A Two-Stage Procedure for Partially Identified Models*

Hiroaki Kaido
Boston University

Halbert White
University of California, San Diego

This version: April, 2013

Abstract

This paper studies a two-stage procedure for estimating partially identified models, based on Chernozhukov, Hong, and Tamer's (2007) theory of set estimation and inference. We consider the case where a sub-vector of parameters or their identified set can be estimated separately from the rest, possibly subject to a priori restrictions. Our procedure constructs the second-stage set estimator and confidence set by taking appropriate level sets of a criterion function, using a first-stage estimator to impose restrictions on the parameter of interest. We give conditions under which the two-stage set estimator is a set-valued random element that is measurable in an appropriate sense. We also establish consistency of the two-stage set estimator.

JEL Classification: C12
Keywords: Partial identification, Set estimation, Two-stage estimation, Effros-measurability.

*Kaido gratefully acknowledges financial support from the Nakajima Foundation and research assistance from Jie Hou. We thank Norman Swanson and two anonymous referees for their helpful comments that have helped improve the paper. We also thank seminar participants at Boston College and Duke University, and participants at the conference in honor of Halbert L. White Jr. for their helpful comments.

1 Introduction

Statistical inference for partially identified economic models, pioneered by Charles Manski (see Manski, 2003, and the references therein), is a growing field in econometrics. In this context, the economic structures of interest are characterized by an *identified set* Θ_I, rather than by a single point in the parameter space Θ. Recent studies of partial identification have shown that consistent set estimators and confidence regions can be constructed for the identified set as a whole or for its elements. In particular, Chernozhukov, Hong, and Tamer (2007) (CHT) propose a general framework based on the extremum estimation approach. Within their framework, the identified set is defined as a set of minimizers of a criterion function. A consistent estimator of the identified set can be constructed by taking a suitable level-set of a sample analogue of the criterion function. This level-set estimator can also be used to compute a critical value to construct a confidence set for the identified set.

The main contribution of this paper is to establish measurability of general level-set estimators allowing for the presence of a sub-vector of parameters or its identified set that can be estimated separately from the rest. Specifically, we consider cases satisfying the following conditions: (i) the parameter vector consists of two subvectors: a "first-stage" parameter and a "second-stage" parameter; (ii) the identification of the first-stage parameter depends on neither the identification nor the value of the second-stage parameter; and (iii) the identified set for the second-stage parameter depends on the first-stage parameter through the criterion function. In practice, we often encounter cases where these conditions are satisfied. For example, Bajari, Benkard, and Levin's (2007) estimation framework for dynamic imperfect competition models has this structure. In financial econometrics, Kaido and White (2009) apply the two-stage procedure developed here to study the set of market risk prices under incomplete markets.

Our procedure constructs a two-stage set estimator by taking appropriate level sets of a criterion function, using a first-stage estimator to impose restrictions on the parameter of interest. The presence of a first-stage parameter introduces some complications when establishing the measurability of the two-stage set estimator. Specifically, when the first-stage parameter is partially identified, our estimator is a level set of a sample criterion function, whose first argument is restricted to a first-stage set estimator. This set is a somewhat complicated object whose measurability is not trivial to show. We show that the measurability of our estimator is related to that of the first-stage set estimator and the infimum of the criterion function over random sets. We then exploit this fact and establish the measurability of the two-stage set estimator using the results of Stinchcombe and White (1992). We also establish Hausdorff consistency of the two-stage set estimator.

For a special case of the two-stage structure, we also discuss an inference procedure based on a quasi-likelihood ratio statistic. Specifically, we assume that the first-stage parameter is point identified and study a procedure for testing hypotheses on the second-stage parameter.

Our analysis *does not* address the general setting where the researcher wishes to conduct inference for particular subcomponents of the whole parameter vector without knowing whether the other components are point or partially identified. The case we study, however, is still of practical interest. For example, the aforementioned applied examples (Bajari, Benkard, and Levin, 2007; Kaido and White, 2009) admit point-identified first-stage parameters when some *a priori* restrictions are available.

The paper is organized as follows. Section 2 summarizes CHT's econometric framework and formalizes the two-stage structure just described. Section 3 gives two illustrative examples. Section 4 provides measurability and consistency results for the two-stage set estimator. Section 5 provides a discussion of inference based on a quasi-likelihood ratio statistic. Section 6 concludes. The appendix contains formal proofs.

2 The Data Generating Process and the Model

2.1 CHT Framework and Two-Stage Structure

Our first assumption describes the data generating process, the parameter space, and the estimation criterion function.

ASSUMPTION 2.1: *Let $d_1, d_2 \in \mathbb{N}$ and $d := d_1 + d_2$. Let $\Theta_1 \subset \mathbb{R}^{d_1}$ and $\Theta_2 \subset \mathbb{R}^{d_2}$ be nonempty compact sets. Let $\Theta := \Theta_1 \times \Theta_2$. (i) For $n = 1, 2, ...$, let $\bar{Q}_n : \Theta \to \bar{\mathbb{R}}_+$ be a continuous function. (ii) Let $(\Omega, \mathfrak{F}, P^0)$ be a complete probability space. For $n = 1, 2, ...$, let $Q_n : \Omega \times \mathbb{R}^d \to \bar{\mathbb{R}}_+$ be such that $Q_n(\cdot, \theta)$ is measurable for each $\theta \in \mathbb{R}^d$ and $Q_n(\omega, \cdot)$ is continuous on Θ for each $\omega \in F \in \mathfrak{F}$, $P^0(F) = 1$, and for all $\omega \in \Omega$ and $\theta \notin \Theta$, $Q_n(\omega, \theta) = \infty$.*

Θ is the finite-dimensional parameter space. Compactness is a standard assumption on Θ for extremum estimation. The parameter of interest $\theta \in \Theta$ consists of two subvectors, $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$. Throughout, we will call θ_1 a *first-stage* parameter and θ_2 a *second-stage* parameter. The probability measure P^0 embodies the data generating process (DGP) and thus governs the stochastic properties of the data. Throughout, we assume that there exists a set $\mathcal{P}$ of complete probability measures on $(\Omega, \mathfrak{F})$ such that $P^0 \in \mathcal{P}$. Consistent with White (1994), we call $\mathcal{P}$ the *model.*

The function Q_n acts as the sample criterion function for estimation; for example,

$$Q_n(\omega, \theta) = n^{-1} \sum_{i=1}^{n} q(X_i(\omega), \theta) - \inf_{\theta \in \Theta} \; n^{-1} \sum_{i=1}^{n} q(X_i(\omega), \theta), \tag{2.1}$$

where $\{X_i : \Omega \to \mathbb{R}\}$ is a sequence of random vectors taking values in $\mathcal{X} \subseteq \mathbb{R}^k, k \in \mathbb{N}$, and q is a suitable function, e.g., $q(x, \theta) = -\ln f(x, \theta)$, where $f(\cdot, \theta)$ is a probability density function for each θ. This example corresponds to the case of quasi-maximum likelihood estimation. The second term ensures that $Q_n(\omega, \theta) \geq 0$. As is common, we may write $Q_n(\theta)$

as a shorthand for $Q_n(\,\cdot\,,\theta)$.

The function $\bar{Q}_n$ is the population criterion function. Without loss of generality, we normalize the minimum value of $\bar{Q}_n$ to 0, i.e. $\inf_\Theta \bar{Q}_n(\theta) = 0$. For example, when the expectations exist, the population analog for the above example is

$$\bar{Q}_n(\theta) = n^{-1}\sum_{i=1}^{n} E[q(X_i(\cdot),\theta)] - \inf_{\theta\in\Theta}\; n^{-1}\sum_{i=1}^{n} E[q(X_i(\cdot),\theta)].$$

CHT defines the *identified set* as the set of minimizers of $\bar{Q}_n$. Examples of studies in which the identified set is defined in this way are those of CHT, Bajari, Benkard, and Levin (2007), Ciliberto and Tamer (2009), Kaido and White (2009), Bugni (2010), and Romano and Shaikh (2010). Under regularity conditions, Q_n eventually reveals the set of minimizers of $\bar{Q}_n$. CHT's approach is to use the level sets of Q_n to construct confidence sets and a set estimator for the identified set. A practical challenge occurs when the identified set has a large dimension. In many cases of interest, this challenge can be addressed by taking advantage of the structure of the optimization problem. Here, we consider identified sets that have "two-stage" structures, defined as follows.

Definition 2.1: *The* unrestricted identified set $\Theta^u_{I,n}$ *is defined as*

$$\Theta^u_{I,n} := \{\theta \in \Theta : \bar{Q}_n(\theta_1,\theta_2) = 0, \qquad \theta_1 \in \Theta^u_{I,1}\}, \tag{2.2}$$

for some $\Theta^u_{I,1} \subseteq \Theta_1$*, where* $\Theta^u_{I,n}$ *is identified without knowledge of* $\bar{Q}_n$*.* $\Theta^u_{I,n}$ *defined above is said to have* two-stage structure.

We will give two examples with this structure in the next section. Observe that $\Theta^u_{I,n}$ has an n index, due to the n index of $\bar{Q}_n$. With stationary data, the n index is unnecessary; with asymptotically stationary data, if $\bar{Q}_n$ converges to a uniform limit, say $\bar{Q}$, then the n index also becomes unnecessary. In what follows, we may suppress the n subscript for notational simplicity and simply write $\bar{Q}$ and Θ^u_I.

