Massachusetts Institute of Technology
Department of Economics
Working Paper Series

ESTIMATION AND CONFIDENCE REGIONS FOR PARAMETER SETS IN ECONOMETRIC MODELS

(Previously called "Inference on Parameter Sets in Econometric Models," id# 909625)

Victor Chernozhukov
Han Hong
Elie Tamer

Working Paper 06-18B
Earlier Version: June 16, 2004
Revision: June 22, 2006

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

ESTIMATION AND CONFIDENCE REGIONS FOR PARAMETER SETS IN ECONOMETRIC MODELS*

VICTOR CHERNOZHUKOV† HAN HONG§ ELIE TAMER‡

ABSTRACT. The paper develops estimation and inference methods for econometric models with partial identification, focusing on models defined by moment inequalities and equalities. Main applications of this framework include analysis of game-theoretic models, revealed preference, regression with missing and mismeasured data, auction models, bounds in structural quantile models, bounds in asset pricing, among many others.

Specifically, this paper provides estimators and confidence regions for minima of an econometric criterion function $Q(\theta)$. In applications, $Q(\theta)$ embodies testable restrictions on economic models. A parameter θ that describes an economic model passes these restrictions if $Q(\theta)$ attains the minimum value normalized to be zero. The interest therefore focuses on the set of parameters Θ_I that minimizes $Q(\theta)$, called the identified set. This paper uses the inversion of the sample analog $Q_n(\theta)$ of the population criterion $Q(\theta)$ to construct the estimators and confidence regions for Θ_I. We develop consistency, rates of convergence, and inference results for these estimators and regions. The results are shown to hold under general yet simple conditions, and practical procedures are provided to implement the approach. In order to derive these results, the paper also develops methods for analyzing the asymptotics of sample criterion functions under set identification.

KEY WORDS: Set estimator, contour sets, moment inequalities, moment equalities

---

This is a revision of a previous paper "Parameter Set Inference in Econometric Models", 2002, which was circulated as an MIT Working Paper and is available at www.ssrn.com.

Date: Original Version: May 2002. This version: June 22, 2006.

We thank Alexandre Belloni, Oleg Rytchkov, Marc Henry, Denis Nekipelov, Igor Makarov, and Whitney Newey for thorough readings of the paper and their valuable comments. We also thank two anonymous referees and Oliver Linton for many valuable comments.

† Department of Economics, MIT, vchern@mit.edu. Research support from the Castle Krob Chair, National Science Foundation, and the Sloan Foundation is gratefully acknowledged.

§ Department of Economics, Duke University, han.hong@duke.edu. Research support from the National Science Foundation is gratefully acknowledged.

‡ Department of Economics, Northwestern University, tamer@northwestern.edu. Research support from the National Science Foundation and the Sloan Foundation is gratefully acknowledged.

1. Introduction

Parameters of interest in econometric models can be defined as values that minimize a population criterion function. If this criterion function is minimized uniquely at a particular parameter vector, then one can obtain confidence regions for this parameter using a sample analog of this function. This paper extends this criterion-based estimation and inference to econometric models where the objective function is minimized on a set of parameters, the identified set. (The terminology follows Manski (2003).) Our goal is to estimate and make inferences directly on the identified set. The development focuses 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 estimators and confidence regions. The resulting estimators and confidence regions are appropriate contour sets of the sample criterion functions. The paper develops consistency, rates of convergence, and inference results for these sets. Specifically, this paper shows that an appropriate lower contour set of the sample criterion function converges in Hausdorff metric to the identified set at (an exact or an arbitrarily close to) $1/\sqrt{n}$ rate, in moment condition problems, and at polynomial rates, more generally. The paper develops a method for determining the appropriate level of the contour set so that it covers the identified set with a prespecified probability. For this purpose, the paper derives the asymptotics of several inferential statistics which quantiles determine the appropriate level of the contour set. The lack of equi-continuous behavior of the sample criterion functions in moment inequality problems poses challenges to this analysis.

The primary applications of the estimation and inference methods developed in this paper are in such areas as (1) empirical game-theoretic models, (2) empirical revealed preference analysis, (2) econometric analysis with missing and mis-measured data, (3) bounds analysis in auction models, (4) structural quantile models and other simultaneous equation models without additivity, (5) bounds analysis in asset pricing models, and (6) the inference on dominance regions in stochastic dominance analysis, among others. In most of these problems, the economic models of interest 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.

In the context of estimation of games and revealed preference analysis, our methods have already been employed by several substantive empirical studies. Bajari, Benkard, and Levin (2006) estimated a dynamic Markov game where the observed action of each player satisfies discrete optimality conditions for equilibrium, which result in moment inequality conditions. Beresteanu and Ellickson (2006) presented a further application to a study of dynamic oligopolistic competition. Ciliberto and Tamer (2003) analyze empirical entry models with multiple equilibria. They do not make the equilibrium selection assumptions, which leads to moment inequality conditions and set identification. Cohen and Manuszak (2006) estimate a game in which firms may enter as different types and which has multiple equilibria in types. They also do not impose equilibrium selection assumptions. Borzekowski and Cohen (2004) estimate a model of strategic complementarity between credit unions in their choice of adopting a technology or outsourcing it. In the context of simultaneous equation models with non-additive disturbances, Chernozhukov and Hansen (2004) estimate a demand model where a partial identification occurs. In fact, they use the pointwise versions of the inference methods developed here. Econometric analysis with missing and mismeasured data is another area of applications of our methods: Molinari (2004) applied our methods to construct a confidence region for the identified set in a causal model with missing treatments. In auction analysis, the very nature of auction mechanisms also often leads to the missing data framework, see Haile and Tamer (2003).

In addition to these existing applications, there appear to be many more potential applications to the revealed preference analysis, e.g. see Varian (1982, 1984), McFadden (2005), Blundell, Browning, and Crawford (2005), and Bajari, Fox, and Ryan (2006). A potentially important application is the inference on the set of asset pricing models that satisfy mean and volatility bounds developed in Hansen, Heaton, and Luttmer (1995). Yet another area of potential applications is in the analysis of stochastic dominance relations, e.g. see Linton, Post, and Wang (2005).

The relationship of this paper to the econometric literature on inference under partial identification is as follows.[1] The concepts of set identification go back to Frisch (1934) and Marschak and Andrews (1944). Marschak and Andrews (1944) constructed the identified set

[1]Manski (2003) for a detailed introduction to partial identifiability.

as a collection of parameters representing different production functions that can not be rejected by the data and that are consistent with functional restrictions the authors consider. Frisch (1934) constructs 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 estimation in a class of nonlinear regression models where the identified set is an interval of parameters that are robust to misspecification of the distribution of the error term. In a different development, Phillips (1989) suggested that multi-collinearity may be a cause for partial identification in a number of econometric models and provided some 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, as well as derived several consistency results. In a previous version of this paper, Chernozhukov, Hong, and Tamer (2002) developed consistency and inference results for linear moment inequality models, using inversion of the econometric criterion functions, and developed an empirical application. Imbens and Manski (2004) investigated a problem of Wald inference in the case of a scalar mean parameter bounded above and below by other scalar means. Recently, Andrews, Berry, and Jia (2004) and Pakes, Porter, Ishii, and Ho (2006) investigated the inference problem using projection methods, which proceeds by constructing a region for point-identified high-dimensional nuisance parameter and then further projecting it with the purpose of obtaining a confidence region for the partially identified functionals of this parameter, such as the identified set. This projection method tends to be conservative relative to the methods developed in this paper.[2] Beresteanu and Molinari (2006) develop inference methods for the linear regression model with interval-censored outcomes, using the Wald statistic that measures the Hausdorff distance between the identified set and a set-valued estimator. These methods differ from the criterion-based inference studied in this paper. Our paper is also related to the literature on the weak identification problem, see notably Dufour

[2]Conservativity of projection methods in the point-identified setting is discussed in Romano and Wolf (2000).

(1997) and Staiger and Stock (1997). However, the problem studied here considerably differs from the latter, as the nature of failure of point identification in our main applications typically can not be approximated by the weak identification framework.

The relationship of this paper to the statistical literature is as follows. Hannan (1982) has pointed out the multi-collinearity (i.e. set-identification) problems in several time series models. Redner (1981) and Hannan and Deistler (1988) showed that a maximum likelihood estimator eventually converges to a point in the identified set Θ_I, though obviously it is not consistent for estimation of Θ_I. Veres (1987), Dacunha-Castelle and Gassiat (1999), and Liu and Shao (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 and ARMA models. These results do not apply or extend in any obvious way to moment condition models analyzed in this paper. Fukumizu (2003) pointed out that the likelihood ratio has an unusually large stochastic order in multi-layer neural networks, which does not apply to the moment condition problems analyzed in this paper. Also, the literature on image processing considers the problem of support estimation of a density, e.g. Korostelëv, Simar, and Tsybakov (1995) and Cuevas and Fraiman (1997), though the structure of such problem is different from that arising in the moment condition models analyzed here.

The rest of the paper is organized as follows. Section 2 presents the moment condition models and several examples that will serve to illustrate the analysis. Section 2 also outlines informally the main results of the paper. Section 3 develops consistency, rates of convergence, and inference results that apply generically. The results require that simple high-level conditions on the econometric criterion functions are met. Section 4 analyzes the moment inequality and moment equality models in detail and verifies the conditions of Section 3. Appendix collects proofs and a definition of notations used in the paper, and also discusses pointwise confidence regions (confidence regions for particular parameter values in the identified set).

2. Problem Definition and Informal Discussion of the Main Results

Consider a nonnegative population criterion function $Q(\theta)$ which attains its minimal value 0 on a set Θ_I, that is $\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\}$. The set Θ_I generally consists of many parameter values, and thus is a singleton. Suppose there is also a sample analog $Q_n(\theta)$ of this function. The parameter θ belongs to the parameter space Θ, which is a compact subset of the

Euclidean space $\mathbb{R}^d$. Every θ in Θ_I indexes an economic model that passes testable empirical restrictions that the criterion $Q(\theta)$ typically embodies in economic applications. This paper investigates the estimators and confidence regions for Θ_I constructed using the contour sets of the sample criterion function Q_n. (Appendix 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. Then, the section gives an informal review of the methods and the results obtained in this paper.

2.1. **Moment Condition Models.** This paper is primarily concerned with applications to two main types of econometric structural models: moment inequalities and moment equalities. In empirical analysis, the moment inequalities, much like moment equalities, represent testable restrictions on economic models. Economic models are described by the finite-dimensional parameters $\theta \in \Theta \subseteq \mathbb{R}^d$, where Θ is the parameter space. We are interested in the set of parameters $\Theta_I \subseteq \Theta$ that satisfy the testable restrictions.

The moment restrictions are computed with respect to the population probability law P of the data and take the form

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

where $m_i(\theta) = m(\theta, w_i)$ is a vector of moment functions parameterized by θ and determined by a vector of real random variables w_i. Therefore the set of parameters θ that pass restrictions (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; and could be a triangle, trapezoid, or a polyhedron, as in Examples 1 and 2 introduced below. When moment functions are non-linear, the set Θ_I is given by an intersection of nonlinear half-spaces which boundaries are defined by nonlinear manifolds.

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

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

[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, the inference on Θ_I may be based on the empirical analog of Q:

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

where $W_n(\theta)$ is a uniformly consistent estimate of $W(\theta)$. In applications $W_n(\theta)$ can be taken to be an identity matrix or chosen to weigh the individual empirical moments by estimates of inverses of their individual variances.

Moment equalities are more traditional in empirical analysis. The economic models, indexed by θ, are assumed to satisfy the set of testable restrictions given by moment equalities:

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

When the moment functions are linear in parameters, the set Θ_I is either a point or a hyperplane intersected with the parameter space Θ. When moment functions are non-linear, 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 as the set of minimizers of the generalized method of moments function

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

where $W(\theta)$ is a continuous and positive-definite matrix for each $\theta \in \Theta$. The inference on Θ_I is based on the conventional generalized method-of-moments function

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

where $W_n(\theta)$ is a uniformly consistent estimate of $W(\theta)$. In applications $W_n(\theta)$ can be an identity matrix or an estimate of the inverse of the asymptotic covariance matrix of empirical moment functions.

In many situations, we can also use the modified objective function for inference:

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

This modification is useful in cases where Q_n does not attain value 0 in finite samples.[4]

[4]In such cases, using the modified objective function typically leads to power improvements, as is well-known in point-identified cases.

2.2. **Motivating Examples.** There are several interesting examples for the moment condition models described above, where the identified set Θ_I is naturally a collection of points, rather than a single point.

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 $Q(\theta) = \|E_P[m_i(\theta)]\|_+^2 = (E_P[Y_{1i}] - \theta)_+^2 + (E_P[Y_{2i}] - \theta)_-^2$, with the 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 linear function $X_i'\theta$. The parameters of this function can be bounded using inequality $E_P[Y_{1i}|X_i] \leq X_i'\theta \leq E_P[Y_{2i}|X_i]$. These conditional restrictions imply the following inequalities are valid:

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

where Z_i is a vector of positive transformations of X_i, for instance, $Z_i = \{1(X_i \leq x_j), j = 1, ..., J\}'$, for a suitable collection of values 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 the 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 – bidder's valuation – by functions of observed bids, Y_1 and Y_2, is very natural and occurs in a variety of settings, see Haile and Tamer (2003). Analogous situations occur in income surveys, where only income brackets are available instead of true income, see Manski and Tamer (2002). Chernozhukov, Hong, and Tamer (2002) analyze this linear moment inequality set up in detail.

Example 3 (Optimal Choice of Economic Agents and Game Interactions). Analysis of the optimal choice behavior of firms and economic agents is another area of applications of (2.1). Suppose that a firm can make 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 $E_P[U_i|X_i] = 0$, for X_i representing information available to make 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

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

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 positive transformations of X_i, 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 are given in the form of moment inequality conditions. It could be noted that this simple example also allows for game-theoretic interactions among economic agents. The moment inequality conditions of the above kind are ubiquitous, and are known to arise in (more realistic) dynamic settings, see Bajari, Benkard, and Levin (2006), Ciliberto and Tamer (2003), and Ryan (2005). Similar principles are used in Blundell, Browning, and Crawford (2005) to analyze bounds on demand functions. Related ideas also appear in the area of stochastic revealed preference analysis, e.g. see Varian (1984) and McFadden (2005).

Example 4 (Structural Equations). Consider the structural instrumental variable estimation of returns to schooling. Suppose that we are interested in the following example where potential income Y is related to education E through a flexible quadratic functional form,

$Y = \theta_0 + \theta_1 E + \theta_2 E^2 + \epsilon = X'\theta + \epsilon$, for $\theta = (\theta_0, \theta_1, \theta_2)$ and $X = (1, E, E^2)'$. Although parsimonious, this simple model is not point-identified in the presence of the standard quarter-of-birth instrument suggested in Angrist and Krueger (1992).[5] In the absence of point identification, all parameter values θ consistent with the instrumental orthogonality restriction $E_P[(Y - \theta' X)Z] = 0$ are of interest for purposes of economic analysis. Phillips (1989) develops a number of related examples. Similar partial identification problems arise in nonlinear moment and instrumental variables problems, see e.g. Demidenko (2000) and Chernozhukov and Hansen (2005). In Chernozhukov and Hansen (2005), the parameters θ of the structural quantile functions for returns to schooling 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 the identification region, in the absence of point identification, is generally given by a nonlinear manifold. Chernozhukov and Hansen (2004) and Chernozhukov, Hansen, and Jansson (2005) analyze an empirical returns-to-schooling example and a structural demand example where such situations arise.

2.3. **Informal Discussion of Results.** The objective of this paper is to construct sets C_n for Θ_I that are consistent estimates of Θ_I, converge to Θ_I at fastest rates, and have the confidence interval property $\liminf_{n\to\infty} P(\Theta_I \subseteq C_n) = \alpha$, for a prespecified confidence level $\alpha \in (0, 1)$.[6] The sets C_n we construct take the form of a contour set $C_n(c)$ of level c of the sample criterion function Q_n:

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

for some appropriate normalization a_n, where $a_n = n$ in Examples 1-4. In order to simplify the discussion, assume $a_n = n$ in this section only.

[5]The instrument is the indicator of the first quarter of birth. Sometimes the indicators of other quarters of birth are used as instruments. However, these instruments are not correlated with education (correlation is extremely small) and thus bring no additional identification information.

[6]Robustness to perturbing P is also discussed in the Addendum to this paper, which obtains the conditions under which coverage holds under contiguous perturbations of P. In addition, Andrews and Guggenberger (2006) establish global robustness of the subsampling confidence regions proposed in this paper in a class of moment inequality problems. Sheikh (2006) establishes global robustness of our subsampling regions in Example 1.

In order to estimate Θ_I consistently, the level $c = \widehat{c}$ (which can be data-dependent) needs to be diverging to infinity slowly; for concreteness, we can set $\widehat{c} = \ln n$. However, in a class of problems, $\widehat{c}$ does not need to be diverging, so we can set $0 \leq \widehat{c} = O_p(1)$ and even $\widehat{c} = 0$. E.g., in Example 1, $\widehat{c} = 0$ gives us $C_n(0) = [E_n[Y_1], E_n[Y_2]]$, which clearly is consistent for the region $[E_P[Y_1], E_P[Y_2]]$. Generally, whether $\widehat{c}$ can be non-diverging in order to maintain consistency depends on whether the sample criterion function $a_n Q_n(\theta)$ has degenerate behavior, i.e. vanishes, over contractions of Θ_I in large samples, as formally stated in Section 3. In particular, the latter property does not hold in Example 4, but typically holds in Examples 1-3, under conditions formally stated in Section 4.

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

$$d_H(A, B) := \max \left[\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \right], \text{ where } d(b, A) := \inf_{a \in A} \|b - a\|,$$

and $d_H(A, B) := \infty$ if either A or B is empty. The motivation for the use of this metric comes from it being a natural generalization of the Euclidean distance and its previous uses by other authors in the consistency analysis in the context of set estimation, see Hansen, Heaton, and Luttmer (1995). The general consistency result

$$d_H(C_n(\widehat{c}), \Theta_I) \to_p 0,$$

obtained in this paper follows from the uniform convergence of the sample function Q_n to the limit continuous function Q over the compact parameter space Θ, where the rate of convergence over set Θ_I is $1/a_n$. Such uniform convergence condition is conventional in econometric literature, and is thus easily verifiable.

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

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

which is very close to $1/\sqrt{n}$ rate of convergence, and is exactly $1/\sqrt{n}$ in many moment inequality problems (such as Examples 1 and 2, where $a_n Q_n$ has degenerate asymptotics over contractions of Θ_I).

