widetilde{\Theta}_I \subseteq C_n(\widehat{c} \vee q_n + \kappa_n)$, the rate and consistency results follow from the rates and consistency results for $C_n(q_n + \kappa_n)$ and $C_n(\widehat{c} \vee q_n + \kappa_n)$ obtained in Theorems 3.1 and 3.2 (for $\kappa_n = 0$ case). *Q.E.D.*

A.13. *Proof of Theorem 5.2*

The proof is a direct corollary of Theorems 4.1 and 4.2.

REFERENCES

AMEMIYA, T. (1985): *Advanced Econometrics*. Cambridge, MA: Harvard University Press. [1252]

ANDERSON, T. W., AND H. RUBIN (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, 20, 46–63. [1268]

ANDREWS, D. W. K. (1997): "Estimation when a Parameter Is on a Boundary: Theory and Applications," Working Paper, Cowles Foundation, Yale University. [1264,1277]

——— (1999): "Estimation when a Parameter Is on a Boundary," *Econometrica*, 67, 1341–1383. [1260]

——— (2001): "Testing when a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734. [1252]

ANDREWS, D. W. K., AND P. GUGGENBERGER (2006): "The Limit of Finite Sample Size and a Problem with Subsampling," Working Paper, Yale University. [1268,1271]

ANDREWS, D. W. K., S. BERRY, AND P. JIA (2004): "Placing Bounds on Parameters of Entry Games in the Presence of Multiple Equilibria," Working Paper, Yale University. [1271]

ANGRIST, J., AND A. KRUEGER (1992): "Age at School Entry," *Journal of the American Statistical Association*, 57, 11–25. [1249]

BAJARI, P., L. BENKARD, AND J. LEVIN (2006): "Estimating Dynamic Models of Imperfect Competition," Working Paper, University of Minnesota; *Econometrica* (forthcoming). [1249]

BERESTEANU, A., AND F. MOLINARI (2006): "Asymptotic Properties for a Class of Partially Identified Models," Working Paper, Cornell University. [1270]

BLUNDELL, R., M. BROWNING, AND I. CRAWFORD (2005): "Best Nonparametric Bounds on Demand Responses," Working Paper, University College London. [1244,1249]

BUGNI, F. (2006): "Bootstrap Methods for Some Partially Identified Models," Working Paper, Northwestern University. [1271]

CHERNOFF, H. (1954): "On the Distribution of the Likelihood Ratio," *The Annals of Mathematical Statistics*, 25, 573–578. [1252,1259,1260,1264]

CHERNOZHUKOV, V., AND I. FERNANDEZ-VAL (2005): "Subsampling Inference on Quantile Regression Processes," *Sankhyā*, 67, 253–276. [1270]

CHERNOZHUKOV, V., AND C. HANSEN (2004): "Instrumental Quantile Regression," Working Paper, MIT. [1250,1268,1271]

——— (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 81, 806–839. [1244,1249]

CHERNOZHUKOV, V., C. HANSEN, AND M. JANSSON (2005): "Finite-Sample Inference for Quantile Regression Models," Working Paper, MIT. [1244,1250,1268,1271]

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2002): "Parameter Set Inference in a Class of Econometric Models," Working Paper, MIT. [1243,1245,1248]

——— (2007): "Supplement to 'Estimation and Confidence Regions for Parameter Sets in Econometric Models'," *Econometrica Supplementary Material*, 75, http://www.econometricsociety.org/ecta/supmat/5210_extensions.pdf. [1251,1271]

CILIBERTO, F., AND E. TAMER (2003): "Market Structure and Multiple Equilibria in Airline Markets," Working Paper, Northwestern University. [1244,1249]

CUEVAS, A., AND R. FRAIMAN (1997): "A Plug-in Approach to Support Estimation," *The Annals of Statistics*, 25, 2300–2312. [1245]

DACUNHA-CASTELLE, D., AND E. GASSIAT (1999): "Testing the Order of a Model Using Locally Conic Parametrization: Population Mixtures and Stationary ARMA Processes," *The Annals of Statistics*, 27, 1178–1209. [1245]

DAVYDOV, Y. A., M. A. LIFSHITS, AND N. V. SMORODINA (1998): *Local Properties of Distributions of Stochastic Functionals*. Providence, RI: American Mathematical Society. [1276,1279]

DEMIDENKO, E. (2000): "Is This the Least Squares Estimate?" *Biometrika*, 87, 437–452. [1249]

DUDLEY, R. M. (1985): "An Extended Wichura Theorem, Definitions of Donsker Class, and Weighted Empirical Distributions," in *Probability in Banach Spaces*, Vol. V, Lecture Notes in Mathematics, Vol. 1153. Berlin: Springer, 141–178. [1262]

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387. [1245,1268]

——— (2006): "Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics," *Journal of Econometrics*, 133, 443–477. [1256]

FRISCH, R. (1934): *Statistical Confluence Analysis by Means of Complete Regression Systems*. Oslo, Norway: University Institute of Economics. [1244]

FUKUMIZU, K. (2003): "Likelihood Ratio of Unidentifiable Models and Multilayer Neural Networks," *The Annals of Statistics*, 31, 833–851. [1245]

GALICHON, A., AND M. HENRY (2006a): "Dilation Bootstrap. A Methodology for Constructing Confidence Regions with Partially Identified Models," Unpublished Manuscript, Harvard and Columbia Universities. [1271]

——— (2006b): "Inference in Incomplete Models," Discussion Paper 0506-28, Columbia University. [1268,1271]

GEYER, C. J. (1994): "On the Asymptotics of Constrained M-Estimation," *The Annals of Statistics*, 22, 1993–2010. [1264]