Definition 2.1 requires $\Theta^u_{I,1}$ is identified separately from the rest of parameters. In other words, we may identify and estimate $\Theta^u_{I,1}$ without using $\bar{Q}$, which can help simplify set estimation. Not all examples studied in the literature, however, have this structure. For example, the identified set for regression coefficients with an interval valued outcome variable can be characterized as the set of minimizers of some criterion function $\bar{Q}$ (see Manski and Tamer, 2002, Section 4.5), yet it does not admit the two-stage structure without additional restrictions.

A special case of the two-stage structure is the setting where the first stage is fully identified, so $\Theta^u_{I,1}$ is simply $\{\theta^0_1\}$, say. From now on, we mainly consider two-stage structures. Formally, we impose

Assumption 2.2: *Θ_I^u has two-stage structure.*

When θ_1^0 satisfies a priori restrictions, these can restrict the first-stage identified set. A common restriction is that θ_1^0 has a known relationship to another point identified parameter ψ^0.

Restriction 2.1: *Let Ψ be a compact subset of a finite dimensional Euclidean space. Let $m_2 \in \mathbb{N}$, and let $s : \Theta_1 \times \Psi \to \mathbb{R}^{m_2}$ be a given jointly measurable function. θ_1^0 satisfies $s(\theta_1^0, \psi^0) = 0$, where $\psi^0 \in \Psi$ is point identified.*

A special case of Restriction 2.1 is the setting where s does not depend on ψ^0, in which case the restriction reduces to $\rho(\theta_1^0) = 0$ for some known function $\rho : \Theta \to \mathbb{R}^{m_2}$. A restriction defines a set Θ_1^r of parameter values satisfying the restrictions. Define

$$\Theta_1^r := \{\theta_1 \in \Theta_1 : s(\theta_1, \psi^0) = 0\}.$$

The set of identified parameter values that satisfy the restrictions is therefore $\Theta_{I,1}^r := \Theta_{I,1}^u \cap \Theta_1^r$. We call this set *the restricted first-stage identified set.*

Most of our results hold even without first-stage restrictions. Accordingly, we state our results in terms of a generic identified set for the first-stage parameter, denoted $\Theta_{I,1}$, whenever the results hold with or without the restrictions. We call $\Theta_{I,1}$ the *first-stage identified set.*

Given $\Theta_{I,1}$, we define the *two-stage identified set* as follows:

Definition 2.2: *The* two-stage identified set *is*

$$\Theta_I := \left\{\theta \in \Theta : \bar{Q}(\theta) = 0 \text{ and } \theta_1 \in \Theta_{I,1}\right\}.$$

Given two-stage structure, Θ_I and Θ_I^u coincide when $\Theta_{I,1} = \Theta_{I,1}^u$. They differ when the first stage is restricted. As a special case, we may achieve full identification of the first-stage parameter. In this case, we can define the *second-stage identified set.*

Definition 2.3: *Let $\Theta_{I,1} = \{\theta_1^0\}$. The* second-stage identified set *is*

$$\Theta_{I,2} := \left\{\theta_2 \in \Theta_2 : \bar{Q}(\theta_1^0, \theta_2) = 0\right\}.$$

In this special case, the identified set for θ is simply $\Theta_I = \{\theta_1^0\} \times \Theta_{I,2}$. We here note that the second-stage identified set, in general, differs from the projection of the set of minimizers of $\bar{Q}$. To see this, for each $\theta_1 \in \Theta_1$ define $\Xi_2(\theta_1) := \{\theta_2 \in \Theta_2 : \bar{Q}(\theta_1, \theta_2) = 0\}$. For each θ_1, $\Xi_2(\theta_1)$ gives a "slice" of the set of minimizers of $\bar{Q}$. The projections of the set of minimizers of $\bar{Q}$ are defined by

$$\Theta^*_{I,1} := \{\theta_1 \in \Theta_1 : \bar{Q}(\theta_1, \theta_2) = 0, \text{ for some } \theta_2 \in \Theta_2\}$$
$$\Theta^*_{I,2} := \{\theta_2 \in \Theta_2 : \bar{Q}(\theta_1, \theta_2) = 0, \text{ for some } \theta_1 \in \Theta_1\}.$$

It is straightforward to show that $\Theta^*_{I,2} = \cup_{\theta_1 \in \Theta^*_{I,1}} \Xi(\theta_1)$, which is in general a superset of the second-stage identified set $\Theta_{I,2} = \Xi_2(\theta_1^0)$. The equality $\Theta^*_{I,2} = \Theta_{I,2}$ holds only if $\bar{Q}$ is minimized on $\{\theta_1^0\} \times \Theta_{I,2}$ and is strictly positive outside this set.

A natural approach to conducting estimation and inference for Θ_I (or $\Theta_{I,2}$ when θ_1^0 is fully identified) is to replace $\Theta_{I,1}$ (or θ_1^0) with its consistent estimator $\hat{\Theta}_{1n}$ (or $\hat{\theta}_{1n}$). We will discuss how to construct the first-stage estimator in Section 4.3. For now, we impose the presence of a possibly set-valued estimator of the first-stage parameter as a high-level assumption. For this, let $\mathcal{F}(\Theta_1)$ be the set of closed subsets of Θ_1.

Assumption 2.3 (First-Stage Estimator): $\hat{\Theta}_{1n} : \Omega \to \mathcal{F}(\Theta_1)$ *is a measurable mapping.*

When θ_1^0 is fully identified, we explicitly denote its estimator by $\hat{\theta}_{1n} : \Omega \to \Theta_1$. The "measurability" imposed here is Effros-measurability. We discuss this in detail in Section 4.

Given a first-stage estimator, we can construct a set estimator or a confidence region for the identified set Θ_I (or $\Theta_{I,2}$) in the second stage based on the CHT framework. Generally, the sample criterion function Q_n approximates $\bar{Q}$ well as n tends to infinity. As CHT show, a level set of Q_n with level decreasing to 0 at a proper rate is a good estimator for the set of minimizers of $\bar{Q}$. As our main focus is to estimate Θ_I, we additionally restrict θ_1 to the estimator $\hat{\Theta}_{1n}$ of $\Theta_{I,1}$. We formally define the two-stage set estimator as follows.

Definition 2.4 (Two-Stage Set Estimator): *For each $n \in \mathbb{N}$, let $\hat{\epsilon}_n : \Omega \to \mathbb{R}_+$ be measurable. Given a sequence $\{a_n\}$ and $\hat{\Theta}_{1n} : \Omega \to \mathcal{F}(\Theta_1)$, the* two-stage set estimator *is*

$$\hat{\Theta}_n := \left\{\theta : a_n Q_n(\theta_1, \theta_2) \le \hat{\epsilon}_n, \quad \theta_1 \in \hat{\Theta}_{1n}\right\}.$$

If $\Theta_{I,1}$ has only one element, then given $\hat{\theta}_{1n} : \Omega \to \Theta_1$, the second-stage set estimator *is*

$$\hat{\Theta}_{2n} := \left\{\theta_2 : a_n Q_n(\hat{\theta}_{1n}, \theta_2) \le \hat{\epsilon}_n\right\}.$$

3 Examples

In this section, we present two examples exhibiting the two-stage structure described in the previous section. The first example is Bajari, Benkard, and Levin's (2007) (BBL) analysis of dynamic models of imperfect competition.

3.1 Dynamic Models of Imperfect Competition

Let $L, M, N \in \mathbb{N}$. Let $i = 1, \cdots, N$ be the player (firm) index. For each period $t \in \mathbb{N}$, let $s_t \in S \subseteq \mathbb{R}^L$ be a vector of commonly observed state variables. Each player observes s_t and a private shock $v_{it} \in \mathcal{V}_i \subseteq \mathbb{R}^M$ and decides their action $a_{it} \in A_i$.

A pure Markov strategy for player i is a measurable function $\sigma_i : S \times \mathcal{V}_i \to A_i$. Let $a_t \in A = A_1 \times \cdots \times A_N$. Given a common subjective discount factor β_0 and a payoff function $\pi_i : A \times S \times \mathcal{V}_i \to \mathbb{R}$, a pure-strategy Markov perfect equilibrium (MPE) is a profile $\sigma = (\sigma_1, \cdots, \sigma_N)$ of Markov strategies such that

$$V_i(s; \sigma) \geq V_i(s; \sigma_i', \sigma_{-i})$$

for all players i, states s and Markov strategies σ_i', where $V_i(s; \sigma)$ is the value function defined recursively as

$$V_i(s; \sigma) := E\left[\pi_i(\sigma(s, v; \alpha_0), s, v_i; \gamma_0) + \beta_0 \int V_i(s'; \sigma) dP(s' | \sigma(s, v; \alpha_0), s; \alpha_0) \middle| s\right].$$

A parameterized version of the Markov process transition probability $P(s'|\ a, s; \alpha_0)$ is $P(s'|a, s; \alpha)$, $\alpha \in \mathbb{R}^{d_1}$. The strategy σ is also assumed to be parameterized by α. The private shock v_{it} is drawn independently across players and over time from a player-specific distribution $G_i(\cdot|s_t; \gamma_0)$, where $\gamma_0 \in \mathbb{R}^{d_2}$. The vector γ_0 also enters the payoff function $\pi_i(a, s, v_i; \gamma_0)$, parameterized as $\pi_i(a, s, v_i; \gamma)$. Assuming that the subjective discount factor β_0 is known, the true parameter vector of interest is $(\alpha_0', \gamma_0')'$.