In order for $C_n(c)$ to have the confidence region property for Θ_I, we need to choose level $c = \widehat{c}(\alpha)$, such that $\widehat{c}(\alpha)$ is a consistent estimate of the α-quantile of the statistic:

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

which is a quasi-likelihood-ratio type quantity. The estimates $\widehat{c}(\alpha)$ can be based on the limit distributions of (2.9) or a generic subsampling method, which are developed, respectively in Section 4 and Section 3.5. For instance, in Example 1, suppose $(\sqrt{n}(E_n[Y_1] - E_P[Y_1]), \sqrt{n}(E_n[Y_2] - E_P[Y_2])) \to_d (W_1, W_2) = N(0, \Omega)$, then Section 3 shows that $\mathcal{C}_n \to_d \mathcal{C} = \max[(W_1)_+^2, (W_2)_-^2]$, where the distribution of $\mathcal{C}$ can be easily obtained by simulation methods (see Section 4). For the cases when the limit distribution of (2.9) is not easily tractable, the paper constructs a generic subsampling estimate $\widehat{c}(\alpha)$, which is based on subsampling an approximation of statistic (2.9), where one uses the consistent estimate $C_n(\widehat{c})$ in place of the unknown set Θ_I in (2.9) (see Section 3.5 for a detailed description of the algorithm).

The paper characterizes the asymptotic behavior and derives the limit distribution of the statistic $\mathcal{C}_n$. The paper also characterizes the limit distribution of related statistics used to determine the probability of false coverage (probability of covering larger sets than Θ_I). (This in turn characterizes the power properties of the testing procedure implicitly defined by the confidence region.) The non-equicontinuous behavior of the empirical process $\theta \mapsto a_n Q_n(\theta)$ in e.g. Examples 1-3 poses a challenge to this analysis, which is addressed through a generalization of epi-convergence and stochastic equi-semi-continuity (Knight 1999) to the set-identified case. The "parameter-on-the-boundary" problem is another challenge in this analysis; e.g. it arises in Example 4, where the identified set Θ_I, defined as an intersection of a hyperplane with Θ, generally has common points with the boundary of Θ, defined relative to $\mathbb{R}^d$. This challenge is addressed through an appropriate generalization of the "Chernoff regularity"– the condition that in the point-identified case requires convergence of the local parameter space to a cone (Chernoff 1954, Andrews 2001). The generalization requires a convergence of a graph of the local parameter space to an appropriate limit graph.

3. General Estimation and Inference in Large Samples

This section defines the estimators and confidence regions formally and develops the basic results on consistency, rates of convergence, and coverage properties of these regions. The section develops general conditions that parallel those used in extremum estimation in point-identified cases (Amemiya 1985, Newey and McFadden 1994, van der Vaart 1998). Section 4 illustrates and verifies these conditions for the moment condition models.

3.1. **Basic Setup.** The parameter space Θ is a non-empty compact subset of $\mathbb{R}^d$ equipped with a subspace topology relative to $\mathbb{R}^d$. Data $w_1, ..., w_n$ are a random vector defined on a complete probability space $(\Omega, \mathcal{F}, P)$. Suppose that the sample criterion function $Q_n(\theta) = Q_n(\theta, w_1, ..., w_n)$ is available, and that Q_n converges uniformly to a continuous criterion function $Q \geq 0$ that attains minimal value 0 on Θ_I. The contour sets of Q_n will be used for estimation and inference on Θ_I. This approach therefore employs the classical duality principle of inverting a likelihood type test statistic to obtain confidence regions.

Regarding notations used in the paper, ϵ-expansion of Θ_I in Θ is defined as $\Theta_I^\epsilon := \{\theta \in \Theta : d(\theta, \Theta_I) \leq \epsilon\}$. Unless an ambiguity arises, $\sup_A f$ is used to denote $\sup_{a \in A} f(a)$. The notions of stochastic convergence, e.g. convergence in (outer) probability, denoted as $\rightarrow_p$, and stochastic order symbols, O_p and o_p are defined with respect to the outer probability P^*, as in van der Vaart and Wellner (1996); wp $\rightarrow 1$ stands for "with (inner) probability approaching 1." For any two numbers a and b, $a \wedge b$ denotes $\min(a, b)$, and $a \vee b$ denotes $\max(a, b)$. For convenience, Appendix A collects other definitions and notations.

3.2. **Consistency and Rates of Convergence in The General Cases.** Let the following assumption hold.

Condition C.1 (Uniform Convergence and Continuity). *(a)* Θ *is a non-empty compact subset of* $\mathbb{R}^d$*, (b)* $Q : \Theta \mapsto \mathbb{R}_+$ *is continuous and* $\min_\Theta Q = 0$*; let* $\Theta_I := \arg\min_\Theta Q$*, (c)* $Q_n(\theta) = Q_n(\theta, w_1, ..., w_n)$ *takes values in* $\mathbb{R}_+$ *and is jointly measurable in the parameter* θ *and the data* $w_1, ..., w_n$ *defined on a complete probability space* $(\Omega, \mathcal{F}, P)$*, (d)* $\sup_\Theta |Q_n - Q| = O_p(1/b_n)$ *for a sequence of constants* $b_n \rightarrow \infty$ *, and (e)* $\sup_{\Theta_I} Q_n = O_p(1/a_n)$ *for a sequence of constants* $a_n \rightarrow \infty$.

Condition C.1 assumes uniform convergence for the criterion function Q_n to the limit Q. It also identifies Θ_I as the minimizer of the limit criterion function Q. The assumptions $Q \geq 0$ and $Q_n \geq 0$ are not restrictive.[7] In C.1(c), measurability is assumed to hold with

[7]Given a function $\widetilde{Q}_n : \Theta \rightarrow \mathbb{R}$ and its continuous uniform limit $\widetilde{Q} : \Theta \rightarrow \mathbb{R}$, we can define $Q_n(\theta) = \widetilde{Q}_n(\theta) - \inf_{\theta' \in \Theta} \widetilde{Q}_n(\theta')$ and $Q(\theta) = \widetilde{Q}(\theta) - \inf_{\theta' \in \Theta} \widetilde{Q}(\theta')$ to reach this assumption.

respect to the product of the sigma-field $\mathcal{F}$ and the Borel sigma-field of Θ. C.1(c) guarantees measurability of $\sup_{\Theta_I} Q_n$ and related statistics; see e.g. van der Vaart and Wellner (1996), p.47.[8]

Condition C.1 also defines the principal quantity $\sup_{\Theta_I} Q_n$ which plays the crucial role in the analysis of inference. Its rate of convergence to zero a_n plays the key role in the analysis of consistency and rates of convergence. Section 4 shows that in the models of Section 2 $a_n = n$.

The contour sets of $a_n Q_n$ form the class of estimates we consider. The level c contour set of $a_n Q_n$ is defined as

$$C_n(c) := \Big\{\theta \in \Theta : a_n Q_n(\theta) \leq c\Big\}, \tag{3.1}$$

where $c \geq 0$. Next let $\widehat{c}$ be a sequence of non-negative random real variables such that $\widehat{c} \rightarrow_p \infty$ slowly so that $\widehat{c}/a_n \rightarrow_p 0$. For instance, when $a_n = n$, we can set $\widehat{c} = \ln n$. We will discuss the choice of $\widehat{c}$ further when we consider inference. The estimates and confidence regions will generally take the form $\widehat{\Theta}_I := C_n(\widehat{c})$.

A condition that determines the rate of convergence of $C_n(c)$ to Θ_I is the following.

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 $\{\theta \in \Theta : d(\theta, \Theta_I) \geq \kappa_\varepsilon / a_n^{1/\gamma}\}$,[9] with probability at least $1 - \varepsilon$.

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

Theorem 3.1 (Coverage, Consistency, and Rates of Convergence of $C_n(\widehat{c})$). *Let $\widehat{\Theta}_I = C_n(\widehat{c})$, where $\widehat{c} \rightarrow_p \infty$ such that $\widehat{c}/a_n \rightarrow_p 0$. Suppose that $\Theta_I \neq \Theta$, then, (1) C.1 implies that $\Theta_I \subseteq \widehat{\Theta}_I$ wp $\rightarrow 1$ and $d_H(\widehat{\Theta}_I, \Theta_I) = o_p(1)$, and (2) C.1 and C.2 imply that $d_H(\widehat{\Theta}_I, \Theta_I) = O_p((\widehat{c}/a_n)^{1/\gamma})$. Suppose that $\Theta_I = \Theta$, then (3) C.1 implies that $d_H(\widehat{\Theta}_I, \Theta_I) = 0$ wp $\rightarrow 1$.*

[8]The condition of joint measurability is only needed to simplify exposition, following a suggestion of a referee. Otherwise, we can easily drop this condition, since we allow for stochastic convergence in the sense of Hoffmann-Jorgensen. In this case, under the other assumptions stated, the primary statistics are asymptotically measurable.

[9]When $\Theta_I = \Theta$, this set is empty, in which case C.2 does not apply.

Parts (1) and (2) address the case of the partial identification, when $\Theta_I \neq \Theta$. In some sense, this is the typical case for applications, and thus 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.[10] Section 4 shows that in the moment-condition models of Section 2, Q_n is locally quadratic, i.e. $\gamma = 2$, and $a_n = n$. It follows by Theorem 3.1 that the convergence rate can be made arbitrarily close to $1/\sqrt{n}$.[11]

Part (3) addresses the case of the complete non-identification, when $\Theta_I = \Theta$, in which 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.

Example 1 (contd.) In Example 1, 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 $(\sqrt{n}(E_n[Y_1] - E_P[Y_1]), \sqrt{n}(E_n[Y_2] - E_P[Y_2]))' \to_d (W_1, W_2)' \sim N(0, \Omega)$. 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, it is simple to verify that C.2 holds with $\gamma = 2$. Hence by Theorem 3.1 the set $C_n(\ln n)$ is consistent exactly at $\sqrt{\ln n/n}$ rate. Note, however, that the set $C_n(0) = [E_n[Y_1], E_n[Y_2]]$ consistently estimates $[E_P[Y_1], E_P[Y_2]]$ at $1/\sqrt{n}$ rate. Hence, in this example and many others, but not all, it is possible to achieve the rate $1/a_n^{1/\gamma}$ exactly. Section 3.3 below develops this point further.

Remark 3.1. (A counter-example) The following example shows why it is not possible to achieve the sharp rate of convergence $1/a_n^{1/\gamma}$ by setting $\widehat{c} = O_p(1)$ in all cases. Setting $\widehat{c} = O_p(1)$ may in general lead to inconsistency. Consider the following trivial example that illustrates the source of the inconsistency. Let $\Theta = [0, 3]$, $Q(\theta) = 0$ for each $\theta \in [0, 2]$ and $Q(\theta) = 1$ for each $\theta \in (2, 3]$, so that $\Theta_I = [0, 2]$; $Q_n(\theta) = \chi^2/n$ for $\theta \in [0, 1]$, $Q_n(\theta) = 0$ for $\theta \in (1, 2]$, and $Q_n(\theta) = 1$ for $\theta \in (2, 3]$, where χ^2 is a chi-square variable. Then Theorem 3.1(1) applies with $a_n = n$ to claim that $C_n(\ln n)$ is consistent. However, $C_n(c)$ for a fixed $c \geq 0$ is not consistent, since $d_H(C_n(c), \Theta_I) = d_H([1, 2], \Theta_I) = 1$ with the asymptotic probability $Pr(\chi^2 > c) > 0$,

[10] 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 in fact show consistency of smaller sets replacing b_n with $a_n \gg b_n$, where $a_n = n$ in regular cases. More generally, b_n and a_n are defined by C.1(d,e).

[11] In other examples like the ones considered in Kim and Pollard (1990), $a_n = n^{2/3}$ and $\gamma = 2$, giving the rate of convergence $n^{1/3}$.

while $d_H(C_n(c), \Theta_I) = d_H([0,2], \Theta_I) = 0$ with the asymptotic probability $Pr(\chi^2 \leq c) < 1$. Therefore, with a positive probability the set $C_n(c)$ does not cover substantial portions of the set Θ_I, and the Hausdorff distance between $C_n(c)$ and Θ_I does not converge to 0. The inconsistency of this kind extends to more general cases such as Example 4. Thus, $\widehat{c} \to_p \infty$ is needed to achieve consistency generally.

3.3. **Consistency and Rates of Convergence with "Degenerate Interior".** In many moment inequality problems, the exact rate of convergence $1/a_n^{1/\gamma}$ can be attained by setting $\widehat{c} = O_p(1)$ or even $\widehat{c} = 0$. The discussion of Example 1 above provides the simplest instance where this is possible. The reason is that in moment-inequality problems, criterion function Q_n can have degenerate asymptotics, i.e. vanish, over subsets of the identified set Θ_I that can approximate Θ_I. Consistency and rate results then follow, because $C_n(c)$ includes these subsets even when $c = 0$.

In order to discuss this property formally, consider the following condition:

Condition C.3 (Degeneracy). *There exists a constant $\eta > 0$ and a collection of subsets $\{\Theta_I^{-\epsilon}, \epsilon \in [0, \eta]\}$ of Θ_I such that (a) $d_H(\Theta_I^{-\epsilon}, \Theta_I) \leq \epsilon$ for all $\epsilon \in [0, \eta]$, (b) for any $\epsilon \in [0, \eta]$, there is n_ϵ such that for all $n > n_\epsilon$, $P\{\sup_{\Theta_I^{-\epsilon}} a_n Q_n = 0\} = 1$, (c) there exists $\gamma > 0$ such that for any $\varepsilon > 0$ there are constants $(\kappa_\varepsilon, n_\varepsilon)$ such that for all $n \geq n_\varepsilon$ $P\{\sup_{\Theta_I^{-\kappa_\varepsilon / a_n^{1/\gamma}}} a_n Q_n = 0\} \geq 1 - \varepsilon$.*

In the remainder of the paper, we take $\Theta_I^{-\epsilon}$ to be an ϵ-contraction of the set Θ_I, that is

$$\Theta_I^{-\epsilon} := \{\theta \in \Theta_I : d(\theta, \Theta \setminus \Theta_I) \geq \epsilon\}, \tag{3.2}$$

where $\epsilon \geq 0$,[12] although, in principle, Condition C.3 does not require the sets $\Theta_I^{-\epsilon}$ to be ϵ-contractions of Θ_I. C.3(a-b) typically arises in the moment inequality models due to all finite-sample moment inequalities satisfied on ϵ-contractions $\Theta_I^{-\epsilon}$ with probability converging to 1, which makes the criterion function vanish on $\Theta_I^{-\epsilon}$. Condition C.3(c) further puts a rate assumption on exactly how this happens. Section 4 verifies this condition in our main applications.

Theorem 3.2 (Consistency and Rates of Convergence of $C_n(\widehat{c})$ with Degenerate Interiors). *Let $\widehat{\Theta}_I$ denote $C_n(\widehat{c})$ where $\widehat{c} \geq 0$ with probability 1 and $\widehat{c} \to_p c \geq 0$. Then, (1) C.2 and C.3(a,b)*

[12]Note that this set is always well-defined, although it may be an empty set.

imply that $d_H(\widehat{\Theta}_I, \Theta_I) = o_p(1)$, *and (2) C.2 and C.3 imply that* $d_H(\widehat{\Theta}_I, \Theta_I) = O_p(1/a_n^{1/\gamma})$. *Moreover, if* $\Theta_I = \Theta$ *and* $\sup_{\Theta_I} a_n Q_n = 0$ *wp* $\to 1$, *then (3)* $d_H(\widehat{\Theta}_I, \Theta_I) = 0$ *wp* $\to 1$.

Parts (1) and (2) contain the results of primary interest, which state that if C.3 holds, the rate $1/a_n^{1/\gamma}$ is achieved exactly. In particular, the smallest contour set, $C_n(0)$, is consistent and converges to Θ_I at the rate $1/a_n^{1/\gamma}$. Section 4 shows that in many moment inequality examples, C.3(b) and C.3(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 non-identification, $\Theta_I = \Theta$, and degenerate behavior of $a_n Q_n$ on Θ_I; in which case the estimator converges to Θ_I in the Hausdorff metric faster than any rate. This case is not of prime interest; we state it for completeness.

Example 1 (contd.) To clarify the role of C.3, recall Example 1. Clearly, for a sufficiently small $\epsilon > 0$, $\Theta_I^{-\epsilon} = [E_P[Y_1] + \epsilon, E_P[Y_2] - \epsilon]$ can approximate $\Theta_I = [E_P[Y_1], E_P[Y_2]]$ in the Hausdorff metric, provided Θ_I is not a singleton. Since $Q_n(\theta) = (E_n[Y_1] - \theta)_+^2 + (E_n[Y_2] - \theta)_-^2$, with probability converging to 1, $Q_n = 0$ on $\Theta_I^{-\epsilon}$. Further, for any $\varepsilon > 0$, a constant κ_ε can be found such that $Q_n = 0$ on $\Theta_I^{-\kappa_\varepsilon/\sqrt{n}} = [E_P[Y_1] + \kappa_\varepsilon/\sqrt{n}, E_P[Y_2] - \kappa_\varepsilon/\sqrt{n}]$ with probability at least $1 - \varepsilon$ in large samples. Thus, $C_n(c)$ is consistent at rate $1/\sqrt{n}$ in this example.

3.4. **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. The inferential properties of sets $C_n(c)$ are determined by the statistic

$$\mathcal{C}_n = \sup_{\theta \in \Theta_I} a_n Q_n(\theta). \tag{3.3}$$

Indeed, Lemma 3.1 below shows that event $\{\mathcal{C}_n \leq c\}$ is equivalent to event $\{\Theta_I \subseteq C_n(c)\}$. If quantiles of $\mathcal{C}_n$ or good upper bounds on them are known, finite-sample inference can be conducted.[13] This paper provides asymptotic estimates of quantiles of $\mathcal{C}_n$, using either a generic subsampling method, developed in Section 3.5, or the asymptotic limits 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 $\mathcal{C}_n$). *Suppose that* $P\{\mathcal{C}_n \leq c\} \to P\{\mathcal{C} \leq c\}$ *for each* $c \in [0, \infty)$, *where the distribution function of* $\mathcal{C}$ *is non-degenerate and continuous on* $[0, \infty)$.

[13]For instance, the upper bounds on quantiles can be obtained using the maximal inequalities for empirical processes.

Section 4 verifies C.4 for moment condition models.

Lemma 3.1 (Basic Large Sample Confidence Regions). *(1) Under C.1 event $\{\mathcal{C}_n \le c\}$ is equivalent to event $\{C_n(c)$ covers $\Theta_I\}$. (2) Suppose that C.4 holds. Then for any $\widehat{c} \to_p c(\alpha) := \inf\{c \ge 0 : P\{\mathcal{C} \le c\} \ge \alpha\}$ for $\alpha \in (0,1)$, such that $\widehat{c} \ge 0$ with probability 1, we have that* $\lim_n P\{\Theta_I \subseteq C_n(\widehat{c})\} = \lim_n P\{\mathcal{C}_n \le \widehat{c}\} = P\{\mathcal{C} \le c(\alpha)\} = \alpha$ *if* $c(\alpha) > 0$, *and* $\liminf_n P\{\Theta_I \subseteq C_n(\widehat{c})\} = \liminf_n P\{\mathcal{C}_n \le \widehat{c}\} \ge P\{\mathcal{C} = 0\} \ge \alpha$ *if* $c(\alpha) = 0$.

3.5. **Generic Estimation of the Critical Value based on 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 $\mathcal{C}_n$). *For $\mathcal{C}_n(\delta_n) := \sup_{\Theta_I^{\delta_n/a_n^{1/\gamma}}} a_n Q_n$, we have that $P[\mathcal{C}_n(\delta_n) \le c] = P[\mathcal{C} \le c] + o(1)$ for any $\delta_n \downarrow 0$ and any $c \ge 0$. If C.3 holds, in addition require that this condition holds for any $\delta_n \uparrow 0$.*

Section 4 verifies Condition C.5 for models of Section 2. C.5 implies that it suffices to apply subsampling to a feasible statistic $\sup_{C_n(\widehat{c})} a_b Q_b$ in place of the infeasible statistic $\sup_{\Theta_I} a_b Q_b$.