GILSTEIN, C. Z., AND E. LEAMER (1983): "Robust Sets of Regression Estimates," *Econometrica*, 51, 555–582. [1244]

HAILE, P., AND E. TAMER (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1–51. [1244,1248]

HANNAN, E. J. (1982): "Testing for Autocorrelation and Akaike's Criterion," *Journal Applied Probability*, 19A, 403–412 (special volume of essays in statistical science). [1245]

HANNAN, E. J., AND M. DEISTLER (1988): *The Statistical Theory of Linear Systems*, Probability and Mathematical Statistics. New York: Wiley. [1245]

HANSEN, L., J. HEATON, AND E. G. LUTTMER (1995): "Econometric Evaluation of Asset Pricing Models," *The Review of Financial Studies*, 8, 237–274. [1244,1245]

HU, L. (2002): "Estimation of a Censored Dynamic Panel Data Model," *Econometrica*, 70, 2499–2517. [1268]

IMBENS, G., AND C. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 74, 1845–1857. [1268,1271]

KLEIBERGEN, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803. [1268]

KLEPPER, S., AND E. LEAMER (1984): "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica*, 52, 163–184. [1244]

KNIGHT, K. (1999): "Epi-Convergence and Stochastic Equisemicontinuity," Technical Report, Department of Statistics, University of Toronto, available at http://www.utstat.toronto.edu/keith/papers/. [1252,1260,1261]

KOROSTELËV, A. P., L. SIMAR, AND A. B. TSYBAKOV (1995): "Efficient Estimation of Monotone Boundaries," *The Annals of Statistics*, 23, 476–489. [1245]

LEHMANN, E. L. (1997): *Testing Statistical Hypotheses* (Second Ed.). New York: Springer-Verlag. [1268]

LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses* (Third Ed.). New York: Springer-Verlag.

LINTON, O., T. POST, AND Y. WHANG (2005): "Testing for Stochastic Dominance Efficiency," Working Paper, London School of Economics. [1244]

LIU, X., AND Y. SHAO (2003): "Asymptotics for Likelihood Ratio Tests under Loss of Identifiability," *The Annals of Statistics*, 31, 807–832. [1245]

MANSKI, C. (2003): *Partial Identification of Probability Distributions*. New York: Springer-Verlag. [1243,1244]

MANSKI, C., AND E. TAMER (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–547. [1244,1245,1248,1254]

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

——— (2007): "Identification for Prediction and Decision," Harvard University Press, forthcoming.

MARSCHAK, J., AND W. H. ANDREWS (1944): "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 812, 143–203. [1244]

MCFADDEN, D. L. (2005): "Revealed Stochastic Preference: A Synthesis," *Economic Theory*, 26, 245–264. [1244,1249]

MIKUSHEVA, A. (2006): "Robust Confidence Sets in the Presence of Weak Instruments," Unpublished Report, Harvard University. [1247]

MOLCHANOV, I. S. (1990): "Empirical Estimation of Quantiles of Distributions of Random Closed Sets," *Teoriya Veroyatnoste i Primeneniya*, 35, 586–592. [1270]

——— (2005): *Theory of Random Sets*. London: Springer-Verlag. [1260,1270]

MOLINARI, F. (2004): "Missing Treatments," Working Paper, Cornell University. [1244]

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. IV. Amsterdam: North-Holland, 2111–2245. [1252]

PAKES, A., J. PORTER, J. ISHII, AND K. HO (2006): "Moment Inequalities and Their Applications," Working Paper, Harvard University. [1271]

PHILLIPS, P. C. B. (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5, 181–240. [1245,1249]

POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. New York: Springer-Verlag. [1257,1270]

PÖTSCHER, B. M., AND I. R. PRUCHA (1997): *Dynamic Nonlinear Econometric Models*. Berlin: Springer-Verlag. [1245]

REDNER, R. (1981): "Note on the Consistency of the Maximum Likelihood Estimate for Nonidentifiable Distributions," *The Annals of Statistics*, 9, 225–228. [1245]

ROMANO, J., AND A. SHAIKH (2006): "Inference for Partially Identified Econometric Models," Working Paper, Stanford University. [1258,1268,1271]

ROSEN, A. (2006): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," Working Paper, UCL. [1268,1271]

RYAN, S. P. (2005): "The Costs of Environmental Regulation in a Concentrated Industry," Working Paper, MIT. [1249]

RYTCHKOV, O. (2006): "Filtering Out Expected Dividends and Expected Returns," Working Paper, MIT. [1271]

SANTOS, A. (2006): "Inference in Nonparametric Instrumental Variables with Partial Identification," Working Paper, Stanford. [1271]

SILVAPULLE, M. J., AND P. K. SEN (2005): *Constrained Statistical Inference*, Probability and Statistics. Hoboken, NJ: Wiley. [1271]

SOARES, G. (2006): "Inference for Partially Identified Models with Inequality Moment Constraints," Working Paper, Yale. [1269,1271]

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586. [1245,1268]

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge, U.K.: University Press. [1245, 1252,1254,1262]

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*, New York: Springer-Verlag. [1252,1253,1262,1272]

VAPNIK, V. N. (1998): *Statistical Learning Theory*, Adaptive and Learning Systems for Signal Processing, Communications, and Control. New York: Wiley. [1256]

VARIAN, H. R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50, 945–973. [1244]

——— (1984): "The Nonparametric Approach to Production Analysis," *Econometrica*, 52, 579–597. [1249]

VERES, S. (1987): "Asymptotic Distributions of Likelihood Ratios for Overparametrized ARMA Processes," *Journal of Time Series Analysis*, 8, 345–357. [1245]

VILLANI, C. (2003): *Topics in Optimal Transportation*, Graduate Studies in Mathematics, Vol. 58. Providence, RI: American Mathematical Society. [1271]