Following BBL, let $x := (i, s, \sigma_i')$. Let $\mathcal{X}$ denote the set of admissible x values, and let

$$g(x, \sigma; \alpha_0, \gamma_0) := V_i(s; \sigma_i', \sigma_{-i}; \alpha_0, \gamma_0) - V_i(s; \sigma_i, \sigma_{-i}; \alpha_0, \gamma_0).$$

Without specifying the equilibrium selection rule, the equilibria consistent with the true parameter vector are characterized by the set of inequalities $g(x, \sigma; \alpha_0, \gamma_0) \leq 0$ for $x \in \mathcal{X}$. BBL show that α_0 can be fully identified in many cases; α_0 is the first-stage parameter. The inequality conditions, however, do not necessarily guarantee the full identification of the second-stage parameter γ_0. Therefore, they consider the case where the parameter is partially identified. Let H be a distribution over $\mathcal{X}$ chosen by the researcher[1]. For $y \in \mathbb{R}$, let $\{y\}_+ := \max\{y, 0\}$. Their population criterion function is defined by

$$\bar{Q}(\alpha, \gamma) := \int_{\mathcal{X}} \{g(x, \sigma; \alpha, \gamma)\}_+^2 \, dH(x).$$

When α_0 is fully identified, the first-stage identified set is $\Theta_{I,1} = \{\alpha_0\}$, and the second-stage

[1] The distribution might be chosen in a variety of ways. BBL considers the possibility that σ_i' is distributed as a slight perturbation of the equilibrium policy σ_i and that σ_i' differs from σ_i at a single state.

identified set is $\Theta_{I,2} = \{\gamma : \bar{Q}(\alpha_0, \gamma) = 0\}$. The identified set for (α, γ) is therefore $\{\alpha_0\} \times \Theta_{I,2}$.

BBL show that α_0 can be estimated from repeated observations on individual actions and states $\{a_{it}, s_t\}$. Let $\hat{\alpha}_n$ be the first-stage estimator of α_0. This gives the first stage estimate $\hat{P} := P(s'|a, s, ; \hat{\alpha}_n)$ and $\hat{\sigma} := \sigma(s, v; \hat{\alpha}_n)$ of the transition probability and the policy function. Now, let $\{x_1, \cdots, x_J\}$ be the set of x values chosen by the researcher[2]. For each $x \in \{x_1, \cdots, x_J\}$ and $\gamma \in \mathbb{R}^{d_2}$, one can estimate the value function V_i by forward simulation (BBL, Section 3.3) given the first stage estimators $\hat{P}$ and $\hat{\sigma}$. This gives an estimator $\hat{V}_i(s, \hat{\sigma}_i, \hat{\sigma}_{-i}; \hat{\alpha}_n, \gamma)$ of the value function. Let $\hat{g}(x, \hat{\sigma}; \hat{\alpha}_n, \gamma) := \hat{V}_i(s, \sigma_i', \hat{\sigma}_{-i}; \hat{\alpha}_n, \gamma) - \hat{V}_i(s, \hat{\sigma}_i, \hat{\sigma}_{-i}; \hat{\alpha}_n, \gamma)$. BBL consider the sample criterion function for γ

$$Q_n(\hat{\alpha}_n, \gamma) := \frac{1}{J} \sum_{j=1}^{J} \{\hat{g}(x_j, \hat{\sigma}; \hat{\alpha}_n, \gamma)\}_+^2 .$$

Using this criterion function and given a sequence $\{\hat{\epsilon}_n\}$, the second-stage estimator for $\Theta_{I,2}$ is $\hat{\Theta}_{2n} = \{\gamma : nQ_n(\hat{\alpha}_n, \gamma) \leq \hat{\epsilon}_n\}$.

BBL focus on how dynamic models of imperfect competition can be estimated, allowing the possibility that the second stage parameter is only partially identified. They implicitly assume measurability, prove the consistency of the second-stage estimator using Manski and Tamer's (2002) result, and describe the construction of a confidence set based on Romano and Shaikh (2010). Our analysis ensures the measurability of BBL's set estimator and extends the consistency of the two-stage set estimator to cases where α is only partially identified.

3.2 Market Price of Risk in Incomplete Markets

The second example is Kaido and White's (2009) study of the market price of risk in incomplete markets. We present one of their main cases. Let $(\Omega, \mathfrak{F}, \{\mathfrak{F}_t\}, P)$ be a filtered probability space. Suppose that there are $d \in \mathbb{N}$ risky assets and that the $\mathbb{R}^d-$valued asset price process $\{S_t\}$ solves the stochastic differential equation

$$dS_t^j = \mu_0^j S_t^j dt + \sigma_0^j S_t^j dW_t, \quad t \in [0, T], \quad j = 1, ..., d,$$

where $\{W_t\}$ is a vector of $N \in \mathbb{N}$ independent standard Brownian motions under P adapted to the filtration $\{\mathfrak{F}_t\}$, $\mu_0 \in \mathbb{R}^d$ has elements μ_0^j, $j = 1, ..., d$, and $\sigma_0 \in \mathbb{R}^{d \times N}$ has $1 \times N$ rows σ_0^j, $j = 1, ..., d$. Let S_t^0 denote the price of the risk-free bond with known rate of return r.

Suppose (i) $\{S_t\}$ does not admit arbitrage; and (ii) extremely good deals (returns with high Sharpe ratios) are not available[3]. The first assumption ensures the existence of a risk-neutral measure and an associated market price of risk λ satisfying $\sigma_0 \lambda = \mu_0 - r\iota$, where ι is a vector of ones. The second ensures that the true market price of risk λ_0 has a finite upper

[2] BBL propose drawing the x-values independently from H.
[3] See Cochrane and Saá-Requejo (2000) for example.

bound M on its norm. Thus, they define the parameter space Θ_2 for λ to be $\Theta_2 = \{\lambda : \|\lambda\| \leq M\}$.

In this example, the diffusion coefficient σ_0 can be (partially or fully) identified and estimated separately from the market price of risk λ_0. The covariance matrix Σ_0 of asset returns satisfies $\Sigma_0 = \sigma_0\sigma_0'$. This is an example of the restriction described in Restriction 2.1. The restricted first-stage identified set is

$$\Theta_{I,1} = \Theta_{I,1}^r = \{\sigma : s_1(\sigma, \Sigma_0) = 0\},$$

where $s_1(x, A) = vec(A - xx')$ for $x \in \mathbb{R}^{d\times N}$ and $A \in \mathbb{R}^{d\times d}$. It can be shown that $\|s_1(\sigma, \hat{\Sigma}_n)\|^2 = O_p(n^{-1})$ on $\Theta_{I,1}$. Then, a first-stage estimator for $\Theta_{I,1}$ can be defined by $\hat{\Theta}_{1n} := \{\sigma : n\|s_1(\sigma, \hat{\Sigma}_n)\|^2 \leq \eta_n\}$, where $\hat{\Sigma}_n$ is the sample covariance matrix of asset returns, and $\{\eta_n\}$ is a non-negative sequence such that $\sup_{\sigma\in\Theta_{I,1}} \|s_1(\sigma, \hat{\Sigma}_n)\|^2 \leq \eta_n/n$ with probability approaching 1. Such a sequence can be constructed by setting $\eta_n = c\kappa_n$, where $c > 0$ and κ_n is a slowly diverging sequence, e.g., $\kappa_n = \log\log n$. Kaido and White (2009) use the population criterion function

$$\bar{Q}_n(\sigma, \lambda) := -E\left[\frac{1}{n}\sum_{i=1}^{n} \ln f(R_{t_i}; \sigma, \lambda)\right] - \bar{q}_n,$$

where for each t_i in the partition $\{t_0, t_1, \cdots, t_n\}$ with $t_0 = 0$ and $t_n = T$, $R_{t_i} \in \mathbb{R}^d$ is the vector of returns over the interval $[t_i, t_{i-1}]$, f is the the multivariate Gaussian density with mean $\sigma\lambda - r\iota$ and covariance Σ, and $\bar{q}_n := \inf_{\sigma,\lambda} -E[n^{-1}\sum_{i=1}^{n} \ln f(R_{t_i}; \sigma, \lambda)]$. The identified set for (σ, λ) is defined by $\Theta_I := \{(\sigma, \lambda) : \bar{Q}_n(\sigma, \lambda) = 0, \text{almost all } n \ (a.a.n), \text{ and } \sigma \in \Theta_{I,1}\}$. The sample criterion function is defined by

$$Q_n(\sigma, \lambda) = -\frac{1}{n}\sum_{i=1}^{n} \ln f(R_{t_i}; \sigma, \lambda) - q_n,$$

where $q_n := \inf_{\sigma,\lambda} -n^{-1}\sum_{i=1}^{n} \ln f(R_{t_i}; \sigma, \lambda)$. Given $\hat{\Theta}_{1n}$ and a sequence $\{\hat{\epsilon}_n\}$, the two-stage set estimator is defined by $\hat{\Theta}_n := \{(\sigma, \lambda) : nQ_n(\sigma, \lambda) \leq \hat{\epsilon}_n, \sigma \in \hat{\Theta}_{1n}\}$.

Under additional assumptions, the diffusion coefficient can be fully identified. Kaido and White (2009) consider examples that satisfy additional restrictions. For example, assuming that each asset is exposed to its own idiosyncratic risk and $n - d$ common risk factors implies that there exists a function $s_2 : \mathbb{R}^{d\times N} \to \mathbb{R}^m$ such that $s_2(\sigma_0) = 0$ for some $m \in \mathbb{N}$. In this case, the first-stage identified set is

$$\Theta_{I,1} = \Theta_{I,1}^r = \{\sigma : s_1(\sigma, \Sigma_0) = 0 \text{ and } s_2(\sigma) = 0\}.$$

Kaido and White (2009) show that these restrictions may fully identify σ_0 in some cases, enabling them to estimate σ_0 using a point estimator $\hat{\sigma}_n$. If σ_0 is fully identified, the second-stage identified set for the market price of risk is simply $\Theta_{I,2} = \{\lambda : \bar{Q}_n(\sigma_0, \lambda) = 0, a.a.n\}$. This set is shown to be a bounded subset of an affine space. The second-stage set estimator is $\hat{\Theta}_{2n} = \{\lambda : nQ_n(\hat{\sigma}_n, \lambda) \leq \hat{\epsilon}_n\}$.