Generic Subsampling Algorithm. At a preliminary stage, for cases when data $\{W_t\}$ is i.i.d. sequence, consider all subsets of size $b \ll n$.[14] Denote the number of subsets by B_n. 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 $\{W_j, ..., W_{j+b-1}\}$. The algorithm has four steps: **(1)** Initialize some starting value $\widehat{c}_0$, which can be data-dependent, such that $\widehat{c}_0 \to_p c_0 \ge 0$. Set $\kappa_n = \ln n$. If C.3 is known to hold, we can also set $\widehat{c}_0 = 0$ and $\kappa_n = 0$. **(2)** Compute $\widehat{c}_1$ as the α-quantile of the sample $\{\widehat{\mathcal{C}}_{j,b,n} := \sup_{\theta \in C_n(\widehat{c})} a_b Q_{j,b,n}(\theta),\ j = 1, ..., B_n\}$, using $\widehat{c} = \widehat{c}_0 + \kappa_n$, where $Q_{j,b,n}$ denotes the criterion function evaluated using j-th subsample. **(3)** (Optional/Asymptotically Equivalent Iterations) Repeat Step 2 for $l = 2, ..., L$ by computing $\widehat{c}_l$ from Step 2 using $\widehat{c} = \widehat{c}_{l-1} + \kappa_n$. **(4)** Report $C_n(\widehat{c}_L + \kappa_n)$ as a consistent estimator and a confidence region. Report $C_n(\widehat{c}_L)$ as a confidence region. (The latter region may be inconsistent as an estimator, if C.3 does not hold).

[14]In applications, since number of such subsets is large, it suffices to consider a smaller number, B_n, of randomly chosen subsets of size b such that $B_n \to \infty$ as $n \to \infty$.

Remark 3.2. Chernozhukov, Hong, and Tamer (2002) discuss implementation and computation in further detail. Using Example 2 as the basis for simulations, they find that a small number of iterations, $L = 1$ or 2, setting c_0 to α-quantile of a χ^2 variable with degrees of freedom equal to the number of moment equations, and using $b = 300$ and $b = 400$, led to good coverage and estimation results for samples of size 1000 and 2000. Politis, Romano, and Wolf (1999) describe calibration methods for choosing b in practice.

Theorem 3.3 (General Validity of Subsampling). *Suppose (a)* $\{W_1, ..., W_n\}$ *is either i.i.d. or a stationary and strongly mixing series, (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 C.1, C.2, C.4, and C.5 hold. Let* $\alpha \in (0,1)$ *denote the desired coverage level. Then, for any finite iteration* L *of the algorithm described above, the following is true: (1)* $\widehat{c}_L \to_p c(\alpha) := \inf\{c \geq 0 : P\{\mathcal{C} \leq c\} \geq \alpha\}$, *(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 $C_n(\widehat{c}_L)$ 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 $C_n(\widehat{c}_L + \kappa_n)$ for κ_n defined above. (When C.3 is known to hold, we do not need to expand them, so we can set $\kappa_n = 0$.)

Remark 3.3. It follows from the proof of Theorem 3.3 that Conditions C.1 and C.4 alone suffice for $C_n(\widehat{c}_L)$ to cover Θ_I with P-probability at least α.

Remark 3.4. If the researcher does not know whether C.3 holds, he can still use the algorithm with the expansion constant $\kappa_n = \ln n$.

Remark 3.5 (Variations). Recently Sheikh (2006) proposed a step-down variant of our algorithm, which is numerically equivalent to our algorithm, except that it employs the choice of constants $c_0 \propto a_n$ and $\kappa_n = 0$, where the very conservative choice $c_0 \propto a_n$ is used to avoid estimation of the set Θ_I. The finitely-iterated step-down algorithm is typically more conservative (hence less powerful) than the original procedure. The infinitely-iterated step-down algorithm has the same asymptotic properties as our algorithm, but it is more computationally expensive. [15]

3.6. **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. Such a task faces two major difficulties: one is the failure of the usual stochastic equicontinuity conditions of the underlying empirical process and another is the parameter on the boundary problem, as defined below. This section outlines a framework for obtaining these limits by relying on concepts of stochastic equi-semi-continuity and generalizations of Chernoff (1954) type conditions on the parameter space Θ.

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

$$\mathcal{C}_n(\delta) \to_d \mathcal{C}(\delta) \text{ in } \mathbb{R}. \tag{3.4}$$

Relation (3.4) implies that the probability that the confidence region for Θ_I covers false local region $\Theta_I^{\delta/a_n^{1/\gamma}}$ satisfies

$$P\{\Theta_I^{\delta/a_n^{1/\gamma}} \subseteq C_n(c(\alpha))\} = P\{\sup_{\Theta_I^{\delta/a_n^{1/\gamma}}} a_n Q_n \leq c(\alpha)\} \to P\{\mathcal{C}(\delta) \leq c(\alpha)\}, \tag{3.5}$$

as long as $c(\alpha)$ is the continuity point of the distribution of $\mathcal{C}(\delta)$. Then asymptotic probability of false coverage satisfies $P\{\mathcal{C}(\delta) \leq c(\alpha)\} \leq P\{\mathcal{C} \leq c(\alpha)\}$, with strict inequality holding if the distribution function of $\mathcal{C}(\delta)$ differs from the distribution function of $\mathcal{C}$ at $c(\alpha)$. From a testing prospective, we can view $\Theta_I^{\delta/a_n^{1/\gamma}}$ as a local alternative to Θ_I, so that statements about false coverage translate in an obvious way to statements about local power.

[15]The starting value $c_0 \propto a_n$ in the step-down algorithm is very conservative. Iteration of the algorithm reduces this critical value; in the limit of iteration, the critical value is essentially the same as the critical value $c(\alpha) + o_p(1)$ produced by our algorithm. Clearly, when the number of iterations is insufficient, the finitely-iterated step-down algorithm provides confidence regions that are more conservative than our confidence regions. Thus, in practice our confidence regions are often smaller than the finitely-iterated step-down regions. More formally, when C.3 holds, starting with $c_0 = 0$ and $\kappa_n = 0$, our algorithm produces the asymptotically valid critical value $\widehat{c}_1 = c(\alpha) + o_p(1)$ in merely a single iteration, and $\widehat{c}_1$ is less than the critical value produced by the step-down variant in any finite number of iterations. When C.3 does not hold, the infinitely-iterated step-down procedure which aggressively sets $\kappa_n = 0$ asymptotically agrees with our critical value $c(\alpha) + o_p(1)$ with probability at least α. However, conditional on the event that the step-down region does not cover Θ_I, which occurs with probability at most $1 - \alpha$, the step-down critical values may be smaller than $c(\alpha) + o_p(1)$. Thus, since the discrepancy between step-down and our regions occurs only when Type I error does, the discrepancy is irrelevant.

The analysis focuses on the asymptotic behavior of the empirical process:

$$\ell_n(\theta, \lambda) := a_n Q_n(\theta + \lambda / a_n^{1/\gamma}), \quad (\theta, \lambda) \in V_n^\delta, \tag{3.6}$$

where

$$V_n^\delta := \{(\theta, \lambda) : \theta \in \Theta_I, \lambda \in V_n^\delta(\theta)\}, \; V_n^\delta(\theta) := a_n^{1/\gamma}(\Theta - \theta) \cap B_\delta, \tag{3.7}$$

where B_δ denotes the closed ball in $\mathbb{R}^d$ of diameter 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 $a_n^{1/\gamma}$.[16] The parameter λ represents the local deviation from θ and ranges over the local parameter space $V_n^\delta(\theta)$. The parameter θ ranges over the identified set Θ_I. The inferential statistics in (3.4) are suprema of the empirical process (8.4) over the set (3.7):

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

The limit properties of $\mathcal{C}_n(\delta)$ will therefore depend on the limit properties of V_n^δ. Observe that V_n^δ is the graph of the correspondence $\theta \rightrightarrows V_n^\delta(\theta)$, defined over domain Θ_I, and let V_∞^δ denote the graph of some other correspondence $\theta \rightrightarrows V_\infty^\delta(\theta)$, also defined over domain Θ_I. The condition below requires that V_n^δ "converges" to V_∞^δ, where the notion of convergence is motivated statistically.

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 $w_1, ..., w_n$ 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 $n \geq n_\varepsilon$, $P\{|\sup_{V_n^\delta} \ell_n - \sup_{V_\infty^\delta} \ell_n| \geq \varepsilon\} \leq \varepsilon$.*

In S.1 (A) is needed to make sure that ℓ_n is well defined over $\Theta_I \times B_\delta$ for large n and hence over V_∞^δ. S.1(B) is obviously satisfied when $\delta = 0$, a case which is relevant for asymptotics of $\mathcal{C}_n = \mathcal{C}_n(0)$, since in this case $V_n^0 = V_\infty^0 = \Theta_I \times \{0\}$.

Next, suppose there exists $\delta > 0$ such that $B_\delta(\theta) \subset \Theta$ for each $\theta \in \Theta_I$, where $B_\delta(\theta)$ is a closed ball in $\mathbb{R}^d$ of radius δ centered at θ. Then,

$$V_n^\delta = V_\infty^\delta = \Theta_I \times B_\delta \text{ for all sufficiently large } n, \tag{3.8}$$

and S.1(B) also holds trivially. This case will be called the *parameter in the interior* case. It appears to be reasonable in many applications, where e.g. Θ is a rectangle or a convex body

[16] That is $\Theta - \theta$ is Minkowski difference of set Θ and set $\{\theta\}$.

in $\mathbb{R}^d$ and Θ_I is in the interior of Θ. The case where the parameter in the interior condition fails to hold will be called *the parameter on the boundary* case. This definition extends the definition of Chernoff (1954) and Andrews (1999) to the present context. In this case the limit graph V_∞^δ will have a form that depends on the structure of Θ.

The parameter on the boundary case arises in many problems. One example is the linear instrumental variable model (Example 4) with Θ given by a rectangular region. There, identified set Θ_I and the boundary of Θ necessarily have common points, so that (3.8) does not hold. Another example is the case where Θ is itself defined by a manifold, so that (3.8) does not hold. Lemmas 4.1 and 4.2 in Section 4 derive V_∞^δ for the moment condition models of Section 2, covering the cases in which the parameter on the boundary problem does occur.

Condition S.2. *(Weak Sup-Convergence) For any finite set* $\Delta \subset [0, \infty)$, $(\sup_{V_\infty^\delta} \ell_n, \delta \in \Delta) \to_d (\sup_{V_\infty^\delta} \ell_\infty, \delta \in \Delta)$ *in* $\mathbb{R}^{|\Delta|}$, *where* $(\theta, \lambda) \mapsto \ell_\infty(\theta, \lambda)$ *is a non-negative stochastic process.*[17]

The process ℓ_∞ will be referred to as the sup limit of ℓ_n. The sup convergence is more general than uniform convergence, namely the convergence in $L^\infty(\Theta_I \times B_\delta)$, which is implied by finite-dimensional convergence and stochastic equicontinuity of ℓ_n.[18] In particular, the uniform convergence fails in the moment inequality model, while the sup convergence does not; see, for instance, discussion of Example 1 below. The following condition is helpful in verifying sup convergence.

Condition S.3. *(Weak Finite-Dimensional Convergence and Approximability)*
A. (Fidi 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 non-negative stochastic process.*
B. (Fidi 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 imply that the finite-dimensional limit and the sup-limit coincide. Otherwise, the two limits may disagree in general. The finite-dimensional approximability is

[17]Here $|A|$ denotes cardinality of the finite set A.

[18]This notion of sup convergence could be modified to yield what is known as weak hypo-convergence, which may then be used to study the convergence of hypo-graphs of ℓ_n to those of ℓ_∞ as random closed sets, e.g. extending the approach in Molchanov (2005) for the present problem. However, the weak convergence of hypo-graphs is not of interest per se in this paper.

motivated and extends the Knight's (1999) notion of stochastic equi-semi-continuity to set-identified models.

Lemma 3.2. *(1) Condition S.1 and S.2 imply (3.4) with the limit variables given by* $\mathcal{C}(\delta) = \sup_{V_\infty^\delta} \ell_\infty$*, in particular* $\mathcal{C} = \mathcal{C}(0) = \sup_{\theta\in\Theta_I} \ell_\infty(\theta, 0)$*. (2) Condition S.3 implies condition S.2.*

Example 1 (contd.) 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 given by

$$\ell_\infty(\theta, \lambda) = (W_1 - \lambda)_+^2 1(\theta = E_P[Y_1]) + (W_2 - \lambda)_-^2 1(\theta = E_P[Y_2]).$$

The limit is not continuous in θ at $\theta = E_P[Y_1]$ and at $\theta = E_P[Y_2]$, hence $\ell_n(\theta, \lambda)$ can not be stochastically equicontinuous and uniform convergence fails. Suppose that $\Theta_I = [E_P[Y_1], E_P[Y_2]]$ is in the interior of Θ. Then S.1 holds, since $V_n^\delta = V_\infty^\delta = \Theta_I \times B_\delta$ for all sufficiently large n. Also, finite-dimensional approximability S.3(B) can be easily verified. Therefore S.2 holds, so that the finite-dimensional limit $\ell_\infty(\theta, \lambda)$ is also the sup-limit of $\ell_n(\theta, \lambda)$. By Lemma 3.2 we have 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

$$\mathcal{C}_n \to_d \mathcal{C} = \sup_{\theta\in\Theta_I} \ell_\infty(\theta, 0) = \max\left((W_1)_+^2, (W_2)_-^2\right).$$

4. Analysis of Moment Condition Models

4.1. **Moment Equalities.** We begin the discussion with moment equalities. Recall the moment-equality set-up in Section 2, where the identification region takes the form $\Theta_I = \{\theta \in \Theta : E_P[m_i(\theta)] = 0]$. Suppose 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 is a partial identification condition, which states that once θ is bounded away from Θ_I, the moment equations are bounded away from zero.

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 $\|\theta^* - \theta\| > 0$, vector $(\theta - \theta^*)$ is orthogonal to the hyperplane $\{v : E_P[ZX']v = 0\}$. Hence if rank $E_P[ZX']$ is non-zero, we have $\|E_P[ZX'](\theta - \theta^*)\| \geq C \cdot \|\theta - \theta^*\|$, where C is the square root of the minimal positive eigenvalue of $E_P[XZ']E_P[ZX']$.

The main stochastic assumption is that $\{m_i(\theta), \theta \in \Theta'\}$ is a P-Donsker class of functions, where Θ' is a neighborhood of Θ in $\mathbb{R}^d$. By this we mean the following: (1) In the metric space $L^\infty(\Theta')$

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

where $\Delta(\theta)$ is a mean zero Gaussian process with a.s. continuous paths and $Var_P[\Delta(\theta)] > 0$ for each $\theta \in \Theta'$. (2) The probability space $(\Omega, \mathcal{F}, P)$ is rich enough (or has been suitably augmented)[19] so that there exists a map $\Delta_n : \Omega \to L^\infty(\Theta')$ such that $\Delta_n(\theta) =_d \mathbb{G}_n[m_i(\theta)]$ and $\Delta_n(\theta) = \Delta(\theta) + o_p(1)$ in $L^\infty(\Theta')$. [20] The second condition does not involve loss of generality due to the Skorohod-Dudley-Wichura Construction, see Theorem 1.10.4 in van der Vaart and Wellner (1996) or Dudley (1985). Other conditions are given in the following assumption.

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

Most of these assumptions are conventional. We needed them to verify C.1, C.2, C.4, C.5, and other main conditions. Condition M.1(b) is a generalization of the Chernoff (1954) condition, which is needed for the analysis of false coverage, as discussed in Section 3.6, and for the second part of Theorem 4.1 below. M.1(b) also holds trivially in the parameter in the interior case, as defined in Section 3.6, in which case $V_\infty^\delta = \Theta_I \times B_\delta$. M.1(b) can be replaced by

[19]We shall use $(\Omega, \mathcal{F}, P)$ to denote the augmented probability space.

[20]Notation $=_d$ means equality in law: given two maps X and Y that map Ω to a metric space $\mathbb{D}$, $X =_d Y$ if $E_{P^*}[f(X)] = E_{P^*}[f(Y)]$ for every bounded $f : \mathbb{D} \mapsto \mathbb{R}$, where E_{P^*} denotes outer expectation with respect to P, see van der Vaart and Wellner (1996), p. 60.

the classical assumption that Θ is convex, in which case V_∞^δ has a very simple form stated in Lemma 4.1. The convergence imposed in M.1(b), known in variational analysis as a graphical convergence of correspondences, is a fairly weak notion of convergence for correspondences, for instance it is considerably weaker than the uniform convergence $\sup_{\theta\in\Theta_I} d_H(V_n^\delta(\theta), V_\infty^\delta(\theta)) = o(1)$ (Rockafellar and Wets 1998). Lemma 4.1 stated below provides further discussion of M.1(b).

Theorem 4.1 (Moment Equations). *(1) Conditions M.1(a,c,d,e) imply C.1, C.2, C.4, C.5 with $\gamma = 2$, $a_n = n$, and $b_n = \sqrt{n}$. If condition M.1(b) also holds, then S.1-S.3 hold, and the sup-limit of $\ell_n(\theta, \lambda) := nQ_n(\theta + \lambda/\sqrt{n})$ is given by: $\ell_\infty(\theta, \lambda) = \| (\Delta(\theta) + G(\theta)\lambda)' W^{1/2}(\theta)\|^2$. In particular,*

$$\mathcal{C} := \sup_{\theta\in\Theta_I} \ell_\infty(\theta, 0) = \sup_{\theta\in\Theta_I} \|\Delta(\theta)' W^{1/2}(\theta)\|^2 \tag{4.3}$$

where $\Delta(\theta)$ is a zero-mean Gaussian process defined in (4.2).

(2) When $\tilde{Q}_n(\theta) = Q_n(\theta) - \inf_{\theta'\in\Theta} Q_n(\theta')$ is used for inference, condition M.1 implies C.1, C.2, C.4, and C.5, with $\gamma = 2$, $a_n = n$, $b_n = \sqrt{n}$, and the sup-limit of $\tilde{\ell}_n(\theta, \lambda) := nQ_n(\theta + \lambda/\sqrt{n}) - n\inf_{\theta'\in\Theta} Q_n(\theta')$ is given by: $\tilde{\ell}_\infty(\theta, \lambda) = \ell_\infty(\theta, \lambda) - \inf_{(\theta',\lambda')\in V_\infty^\infty} \ell_\infty(\theta', \lambda')$, where $V_\infty^\infty := \lim_{\delta\uparrow\infty} V_\infty^\delta$. In particular,

$$\mathcal{C} := \sup_{\theta\in\Theta_I} \tilde{\ell}_\infty(\theta, 0) = \sup_{\theta\in\Theta_I} \|\Delta(\theta)' W^{1/2}(\theta)\|^2 - \inf_{(\theta,\lambda)\in V_\infty^\infty} \|(\Delta(\theta) + G(\theta)\lambda)' W^{1/2}(\theta)\|^2. \tag{4.4}$$

The most basic implications are that, for $\widehat{c} \to_p c(\alpha)$, the confidence region $C_n(\widehat{c})$ has asymptotic coverage α (it need not be consistent). The estimator $C_n(\widehat{c} + \ln n)$ is consistent at $\sqrt{\ln n/n}$ rate with respect the Hausdorff distance, and has asymptotic coverage of 1. The sup-limit ℓ_∞ of the empirical process ℓ_n obtained by the theorem describes the limit behavior of related inferential statistics, following Section 3.6. Lastly, note that compactness of Θ_I insures that the limit variable $\mathcal{C}$ is finite.