4 Measurability and Consistency of the Two-Stage Estimator

In this section, we establish the measurability and consistency of the two-stage set estimator. We first show that our estimator is a set-valued random element that is measurable in an appropriate sense. This ensures that various functionals, including the distance between the set estimator and the identified set, are measurable, which in turn implies that standard asymptotic theory can be used to establish consistency. An alternative to our approach here would be to work with outer probability and modified definitions of convergence (van der Vaart and Wellner, 2000). The consistency result is a straightforward extension of CHT.

4.1 Effros-measurability

The measurability of set estimators is defined for mappings from Ω to the space of closed subsets of a Euclidean space. We first briefly review useful concepts and results in the theory of random sets. Details can be found in Molchanov (2005).

For $\mathcal{A} \subset \mathbb{R}^d$, let $\mathcal{F}(\mathcal{A})$ denote the collection of all closed subsets of $\mathcal{A}$. A useful measurability concept for set-valued functions is Effros-measurability.

Definition 4.1 (Effros-measurability): *A map $X : \Omega \to \mathcal{F}(\mathcal{A})$ is* Effros-measurable *with respect to $\mathfrak{F}$ if for each closed set $F \in \mathcal{F}(\mathcal{A})$*

$$X^-(F) := \{\omega : X(\omega) \cap F \neq \varnothing\} \in \mathfrak{F}. \tag{4.1}$$

Effros-measurability ensures that many functionals of interest, such as the distance between random sets, become random variables; it is also flexible, handling as many random elements as one typically requires[4]. To show the measurability of the set estimators defined in Definition 2.4, we impose mild conditions on the sample criterion function. For this, we require the criterion function to be a *proper normal integrand*, defined below. Recall that a function $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ is *lower semicontinuous* (*lsc*) if $\liminf_{x \to \bar{x}} f(x) \geq f(\bar{x})$ for every $\bar{x} \in \mathbb{R}^d$.

Definition 4.2 (Epigraph and Proper Normal Integrand): *If $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ is lsc, then*

$$epi\ f = \{(x, \alpha) \in \mathbb{R}^d \times \mathbb{R} : f(x) \leq \alpha\}$$

[4] Molchanov (2005) defines an Effros-measurable map as one such that $X^-(G) \in \mathfrak{F}$ for each open set $G \subset \mathcal{A}$. When $\mathcal{A}$ is a Polish space and $(\Omega, \mathfrak{F}, P^0)$ is complete, this is equivalent to our definition (Theorem 2.2.3. in Molchanov, 2005).

is called the epigraph *of f.*

A function $\zeta : \Omega \times \mathbb{R}^d \to \bar{\mathbb{R}}$, is called a normal integrand *if its epigraph $X(\omega) = epi\ \zeta(\omega, \cdot)$ defines a closed set that is Effros-measurable with respect to $\mathfrak{F}$. A normal integrand is said to be* proper *if it does not take the value $-\infty$ and is not identically equal to $+\infty$.*

The following are useful facts about normal integrands.

FACT 4.1 (Proposition 3.6 in Molchanov, 2005): *Let $\zeta : \Omega \times \mathbb{R}^d \to \bar{\mathbb{R}}$ be such that $\zeta(\omega, \cdot)$ is lsc on $\mathbb{R}^d$ for each $\omega \in \Omega$. If ζ is jointly measurable on $\Omega \times \mathbb{R}^d$, then ζ is a normal integrand.*

FACT 4.2 (Proposition 3.10 (i) in Molchanov, 2005): *If ζ is a normal integrand, then for every random variable $\hat{\alpha} : \Omega \to \bar{\mathbb{R}}$, $\{\zeta \leq \hat{\alpha}\} = \{x \in \mathbb{R}^d : \zeta(\cdot, x) \leq \hat{\alpha}(\cdot)\}$ is a random closed set, i.e. it is Effros-measurable.*

Recall that when the first-stage parameter is fully identified, the second-stage set estimator $\hat{\Theta}_{2n}$ is defined as a level set of a random continuous function. Therefore, to ensure the measurability of the second-stage set estimator, it suffices to require that the criterion function is jointly measurable in (ω, θ).

For the two-stage set estimator $\hat{\Theta}_n$, however, we need a somewhat more careful treatment. $\hat{\Theta}_n$ is a level set of a random criterion function whose first argument is restricted to the first-stage set estimator $\hat{\Theta}_{1n}$. As $\hat{\Theta}_{1n}$ is also a random set, this introduces some complications.

As we show below, the measurability of $\hat{\Theta}_n$ is related to the Effros-measurability of the first-stage set estimator and the measurability of the criterion function over random sets. For this, we make use of results from Stinchcombe and White (1992) (SW)[5]. We can now state a general result for Effros-measurability of two-stage set estimators:

THEOREM 4.1: *(i) Let $(\Omega, \mathfrak{F}, P)$ be a complete probability space, and let $\Theta = \Theta_1 \times \Theta_2$ where Θ_1 and Θ_2 are nonempty compact subsets of finite-dimensional Euclidean space; (ii) Let $\zeta : \Omega \times \Theta_1 \times \Theta_2 \to \bar{\mathbb{R}}_+$ be such that $\zeta(\cdot, \theta_1, \theta_2)$ is measurable for each (θ_1, θ_2) in $\Theta_1 \times \Theta_2$ and $\zeta(\omega, \cdot, \cdot)$ is continuous on $\Theta_1 \times \Theta_2$ for each ω in $F \in \mathfrak{F}$ with $P(F) = 1$; (iii) Let $\hat{\Theta}_1 : \Omega \to \mathcal{F}(\Theta_1)$ be Effros-measurable.*

Then, for any measurable $\hat{\epsilon} : \Omega \to \mathbb{R}_+$, the $\hat{\epsilon}$-level set $\hat{\Theta} : \Omega \to \mathcal{F}(\Theta)$, defined by

$$\hat{\Theta}(\omega, \hat{\epsilon}(\omega)) = \{\theta : \zeta(\omega, \theta_1, \theta_2) \leq \hat{\epsilon}(\omega), \theta_1 \in \hat{\Theta}_1(\omega)\}.$$

is Effros-measurable with respect to $\mathfrak{F}$.

Suppose instead (iii′) $\hat{\theta}_1 : \Omega \to \Theta_1$ is measurable. Then

(a) For each $\theta_2 \in \Theta_2$, $\tilde{\zeta}(\cdot, \theta_2) := \zeta(\cdot, \hat{\theta}_1(\cdot), \theta_2)$ is a measurable function on Ω and for each $\omega \in F$, $\tilde{\zeta}(\omega, \cdot)$ is a continuous function on Θ_2;

[5] Details on the results of SW are given in the appendix.

(*b*) *For any measurable* $\hat{\epsilon} : \Omega \to \mathbb{R}_+$*, the* $\hat{\epsilon}$*-level set* $\hat{\Theta}_2 : \Omega \to \mathcal{F}(\Theta_2)$*, defined by*

$$\hat{\Theta}_2(\omega, \hat{\epsilon}(\omega)) = \{\theta_2 : \zeta(\omega, \hat{\theta}_1(\omega), \theta_2) \leq \hat{\epsilon}(\omega)\}$$

is Effros-measurable with respect to $\mathfrak{F}$*.*

These results yield Effros-measurability for our two- and second-stage estimators.

Corollary 4.1: *Suppose Assumptions 2.1 and 2.3 (Effros-measurability of* $\hat{\Theta}_{1n}$*) hold. Then for any measurable* $\hat{\epsilon}_n : \Omega \to \mathbb{R}_+$*, the two-stage estimator* $\hat{\Theta}_n$ *and the second-stage estimator* $\hat{\Theta}_{2n}$ *of Definition 2.4 are Effros-measurable.*

4.2 Consistency

In this section, we show that the two-stage set estimators of Definition 2.4 converge in probability to the identified set. The consistency is in terms of Hausdorff metric. For two closed sets A and B in $\mathcal{F}(\Theta)$, the Hausdorff metric is defined as

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

where $d(b, A) := \inf_{a \in A} \|b - a\|$ and $d_H(A, B) := \infty$ if either A or B is empty. The Hausdorff metric is standard in this context. The following theorem establishes the consistency of two-stage set estimators generally. For this, recall that $\{x\}_+ := \max\{x, 0\}$ for any $x \in \mathbb{R}$.

Theorem 4.2: *(i) Let* $(\Omega, \mathfrak{F}, P)$ *and* $\Theta = \Theta_1 \times \Theta_2$ *satisfy the conditions of Theorem 4.1, and suppose that for* $n = 1, 2, \cdots$, Q_n *and* $\hat{\Theta}_{1n}$ *satisfy the conditions on* ζ *and* $\hat{\Theta}_1$ *in Theorem 4.1; (ii) Suppose there exists* $\bar{Q} : \Theta \to \bar{\mathbb{R}}_+$ *such that* $\sup_\Theta\{\bar{Q}(\theta) - Q_n(\theta)\}_+ = o_p(1)$*. Let* $\Theta_{I,1} \in \mathcal{F}(\Theta_1)$ *and define*

$$\bar{\Theta}_I := \underset{\Theta_{I,1} \times \Theta_2}{\arg\min}\, \bar{Q}(\theta).$$

(iii) Let $\{a_n\}$ *be a sequence of normalizing constants such that* $a_n \to \infty$*; (iv) Let* $\hat{\epsilon}_n$ *be such that* $\hat{\epsilon}_n / a_n = o_p(1)$ *and*

$$\lim_{n \to \infty} P\left(\omega : \sup_{\theta \in \bar{\Theta}_I} Q_n(\omega, \theta) \leq \hat{\epsilon}_n(\omega)/a_n\right) = 1.$$

(v) Suppose further that $d_H(\hat{\Theta}_{1n}, \Theta_{I,1}) = o_p(1)$*,* $P(\Theta_{I,1} \subseteq \hat{\Theta}_{1n}) \to 1$*, and let*

$$\hat{\Theta}_n(\omega) = \{\theta : a_n Q_n(\omega, \theta_1, \theta_2) \leq \hat{\epsilon}_n(\omega), \theta_1 \in \hat{\Theta}_{1n}(\omega)\}.$$

Then $\hat{\Theta}_n$ *is Effros-measurable with respect to* $\mathfrak{F}$*, and* $d_H(\hat{\Theta}_n, \bar{\Theta}_I) = o_p(1)$*.*

Note that the estimated set $\bar{\Theta}_I$ need not correspond to the identified set Θ_I, as this result does not assume two-stage structure. Nevertheless, $\bar{\Theta}_I = \Theta_I$ when $\bar{Q}$ does have two-stage structure. In the absence of two-stage structure, the result above could be useful for partial identification analysis of profile estimators, like the EM algorithm (Dempster, Laird, Rubin, et al., 1977), which iterate between estimating one subset of parameters and another.

When θ_1^0 is point identified, the second-stage set estimator $\hat{\Theta}_{2n}$ is consistent for the identified set $\Theta_{I,2}$:

Corollary 4.2: *(i) Let the conditions of Theorem 4.2 hold, and suppose that $\Theta_{I,1}$ is a singleton, $\Theta_{I,1} = \{\theta_1^0\}$; (ii) Let*

$$\Theta_{I,2} := \underset{\theta_2 \in \Theta_2}{\arg\min}\ \bar{Q}(\theta_1^0, \theta_2);$$

(iii) Let $\hat{\theta}_{1n} : \Omega \to \Theta_1$ be $\mathfrak{F}$-measurable such that $\hat{\theta}_{1n} - \theta_1^0 = o_p(1)$, and let

$$\hat{\Theta}_{2n}(\omega) := \left\{ \theta_2 : a_n Q_n(\omega, \hat{\theta}_{1n}(\omega), \theta_2) \leq \hat{\epsilon}_n(\omega) \right\}.$$

Then $\hat{\Theta}_{2n}$ is Effros-measurable with respect to $\mathfrak{F}$, and $d_H(\hat{\Theta}_{2n}, \Theta_{I,2}) = o_p(1)$.

To apply these results, we add a one-sided uniform convergence assumption.

Assumption 4.1: $\sup_{\Theta}\{\bar{Q}(\theta) - Q_n(\theta)\}_+ = o_p(1)$.

This assumption holds by any of a variety of uniform laws of large numbers. We can now obtain the Hausdorff consistency of our two- and second-stage estimators.

Corollary 4.3: *Suppose Assumptions 2.1 (i) and 2.2 hold. Then $\bar{\Theta}_I = \Theta_I$. If Assumptions 2.1 (ii), 2.3, and 4.1 also hold and $d_H(\hat{\Theta}_{1n}, \Theta_{I,1}) = o_p(1)$, then for any measurable $\hat{\epsilon}_n : \Omega \to \mathbb{R}_+$ such that $\hat{\epsilon}_n / a_n = o_p(1)$ and $\lim_{n\to\infty} P(\sup_{\Theta_I} Q_n(\theta) \leq \hat{\epsilon}_n / a_n) = 1$, the two-stage estimator $\hat{\Theta}_n$ of Definition 2.4 is consistent: $d_H(\hat{\Theta}_n, \Theta_I) = o_p(1)$. If $\Theta_{I,1} = \{\theta_1^0\}$ and $\hat{\theta}_{1n} : \Omega \to \Theta_1$ is $\mathfrak{F}$-measurable such that $\hat{\theta}_{1n} - \theta_1^0 = o_p(1)$, then the second-stage estimator $\hat{\Theta}_{2n}$ of Definition 2.4 is consistent: $d_H(\hat{\Theta}_{2n}, \Theta_{2,I}) = o_p(1)$.*

4.3 First-Stage Set Estimation

In this section, we explicitly consider the estimation of $\Theta_{I,1}$. If $\Theta_{I,1} = \Theta_{I,1}^u$, then one can estimate the first-stage identified set using existing set-estimation techniques such as that of CHT. It remains, however, to study set estimation with the *a priori* restrictions considered in Section 2. We therefore take an estimator $\hat{\Theta}_{1n}^u$ of $\Theta_{I,1}^u$ as given and show that the restrictions can be also incorporated into the extremum estimation framework. The next theorem establishes the Effros-measurability and consistency of the first-stage set estimator.

THEOREM 4.3: *(i) Let* $(\Omega, \mathfrak{F}, P)$ *be a complete probability space, and let* Θ_1 *be a nonempty compact subset of finite-dimensional Euclidean space; (ii)* $s : \Theta_1 \times \Psi \to \mathbb{R}^{m_2}$ *is a continuous function; (iii)* $\Theta_1^r = \{\theta_1 \in \Theta_1 : s(\theta_1, \psi_0) = 0\}$*; (iv) there is a point estimator* $\hat{\psi}_n : \Omega \to \Psi$ *that is* $\mathfrak{F}$*-measurable; (v) For a positive sequence* $\{\alpha_n\}$ *and a non-negative sequence* $\{\eta_n\}$*,* $\hat{\Theta}_{1n}^r = \{\theta_1 \in \Theta_1 : \|s(\theta_1, \hat{\psi}_n)\|^2 \leq \eta_n/\alpha_n\}$*; (vi)* $\hat{\Theta}_{1n}^u$ *is Effros-measurable; (vii) Restriction 2.1 holds; (viii)* $\{\alpha_n\}$ *and* $\{\eta_n\}$ *satisfy* $\sup_{\theta_1 \in \Theta_1} \|s(\theta_1, \hat{\psi}_n)\|^2 = O_p(1/\alpha_n)$ *and* $P(\sup_{\theta_1 \in \Theta_1} \alpha_n \|s(\theta_1, \hat{\psi}_n)\|^2 \leq \eta_n) \to 1$*; (ix)* $\sup_{\theta_1 \in \Theta_1} \|s(\theta_1, \hat{\psi}_n) - s(\theta_1, \psi_0)\| = o_p(1)$*; (x)* $d_H(\hat{\Theta}_{1n}^u, \Theta_{I,1}^u) = o_p(1)$ *and* $P(\Theta_{I,1}^u \subseteq \hat{\Theta}_{1n}^u) \to 1$*. Then*

(a) If conditions (i)-(vi) hold, $\hat{\Theta}_{1n} := \hat{\Theta}_{1n}^u \cap \hat{\Theta}_{1n}^r$ *is Effros-measurable, i.e. Assumption 2.3 holds.*

(b) If conditions (i)-(x) hold, $d_H(\hat{\Theta}_{1n}, \Theta_{I,1}) = o_p(1)$ *and* $P(\Theta_{I,1} \subseteq \hat{\Theta}_{1n}) \to 1$*.*

Here, we also employ CHT's framework to construct the set estimator $\hat{\Theta}_{1n}^r$. Conditions (vii)-(ix) are imposed so that $\hat{\Theta}_{1n}^r$ is consistent in Hausdorff metric to Θ_1^r. Condition (ix) is satisfied, for example, if $\hat{\psi}_n$ is a consistent estimator of ψ_0 and s is Lipschitz continuous in ψ, i.e., for each $\theta_1 \in \Theta_1$, $|s(\theta_1, \psi) - s(\theta_1, \psi')| \leq M(\theta_1)\|\psi - \psi'\|$ for a bounded function $M : \Theta_1 \to \mathbb{R}$.

5 Two-Stage Inference using the Quasi-Likelihood Ratio Statistic

Set estimation is useful when interest focuses on the properties of the identified set. If instead one wishes to test hypotheses regarding the identified set, it is not necessary to estimate it. In this section, we discuss hypothesis testing based on a quasi-likelihood ratio statistic when the first-stage parameter is point identified[6].

Let R be a closed subset of Θ_2, where R is a set of parameter values that satisfy the restrictions of interest. As the true second-stage parameter value θ_2^0 is in the identified set, if θ_2^0 satisfies the given restrictions, the identified set has nonempty intersection with R. We can thus consider the hypothesis

$$H_0^{\Theta_2} : \quad \Theta_{I,2} \cap R \neq \varnothing \text{ versus } \Theta_{I,2} \cap R = \varnothing. \tag{5.1}$$

Because R is a closed subset of the compact parameter space, this null hypothesis is equivalent to

$$H_0^{\Theta_2} : \inf_{\theta_2 \in \Theta_2 \cap R} \bar{Q}_n(\theta_1^0, \theta_2) = 0.$$

Such hypothesis is considered for the partially identified case in the single-stage context by Romano and Shaikh (2010) for parametric inference and by Santos (2012) for nonparametric

[6]More detailed discussion of the results in this section is available from the authors upon request.

inference.