The quantiles of $\mathcal{C}$ in (4.3) can be estimated by the generic subsampling method of Section 3.5 or by simulating the limit distribution. The latter method is generally more accurate than subsampling.

Remark 4.1. (Quantiles of (4.3) by Simulation) For instance, if the data are i.i.d., we can estimate the distribution of $\mathcal{C}$ by making the simulation draws of

$$\mathcal{C}_n^* := \sup_{\theta\in\widehat{\Theta}_I} \mathcal{C}_n^*(\theta), \quad \mathcal{C}_n^*(\theta) := \|\Delta_n^*(\theta)' W_n^{1/2}(\theta)\|^2,$$

where $\Delta_n^*(\theta) = n^{-1/2}\sum_{i=1}^n[m_i(\theta)z_i]$, and $(z_i, i \leq n)$ is a n-vector of i.i.d. $N(0,1)$ variables. Note that $\Delta_n^*(\theta)$ is a zero-mean Gaussian process in $L^\infty(\Theta)$ with covariance function $E_n[m_i(\theta)m_i(\theta')']$. Then $E_n[m_i(\theta)m_i(\theta')'] = E_P[m(\theta)m_i(\theta')'] + o_p(1)$ uniformly in $(\theta,\theta') \in \Theta \times \Theta$. Thus the distance between the law of $\Delta_n^*(\theta)$ and the law of $\Delta(\theta)$, in the weak convergence metric, converges in probability to zero. Since $\Delta_n^*(\theta)$ is stochastically equicontinuous, $d_H(\widehat{\Theta}_I, \Theta_I) = o_p(1)$, and $\sup_{\theta\in\Theta}\|W_n(\theta) - W(\theta)\| = o_p(1)$, the distance between the law of $\mathcal{C}_n^*$ and the law of $\mathcal{C}$, in the weak convergence metric, converges in probability to zero. The same argument applies if the distribution of $\Delta(\theta)$ is estimated by the nonparametric bootstrap with recentering.

Remark 4.2 (Quantiles of (4.4) by Simulation). We can estimate the quantiles of $\mathcal{C}$ in (4.4) by simulating the distribution of the variable

$$\mathcal{C}_n^* := \sup_{\theta\in\widehat{\Theta}_I} \mathcal{C}_n^*(\theta), \quad \mathcal{C}_n^*(\theta) := \|\Delta_n^*(\theta)'W_n^{1/2}(\theta)\|^2 - \inf_{\theta+\frac{\lambda}{\sqrt{n}}\in\Theta} \|(\Delta^*(\theta) + \widehat{G}(\theta)\lambda)'W_n^{1/2}(\theta)\|^2,$$

where $\widehat{G}(\theta)$ is a uniformly consistent estimate of $\nabla_\theta E_P[m_i(\theta)]$.

The form of Θ plays an important role as it determines the limit form of local parameter spaces and statistics $\mathcal{C}_n(\delta)$ which behavior determines the probability of false coverage.

Lemma 4.1 (Chernoff Regularity for Moment Equations). *Sufficient conditions for V_n^δ to converge in the Hausdorff metric to some set V_∞^δ, which is non-decreasing in $\delta \geq 0$, include either one of the following: (1) Suppose there exists $\delta > 0$ such that $B_\delta(\theta) \subset \Theta$ for each $\theta \in \Theta_I$. Then, $V_n^\delta = V_\infty^\delta = \Theta_I \times B_\delta$ for all sufficiently large n. (2) Suppose $\Theta = \Theta_g \cap_{r=1}^R \{\theta \in \mathbb{R}^d : g_r(\theta) = 0\}$, where Θ_g is a compact and convex set, $g_r : \Theta_g' \to \mathbb{R}^{d_r}$ has a continuous Jacobian $\nabla g_r(\theta)$ with a constant row rank over Θ_g', a neighborhood of Θ_g in $\mathbb{R}^d$. Then the above convergence holds with V_∞^δ that has $V_\infty^\delta(\theta) = \{\lambda \in B_\delta : \lambda \in \sqrt{n'}(\Theta_g - \theta) \text{ for some } n' \geq 1, \nabla_\theta g_r(\theta)\lambda = 0, r = 1, ..., R\}$.*

Lemma 4.1 provides the sufficient condition for S.1 to hold. Case (1) is the parameter in the interior case that may arise e.g. when Θ_I is a collection of isolated points that lie in the interior of Θ (defined relative to $\mathbb{R}^d$), in which case V_∞^δ has a trivial form. Case (2) addresses the parameter on the boundary case, and covers the convex parameter space Θ_g as well as the parameter space generated by an intersection of Θ_g with several manifolds representing various restrictions imposed on the parameter space; in this case, the limit local parameter space $V_\infty^\delta(\theta)$

is given (up to a closure) by a tangent cone of Θ at θ intersected with the ball B_δ. This extends the results obtained for the point-identified cases (Chernoff 1954, Geyer 1994, Andrews 1997).

4.2. **Moment Inequalities.** Recall the setup of the moment-inequality model in Section 2. We have $\Theta_I = \{\theta \in \Theta : \|E_P[m_i(\theta)]\|_+ = 0\}$. Assume there are positive constants (C, δ, η) such that for all $\theta \in \Theta$

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

and, recalling that $m_i(\theta)$ is a J-vector with components $(m_{ij}(\theta), j = 1, ..., J)$,

$$\begin{aligned} &\max_j E_P[m_{ij}(\theta)] \leq -C \cdot (d(\theta, \Theta \setminus \Theta_I) \wedge \delta), \text{ for all } \theta \in \Theta_I : \\ &d(\Theta_I^{-\epsilon}, \Theta_I) \leq \epsilon \text{ for all } \epsilon \in [0, \eta]. \end{aligned} \tag{4.6}$$

Equation (4.5) is the partial identification condition. 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. Equation (4.6) needs not hold generally, but it is satisfied in many empirical examples listed in Section 2.[21]

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 (non-empty) subset of $\{1, ..., J\}$ and $\mathcal{J}^c$ is the complement of $\mathcal{J}$ relative to $\{1, ..., J\}$.

Condition M.2. *Suppose the following conditions hold for the moment inequality model of Section 2: (a) Θ is a non-empty 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 data $w_1, ..., w_n$ defined on a complete probability space $(\Omega, \mathcal{F}, P)$, (b) Θ is such that the graph of the local parameter space[22] $V_n^\delta | \theta \in \Theta_{\mathcal{J}}$ converges to some set $V_\infty^\delta | \theta \in \Theta_{\mathcal{J}}$ in the Hausdorff metric, where $V_\infty^\delta | \theta \in \Theta_{\mathcal{J}}$ is non-decreasing in $\delta \geq 0$, for each $\mathcal{J}$, (c) $\{m_i(\theta), \theta \in \Theta'\}$ satisfies P-Donsker condition stated in Section 4.1, (d) $E_P[m_i(\theta)]$ satisfies partial identification condition (4.5) and has continuous Jacobian $G(\theta) = \nabla_\theta E_P[m_i(\theta)]$ for each $\theta \in \Theta'$, (e) $W_n(\theta) = W(\theta) + o_p(1)$ uniformly in $\theta \in \Theta'$, where $W(\theta)$ is a diagonal matrix with positive diagonal elements and is continuous for all $\theta \in \Theta'$, and (f) condition (4.6) holds.*

[21] A detailed illustration and verification of this condition for the linear moment inequality framework has been provided in the previous version of this paper (Chernozhukov, Hong, and Tamer 2002).

[22] The set $V_n^\delta | \theta \in \Theta_{\mathcal{J}}$ is defined as $\{(\theta, \lambda) \in V_n^\delta : \theta \in \Theta_{\mathcal{J}}\}$.

Most of these assumptions are conventional. We need them to verify C.1, C.2, C.4, C.5 and other main conditions. Condition M.2(b) is an assumption of Chernoff type, which is needed for the analysis of false coverage, as discussed in Section 3.6, and for the second part of Theorem 4.2 below. Lemma 4.2 stated below provides further discussion of M.2(b). Condition M.2(b) can be replaced by the classical assumption that Θ is convex, in which case V_∞^δ has a very simple form stated in Lemma 4.1. Condition M.2(f) is needed to verify the "degenerate interior" condition C.3.

Theorem 4.2 (Moment Inequalities). *(1) Conditions M.2(a,c,d,e) imply C.1, C.2, C.4, C.5 with* $\gamma = 2$, $a_n = n$, *and* $b_n = \sqrt{n}$. *If further condition M.2 (f) holds, then C.3 holds. If further condition M.2(b) holds, then S.1-S.3 holds, and the sup-limit of* $\ell_n(\theta, \lambda) := nQ_n(\theta + \lambda/\sqrt{n})$ *is given by:* $\ell_\infty(\theta, \lambda) = \| (\Delta(\theta) + G(\theta)\lambda + \xi(\theta))' W^{1/2}(\theta)\|_+^2$. *In particular,*

$$\mathcal{C} = \sup_{\theta \in \Theta_I} \ell_\infty(\theta, 0) = \sup_{\theta \in \Theta_I} \|(\Delta(\theta) + \xi(\theta))' W^{1/2}(\theta)\|_+^2, \tag{4.7}$$

where $\Delta(\theta)$ *is a zero-mean Gaussian process defined in (4.2) and* $\xi(\theta) = (\xi_j(\theta), j \leq J)$ *with*

$$\xi_j(\theta) = -\infty \text{ if } E_P[m_{ij}(\theta)] < 0 \text{ and } \xi_j(\theta) = 0 \text{ if } E_P[m_{ij}(\theta)] = 0.$$

(2) When $Q_n(\theta) - \inf_{\theta' \in \Theta} Q_n(\theta')$ *is used for inference, conditions M.2(a,b,c,d,e) imply C.1, C.2, C.4, C.5, and S.1-S.3. In particular,* $\gamma = 2$, $a_n = n$, $b_n = \sqrt{n}$, *and the sup-limit of* $\tilde{\ell}_n(\theta, \lambda) := nQ_n(\theta + \lambda/\sqrt{n}) - n \inf_{\theta' \in \Theta} Q_n(\theta')$ *is given by:* $\tilde{\ell}_\infty(\theta, \lambda) = \ell_\infty(\theta, \lambda) - \inf_{(\theta', \lambda') \in V_\infty^\infty} \ell_\infty(\theta', \lambda')$, *where* $V_\infty^\infty := \lim_{\delta \uparrow \infty} V_\infty^\delta$. *In particular* $\mathcal{C} = \sup_{\theta \in \Theta_I} \tilde{\ell}_\infty(\theta, 0)$, *i.e.*

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

where the second term equals zero if M.2(f) holds.

Therefore, for $\widehat{c} \to_p c(\alpha)$, the region $C_n(\widehat{c})$ is consistent at $1/\sqrt{n}$ rate with respect to the Hausdorff distance as an estimator, and has asymptotic coverage α as a confidence region. The theorem also obtains the sup-limit ℓ_∞ of the empirical process ℓ_n, which describes the limit behavior of the related inferential statistics. Following Section 3.6, the latter results are needed to describe the probability of false coverage.

The quantiles of $\mathcal{C}$ in (4.7) can be estimated by either the generic subsampling method of Section 3.5 or simulating the limit distribution. The latter method is generally more accurate than subsampling.

Remark 4.3 (Quantiles of (4.7) by Simulation). If the data are i.i.d., we can simulate the limit distribution of $\mathcal{C}_n$ by making the simulation draws of

$$\mathcal{C}_n^* := \sup_{\theta \in \widehat{\Theta}_I} \mathcal{C}_n^*(\theta), \quad \mathcal{C}_n^*(\theta) := \|(\Delta_n^*(\theta) + \widehat{\xi}(\theta))' W_n^{1/2}(\theta)\|_+^2,$$

where $\Delta_n^*(\theta) = n^{-1/2} \sum_{i=1}^n [m_i(\theta) z_i]$, and $(z_i, i \leq n)$ is a n-vector of i.i.d. $N(0,1)$ variables. Note that $\Delta_n^*(\theta)$ is a zero-mean Gaussian process in $L^\infty(\Theta)$ with covariance function $E_n[m_i(\theta) m_i(\theta')']$, as discussed in Remark 4.1. $\widehat{\xi}(\theta) := (\widehat{\xi}_j(\theta), j = 1, ..., J)'$ with $\widehat{\xi}_j(\theta) := -\infty$ if $E_n[m_{ij}(\theta)] \leq -c_j \log n/\sqrt{n}$, and $\widehat{\xi}_j(\theta) := 0$ if $E_n[m_{ij}(\theta)] > -c_j \log n/\sqrt{n}$, for some positive constants $c_j > 0$.

Remark 4.4 (Quantiles of (4.8) by Simulation). If the data are i.i.d., we can simulate the limit distribution of $\mathcal{C}$ by making the simulation draws of

$$\mathcal{C}_n^* := \sup_{\theta \in \widehat{\Theta}_I} \mathcal{C}_n^*(\theta), \quad \mathcal{C}_n^*(\theta) := \|(\Delta_n^*(\theta) + \widehat{\xi}(\theta))' W_n^{1/2}(\theta)\|_+^2 - \inf_{\theta + \frac{\lambda}{\sqrt{n}} \in \Theta} \|(\Delta_n^*(\theta) + \widehat{G}(\theta)\lambda + \widehat{\xi}(\theta))' W_n^{1/2}(\theta)\|_+^2$$

where $\widehat{G}(\theta)$ is a uniformly consistent estimate of $\nabla_\theta E_P[m_i(\theta)]$.

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

Lemma 4.2 (Chernoff Regularity for Moment Inequalities). *Sufficient conditions for the graph of the local parameter space $V_n^\delta|\theta \in \Theta_{\mathcal{J}}$ to converge in the Hausdorff metric to some set $V_\infty^\delta|\theta \in \Theta_{\mathcal{J}}$ that is non-decreasing in $\delta \geq 0$ include either one of the following: **(1)** Suppose there exists $\delta > 0$ such that $B_\delta(\theta) \subset \Theta$ for each $\theta \in \Theta_I$. Then, $V_n^\delta = V_\infty^\delta = \Theta_I \times B_\delta$ for all sufficiently large n. **(2)** Suppose $\Theta = \Theta_g \cap_{r=1}^R \{\theta \in \mathbb{R}^d : g_r(\theta) = 0\}$, where Θ_g is a compact and convex set, $g_r : \Theta_g' \to \mathbb{R}^{d_r}$ has continuous Jacobian $\nabla g_r(\theta)$ with a constant row rank over Θ_g', a neighborhood of Θ_g in $\mathbb{R}^d$. Then the above convergence holds with V_∞^δ that has $V_\infty^\delta(\theta) = \{\lambda \in B_\delta : \lambda \in \sqrt{n'}(\Theta_g - \theta) \text{ for some } n' \geq 1, \nabla_\theta g_r(\theta)\lambda = 0, r = 1, ..., R\}$.*

Lemma 4.2 is similar to Lemma 4.1 and the comments that are similar to those stated after Lemma 4.1 apply here.

5. Appendix A: Notation

The following standard notation for empirical processes will be used:

$$E_n[f_i] := \frac{1}{n} \sum_{t=1}^n f(w_t), \quad \mathbb{G}_n[f_i] := \frac{1}{\sqrt{n}} \sum_{t=1}^n (f(w_t) - E_P[f(w_i)]).$$

The notions of convergence and outer and inner probabilities, P^* and P_*, are defined as in van der Vaart and Wellner (1996). For instance, $\to_p$ denotes convergence in outer probability, wp $\to 1$ means "with the inner probability approaching 1"; the stochastic order notations $O_p(1)$ and $o_p(1)$ are with respect to P^*, unless otherwise stated. Notation $=_d$ means equality in law: given two elements X and Y that map Ω to a metric space $\mathbb{D}$, $X =_d Y$ if $E_{P^*}[f(X)] = E_{P^*}[f(Y)]$ for every bounded $f : \mathbb{D} \mapsto \mathbb{R}$, where E_{P^*} denotes outer expectation with respect to P. Let $\|x\|_+ = \|\max(x,0)\|$ and $\|x\|_- = \|\max(-x,0)\|$, where in the case of vectors the max operations are elementwise. B_δ denotes a closed ball of diameter δ centered at the origin. In many instances, we use abbreviated notation $\sup_A f$ to mean $\sup_{a\in A} f(a)$, unless an ambiguity arises, in which case the latter notation is used. The Hausdorff distance between sets is defined as

$$d_H(A,B) := \max\left[\sup_{a\in A} d(a,B), \sup_{b\in B} d(b,A)\right], \text{ where } d(b,A) := \inf_{a\in A}\|b-a\|,$$

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 as $\Theta_I^{-\epsilon} := \{\theta\in\Theta_I : d(\theta,\Theta\setminus\Theta_I)\ge\epsilon\}$, where $\epsilon\ge 0$.

6. Appendix B: Proofs

6.1. **Proof of Theorem 3.1:** Proof of Part (1). Step (a). Wp $\to 1$ by C.1(e) and by $\widehat{c}\to_p\infty$, $\sup_{\Theta_I} Q_n = O_p(1/a_n) < \widehat{c}/a_n$, 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 $\epsilon>0$, $\inf_{\Theta\setminus\Theta_I^\epsilon} Q_n =_{(i)} \inf_{\Theta\setminus\Theta_I^\epsilon} Q + o_p(1) \ge_{(ii)} \delta(\epsilon)+o_p(1)$ for some $\delta(\epsilon)>0$, where (i) follows from uniform convergence as assumed in C.1(d) and (ii) from Q being minimized on Θ_I as assumed in C.1(b). Similarly, $\sup_{\widehat{\Theta}_I} Q = \sup_{\widehat{\Theta}_I} Q_n + o_p(1) \le_{(i)} \widehat{c}/a_n + o_p(1) =_{(ii)} o_p(1)$, where (i) holds by construction of $\widehat{\Theta}_I$ and (ii) holds by $\widehat{c}/a_n\to_p 0$. Hence $\sup_{\widehat{\Theta}_I} Q < \delta(\epsilon) = \inf_{\Theta\setminus\Theta_I^\epsilon} Q$ where $\delta(\epsilon)>0$, wp $\to 1$. Hence $\widehat{\Theta}_I\cap(\Theta\setminus\Theta_I^\epsilon)=\emptyset$ wp $\to 1$, which implies $\widehat{\Theta}_I\subseteq\Theta_I^\epsilon$. Given Step (a), this implies $\sup_{\theta\in\widehat{\Theta}} d(\theta,\Theta_I)\le\epsilon$.