To test the hypothesis in our two-stage framework, we replace $\bar{Q}_n$ and θ_0^1 with their sample analogs Q_n and $\hat{\theta}_{1n}$, which leads to the test statistic

$$\hat{T}_n(\Theta_2, R) = \inf_{\theta_2 \in \Theta_2 \cap R} a_n Q_n(\hat{\theta}_{1n}, \theta_2).$$

Below, we focus on the cases where Q_n takes the quasi-maximum likelihood form:

$$Q_n(\omega, \theta) = n^{-1} \sum_{i=1}^{n} q(X_i(\omega), \theta) - \inf_{\theta \in \Theta} \; n^{-1} \sum_{i=1}^{n} q(X_i(\omega), \theta). \tag{5.2}$$

When $q(x, \theta) = -\ln f(x, \theta)$, where $f(\cdot, \theta)$ is a probability density function for each θ, $\hat{T}_n(\Theta_2, R)$ can be viewed as a quasi-likelihood ratio statistic. We note here that this statistic can be used to conduct inference for both correctly specified and misspecified parametric models. Let $\mathcal{P} = \{P : dP/d\mu = f(\cdot, \theta), \theta \in \Theta\}$ be a parametric model, where μ is a σ-finite measure on $\mathcal{X}$. If the model is correctly specified, i.e., $P^0 \in \mathcal{P}$, the problem then reduces to inference based on a likelihood-ratio statistic with a partially identified parameter. When the first-stage parameter is not present, this type of problem has been studied in Liu and Shao (2003). Recently, Chen, Tamer, and Torgovitsky (2011) have further extended the framework of Liu and Shao (2003) to account for infinite dimensional parameters. On the other hand, if the model is misspecified, the set of minimizers of $\bar{Q}$ should be interpreted as the set of points that have the best approximation property to P^0 as discussed in Ponomareva and Tamer (2010). Here, the approximation is in terms of the Kullback-Leibler divergence (see White, 1994). Although we do not explicitly study here, another type of misspecification, namely local violations of model restrictions, may have impacts on inference. The asymptotic distortion of inference methods for partially identified models under such local misspecification has been studied in Bugni, Canay, and Guggenberger (2012).

Let $a_n = n$, a typical case. Then the statistic can be written as

$$\hat{T}_n(\Theta_2, R) = \inf_{\theta_2 \in R} \sum_{i=1}^{n} q(X_i(\omega), \hat{\theta}_{1n}, \theta_2) - \inf_{\theta_2 \in \Theta_2} \; \sum_{i=1}^{n} q(X_i(\omega), \hat{\theta}_{1n}, \theta_2). \tag{5.3}$$

When the first-stage parameter is not point identified, we may still be able to define a qusi-likelihood ratio statistic using a first-stage set estimator $\hat{\Theta}_{1n}$. However, the impact of the first-stage set estimation on the distribution of the test statistic is complicated for this general case, and its analysis is beyond the scope of this paper. This is why we focus on the statistic $\hat{T}_n(\Theta_2, R)$, where the first-stage parameter is fully identified. In this setting, we can often exploit a two-term mean-value expansion of the test statistic and write it as

$$\hat{T}_n(\Theta_2, R) = T_n(\Theta_2, R; \theta_1^0) + n g_n' \left(\hat{\theta}_{1n} - \theta_0^1\right) + n \left(\hat{\theta}_{1n} - \theta_0^1\right)' H_n \left(\hat{\theta}_{1n} - \theta_0^1\right)/2 + o_p(1),$$

where $T_n(\Theta_2, R; \theta_1^0)$ replaces $\hat{\theta}_{1n}$ in $\hat{T}_n(\Theta_2, R)$ with θ_0^1, $g_n = n^{-1}(\partial/\partial\theta_1)T_n(\Theta_2, R; \theta_1^0)$, and $H_n = n^{-1}(\partial^2/\partial\theta_1\partial\theta_1')T_n(\Theta_2, R; \theta_1^0)$[7]. Then, $\hat{T}_n(\Theta_2, R)$ converges weakly to a non-degenerate random variable T under regularity conditions.

Subsampling can be then used to estimate an appropriate critical value. For each $x \in \mathbb{R}$, let $F(x) := P(T \leq x)$ and let $c_{1-\alpha} := \inf\{x : F(x) \geq 1 - \alpha\}$. Let $b := b_n < n$ be a positive integer. Let $N_{n,b} = \binom{n}{b}$ denote the number of subsamples of size b from a sample of size n. For each $1 \leq k \leq N_{n,b}$, let $\hat{T}_{b,k}(\Theta_2, R)$ be the statistic computed from the k-th subsample of size b. The subsampling estimator $\hat{c}_{n,b,1-\alpha}$ of $c_{1-\alpha}$ is defined by

$$\hat{c}_{n,b,1-\alpha} := \inf\left\{x : \hat{F}_{n,b}(x) \geq 1 - \alpha\right\},$$

where $\hat{F}_{n,b}(x) := N_{n,b}^{-1} \sum_{1 \leq k \leq N_{n,b}} 1\{\hat{T}_{b,k}(\Theta_2, R) \leq x\}$. If the distribution of T is continuous at $c_{1-\alpha}$ and $b \to \infty$ and $b/n \to 0$ as $n \to \infty$, then $\hat{c}_{n,b,1-\alpha}$ can be shown to be consistent for $c_{1-\alpha}$, which ensures asymptotic validity of the test. Specifically, one can show that, under the null hypothesis,

$$\limsup_{n\to\infty} P(\hat{T}_n(\Theta_2, R) > \hat{c}_{n,b,1-\alpha}) \leq \alpha;$$

and the test is also consistent against any fixed alternative.

It should be noted, however, that this asymptotic validity is ensured only pointwise but not uniformly in the underlying distributions. Uniform validity of inference would be a desirable property in this context because the limiting distribution of $\hat{T}_n(\Theta_2, R)$ may be discontinuous in the underlying distribution. In some examples, the quasi-likelihood ratio statistic can be reduced to a statistic based on moment inequalities. In such a setting, it is well-known that inference procedures based on pointwise asymptotics are not guaranteed to provide a good approximation to the finite sample distribution of the test statistic. For this class of models, uniformly valid procedures have been recently developed; see Andrews and Soares (2010), Bugni (2010), and Andrews and Barwick (2012) among others. Further, our simulation results (not reported here) suggest that the test based on the subsampling critical value has severe size distortion when the identified set is close to a singleton. Therefore, care must be taken in applying our testing procedure. An important avenue for future work is to extend our inference framework and develop a uniformly valid procedure.

6 Concluding Remarks

This paper studies estimation and inference for a parameter vector that has a two-stage structure. Our procedure constructs a two-stage set estimator by taking an appropriate level

[7] This expansion is valid when $\Theta_{I,1} = \{\theta_0^1\}$, in which case the identified set $\Theta_I = \{\theta_1^0\} \times \Theta_{I,2}$ does not have non-empty interior. This implies that our setting may not belong to the class of problems where CHT's degeneracy condition (Condition C.3) holds.

set of a criterion function, using a first-stage estimator to impose restrictions on the parameter of interest. A special case of this estimator where the first-stage parameter is fully identified was considered, for example, in Bajari, Benkard, and Levin (2007), but its measurability has not been previously studied. We give conditions for the measurability of the two-stage set estimator and establish consistency of the two-stage estimator based on the results of Chernozhukov, Hong, and Tamer (2007).

A Mathematical Appendix

This appendix collects proofs for Theorems 4.1, 4.2, and 4.3. The proofs for Corollaries 4.1, 4.2, and 4.3 are omitted for brevity.

A.1 Proof of Theorem 4.1

To prove Theorem 4.1, we make use of some results from SW. We therefore introduce some of the concepts used in SW and give two lemmas for easy reference.

Let $(\Omega, \mathfrak{F})$ be a measurable space. A set $A \subset \Omega$ is said to be an *analytic subset* of Ω if there exists a compact metric space E, with its Borel σ-field $\mathfrak{E}$, such that A is the projection of some $B \in \mathfrak{F} \otimes \mathfrak{E}$. The collection of all analytic subsets of Ω is denoted by $\mathbf{A}(\mathfrak{F})$. A function $f : \Omega \to \bar{\mathbb{R}}$ is said to be *analytic* if for all $r \in \mathbb{R}$, $\{\omega | f(\omega) > r\} \in \mathbf{A}(\mathfrak{F})$. By Fact 2.9 in SW, all measurable functions are analytic. Conversely, all analytic functions are measurable with respect to every completion of $\mathfrak{F}$.

Lemma A.1 (Lemma 2.15 in SW): *Let $(\Omega, \mathfrak{F})$ be a measurable space and (H, d) a separable metric space with its Borel σ-algebra $\mathcal{H}$. If ζ is measurable on Ω and continuous on H, that is, for every $\omega \in \Omega, \zeta(\omega, \cdot) : H \to \bar{\mathbb{R}}$ is continuous and for every $h \in H, \zeta(\cdot, h) : \Omega \to \bar{\mathbb{R}}$ is measurable, then $\zeta : \Omega \times H \to \bar{\mathbb{R}}$ is $\mathfrak{F} \otimes \mathcal{H}$-measurable.*

Lemma A.2 (Theorem 2.17, a in SW): *(i) Let $(\Omega, \mathfrak{F})$ be a measurable space; (ii) Let $(H, \mathcal{H})$ be a Souslin measurable space, i.e. a space that is measurably isomorphic to an analytic subset of a compact metric space. (iii) Suppose $\zeta : \Omega \times H \to \bar{\mathbb{R}}$ is $\mathfrak{F} \otimes \mathcal{H}$-measurable; (iv) $S : \Omega \Rightarrow H$ is a correspondence from Ω to H with $grS \in \mathfrak{F} \otimes \mathcal{H}$, where grS is the graph of S*[8]*. Then the function $\zeta^* : \Omega \to \bar{\mathbb{R}}$ defined by*

$$\zeta^*(\omega) := \sup_{h \in S(\omega)} \zeta(\omega, h)$$

is analytic.