Combining Steps (a) and (b), $d_H(\widehat{\Theta}_I,\Theta)\le\epsilon$ wp $\to 1$. Since $\epsilon>0$ is arbitrary, the result is proven. □

Proof of Part (2). For any $\varepsilon>0$ there exist positive constants $(n_\varepsilon,\kappa_\epsilon,\kappa)$ such that for all $n>n_\varepsilon$ we have $\widehat{c}/a_n<\delta$ and $\widehat{c}/a_n>\kappa_\varepsilon/a_n$, by $\widehat{c}/a_n\to_p 0$ and $\widehat{c}\to_p\infty$; so that, with probability larger than $1-\varepsilon$,

$$\inf_{\Theta\setminus\Theta_I^{[\widehat{c}/(a_n\kappa)]^{1/\gamma}}} a_nQ_n(\theta) \ge_{(i)} \kappa\cdot a_n\cdot\left([\widehat{c}/(a_n\kappa)]^{1/\gamma}\wedge\delta\right)^\gamma =_{(ii)} \widehat{c},$$

where (i) follows by C.2 and (ii) follows by $\widehat{c}/a_n\to_p 0$. By construction of $\widehat{\Theta}_I$, we have that $\sup_{\widehat{\Theta}_I} a_nQ_n\le\widehat{c}$. Hence $\widehat{\Theta}_I\subseteq\Theta_I^{[\widehat{c}/(a_n\kappa)]^{1/\gamma}}$. Hence, combining with Step (a) of the Proof of Part (1), we have that $d_H(\widehat{\Theta}_I,\Theta_I)\le[\widehat{c}/(a_n\kappa)]^{1/\gamma}$. Therefore $d_H(\widehat{\Theta}_I,\Theta_I)=O_p([\widehat{c}/a_n]^{1/\gamma})$. □

Proof of Part (3). When $\Theta=\Theta_I$, by Step (a) of Proof of Part (1), $\widehat{\Theta}_I=\Theta$ wp $\to 1$, so $d_H(\widehat{\Theta}_I,\Theta)=0$ wp $\to 1$. □

6.2. **Proof of Theorem 3.2.** . PROOF OF PART (1). Fix any $\epsilon \in (0, \eta]$. It follows that wp $\to 1$, $\Theta_I^{-\epsilon} \subseteq_{(i)} C_n(\widehat{c}') \subseteq_{(ii)} C_n(\widehat{c}) \subseteq_{(iii)} \Theta_I^{\epsilon}$, where $\widehat{c}$ is from Theorem 3.1. Inclusion (i) follows since $\sup_{\Theta_I^{-\epsilon}} a_n Q_n = 0 \leq \widehat{c}'$ wp $\to 1$, by C.3(b), (ii) follows from $\widehat{c} > \widehat{c}'$ wp $\to 1$, and (iii) follows from Part 1 of Theorem 3.1. Since $d_H(\Theta_I^{-\epsilon}, \Theta_I) \leq \epsilon$ by Condition C.3(a) and $d_H(\Theta_I^{\epsilon}, \Theta_I) \leq \epsilon$ by definition of Θ_I^{ϵ}, it follows that $d_H(\Theta_I, C_n(\widehat{c}')) \leq \epsilon$. Part (1) follows. □

PROOF OF PART (2). Let $\epsilon_n = \inf\{\epsilon : \sup_{\Theta_I^{-\epsilon}} a_n Q_n = 0\}$. By Condition C.3(c) ϵ_n exists and $\epsilon_n = O_p(a_n^{-1/\gamma})$. Hence by C.3(a) and C.3(c), $d_H(\Theta^{-\epsilon_n}, \Theta_I) = O_p(a_n^{-1/\gamma})$. Then we have that $\Theta_I^{-\epsilon_n} \subseteq C_n(\widehat{c}') \subseteq C_n(c')$ wp $\to 1$, where $c' > c$ and $\widehat{c}' = c + o_p(1)$. It can be shown, similarly to the Proof of Part (2) of Theorem 3.1 that for any $\varepsilon > 0$, there exist $(\delta_\varepsilon, n_\varepsilon)$ such that for all $n \geq n_\varepsilon$, we have that $C_n(c') \subseteq \Theta_I^{(\delta_\varepsilon/a_n)^{1/\gamma}}$. Conclude that $d_H(C_n(\widehat{c}), \Theta_I) = O_p(a_n^{-1/\gamma})$. □

PROOF OF PART (3). Under the stated condition, it is immediate that $\widehat{\Theta}_I = \Theta_I = \Theta$ wp $\to 1$. Hence $d_H(\widehat{\Theta}_I, \Theta)$ wp $\to 1$. □

6.3. **Proof of Lemma 3.1.** PROOF OF PART (1). Clearly, $\mathcal{C}_n = \sup_{\theta \in \Theta_I} a_n Q_n(\theta) \leq c$ implies $\Theta_I \subseteq C_n(c) = \{\theta \in \Theta : a_n Q_n(\theta) \leq c\}$. Conversely, $\Theta_I \subseteq C_n(c)$ implies $\mathcal{C}_n = \sup_{\theta \in \Theta_I} a_n Q_n(\theta) \leq c$ by compactness of Θ_I. □

PROOF OF PART (2). The result is elementary and its proof is therefore omitted. □

6.4. **Proof of Theorem 3.3.** PROOF OF PART (1). It suffices to prove the result for $\widehat{c}_1$ only. The proof for any subsequent step is identical to this proof, since $\widehat{c}_1$ is allowed to be data-dependent. Step 1 is special to our problem, while Step 2 is standard for subsampling.

STEP 1. By Theorem 3.1 or Theorem 3.2 wp $\to 1$, we have that $\Theta_I^{\eta_n} \subseteq C_n(\widehat{c}) \subseteq \Theta_I^{\epsilon_n}$, where $\epsilon_n := (\ln n/a_n)^{1/\gamma}$ and $\eta_n := -\epsilon_n$, if C.3 holds, and $\eta_n := 0$, if C.3 does not hold. Hence wp $\to 1$,

$$\underline{\mathcal{C}}_{j,b,n} := \sup_{\Theta_I^{\eta_n}} a_b Q_{j,b,n} \leq \widehat{\mathcal{C}}_{j,b,n} := \sup_{C_n(\widehat{c})} a_b Q_{j,b,n} \leq \overline{\mathcal{C}}_{j,b,n} := \sup_{\Theta_I^{\epsilon_n}} a_b Q_{j,b,n}, \qquad \text{for all } j \leq B_n,$$

where index j denotes that the statistic was computed using j-th subsample; total number of subsamples is B_n. Define $\widehat{G}_{b,n}(x) := B_n^{-1} \sum_{j=1}^{B_n} 1\{\widehat{\mathcal{C}}_{j,b,n} \leq x\}$. Hence wp $\to 1$

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

By Step 2 below $\underline{G}_{b,n}(x) \to_p G(x) = P\{\mathcal{C} \leq x\}$ and $\overline{G}_{b,n}(x) \to_p G(x) = P\{\mathcal{C} \leq x\}$, for each $x \geq 0$. This proves that

$$\widehat{G}_{b,n}(x) \to_p G(x) = P\{\mathcal{C} \leq x\} \text{ for each } x \geq 0. \tag{6.1}$$

Convergence of the distribution function at continuity points implies convergence of the quantile function at continuity points. By C.4 $c(\alpha) := G^{-1}(\alpha)$ is continuous in $\alpha \in (0,1)$. Hence, (6.1) implies that $\widehat{c} := \widehat{G}_{b,n}^{-1}(\alpha) \to_p G^{-1}(\alpha)$ for each $\alpha \in (0,1)$.

STEP 2. Define $\underline{\mathcal{C}}_b := \sup_{\Theta_I^{\eta_n}} a_b Q_b$ and $\overline{\mathcal{C}}_b := \sup_{\Theta_I^{\epsilon_n}} a_b Q_b$. Write $\underline{G}_{b,n}(x) \overset{(1)}{=} E_P[\underline{G}_{b,n}(x)] + o_p(1) = P\{\overline{\mathcal{C}}_b \leq x\} + o_p(1) \overset{(2)}{=} P\{\mathcal{C} \leq x\} + o_p(1)$ at each $x \geq 0$. Conclusion (1) follows by $\mathrm{Var}_P(B_n^{-1} \sum_{j=1}^{B_n} 1\{\overline{\mathcal{C}}_{j,b,n} \leq x\}) = o(1)$. For i.i.d. data, this follows from $B_n \to \infty$ and the Hoeffding inequality for bounded U-statistics; for stationary α-mixing series, this follows from $B_n \to \infty$ and an upper bound on covariance given in the proof of Theorem 3.2.1 in Politis, Romano, and Wolf (1999). Conclusion (2) follows by C.5 and C.4 and by $\epsilon_n = o(1/a_b^{1/\gamma})$ and $\eta_n = o(1/a_b^{1/\gamma})$ arising due to restrictions on the subsample size b and the rate a_n stated in conditions (b,c) of this theorem. Likewise, conclude $\overline{G}_{b,n}(x) \to_p G(x) = P\{\mathcal{C} \leq x\}$. □

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

6.5. **Proof of Lemma 3.2.** PROOF OF PART (1). Conditions S.1 and S.2 immediately imply (3.4). □

PROOF OF PART (2). This part shows that S.3 implies 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

$$\limsup_{n\to\infty} P\{\sup_{V_\infty^\delta} \ell_n \leq r\} \leq_{(i)} \limsup_{n\to\infty} P\{\max_{M(\varepsilon)} \ell_n \leq r\} \leq_{(ii)} P\{\max_{M(\varepsilon)} \ell_\infty \leq r\} \leq_{(iii)} P\{\sup_{V_\infty^\delta} \ell_\infty \leq r + \varepsilon\} + \varepsilon,$$

where inequality (i) follows from $\sup_{V_\infty^\delta} \ell_n \geq \sup_{M(\varepsilon)} \ell_n$, (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

$$\begin{aligned}\liminf_{n\to\infty} P\{\sup_{V_\infty^\delta} \ell_n < r\} &\geq_{(i)} \liminf_{n\to\infty} P\{\max_{M(\varepsilon)} \ell_n < r - \varepsilon\} - \varepsilon \\ &\geq_{(ii)} P\{\max_{M(\varepsilon)} \ell_\infty < r - \varepsilon\} - \varepsilon \geq_{(iii)} P\{\sup_{V_\infty^\delta} \ell_\infty < r - \varepsilon\} - \varepsilon,\end{aligned}$$

where inequality (i) follows from the finite-dimensional approximability condition S.3(B), (ii) from finite-dimensional convergence condition S.3(A), and (iii) from $\sup_{V_\infty^\delta} \ell_\infty \geq \sup_{M(\varepsilon)} \ell_\infty$. Since ε is arbitrary, $\liminf_{n\to\infty} P\{\sup_{V_\infty^\delta} \ell_n < r\} \geq P\{\sup_{V_\infty^\delta} \ell_\infty < r\}$. 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 finite set Δ in S.2 follows similarly. □

6.6. **Proof of Theorem 4.1.** PROOF OF PART (1). The proof is organized in the following steps. Step 1 verifies C.1 and C.2. Step 2 gives an auxiliary basic approximation for ℓ_n. Using Step 2, Step 3 verifies C.4, Step 4 verifies C.5, and Step 5 verifies S.1-S.3.

STEP 1. (C.1 and C.2: Uniform Convergence and Quadratic Minorants) Condition C.1 is immediate from condition M.1(a,c,d,e). In particular, uniform convergence and the rates of convergence $a_n = n$ and $b_n = \sqrt{n}$ in 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 C.2 observe that wp $\to$ 1, uniformly in $\theta \in \Theta$

$$\begin{aligned}nQ_n(\theta) &= \|(\mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)])' W_n^{1/2}(\theta)\|^2 \quad \text{by definition} \\ &\geq \zeta \cdot \|\mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)]\|^2 \quad \text{by } \inf_{\theta\in\Theta} \text{mineig } W_n(\theta) \geq \zeta > 0, \text{ wp} \to 1, \text{ by M.1(e)} \\ &\geq \zeta \cdot |\sqrt{n}\|E_P[m_i(\theta)]\| - \|\mathbb{G}_n[m_i(\theta)]\||^2 \text{ by inequality } \|x+y\| \geq |\|y\| - \|x\|| \\ &\geq \zeta \cdot |C \cdot \sqrt{n}(d(\theta, \Theta_I) \wedge \delta) - O_p(1)|^2, \text{ by } \sup_{\theta\in\Theta} \|\mathbb{G}_n[m_i(\theta)]\| = O_p(1) \text{ and M.1(d)},\end{aligned} \tag{6.2}$$

where $\sup_{\theta\in\Theta}\|\mathbb{G}_n[m_i(\theta)]\| = O_p(1)$ follows from P-Donskerness. 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 $1-\varepsilon$

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

This verifies C.2.

STEP 2 (An Auxiliary Expansion). Write $\ell_n(\theta,\lambda) = \|\sqrt{n}E_n[m_i(\theta+\lambda/\sqrt{n})]'W_n^{1/2}(\theta+\lambda/\sqrt{n})\|^2 = \|\left(\mathbb{G}_n[m_i(\theta+\lambda/\sqrt{n})] + \sqrt{n}E_P[m_i(\theta+\lambda/\sqrt{n})]\right)' W_n^{1/2}(\theta+\lambda/\sqrt{n})\|^2$. For any non-empty compact subset K of $\mathbb{R}^d$, we have uniformly in $(\theta,\lambda)\in\Theta\times K$: (1) $\mathbb{G}_n[m_i(\theta+\lambda/\sqrt{n})] = \mathbb{G}_n[m_i(\theta)] + o_p(1)$, by the stochastic equicontinuity arising due to P-Donskerness, (2) $W_n(\theta+\lambda/\sqrt{n}) = W(\theta)+o_p(1)$, by M.1(e), and (3) $\mathbb{G}_n[m_i(\theta)] =_d \Delta(\theta) + o_p(1)$ in $L^\infty(\Theta)$, by P-Donskerness, where $\Delta(\theta)$ is the Gaussian process defined in the statement of the theorem, and (4) $\sqrt{n}E_P[m_i(\theta+\lambda/\sqrt{n})] = G(\theta)\lambda + o(1)$, by M.1(d) and by $E_P[m_i(\theta)] = 0$ for all $\theta\in\Theta_I$. These results imply that

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

Note that $\ell_\infty(\theta,\lambda)$ is stochastically equicontinuous in $L^\infty(\Theta_I\times K)$, because $(\theta,\lambda)\mapsto(\Delta(\theta), G(\theta)\lambda, W(\theta))$ is stochastically equicontinuous in $L^\infty(\Theta_I\times K)$.

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

STEP 4 (C.5: Approximability of $\mathcal{C}_n$). By expansions in Step 2

$$\mathcal{C}_n(\delta_n) = \sup_{\theta\in\Theta_I^{\delta_n/\sqrt{n}}} nQ_n(\theta) =_d \sup_{\theta\in\Theta_I^{\delta_n/\sqrt{n}}}\|\Delta(\theta)'W^{1/2}(\theta)\| + o_p(1) =_d \underbrace{\sup_{\theta\in\Theta_I}\|\Delta(\theta)'W^{1/2}(\theta)\|}_{\mathcal{C}} + o_p(1),$$

where the last equality follows by stochastic equicontinuity of $\theta\mapsto\Delta(\theta)'W^{1/2}(\theta)$. This verifies C.5.

STEP 5 (S.1-S.3: Limits of Related Statistics) This step shows that if M.1(b) holds in addition to M.1(a,c,d,e), then S.1 and S.3 hold. S.3 implies S.2 by Lemma 3.2. M.1(b) states $d_H(V_n^\delta, V_\infty^\delta) = o(1)$. Then, for some $\epsilon_n \downarrow 0$, $|\sup_{V_n^\delta}\ell_n - \sup_{V_\infty^\delta}\ell_n| \leq \sup_{\|(\theta,\lambda)-(\theta',\lambda')\|\leq\epsilon_n}|\ell_n(\theta,\lambda) - \ell_n(\theta',\lambda')| = \sup_{\|(\theta,\lambda)-(\theta',\lambda')\|\leq\epsilon_n}|\ell_\infty(\theta,\lambda) - \ell_\infty(\theta',\lambda') + o_p(1)| + o_p(1) = o_p(1)$ by Step 2 and stochastic equicontinuity of $\ell_\infty(\theta,\lambda)$. This verifies S.1(B). Condition M.1(a) implies S.1(A).

By Step 2, the finite-dimensional limit of $\ell_n(\theta,\lambda)$ equals $\ell_\infty(\theta,\lambda) = \|(\Delta(\theta)+G(\theta)\lambda)'W^{1/2}(\theta)\|^2$. This verifies S.3(A).

Finally, note that by stochastic equicontinuity of $\ell_\infty(\theta,\lambda)$ and Step 2, finite-dimensional approximability condition S.3(B) is trivially satisfied. □

PROOF OF PART (2). The proof is similar to the proof of Part (1), and it is therefore omitted. In particular, we have that $n\tilde{Q}_n(\theta) = nQ_n(\theta) - n\inf_{\theta'\in\Theta}Q_n(\theta')$, 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 $\inf_{(\theta,\lambda)\in V_n^\delta}nQ_n(\theta+\lambda/\sqrt{n}) =_d \inf_{(\theta,\lambda)\in V_\infty^\delta}\ell_\infty(\theta,\lambda) + o_p(1)$. Setting δ arbitrarily large gives 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 $\inf_{(\theta,\lambda)\in V_\infty^\delta} \ell_\infty(\theta,\lambda) \downarrow \inf_{(\theta,\lambda)\in V_\infty^\infty} \ell_\infty(\theta,\lambda) \geq 0$ a.s. □

6.7. **Proof of Lemma 4.1.** PROOF OF PART (1). This part holds trivially. □

PROOF OF PART (2). Consider the simplest case where $\Theta = \Theta_g$ is convex and compact. Define $V_\infty^\delta(\theta) = \{\lambda \in B_\delta : \lambda \in \sqrt{n'}(\Theta_g - \theta) \text{ for some } n'\}$. Define $V_n^\delta = \{(\theta,\lambda) : \theta \in \Theta_I, \lambda \in \sqrt{n}(\Theta_g - \theta)\}$. Note that $V_n^\delta \subseteq V_\infty^\delta$ by convexity of Θ_g and $V_n^\delta \uparrow V_\infty^\delta$ monotonically in the set-theoretic sense. This implies convergence in the Hausdorff distance because V_n^δ and V_∞^δ are subsets of a compact set.

Further, let $V_n^{1\delta}$ denote V_n^δ from the convex case. Define $V_n^\delta := V_n^{1\delta} \cap_{r=1}^R \mathcal{M}_{nr}^\delta$, $\mathcal{M}_{nr}^\delta = \{(\theta,\lambda) : \theta \in \Theta_I, \lambda \in B_\delta, g_r(\theta + \lambda/\sqrt{n}) = 0\}$, $V_\infty^\delta := V_\infty^{1\delta} \cap_{r=1}^R \mathcal{M}_{\infty r}^\delta$, $\mathcal{M}_{\infty r}^\delta := \{(\theta,\lambda) : \theta \in \Theta_I, \lambda \in B_\delta, \nabla_\theta g_r(\theta)\lambda = 0\}$. We have $d_H(V_n^\delta, V_\infty^\delta) \leq d_H(V_n^{1\delta}, V_\infty^{1\delta}) + \sum_{r=1}^R d_H(\mathcal{M}_{nr}^\delta, \mathcal{M}_{\infty r}^\delta) = o(1)$,[23] where the first term is $o(1)$ by the argument for the convex case and the second term is bounded by $\sum_{r=1}^R \sup_{\theta\in\Theta_I} d_H(\mathcal{M}_{nr}^\delta(\theta), \mathcal{M}_{\infty r}^\delta(\theta))$, which is $o(1)$ by an argument similar to that in Lemma 2 in Andrews (1997). □

6.8. **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 6.1 gives another approximation. Using Step 2 and Lemma 6.1, 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 C.1, C.2, and C.3). C.1 is immediate from M.2(a,c,d,e). In particular, uniform convergence and the rates of convergence $a_n = n$ and $b_n = \sqrt{n}$ in C.1 follow from $\{m_i(\theta), \theta \in \Theta\}$ being P-Donsker and $E_P[m_i(\theta)] \leq 0$ on Θ_I. To verify C.2 observe that wp $\to 1$, uniformly in $\theta \in \Theta$

$$\begin{aligned} nQ_n(\theta) &= \|(\mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)])' W_n^{1/2}(\theta)\|_+^2 \quad \text{by definition} \\ &\geq \zeta \cdot \|\mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)]\|_+^2 \\ &\quad \text{by } \inf_{\theta\in\Theta} \text{mineig } W_n(\theta) \geq \zeta > 0 \text{ wp} \to 1, \text{ by M.2(e)} \\ &= \zeta \cdot \|\sqrt{n}E_P[m_i(\theta)]\|_+^2 \cdot (\|\mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)]\|_+^2 / \|\sqrt{n}E_P[m_i(\theta)]\|_+^2). \end{aligned} \tag{6.3}$$

By M.2(d), $\|\sqrt{n}E_P[m_i(\theta)]\|_+^2 \geq C \cdot n \cdot (d(\theta,\Theta_I) \wedge \delta)^2$ 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 $1 - \varepsilon$

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

This follows by (6.3), by $\|y+x\|_+/\|x\|_+ \to 1$ as $\|x\|_+ \to \infty$ for any $y \in \mathbb{R}^J$, and by $\sup_{\theta\in\Theta} \|\mathbb{G}_n[m_i(\theta)]\| = O_p(1)$, where the latter holds by the P-Donsker property. This verifies condition C.2.

[23]This follows by the elementary inequality $d_H(A \cap B, C \cap D) \leq d_H(A \cap B, C \cap B) + d_H(C \cap D, C \cap B) \leq d_H(A,C) + d_H(B,D)$.

To verify C.3 observe that wp $\to 1$, uniformly in $\theta \in \Theta_I$

$$
\begin{aligned}
nQ_n(\theta) &= \|(\mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)])' W_n^{1/2}(\theta)\|_+^2 \quad \text{by definition} \\
&\le \zeta' \cdot \|\mathbb{G}_n[m_i(\theta)] + \sqrt{n}E_P[m_i(\theta)]\|_+^2 \quad \text{by } \sup_{\theta\in\Theta} \text{maxeig } W_n(\theta) \le \zeta' < \infty \text{ wp} \to 1 \text{, by M.2(e)} \\
&\le \zeta' \cdot \sum_{j\le J} |\mathbb{G}_n[m_{ij}(\theta)] + \sqrt{n}E_P[m_{ij}(\theta)]|_+^2, \quad \text{where subscript } j \text{ denotes } j\text{-th element of vector } m_i(\theta) \\
&\le \zeta' \cdot \sum_{j\le J} |O_p(1) - \sqrt{n}\cdot C \cdot (d(\theta, \Theta \setminus \Theta_I) \wedge \delta)|_+^2 \text{ for some } C > 0 \text{ and } \delta > 0 \text{ by M.2(f).}
\end{aligned}
$$

Therefore, for any $\varepsilon > 0$ we can choose κ_ε large enough so that for all $n \ge n_\varepsilon$ with probability at least $1 - \varepsilon$

$$
Q_n(\theta) = 0 \text{ uniformly on } \Theta_I^{-\kappa_\varepsilon/\sqrt{n}} = \{\theta \in \Theta_I : d(\theta, \Theta \setminus \Theta_I) \ge \kappa_\varepsilon / n^{1/2}\}.
$$

This verifies Condition C.3.

STEP 2. (A Basic Approximation). Write $\ell_n(\theta, \lambda) = \|\sqrt{n}E_n[m_i(\theta+\lambda/\sqrt{n})]' W_n^{1/2}(\theta+\lambda/\sqrt{n})\|_+^2 \equiv \|(\mathbb{G}_n[m_i(\theta+\lambda/\sqrt{n})] + \sqrt{n}E_P[m_i(\theta+\lambda/\sqrt{n})])' W_n^{1/2}(\theta+\lambda/\sqrt{n})\|_+^2$. We have for any $\delta > 0$, uniformly in $(\theta, \lambda) \in (\Theta \times B_\delta)$: (1) $\mathbb{G}_n[m_i(\theta+\lambda/\sqrt{n})] = \mathbb{G}_n[m_i(\theta)] + o_p(1)$, by the stochastic equicontinuity implied by the P-Donsker property, (2) $W_n(\theta + \lambda/\sqrt{n}) = W(\theta) + o_p(1)$, by M.2(e), (3) $\mathbb{G}_n[m_i(\theta)] =_d \Delta(\theta) + o_p(1)$ in $L^\infty(\Theta)$, by P-Donskerness property, where $\Delta(\theta)$ is the Gaussian process defined in the statement of the theorem, (4) $\sqrt{n}E_P[m_i(\theta+\lambda/\sqrt{n})] - \sqrt{n}E_P[m_i(\theta)] = G(\theta)\lambda + o(1)$, by M.2(d). Therefore

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

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

Lemma 6.1. *The following approximation is true:*

$$
\begin{aligned}
\sup_{V_n^\delta} \ell_n(\theta, \lambda) &=_d \sup_{V_n^\delta} \|(\Delta(\theta) + G(\theta)\lambda + \sqrt{n}E_P[m_i(\theta)])' W^{1/2}(\theta) + o_p(1)\|_+^2 \\
&\overset{(*)}{=} \sup_{V_\infty^\delta} \|(\Delta(\theta) + G(\theta)\lambda + \xi(\theta))' W^{1/2}(\theta) + o_p(1)\|_+^2 \\
&= \max_{\mathcal{J}} \sup_{V_\infty^\delta | \theta\in\Theta_{\mathcal{J}}} \sum_{j\in\mathcal{J}} |(\Delta_j(\theta) + G_j(\theta)'\lambda) W_{jj}^{1/2}(\theta) + o_p(1)|_+^2,
\end{aligned}
\tag{6.4}
$$

where $\Theta_{\mathcal{J}} := \{\theta \in \Theta_I : E_P[m_{ij}(\theta)] = 0 \ \forall j \in \mathcal{J}, E_P[m_{ij}(\theta)] < 0 \ \forall j \in \mathcal{J}^c\}$, $\mathcal{J}$ *denotes any non-empty subset of* $\{1, ..., J\}$, *and*

$$
\xi_j(\theta) := 0 \textit{ if } E_P[m_{ij}(\theta)] = 0 \textit{ and } \xi_j(\theta) := -\infty \textit{ if } E_P[m_{ij}(\theta)] < 0. \tag{6.5}
$$

The proof of this lemma is given below, immediately after the proof of this theorem.

STEP 3. (C.4: Convergence of $\mathcal{C}_n$) Application of Lemma 6.1 for $V_n^0 = V_\infty^0 = \Theta \times \{0\}$ yields

$$
\begin{aligned}
\mathcal{C}_n = \sup_{\theta\in\Theta_I} nQ_n(\theta) &=_d \sup_{\theta\in\Theta_I} \|(\Delta(\theta) + \xi(\theta))' W^{1/2}(\theta) + o_p(1)\|_+^2 \\
&= \max_{\mathcal{J}} \sup_{\theta\in\Theta_{\mathcal{J}}} \sum_{j\in\mathcal{J}} |\Delta_j(\theta)' W_{jj}^{1/2}(\theta) + o_p(1)|_+^2.
\end{aligned}
\tag{6.6}
$$

Hence $P[\mathcal{C}_n \leq c] \to P[\mathcal{C} \leq c]$ for each $c > 0$, for $\mathcal{C}$ defined in the statement of the theorem. By Theorem 11.1 of Davydov, Lifshits, and Smorodina (1998), non-degeneracy of the covariance function of $\Delta(\theta)$ implies that $\mathcal{C}$ has 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 non-degeneracy implies that $Y = \max_{\mathcal{J}} \max_{j \in \mathcal{J}} \sup_{\theta \in \Theta_{\mathcal{J}}} [\Delta_j(\theta) W_{jj}^{1/2}(\theta)]$ has a continuous distribution function on $\mathbb{R}$. Then by (6.6) $P[\mathcal{C}_n \leq 0]$ is bounded above (below) by $P[Y \leq \epsilon_n]$ with some $\epsilon_n \downarrow 0$ ($\epsilon_n \uparrow 0$), and $P[Y \leq \epsilon_n] \to P[Y \leq 0] = P[\mathcal{C} \leq 0]$. This verifies Condition C.4.

STEP 4. (C.5: Approximability of $\mathcal{C}_n$) By Step 2

$$\sup_{\theta \in \Theta_I^{\delta_n/\sqrt{n}}} nQ_n(\theta) =_d \sup_{\theta \in \Theta_I^{\delta_n/\sqrt{n}}} \|(\Delta(\theta) + \sqrt{n} E_P[m_i(\theta)])' W^{1/2}(\theta) + o_p(1)\|_+^2$$

and by stochastic equicontinuity of $\theta \mapsto (\Delta(\theta), W^{1/2}(\theta))$ and by $\sup_{\|\theta' - \theta\| \leq \delta_n/\sqrt{n}} \|\sqrt{n}(E_P[m_i(\theta)] - E_P[m_i(\theta')])\| = o(1)$ it follows that for any $\delta_n \downarrow 0$ or $\delta_n \uparrow 0$

$$\sup_{\theta \in \Theta_I^{\delta_n/\sqrt{n}}} \|(\Delta(\theta) + \sqrt{n} E_P[m_i(\theta)])' W^{1/2}(\theta) + o_p(1)\|_+^2 = \sup_{\theta \in \Theta_I} \|(\Delta(\theta) + \sqrt{n} E_P[m_i(\theta)])' W^{1/2}(\theta) + o_p(1)\|_+^2.$$

Then it follows as in Step 3 that $P[\mathcal{C}_n(\delta_n) \leq c] \to P[\mathcal{C} \leq c]$ for each $c \geq 0$. This verifies condition C.5.

STEP 5. (Verification of S.1-S.3) S.1(A) follows from M.2(a). S.1(B) and S.2 follow from Lemma 6.1. Further, in equation (6.4), for each $\mathcal{J}$, $\sup_{V_\infty^\delta | \theta \in \Theta_{\mathcal{J}}} \sum_{j \in \mathcal{J}} |(\Delta_j(\theta) + G_j(\theta)'\lambda) W_{jj}^{1/2}(\theta) + o_p(1)|_+^2$ admits finite-dimensional approximation by stochastic equicontinuity of $(\theta, \lambda) \mapsto (\Delta(\theta), G(\theta)\lambda, W(\theta))$ in $L^\infty(\Theta \times B_\delta)$, which implies 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 S.3(A). □

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

6.9. **Proof of Lemma 6.1.** The first equality in (6.4) is immediate by Step 2 of the proof of Theorem 4.2. Equality (*) in (6.4), the main claim of the lemma, is proven as follows. Define

$$\begin{aligned} f_n(\theta, \lambda, x) &:= \|(\Delta(\theta) + G(\theta)\lambda + \sqrt{n} E_P[m_i(\theta)])' W^{1/2}(\theta) + x\|_+^2, \\ g_n(\theta, \lambda, x) &:= \|(\Delta(\theta) + G(\theta)\lambda + \xi(\theta))' W^{1/2}(\theta) + x\|_+^2. \end{aligned} \quad (6.7)$$

STEP 1. Wp $\to 1$, for some $\epsilon_n \downarrow 0$, $g_n(\theta, \lambda, -\epsilon_n) \leq_{(i)} f_n(\theta, \lambda, -\epsilon_n) \leq_{(ii)} \ell_n(\theta, \lambda) \leq_{(iii)} f_n(\theta, \lambda, \epsilon_n)$. Here (i) follows by $\sqrt{n} E_P[m_i(\theta)] \geq \xi(\theta)$ for each $\theta \in \Theta_I$ and by monotonicity: $x_1 \geq x_2$ implies $\|(\Delta(\theta) + G(\theta)\lambda + x_1)' W^{1/2}(\theta)\|_+^2 \geq \|(\Delta(\theta) + G(\theta)\lambda + x_2)' W^{1/2}(\theta)\|_+^2$, 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, wp $\to 1$, for some $\epsilon_n \downarrow 0$

$$\sup_{V_n^\delta} g_n(\theta, \lambda, -\epsilon_n) \leq \sup_{V_n^\delta} \ell_n(\theta, \lambda) \leq \sup_{V_n^\delta} f_n(\theta, \lambda, \epsilon_n).$$

STEP 2. Furthermore, for any $\epsilon_n \uparrow 0$ or $\epsilon_n \downarrow 0$

$$\sup_{V_n^\delta} g_n(\theta, \lambda, \epsilon_n) = \sup_{V_\infty^\delta} g_n(\theta, \lambda, o_p(1)) = \sup_{\overline{V_\infty^\delta}} g_n(\theta, \lambda, o_p(1)), \quad (6.8)$$

where $\overline{V^\delta_\infty}$ is the closure of V^δ_∞. To show this, write

$$\sup_{V^\delta_n} g_n(\theta,\lambda,\epsilon_n) = \max_{\mathcal{J}} \sup_{V^\delta_n|\theta\in\Theta_\mathcal{J}} \sum_{j\in\mathcal{J}} |(\Delta_j(\theta)+G_j(\theta)'\lambda)W^{1/2}_{jj}(\theta)+\epsilon_n|^2_+.$$

Then by M.2(b) for every $\mathcal{J}$, $d_H(V^\delta_n|\theta\in\Theta_\mathcal{J}, V^\delta_\infty|\theta\in\Theta_\mathcal{J}) = o(1)$, which implies by stochastic equicontinuity of $(\theta,\lambda)\mapsto(\Delta(\theta),G(\theta)\lambda,W(\theta))$ in $L^\infty(\Theta\times B_\delta)$ that

$$\begin{aligned}
&\sup_{V^\delta_n|\theta\in\Theta_\mathcal{J}} \sum_{j\in\mathcal{J}} |(\Delta_j(\theta)+G_j(\theta)'\lambda)W^{1/2}_{jj}(\theta)+\epsilon_n|^2_+ \\
&\quad = \sup_{V^\delta_\infty|\theta\in\Theta_\mathcal{J}} \sum_{j\in\mathcal{J}} |(\Delta_j(\theta)+G_j(\theta)'\lambda)W^{1/2}_{jj}(\theta)+o_p(1)|^2_+ \\
&\quad = \sup_{\overline{V^\delta_\infty}|\theta\in\Theta_\mathcal{J}} \sum_{j\in\mathcal{J}} |(\Delta_j(\theta)+G_j(\theta)'\lambda)W^{1/2}_{jj}(\theta)+o_p(1)|^2_+,
\end{aligned}$$

so relation (6.8) follows.

STEP 3. This step shows that for some $\epsilon'_n \downarrow 0$,

$$\sup_{V^\delta_n} f_n(\theta,\lambda,\epsilon_n) \le \sup_{V^\delta_\infty} g_n(\theta,\lambda,\epsilon'_n) \quad \text{wp} \to 1. \tag{6.9}$$

Observe that for any $\theta_n\in\Theta_I$ converging to $\theta\in\Theta_I$,

$$\limsup_n \sqrt{n}E_P[m_{ij}(\theta_n)] \le \xi_j(\theta) \text{ if } \xi_j(\theta)=0, \quad \limsup_n \sqrt{n}E_P[m_{ij}(\theta_n)] = \xi_j(\theta) \text{ if } \xi_j(\theta)=-\infty. \tag{6.10}$$

Let $\Omega_{n,\varepsilon} = \{\omega\in\Omega : \sup_{\theta\in\Theta_I,\lambda\in B_\delta}\|\Delta(\theta)\| \le K_\varepsilon\}$. For any $\varepsilon>0$, there exists K_ε such that $P(\Omega_{n,\varepsilon}) \ge 1-\varepsilon$ for all $n\ge n_\varepsilon$. Suppose that relation (6.9) 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

$$\lim_k [f_{n(k)}(\theta_{n(k)},\lambda_{n(k)},\epsilon_{n(k)}) - \sup_{V^\delta_\infty} g_{n(k)}(\theta,\lambda,\epsilon)](\omega_{n(k)}) > 0. \tag{6.11}$$

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 $d_H(V^\delta_n,V^\delta_\infty)\to 0$ and by $V^\delta_\infty\subseteq\Theta_I\times B_\delta$. As in Step 2 conclude that

$$\sup_{V^\delta_\infty} g_n(\theta,\lambda,\epsilon) = \sup_{\overline{V^\delta_\infty}} g_n(\theta,\lambda,\epsilon) \ge g_n(\theta^*,\lambda^*,\epsilon/2) \text{ wp} \to 1\ ,$$

which together with (6.11) gives that $\lim_l[f_{n(k(l))}(\theta_{n(k(l))},\lambda_{n(k(l))},0) - g_{n(k(l))}(\theta^*,\lambda^*,\epsilon/2)](\omega_{n(k(l))}) > 0$. Given the definition of f_n and g_n stated in (6.7), this inequality can occur 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 (6.10). Therefore, the claim of Step 3 is correct.

Combining Steps 1, 2, and 3 implies the result of the lemma. □

6.10. **Proof of Lemma 4.2.** Define $V^\delta_{n,\mathcal{J}} := V^\delta_n|\theta\in\Theta_\mathcal{J}$. Apply the proof of Lemma 4.1 for V^δ_n to $V^\delta_{n,\mathcal{J}}$. □

Suppose that one is interested in a particular parameter θ^* inside Θ_I. The inference about some θ^* in Θ_I is well motivated, when there is a sense in which θ^* is the true parameter. The latter is typically the case when it is maintained that the economic models are correct representations of data-generating processes of real data for some parameter value θ^*.[24] In this scenario, Θ_I is not of interest per se, but rather θ^* is.

In order to facilitate inference about θ^* we make the following assumption.

Condition C.6. *Suppose there exists $a_n \to \infty$ such that, for $\mathcal{C}_n(\theta) := a_n Q_n(\theta)$, $P(\mathcal{C}_n(\theta) \leq c) \to P(\mathcal{C}(\theta) \leq c)$ for each $c \geq 0$ 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.*

Using the fact that $Q(\theta) = 0$ at $\theta = \theta^*$, we construct a confidence region for θ^* as follows. We test whether whether $Q(\theta) = 0$ for each $\theta \in \Theta$. Then we collect all $\theta \in \Theta$ that pass the test to form a confidence region for θ^*. More precisely, we collect all $\theta \in \Theta$ such that $a_n Q_n(\theta) \leq c(\alpha,\theta)$.