The result of Lemma A.2 is a bit more general than strictly necessary for our purposes. If $\mathfrak{F}$ is completed with respect to the probability measure P, then the results above imply the *measurability* of ζ^* with respect to the completed σ-algebra. This is the result we exploit to establish Effros-measurability. A closely related result was established by Debreu (1967) for mappings S that are non-empty and compact for all $\omega \in \Omega$ and for functions ζ such that $\zeta(\omega, \cdot)$ is lower semicontinuous for all $\omega \in \Omega$. Therefore, if the first stage set estimator is almost surely non-empty and if a criterion function Q_n is jointly measurable, we can relax the continuity assumption on Q_n and allow Q_n to be only lower semicontinuous.

[8] This condition is equivalent to the Effros-measurability of S when the parameter space H is a Polish space and the probability space is complete. See Theorem 1.2.3 in Molchanov (2005).

Proof of Theorem 4.1. For any $E \subseteq \Theta$, let $\hat{\Theta}^{-1}(E) := \{\omega : \hat{\Theta}(\omega, \hat{\epsilon}(\omega)) \cap E \neq \varnothing\}$. Recall that $\mathcal{F}(\Theta)$ is the collection of all closed subsets of Θ. Below, we establish the Effros-measurability of $\hat{\Theta}$ by showing

$$\hat{\Theta}^{-1}(F) \in \mathfrak{F}, \quad \forall F \in \mathcal{F}(\Theta). \tag{A.1}$$

Let $F \in \mathcal{F}(\Theta)$. If $F = \varnothing$, then trivially $\hat{\Theta}^{-1}(F) = \varnothing \in \mathfrak{F}$. Now suppose $F \neq \varnothing$. For any $\omega \in \Omega$, it follows that

$$\begin{aligned}
&\hat{\Theta}(\omega, \hat{\epsilon}(\omega)) \cap F \neq \varnothing \\
&\qquad\qquad \Leftrightarrow (\hat{\Theta}_1(\omega) \times \Theta_2) \cap F \neq \varnothing \\
&\qquad\qquad\quad \text{and } \exists (\theta_1, \theta_2) \in (\hat{\Theta}_1(\omega) \times \Theta_2) \cap F \text{ such that } \zeta(\omega, \theta_1, \theta_2) \leq \hat{\epsilon}(\omega) \\
&\qquad\qquad \Leftrightarrow \inf_{(\hat{\Theta}_1(\omega) \times \Theta_2) \cap F} \zeta(\omega, \theta_1, \theta_2) \leq \hat{\epsilon}(\omega),
\end{aligned}$$

where the second equivalence follows from the compactness of $(\hat{\Theta}_1 \times \Theta_2) \cap F$ and the continuity of $\zeta(\omega, \cdot, \cdot)$. For each $\omega \in \Omega$, let $R(\omega) := \hat{\Theta}_1(\omega) \times \Theta_2$ and $R_F(\omega) := R(\omega) \cap F$. Thus, we may write

$$\hat{\Theta}^{-1}(F) = \left\{\omega : \inf_{R_F(\omega)} \zeta(\omega, \theta_1, \theta_2) \leq \hat{\epsilon}(\omega)\right\}. \tag{A.2}$$

Therefore, it suffices to show that the infimum of the random function ζ over the random closed set R_F is measurable in usual sense. For this, we first show that R is Effros-measurable. Let $F_1 := \{\theta_1 \in \Theta_1 : (\theta_1, \theta_2) \in F \text{ for some } \theta_2 \in \Theta_2\}$ and note that F_1 is closed. Then, it follows that

$$\{\omega : R(\omega) \cap F \neq \varnothing\} = \{\omega : \hat{\Theta}_1(\omega) \cap F_1 \neq \varnothing\} \in \mathfrak{F}, \tag{A.3}$$

where the last inclusion holds because $\hat{\Theta}_1$ is Effros-measurable by assumption and the fact that the projection F_1 of F is also closed. This ensures that R is Effros-measurable. A similar argument shows R_F is also Effros-measurable.

For the measurability of $\inf_{R_F(\omega)} \zeta(\omega, \theta_1, \theta_2)$, we apply Lemma A.2. Condition (i) of Lemma A.2 is satisfied by our hypotheses. By assumption, Θ is a compact subset of a finite dimensional Euclidean space. Hence, it is closed and is therefore Borel measurable. This implies that Θ is a Lusin space. Since every Lusin space is also a Souslin space (p.498 in SW), Condition (ii) of Lemma A.2 is also satisfied. Assumption (ii) of Theorem 4.1 ensures that ζ is jointly measurable by Lemma A.1, which ensures condition (iii) of Lemma A.2. Condition (iv) of Lemma A.2 is equivalent to Effros-measurability of S by Theorem 1.2.3 in Molchanov (2005). Thus, for any $F \in \mathcal{F}(\Theta)$, R_F satisfies this condition. Lemma A.2 then implies the measurability of $\inf_{R_F(\cdot)} \zeta(\cdot, \theta_1, \theta_2)$.

For (a), this is immediate from Theorem 2.14 (i) in White (1994). Given this result, we can apply Lemma A.1 to establish the jointly measurability of $\tilde{\zeta}$. By Fact 4.1, $\tilde{\zeta}$ is a normal integrand. This ensures the measurability of its level sets. Therefore $\{\omega : \inf_{R_F(\omega)} \zeta(\omega, \theta_1, \theta_2) \leq \hat{\epsilon}(\omega)\} \neq \varnothing$. This and Eq. (A.3) imply that the right hand side of (A.2) is an element of $\mathfrak{F}$. Since F was arbitrary, the conclusion follows. □

A.2 Proof of Theorem 4.2

Proof of Theorem 4.2. To show $d_H(\hat{\Theta}_n, \bar{\Theta}_I) = o_p(1)$, we need both (a) $\sup_{\theta\in\bar{\Theta}_I} d(\theta, \hat{\Theta}_n) = o_p(1)$ and (b) $\sup_{\theta\in\hat{\Theta}_n} d(\theta, \bar{\Theta}_I) = o_p(1)$.

We first show (a). By the definition of $\hat{\Theta}_n$, it follows that

$$P(\bar{\Theta}_I \not\subseteq \hat{\Theta}_n) \leq P(\sup_{\theta\in\bar{\Theta}_I} a_n Q_n(\theta) > \hat{\epsilon}_n) + P(\Theta_{I,1} \not\subseteq \hat{\Theta}_{1n}). \tag{A.4}$$

Conditions (iv) and (v) ensure that the probabilities on the right hand side of (A.4) tend to 0 as $n \to \infty$. Thus, $\bar{\Theta}_I \subseteq \hat{\Theta}_n$ with probability approaching 1. This ensures that $\sup_{\theta\in\bar{\Theta}_I} d(\theta, \hat{\Theta}_n) = o_p(1)$.

For (b), we need to show that for any $\epsilon > 0$, $\sup_{\theta\in\hat{\Theta}_n} d(\theta, \bar{\Theta}_I) \leq \epsilon$ with probability approaching 1. This can be established by the uniform convergence of Q_n and the convergence of the first stage set estimator in Hausdorff metric. For this, let

$$\bar{\zeta}(\theta) := \bar{Q}(\theta) + d(\theta_1, \Theta_{I,1}),$$

where the second term takes a positive value if the first stage restriction $\theta_1 \in \Theta_{I,1}$ is violated. Note that $\bar{\Theta}_I = \arg\min_{\theta\in\Theta} \bar{\zeta}(\theta)$. Let

$$\zeta_n(\theta) := Q_n(\theta) + d(\theta_1, \hat{\Theta}_{1n}).$$

Define the $\hat{\epsilon}_n$-level set $\tilde{\Theta}_n = \{\theta : a_n\zeta_n(\theta) \leq \hat{\epsilon}_n\}$. We first show that $\sup_{\tilde{\Theta}_n} d(\theta, \bar{\Theta}_I) \leq \epsilon$.

Note that for any $x, y \in \mathbb{R}$, $\{x + y\}_+ \leq \{x\}_+ + \{y\}_+$. This implies

$$\begin{aligned}\sup_{\theta\in\Theta} \left\{\bar{\zeta}(\theta) - \zeta_n(\theta)\right\}_+ &\leq \sup_{\theta\in\Theta} \left\{\bar{Q}(\theta) - Q_n(\theta)\right\}_+ + \sup_{\theta_1\in\Theta_1} \{d(\theta_1, \Theta_{I,1}) - d(\theta_1, \hat{\Theta}_{1n})\}_+ \\ &\leq \sup_{\theta\in\Theta} \left\{\bar{Q}(\theta) - Q_n(\theta)\right\}_+ + d_H(\Theta_{I,1}, \hat{\Theta}_{1n}) = o_p(1),\end{aligned}$$

where the second inequality holds since $\Theta_{I,1}$ and $\hat{\Theta}_{1n}$ are closed under our assumptions and by Proposition C.7 of Molchanov (2005).