Towards the construction of confidence regions, suppose the estimate $\widehat{c}(\theta)$ is available such that $\widehat{c}(\theta) \to_p c(\alpha,\theta)$ for each $\theta \in \Theta_I$. Consistent estimates $\widehat{c}(\theta)$ can be obtained by subsampling or, for the moment condition models, through the use of the limit distributions obtained in Theorem 7.3. Consistency of the subsampling estimate $\widehat{c}(\theta)$ follows by the standard argument, e.g. the one given in Step 2 of the Proof of Theorem 3.3. It should be noted that subsampling is generally less accurate than the use of the limit distributions.

Let $\widehat{\Theta}_I$ be an estimator of Θ_I so that $\Theta_I \subseteq \widehat{\Theta}_I$ wp $\to 1$. We also want $\widehat{\Theta}_I$ to be a sharp estimate, for instance, we can set $\widehat{\Theta}_I = C_n(\log n)$, which under C.1 and C.2 is consistent and converges at rate $(\log n/n)^{1/\gamma}$. Let also $\widehat{c}$ be any consistent estimate of the α-quantile of $\mathcal{C}$ defined in C.4. Recall that we used $\widehat{c}$ for the construction of the region-wise critical value.

The following two regions will be considered. The first region is a simple region defined by a single critical value:

$$C_n(\widehat{c}^* \wedge \widehat{c}) = \{\theta \in \Theta : a_n Q_n(\theta) \leq \widehat{c}^* \wedge \widehat{c}\}, \text{ where } \widehat{c}^* = \sup_{\theta \in \widehat{\Theta}_I} \widehat{c}(\theta). \tag{7.1}$$

The second regions is a region that employs critical values that depend on θ:

$$C_n(\widehat{c}(\cdot) \wedge \widehat{c}) = \{\theta \in \Theta : a_n Q_n(\theta) \leq \widehat{c}(\theta) \wedge \widehat{c}\}. \tag{7.2}$$

Remark 7.1. The two constructions are equivalent in many cases, since the objective functions can be transformed to have equal quantiles.[25] The fist construction is more parsimonious, easier to compute, and report. Clearly, either region is a subset of, and hence is no larger than, the confidence region $C_n(\widehat{c})$ for Θ_I.

[24]There is $\theta^* \in \Theta$ such that the model law P_θ agrees with the actual law of data P.

[25]This can be seen by defining the new criterion function $\widetilde{Q}_n(\theta) := Q_n(\theta)/\max[c(\alpha,\theta),\epsilon]$ for all $\theta \in \Theta_I$. In many examples this is unnecessary, as criterion functions have the equi-quantile property by using optimal weights.

Remark 7.2. If $\widehat{c}(\theta)$ is obtained by subsampling, the truncation of critical values by $\widehat{c}$ improves Bahadur efficiency: Indeed, if $\theta \notin \Theta_I$, we have that $\widehat{c}(\theta) \to_p +\infty$, typically at the rate a_b, but $\widehat{c}(\theta) \wedge \widehat{c} \to_p c(\alpha) < \infty$. Therefore, subsampling implementations that, in contrast to our construction, do not truncate $\widehat{c}(\theta)$ by $\widehat{c}$ suffer from the loss of power in finite-samples. Chernozhukov and Fernandez-Val (2005) show, in a different situation, that the Bahadur inefficiency of canonical (untruncated) subsampling leads to a substantial loss of power in finite samples.

Remark 7.3. The construction of either region employs the pointwise inversions of tests of point hypotheses $Q(\theta) = 0$. This follows the Anderson and Rubin (1949) construction of confidence regions for the case of simultaneous equations. In the case of weakly identified and unidentified linear instrumental variable models, the construction was used by Dufour (1997) and Staiger and Stock (1997), among others. In a partially identified dynamic censored regression model, Hu (2002) also employed region (7.2) for inference. In partially identified instrumental variable quantile regression model, Chernozhukov and Hansen (2004) also use the region (7.2). A previous version of the paper, Chernozhukov, Hong, and Tamer (2002), Appendix G, also gave pointwise constructions. Imbens and Manski (2004) investigate the Wald type inference about θ^* for the special case where θ^* is a real parameter known to belong to an interval which endpoints can be consistently estimated. The analysis here applies to a considerably more general setting.

Remark 7.4. The more recent developments in the literature include Andrews and Guggenberger (2006) and Sheikh (2006) who show that the confidence regions of the type proposed here, with critical values obtained by subsampling, have important robustness (uniform coverage) properties. Note, however, that in moment condition models, we can construct the critical values using limit distributions, e.g. see Remark 7.6, which should be preferable to subsampling due to higher accuracy.

Remark 7.5. Due to reasons given in Remark 7.2, our regions (7.2) constructed using subsampling will be less conservative than the regions studied by Andrews and Guggenberger (2006) and Sheikh (2006). The latter are constructed using canonical (untruncated) subsampling critical value $\widehat{c}(\theta)$. In contrast, regions (7.2) use the truncated critical value $\widehat{c}(\theta) \wedge \widehat{c}$.

Theorem 7.1. *Suppose that (a) Conditions C.4 and C.6 hold, and (b) for each $\theta \in \Theta_I$ we have $\widehat{c}(\theta) \to_p c(\alpha, \theta)$ and $\widehat{c} \to_p c(\alpha) \geq \sup_{\theta \in \Theta_I} c(\alpha, \theta)$, where $\widehat{c}(\theta) \geq 0$ and $\widehat{c} \geq 0$ with probability 1. Then, (1) for any $\theta^* \in \Theta_I$, $\liminf_{n\to\infty} P\{\theta^* \in C_n(\widehat{c}^* \wedge \widehat{c})\} \geq \alpha$, and (2) $\liminf_{n\to\infty} P\{\theta^* \in C_n(\widehat{c}(\cdot) \wedge \widehat{c})\} \geq \alpha$.*

Proof of Theorem 7.1: Part **(1)**: $\liminf_{n\to\infty} P\{\theta^* \in C_n(\widehat{c}^* \wedge \widehat{c})\} = \liminf_{n\to\infty} P\{a_n Q_n(\theta^*) \leq \widehat{c}^* \wedge \widehat{c}\} \geq_{(i)} \liminf_{n\to\infty} P\{a_n Q_n(\theta^*) \leq \widehat{c}(\theta^*) \wedge \widehat{c}\} \geq_{(ii)} \liminf_{n\to\infty} P\{a_n Q_n(\theta^*) \leq (c(\theta^*) + o_p(1)) \vee 0\}$ $\geq_{(iii)} P\{\mathcal{C}(\theta) \leq c(\theta^*)\} =_{(iv)} \geq \alpha$, where (i) follows by construction, (ii) follows by the assumptions on $\widehat{c}$ and $\widehat{c}(\theta)$, (iii) follows by Condition C.6, and (iv) follows by Condition C.6. Part **(2)**: This part trivially follows from inequality (i) in the proof of Part **(1)**. □

The following theorem provides consistency and rates of convergence of the sets constructed above.

Theorem 7.2 (Consistency and rates of convergence). *Suppose C.1, C.2, and conditions of Theorem 7.1 hold. Consider estimators $\widehat{\Theta}_I := \{\theta \in \Theta : a_n Q_n(\theta) \leq \widehat{c}(\theta) \wedge \widehat{c} + \kappa_n\}$ and $\widetilde{\Theta}_I := \{\theta \in \Theta :$*

$a_nQ_n(\theta) \leq \widehat{c}^* \wedge \widehat{c} + \kappa_n\}$, where $\kappa_n := 0$ when C.3 is known to hold, and $\kappa_n := \log n$ otherwise. Then, $d_H(\widehat{\Theta}_I, \Theta_I) = O_p([1/a_n]^{1/\gamma}) = o_p(1)$ if C.3 is known to hold, and $d_H(\widehat{\Theta}_I, \Theta_I) = O_p([\log n/a_n]^{1/\gamma}) = o_p(1)$, otherwise. The same results apply to $\widetilde{\Theta}_I$.

PROOF OF THEOREM 7.2: Since $C_n(\kappa_n) \subseteq \widehat{\Theta}_I \subseteq \widetilde{\Theta}_I \subseteq C_n(\widehat{c} + \kappa_n)$, the rate and consistency results follows from the rates and consistency results for $C_n(\kappa_n)$ and $C_n(\widehat{c} + \kappa_n)$ obtained in Theorem 3.1 and Theorem 3.2. □

Theorem 7.3. *(Limits of $C_n(\theta)$ in Moment Condition Models) (1) Suppose Condition M.1 holds for the moment equality model. In particular, the P-Donsker condition on moment functions implies $n^{-1/2}\left(\sum_{t=1}^n (m_i(\theta) - E_P[m_i(\theta)])\right) \to_d \Delta(\theta) = N(0, E_P[\Delta(\theta)\Delta(\theta)'])$. Then, Condition C.6 holds with*

$$\mathcal{C}(\theta) := \|\Delta(\theta)'W^{1/2}(\theta)\|^2, \quad (7.3)$$

$$\mathcal{C}(\theta) := \|\Delta(\theta)'W^{1/2}(\theta)\|^2 - \inf_{(\theta',\lambda)\in V_\infty^\infty} \|(\Delta(\theta') + G(\theta')\lambda)'W^{1/2}(\theta')\|^2, \quad (7.4)$$

for the case when $Q_n(\theta)$ and $\widetilde{Q}_n(\theta) = Q_n(\theta) - \inf_{\theta'\in\Theta} Q_n(\theta')$ are used for inference, respectively.
(2) Suppose Condition M.2 holds for the moment inequality model. Then, Condition C.6 holds with

$$\mathcal{C}(\theta) := \|(\Delta(\theta) + \xi(\theta))'W^{1/2}(\theta)\|_+^2, \quad (7.5)$$

$$\mathcal{C}(\theta) := \|(\Delta(\theta) + \xi(\theta))'W^{1/2}(\theta)\|_+^2 - \inf_{(\theta',\lambda)\in V_\infty^\infty} \|(\Delta(\theta') + G(\theta')\lambda + \xi(\theta))'W^{1/2}(\theta')\|_+^2, \quad (7.6)$$

for the case when $Q_n(\theta)$ and $\widetilde{Q}_n(\theta) = Q_n(\theta) - \inf_{\theta'\in\Theta} Q_n(\theta')$ are used for inference, respectively.

PROOF OF THEOREM 7.3. Part (1) follows from the proof of Theorem 4.1. Part (2) follows from the proof of Theorem 4.2. □

Remark 7.6. (Quantiles of $\mathcal{C}(\theta)$ by Simulation) The quantiles of $\mathcal{C}(\theta)$, specified in (7.3), (7.4), (7.5), and (7.6), can be obtained by simulating variable $\mathcal{C}_n^*(\theta)$ specified respectively in Remarks 4.1, 4.2, 4.3, and 4.4. □

REFERENCES

AMEMIYA, T. (1985): *Advanced Econometrics.* Harvard University Press.

ANDERSON, T. W., AND H. RUBIN (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Ann. Math. Statistics*, 20, 46–63.

ANDREWS, D. W. K. (1997): "Estimation when a parameter is on a boundary: Theory and Applications," Cowles Foundation Working Paper, Yale University.

——— (1999): "Estimation when a parameter is on a boundary," *Econometrica*, 67(6), 1341–1383.

——— (2001): "Testing when a parameter is on the boundary of the maintained hypothesis," *Econometrica*, 69(3), 683–734.

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.

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

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

BAJARI, P., L. BENKARD, AND J. LEVIN (2006): "Estimating Dynamic Models of Imperfect Competition," Working Paper, University of Minnesotta, forthcoming in Econometrica.

BAJARI, P., J. FOX, AND S. RYAN (2006): "Evaluating Wireless Carrier Consolidation Using Semiparametric Demand Estimation," Working Paper, MIT.

BERESTEANU, A., AND B. ELLICKSON (2006): "Dynamics of the Retail Olygopoly," Working Paper, Duke University.

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

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

BORZEKOWSKI, R., AND A. COHEN (2004): "Estimating Strategic Complementarities in Credit Unions' Outsourcing Decisions," Working Paper, New York Federal Reserve Bank.

CHERNOFF, H. (1954): "On the distribution of the likelihood ratio," *Ann. Math. Statistics*, 25, 573–578.

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

CHERNOZHUKOV, V., AND C. HANSEN (2004): "Instrumental Quantile Regression," Working Paper, MIT.

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

CHERNOZHUKOV, V., C. HANSEN, AND M. JANSSON (2005): "Finite-Sample Inference for Quantile Regression Models," Wroking Paper, MIT.

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

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

COHEN, A., AND M. MANUSZAK (2006): "Endogenous Market Structure with Discrete Product Differentiation and Multiple Equilibria: An Empirical Analysis of Competition Between Banks and Thrifts," Working Paper, New York Federal Reserve Bank.

CUEVAS, A., AND R. FRAIMAN (1997): "A plug-in approach to support estimation," *Ann. Statist.*, 25(6), 2300–2312.

DACUNHA-CASTELLE, D., AND E. GASSIAT (1999): "Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes," *Ann. Statist.*, 27(4), 1178–1209.

DAVYDOV, Y. A., M. A. LIFSHITS, AND N. V. SMORODINA (1998): *Local properties of distributions of stochastic functionals*. American Mathematical Society, Providence, RI.

DEMIDENKO, E. (2000): "Is this the least squares estimate?," *Biometrika*, 87(2), 437–452.

DUDLEY, R. M. (1985): "An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions," in *Probability in Banach spaces, V (Medford, Mass., 1984)*, vol. 1153 of *Lecture Notes in Math.*, pp. 141–178. Springer, Berlin.

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

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

FUKUMIZU, K. (2003): "Likelihood ratio of unidentifiable models and multilayer neural networks," *Ann. Statist.*, 31(3), 833–851.

GEYER, C. J. (1994): "On the asymptotics of constrained M-estimation," *Ann. Statist.*, 22(4), 1993–2010.

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

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

HANNAN, E. J. (1982): "Testing for autocorrelation and Akaike's criterion," *J. Appl. Probab.*, (Special Vol. 19A), 403–412, Essays in statistical science.

HANNAN, E. J., AND M. DEISTLER (1988): *The statistical theory of linear systems*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York.

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

HU, L. (2002): "Estimation of a censored dynamic panel data model," *Econometrica*, 70(6), 2499–2517.

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

KIM, J., AND D. POLLARD (1990): "Cube root asymptotics," *Ann. Statist.*, 18(1), 191–219.

KLEPPER, S., AND E. LEAMER (1984): "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica*, 52(1), 163–184.

KNIGHT, K. (1999): "Epi-convergence and Stochastic Equisemicontinuity," Technical Report, University of Toronto, Department of Statistics (http://www.utstat.toronto.edu/keith/papers/).

KOROSTELËV, A. P., L. SIMAR, AND A. B. TSYBAKOV (1995): "Efficient estimation of monotone boundaries," *Ann. Statist.*, 23(2), 476–489.

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

LIU, X., AND Y. SHAO (2003): "Asymptotics for likelihood ratio tests under loss of identifiability," *Ann. Statist.*, 31(3), 807–832.

MANSKI, C. F. (2003): *Partial identification of probability distributions*, Springer Series in Statistics. Springer-Verlag, New York.

MANSKI, C. F., AND E. TAMER (2002): "Inference on regressions with interval data on a regressor or outcome," *Econometrica*, 70(2), 519–546.

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

MCFADDEN, D. L. (2005): "Revealed stochastic preference: a synthesis," *Economic Theory*, 26, 245–264.

MOLCHANOV, I. S. (2005): *Theory of Random Sets.* Springer, London.

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

NEWEY, W. K., AND D. MCFADDEN (1994): "Large sample estimation and hypothesis testing," in *Handbook of econometrics, Vol. IV*, vol. 2 of *Handbooks in Econom.*, pp. 2111–2245. North-Holland, Amsterdam.

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

PHILLIPS, P. C. B. (1989): "Partially identified econometric models," *Econometric Theory*, 5(2), 181–240.

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

PÖTSCHER, B. M. (1991): "Effects of model selection on inference," *Econometric Theory*, 7, 163 – 185.

REDNER, R. (1981): "Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions," *Ann. Statist.*, 9(1), 225–228.

ROCKAFELLAR, R. T., AND R. J.-B. WETS (1998): *Variational analysis.* Springer-Verlag, Berlin.

Romano, J. P., and M. Wolf (2000): "Finite sample nonparametric inference and large sample efficiency," *Ann. Statist.*, 28(3), 756–778.

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

Sheikh, A. (2006): "Inference for Partially Identified Econometric Models," Working Paper, Stanford University.

Staiger, D., and J. H. Stock (1997): "Instrumental variables regression with weak instruments," *Econometrica*, 65(3), 557–586.

van der Vaart, A. W. (1998): *Asymptotic statistics*. Cambridge University Press, Cambridge.

van der Vaart, A. W., and J. A. Wellner (1996): *Weak convergence and empirical processes*. Springer-Verlag, New York.

Varian, H. R. (1982): "The nonparametric approach to demand analysis," *Econometrica*, 50(4), 945–973.

——— (1984): "The nonparametric approach to production analysis," *Econometrica*, 52(3), 579–597.

Veres, S. (1987): "Asymptotic distributions of likelihood ratios for overparametrized ARMA processes," *J. Time Ser. Anal.*, 8(3), 345–357.

ADDENDUM FOR "ESTIMATION AND CONFIDENCE REGIONS FOR PARAMETER SETS IN ECONOMETRIC MODELS"

8. ROBUSTNESS TO CONTIGUOUS PERTURBATIONS OF P

In this paper P, the true probability measure, is the nuisance parameter. The goal is to examine which contiguous perturbations of the original fixed P preserve or do not preserve the estimation and coverage properties of the confidence regions. The idea of focusing on the local perturbations follows its uses in the confidence interval literature, see notably Dufour (1997), Pötscher (1991), and Andrews and Guggenberger (2006). Intuitively, contiguous perturbations of P can not be statistically detected with certainty, and we therefore want to make sure that contiguous changes in P do not affect the coverage properties of confidence regions. An alternative motivation is that, in the asymptotic context, the relevant parameter space for nuisance parameters consists of contiguous parameter values, which is a standard approach in asymptotic efficiency analysis, see van der Vaart (1998), Chapter 8.7. In fact, finding minimal coverage under contiguous sequences is equivalent to establishing local uniform coverage, when the local nuisance parameters are allowed to vary over a compact set.[26]

We focus on examining the robustness of the main estimation and inferential results, the ones stated in Theorem 3.1 and Theorem 3.3.

8.1. **Regular Cases.** Consider a triangular sequence of probability measures $\{P_{n,\gamma}, n = 1, 2, ...\}$, where γ is an index of a sequence in Γ and $\{P_{n,\gamma}, \gamma \in \Gamma, n = 1, ...\} \subseteq \mathcal{P}$. Let $P^n_{n,\gamma}$ denote the law of data $w_1, ..., w_n$ under $P_{n,\gamma}$. Each $\gamma \in \Gamma$ is such that $P^n_{n,\gamma}$ is contiguous to P^n, the law of data $w_1, ..., w_n$ under P, namely $P^n(A_n) = o(1)$ implies $P^n_{n,\gamma}(A_n) = o(1)$ for any sequence of measurable events A_n.[27] In what follows, notation $\Theta_I(P)$ is used to reflect that identification region Θ_I depends on the law of the data P. Similarly, notation $c(\alpha, P)$ is used to denote that the α-quantile of $\mathcal{C}$ depends on P.

Lemma 8.1. *[Conditions for Maintaining Consistency, Rates of Convergence, and Coverage] (1) Assume that Conditions C.1 and C.2 hold with $\{P_{n,\gamma}\}$ replacing $\{P\}$, for each $\gamma \in \Gamma$. Then so do conclusions of Theorem 3.1. (2) Assume that Conditions C.1, C.2, and C.4 hold under $\{P_{n,\gamma}\}$ in place of $\{P\}$, for any $\gamma \in \Gamma$, as well as hold under $\{P\}$, with the common limit real random variable $\mathcal{C}$, distribution of which does not depend on γ. Take any estimate $\widehat{c} \to_p c(\alpha, P)$ under $\{P\}$, for instance, that provided in Sections 3 or 4. Then for each $\gamma \in \Gamma$,*

$$\liminf_{n\to\infty} P_{n,\gamma}\{\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})\} \geq \alpha \textit{ and } = \alpha \textit{ if } P\{\mathcal{C} > 0\} \geq \alpha.$$

The first result states that consistency and rates of convergence will be preserved under sequences as long as C.1 and C.2 hold under sequences (replacing P with $P_{n,\gamma}$ and Θ_I with $\Theta_I(P_{n,\gamma})$ should cause no ambiguity in the re-statement of C.1 and C.2). The second result of the lemma addresses

[26]The weak IV example presented below clarifies this statement; see, specifically, equations (8.8)- (8.9).

[27]Throughout this section, measurable events A_n are events that are measurable with respect to $(\Omega, \mathcal{F})$ completed with respect to both P^n and $P^n_{n,\gamma}$.

coverage properties in the *regular case* – when the limit of $\mathcal{C}_n$ does *not* depend on the local sequence.[28] Note that the coverage result is *independent* of the way the critical value is estimated.

Note that if $\mathcal{C}_n$ is *non-regular* – that is, its limit distribution under $\{P_{n,\gamma}\}$ depends on γ – the coverage under sequence depends on whether the distribution of $\mathcal{C}_n$ under $P_{n,\gamma}$ is stochastically dominated in large samples by the distribution under fixed sequence $\{P\}$, as stated in Lemma 8.3 below.

Conditions of Lemma 8.1 are verified in our principal applications as follows:

Condition M.3. *(Moment Equalities) Suppose that M.1 holds for each $P \in \mathcal{P}$ and that (a) the partial identification conditions (4.1) holds uniformly in $\mathcal{P}$, (b) $G(\theta) = \lim_n \nabla_\theta E_{P_{\gamma,n}}[m_i(\theta)]$ exists and is continuous over a neighborhood of Θ, for each $\gamma \in \Gamma$, (c) the Donsker condition (4.2) holds under $\{P_{n,\gamma}\}$ in place of $\{P\}$ for each $\gamma \in \Gamma$, with the common limit Gaussian process $\Delta(\theta)$, (d) $E_{P_{n,\gamma}}[m_i(\theta)] = E_P[m_i(\theta)] + o(1)$ for each $\gamma \in \Gamma$, (e) $d_H(\Theta_I(P_{n,\gamma}), \Theta_I(P)) = o(1)$ for each $\gamma \in \Gamma$.*

Condition M.4. *(Moment Inequalities) Suppose that M.1 holds for each $P \in \mathcal{P}$ and that (a) the partial identification conditions (4.5) holds uniformly in $\mathcal{P}$, (b) $G(\theta) = \lim_n \nabla_\theta E_{P_{\gamma,n}}[m_i(\theta)]$ exists and is continuous over a neighborhood of Θ, for each $\gamma \in \Gamma$, (c) the Donsker condition (4.2) holds under $\{P_{n,\gamma}\}$ in place of $\{P\}$ for each $\gamma \in \Gamma$, with the common limit Gaussian process $\Delta(\theta)$, (d) $E_{P_{n,\gamma}}[m_i(\theta)] = E_P[m_i(\theta)] + o(1)$ for each $\gamma \in \Gamma$, (e) and $d_H(\Theta_I(P_{n,\gamma}), \Theta_I(P)) = o(1)$ and $d_H(\Theta_{\mathcal{J}}(P_{n,\gamma}), \Theta_{\mathcal{J}}(P)) = o(1)$ for each $\mathcal{J}$ and each $\gamma \in \Gamma$.*

Condition (a) is a locally uniform partial identification condition. Sufficient condition for condition (c) are well known and are given in van der Vaart and Wellner (1996), p.173, including a quadratic-mean-differentiability condition, p. 406. The principal condition is Condition (e), which requires that the perturbations of P affect the identification region smoothly.

Lemma 8.2. *(Coverage, Consistency, Rates under Regular Sequences in Moment Condition Models) (1) Condition M.3 implies conditions of Lemma 8.1. (2) Condition M.4 implies conditions of Lemma 8.1.*

Example 1 (contd.) It is helpful to illustrate conditions M.4(a)-(e) via a simple example. Recall the example of interval censored Y without covariates, in which case $\Theta_I(P) = [E_P[Y_1], E_P[Y_2]]$ and suppose $Y_1 \leq Y_2$ P-a.s. for all $P \in \mathcal{P}$ and that (Y_1, Y_2) are uniformly Donsker in $\mathcal{P}$.[29] Then condition M.4(a)-(d) easily follow. To verify M.4(e) note that by contiguity and uniform integrability implied by the uniform in $\mathcal{P}$ Donskerness,

$$(E_{P_{n,\gamma}}[Y_1], E_{P_{n,\gamma}}[Y_2]) \to (E_P[Y_1], E_P[Y_2]),$$

including the case of $[E_P[Y_1], E_P[Y_2]]$ being a singleton. The last point is noteworthy, since Imbens and Manski (2004) used precisely the case of identification region shrinking to a singleton at a $1/\sqrt{n}$ rate as a counterexample to the coverage of certain types of confidence regions.

---

[28]The definition of regularity follows that given by van der Vaart and Wellner (1996), p. 413

[29]Conditions for the Donskerness uniformly in $\mathcal{P}$ is well known, see van der Vaart and Wellner (1996), p.168-170.

Conditions M.3 and M.4 are reasonable in many examples we have considered, provided the boundary of Θ_I (in $\mathbb{R}^d$) is strongly-identified. Conditions M.3 and M.4 are not expected to hold otherwise. Therefore, the models with weak identification, cf. Dufour (1997), that are local to non-identification, are not covered by the framework of regular sequences. Weak identification is not our focus in any case. However, Section 8.2 provides a general condition under which the proposed inference methods will work. The condition is illustrated with a weak IV framework of Dufour (1997) and Staiger and Stock (1997).

8.2. **Non-regular cases.** The following lemma addresses non-regular cases mentioned earlier, and shows that coverage results will be preserved in much greater generality.

Lemma 8.3 (Maintaining Partial Consistency and Minimal Coverage under Non-Regular Sequences). *(1) Suppose that* $\sup_{\Theta_I(P_{n,\gamma})} Q_n = O_{p_{n,\gamma}}(1/a_n)$ *under* $\{P_{n,\gamma}\}$. *Then* $\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})$ *wp* $\to 1$, *provided* $\widehat{c} \to_p \infty$, *under* $\{P_{n,\gamma}\}$. *(2) Let there be any estimate* $\widehat{c} \to_p c(\alpha, P)$ *under* $\{P\}$. *Suppose that Condition C.4 holds under fixed* P *with the limit real variable* $\mathcal{C}$ *that has* α*-quantile* $c(\alpha, P)$. *Suppose that for each* $\gamma \in \Gamma$ *and any sequence* $\epsilon_n \downarrow 0$, *we have*

$$\liminf_{n\to\infty} P_{n,\gamma}[\mathcal{C}_n \leq (c(\alpha, P) - \epsilon_n) \vee 0] \geq \alpha. \tag{8.1}$$

Consider any estimate $\widehat{c} \to_p c(\alpha, P)$ *under* $\{P\}$, *for instance, that provided in Section 3 or 4. Then for each* $\gamma \in \Gamma$

$$\liminf_{n\to\infty} P_{n,\gamma}\{\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})\} \geq \alpha. \tag{8.2}$$

Note again that the result is independent of the way the critical value is estimated.

Example (Weak IV). The point of this lemma can be illustrated using a very simple IV example with one regressor:

$$Y = \theta_0 X + \epsilon,\ \theta_0 \in \Theta(\text{compact}) \subset \mathbb{R},\ X = 0 \cdot Z + v,\ \text{and } (\epsilon, v)|Z \sim N(0, \Omega), Z \sim N(\mu, \sigma_Z^2). \tag{8.3}$$

The identification region is $\Theta_I(P) = \Theta$, that is, we have complete non-identification. Assume i.i.d. sampling and other conditions as in Section 4 hold under P.

Now consider a sequence of models where

$$Y = \theta_0 X + \epsilon,\ X = (\rho/\sqrt{n})Z + v,\ \text{and } (\epsilon, v)|Z \sim N(0, \Omega), Z \sim N(0, \sigma_Z^2). \tag{8.4}$$

Let γ index the parameter sequence $\{\rho\}$. Let $P^n_{n,\gamma}$ denote the law of vector $(Y_i, X_i, Z_i, i \leq n)$ in (8.4); it is contiguous to law P^n. Let $P_{n,\gamma}$ denote the law of the infinite sequence $(Y_i, X_i, Z_i, i < \infty)$ generated according to (8.4).

Note

$$\Theta_I(P_{n,\gamma}) = \Theta_I(P) = \Theta \text{ if } \rho = 0 \ \text{ and } \ \Theta_I(P_{n,\gamma}) = \theta_0 \in \Theta_I(P) \text{ if } \rho \neq 0. \tag{8.5}$$

This implies that the weak limit of $\mathcal{C}_n$ under $P_{n,\gamma}$ with $\rho \neq 0$, is stochastically smaller than the weak limit of $\mathcal{C}_n$ under $P_{n,\gamma}$ with $\rho = 0$, since

$$\sup_{\theta_0} \|\Delta(\theta)' W^{1/2}(\theta)\|^2 \leq \sup_{\Theta} \|\Delta(\theta)' W^{1/2}(\theta)\|^2. \tag{8.6}$$

Therefore α-quantile of the right side is bigger than α-quantile of the left side, and we have that (8.1) is satisfied. Note, compactness of Θ is important in insuring that the right-hand side is finite a.s. Therefore, for each $\rho \in \mathbb{R}^d$ and $\gamma = \{\rho\}$

$$\liminf_{n\to\infty} P_{n,\gamma}\{\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})\} \geq \alpha.$$

Next we consider more general local parameter sequences $\gamma = \{\rho_n\}$ with $\rho_n \in K$ for each n, where K is a compact subset of $\mathbb{R}$; let Γ denote the set of all these sequences. The limit under each convergent subsequence $\rho_n \to \rho$ is either the left or right side of (8.6). Hence, for each sequence $\{\gamma_n\}$ in Γ and each sequence $\epsilon_n \searrow 0$,

$$\begin{aligned} &\liminf_{n\to\infty} P_{n,\gamma_n}[\sup_{\Theta_I(P_{n,\gamma_n})} \|\Delta(\theta)'W^{1/2}(\theta)\|^2 \leq (c(\alpha,P)-\epsilon_n)\vee 0] \\ &\geq \liminf_{n\to\infty} P[\sup_{\Theta} \|\Delta(\theta)'W^{1/2}(\theta)\|^2 \leq (c(\alpha,P)-\epsilon_n)\vee 0] \geq \alpha. \end{aligned} \tag{8.7}$$

This implies by Lemma 8.3 that

$$\liminf_{n\to\infty} P_{n,\gamma_n}\{\Theta_I(P_{n,\gamma_n}) \subseteq C_n(\widehat{c})\} \geq \alpha. \tag{8.8}$$

Equivalently, for K denoting any non-empty compact subset of $\mathbb{R}$

$$\inf_K \liminf_{n\to\infty} \inf_{\rho\in K} P^n_{n,\rho}\{\Theta_I(P_{n,\rho}) \subseteq C_n(\widehat{c})\} \geq \alpha, \tag{8.9}$$

where $P^n_{n,\rho}$ denotes the law of vector $(Y_i, X_i, Z_i, i \leq n)$ in (8.4), and $P_{n,\rho}$ denotes the law of infinite sequence $(Y_i, X_i, Z_i, i < \infty)$ generated according to (8.4). This coverage property is in the spirit of local asymptotic minimax analysis of estimation, see van der Vaart (1998), Chapter 8.7.

8.3. **Proof of Lemma 8.1.** Proof of Part (**1**). The proof is straightforward by substituting $\{P_{n,\gamma}\}$ in place of the fixed sequence P in the proof of Theorems 3.1. □

Proof of Part (**2**). We have that $\widehat{c} \to_p c(\alpha,P)$ under $\{P\}$. By contiguity, $\widehat{c} \to_p c(\alpha,P)$ under $\{P_{n,\gamma}\}$. Therefore $P^n_{n,\gamma}\{\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})\} \geq P_{n,\gamma}[\mathcal{C}_n \leq \widehat{c}] = P_{n,\gamma}[\mathcal{C}_n \leq c(\alpha,P) + o_{P_{n,\gamma}}(1)] = P[\mathcal{C} \leq c(\alpha,P)] + o(1)$, by assumption that $P_{n,\gamma}[\mathcal{C}_n \leq c] \to P[\mathcal{C} \leq c]$ for all $c \geq 0$, by $\widehat{c} \geq 0$, and by continuity of the distribution function $c \mapsto P[\mathcal{C} \leq c]$ on $[0,\infty)$. □

8.4. **Proof of Lemma 8.2.** Proof of Part (**1**). The proof is straightforward by repeating Steps 1-4 in the Proof of Theorem 4.1, having replaced P with $P_{n,\gamma}$, Θ_I with $\Theta_I(P_{n,\gamma})$, $o_p(1)$ with $o_{P_{n,\gamma}}(1)$, etc., and then noting that $\sup_{\Theta_I(P_{n,\gamma})} \|\Delta(\theta)'W^{1/2}(\theta)\|^2 = \sup_{\Theta_I(P)} \|\Delta(\theta)'W^{1/2}(\theta)\|^2 + o_{P_{n,\gamma}}(1)$, which follows by equicontinuity of $\theta \mapsto \Delta(\theta)'W^{1/2}(\theta)$ and by $d_H(\Theta_I(P_{n,\gamma}), \Theta_I(P)) = o(1)$ imposed in M.3(e). By M.3(b) $\Delta(\theta)$ does not depend on γ, and by contiguity $W(\theta)$ does not either. Hence the limit variable $\mathcal{C} := \sup_{\Theta_I(P)} \|\Delta(\theta)'W^{1/2}(\theta)\|^2$ does not depend on γ. □

Proof of Part (**2**). The proof is straightforward by repeating Steps 1-4 in the Proof of Theorem 4.1, having replaced P with $P_{n,\gamma}$, Θ_I with $\Theta_I(P_{n,\gamma})$, and $o_p(1)$ with $o_{P_{n,\gamma}}(1)$. The exception is that in Step 2, we need to define $\xi(\theta) = \lim_n \sqrt{n}E_P[m_i(\theta)]$ under fixed sequence $\{P\}$. Note that the key inequality (6.10) in Lemma 6.2 on which the proof is based will be preserved under sequences $\{P_{n,\gamma}\}$. In the proof of Lemma 6.2, the convergent subsequence $\{\theta_n\}$ in $\Theta_I(P)$ is replaced by the convergent subsequence $\{\theta_n\}$ in $\Theta_I(P_{n,\gamma})$, where convergent means $\theta_n \to \theta \in \Theta_I(P)$. Since we care only about

$\mathcal{C}_n$ in this Lemma, in repeating the proof of Lemma 6.2, we have a drastic simplification from setting $\lambda = 0$. [That is, we only need to consider the set $V_n^0 = V_\infty^0 = \Theta_I(P_{n,\gamma}) \times \{0\}$.] In addition, we note that for every $\mathcal{J}$, $d_H(\Theta_\mathcal{J}(P_{n,\gamma}), \Theta_\mathcal{J}(P)) = o(1)$ by M.4(e), so that $\max_\mathcal{J} \sup_{\Theta_\mathcal{J}(P_{n,\gamma})} \sum_{j\in\mathcal{J}} |(\Delta_j(\theta) + G_j(\theta)'\lambda)W_{jj}^{1/2}(\theta) + o_{p_{n,\gamma}}(1)|_+^2 = \max_\mathcal{J} \sup_{\Theta_\mathcal{J}(P)} \sum_{j\in\mathcal{J}} |(\Delta_j(\theta) + G_j(\theta)'\lambda)W_{jj}^{1/2}(\theta) + o_{p_{n,\gamma}}(1)|_+^2$. The last observation utilized equicontinuity of $\theta \mapsto \Delta(\theta)'W^{1/2}(\theta)$ and the fact that by M.4(b) $\Delta(\theta)$ does not depend on γ, and by contiguity $W(\theta)$ does not depend on γ either. The result of the modified Step 2 can be stated then as

$$\sup_{\Theta_I(P_{n,\gamma})} \ell_n(\theta, 0) =_d \max_{\mathcal{J}} \sup_{\theta\in\Theta_\mathcal{J}(P)} \sum_{j\in\mathcal{J}} |(\Delta_j(\theta))W_{jj}^{1/2}(\theta) + o_{p_{n,\gamma}}(1)|_+^2.$$

Hence $\mathcal{C} = \max_\mathcal{J} \sup_{\theta\in\Theta_\mathcal{J}(P)} \sum_{j\in\mathcal{J}} |(\Delta_j(\theta))W_{jj}^{1/2}(\theta)|_+^2$, which does not depend on γ. □

8.5. **Proof of Lemma 8.3.** Part (1). Under $\{P_{n,\gamma}\}$, wp $\to 1$, by construction of $\widehat{c}$, $\sup_{\Theta_I(P_{n,\gamma})} Q_n = O_{p_{n,\gamma}}(1/a_n) < \widehat{c}/a_n$, which implies $\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})$. □

Part (2). We have that $\widehat{c} \to_p c(\alpha, P)$ under $\{P\}$. By contiguity, $\widehat{c} \to_p c(\alpha, P)$ under $\{P_{n,\gamma}\}$. Hence $P_{n,\gamma}\{\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})\} \geq P_{n,\gamma}\{\mathcal{C}_n \leq \widehat{c}\} \geq P_{n,\gamma}\{\mathcal{C}_n \leq (c(\alpha, P) - \epsilon_n) \vee 0\} = P_{n,\gamma}\{\mathcal{C} \leq (c(\alpha, P) - \epsilon_n) \vee 0\}$, for some $\epsilon_n \downarrow 0$. The conclusion follows from the assumption that $\liminf_{n\to\infty} P_{n,\gamma}\{\mathcal{C}_n \leq (c(\alpha, P) - \epsilon_n) \vee 0\} \geq \alpha$. □