Let $\delta > 0$ and $A_n := \{\omega : \sup_{\theta\in\Theta}\{\bar{\zeta}(\theta) - \zeta_n(\theta)\}_+ < \delta/2, \text{ and } \hat{\epsilon}_n/a_n < \delta/2\}$. Note that $P(A_n) \to 1$ as $n \to \infty$. Let $\omega \in A_n$. Then, for any $\theta \in \Theta$, $\bar{\zeta}(\theta) < \zeta_n(\omega, \theta) + \delta/2$. Taking the supremum over $\tilde{\Theta}_n(\omega)$, we obtain

$$\begin{aligned}\sup_{\theta\in\tilde{\Theta}_n(\omega)} \bar{\zeta}(\theta) &< \sup_{\theta\in\tilde{\Theta}_n(\omega)} \zeta_n(\omega,\theta) + \delta/2 \\ &\leq \hat{\epsilon}_n/a_n + \delta/2 \\ &\leq \delta.\end{aligned}$$

Recall that $\bar{\zeta} > 0$ outside $\bar{\Theta}_I$. Therefore, for any $\epsilon > 0$, there exists N_ϵ such that

$$P\Big(\sup_{\tilde{\Theta}_n(\omega)} \bar{\zeta}(\theta) < \delta < \inf_{\Theta\setminus\bar{\Theta}_I^\epsilon} \bar{\zeta}(\theta)\Big) \geq 1-\epsilon, \quad \forall n \geq N_\epsilon.$$

This implies $\tilde{\Theta}_n \cap (\Theta \setminus \bar{\Theta}_I^\epsilon) = \varnothing$ with probability approaching 1. Therefore, for any $\epsilon > 0$, $\sup_{\theta\in\tilde{\Theta}_n} d(\theta,\bar{\Theta}_I) \leq \epsilon$ with probability approaching 1. Note that for each $\theta \in \hat{\Theta}_n$, $d(\theta_1,\hat{\Theta}_{1n}) = 0$, which implies $\hat{\Theta}_n \subseteq \tilde{\Theta}_n$ for any $\omega \in \Omega$. Therefore, $\sup_{\theta\in\hat{\Theta}_n} d(\theta,\bar{\Theta}_I) \leq \epsilon$ with probability approaching 1.

Combining steps (a) and (b), we conclude that $d_H(\hat{\Theta}_n,\bar{\Theta}_I) = o_p(1)$. □

A.3 Proof of Theorem 4.3

In order to prove Theorem 4.3 (ii), we use the fact that convergence in Hausdorff metric is equivalent to the general notion of set convergence called Painlevé-Kuratowski convergence (PK-convergence) when the parameter space is bounded. We also use a lemma that is instrumental for checking whether a given sequence of sets converges in the Painlevé-Kuratowski sense. For easy reference we give the definition of PK-convergence and the lemma below. Please see section 4.C in Rockafellar and Wets (2005) (RW) for details.

Definition A.1 (PK convergence): *A sequence $\{F_n, n \geq 1\}$ of subsets of $\mathbb{E}$ is said to converge to F in the Painlevé-Kuratowski sense if*

$$\liminf_{n\to\infty} F_n = \limsup_{n\to\infty} F_n = F,$$

where

$$\begin{aligned}\liminf_{n\to\infty} F_n &:= \{x \in \mathbb{E} : \exists\{x_n\}, x_n \to x \text{ and } x_n \in F_n, \forall n\} \\ \limsup_{n\to\infty} F_n &:= \{x \in \mathbb{E} : \exists\{x_{n_k}, F_{n_k}\}, x_{n_k} \to x \text{ and } x_{n_k} \in F_{n_k}, \forall k\}.\end{aligned}$$

We write[9] $F_n \overset{PK}{\to} F$ *or* $PK - \lim F_n = F$.

[9]Since we always have $\liminf F_n \subseteq \limsup F_n$, the condition for PK convergence can be restated as $\limsup_{n\to\infty} F_n \subseteq A \subseteq \liminf_{n\to\infty} F_n$

Lemma A.3 (Hit or miss criteria: Theorem 4.5 in RW): *Let $\mathbb{E}$ be a locally compact Hausdorff second countable space. For $F_n, F \subseteq \mathbb{E}$ with $F \in \mathcal{F}(\mathbb{E})$, one has*

1. *$F \subseteq \liminf_{n\to\infty} F_n$ iff for every open set G with $F \cap G \neq \varnothing$, one has $F_n \cap G \neq \varnothing$ for all sufficiently large n.*

2. *$\limsup_{n\to\infty} F_n \subseteq F$ iff for every compact set K with $F \cap K = \varnothing$, one has $F_n \cap K = \varnothing$ for all sufficiently large n.*

Proof of Theorem 4.3. For (a), we again consider intersections with closed sets. Let ω be given. For each $c \geq 0$ and $F \in \mathcal{F}(\Theta_1)$, let

$$G_{c,F}(\omega) := \{\theta_1 \in \Theta_1 : \|s(\theta_1, \hat{\psi}_n(\omega))\|^2 \leq c\} \cap F.$$

Let $\mathcal{Q}_n : \Omega \times \Theta_1 \to \mathbb{R}$ be the measurable map defined by $(\omega, \theta_1) \mapsto \|s(\theta_1, \hat{\psi}_n(\omega))\|^2$. By conditions (i), (ii), (iv), and Lemma A.1, $\mathcal{Q}_n$ is jointly measurable; following the proof of Example 3.1 in Stinchcombe and White (1992), we can show $gr(G_{\eta_n/\alpha_n,F}) = \mathcal{Q}_n^{-1}((-\infty, \eta_n/\alpha_n]) \cap (\Omega \times F)$. Therefore, $gr(G_{\eta_n/\alpha_n,F}) \in \mathfrak{F} \otimes \mathcal{B}_{\Theta_1}$ for any $F \in \mathcal{F}(\Theta_1)$. This is equivalent to the Effros-measurability of $\hat{\Theta}^r_{1n}$ by Theorem 1.2.3 in Molchanov (2005). This implies the Effros-measurability of $\hat{\Theta}_{1n}$.

We now show (b). For Hausdorff consistency of $\hat{\Theta}_{1n}$, we first show that $d_H(\hat{\Theta}^r_{1n}, \Theta^r_1) = o_p(1)$ and $P(\hat{\Theta}^r_{1n} \subseteq \Theta^r_1) \to 1$. For this, we invoke Theorem 3.1 in CHT. Let $\mathcal{Q} : \Theta_1 \to \mathbb{R}$ be defined by $\mathcal{Q}(\theta_1) := \|s(\theta_1, \psi_0)\|^2$. Conditions (i)-(iii) ensure that Θ_1 is compact, $\mathcal{Q}$ is lower semicontinuous, and Θ^r_1 is the set of minimizers of $\mathcal{Q}$. The joint measurability of $\mathcal{Q}_n$ is established above. Condition (ix) implies $\sup_{\theta_1 \in \Theta_1}\{\mathcal{Q}(\theta_1) - \mathcal{Q}_n(\theta_1)\}_+ = o_p(1)$. Condition (viii) then ensures the rest of conditions required by Theorem 3.1 (1) in CHT. Hence, $d_H(\hat{\Theta}^r_{1n}, \Theta^r_1) = o_p(1)$ and $P(\hat{\Theta}^r_{1n} \subseteq \Theta^r_1) \to 1$.

We now show that $\hat{\Theta}_{1n}$ is Hausdorff consistent using Definition A.1 and Lemma A.3. Let $\{n'\}$ be a subsequence of $\{n\}$. Since $\hat{\Theta}^u_{1n}$ is Hausdorff consistent for $\Theta^u_{I,1}$, there is a further subsequence $\{n''\}$ of $\{n'\}$ such that $\hat{\Theta}^u_{1n''} \stackrel{PK}{\to} \Theta^u_{I,1}$ almost surely. By Lemma A.3, for every open set G with $\Theta^u_{I,1} \cap G \neq \varnothing$, it then holds that $\hat{\Theta}^u_{1n''} \cap G \neq \varnothing$ for n'' sufficiently large with probability 1. Similarly, by Hausdorff consistency of $\hat{\Theta}^r_{1n}$ established above and Lemma A.3, we can then find a further subsequence $\{n'''\}$ of $\{n''\}$ such that for every open set G with $\Theta^r_1 \cap G \neq \varnothing$, $\hat{\Theta}^r_{1n'''} \cap G \neq \varnothing$ for n''' sufficiently large. Then along this subsequence, for every open set G with $(\Theta^u_{I,1} \cap \Theta^r_1) \cap G \neq \varnothing$, we have $\hat{\Theta}^u_{1n'''} \cap \hat{\Theta}^r_{1n'''} \cap G \neq \varnothing$ for n''' sufficiently large with probability 1. A similar argument ensures that there is a further subsequence $\{n'''\}$ of $\{n'\}$ such that, for every compact set K with $(\Theta^u_{I,1} \cap \Theta^r_1) \cap K = \varnothing$, we have $(\hat{\Theta}^u_{1n'''} \cap \hat{\Theta}^r_{1n'''}) \cap K = \varnothing$ for n''' sufficiently large with probability 1.

Now we have shown that every subsequence of $\hat{\Theta}_{1n}$ has a further subsequence $\{\hat{\Theta}_{1n'''}\}$, which satisfies the hit-or-miss criteria almost surely. Therefore $d_H(\hat{\Theta}_{1n'''}, \Theta_{I,1}) = o_{as}(1)$; but this also implies $d_H(\hat{\Theta}_{1n}, \Theta_{I,1}) = o_p(1)$, see e.g. Lukacs (1975) Theorem 2.4.4.

The claim $P(\hat{\Theta}_{I,1} \subseteq \hat{\Theta}_{1n}) \to 1$ follows from condition (x), $P(\hat{\Theta}^r_{1n} \subseteq \Theta^r_1) \to 1$ established above, and Bonferroni's inequality. $\square$

References

Andrews, D. W. K., and P. J. Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80(6), 2805–2826.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters defined by Moment Inequalities using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Bajari, P., C. L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75(5), 1331–1370.

Bugni, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78(2), 735–753.

Bugni, F. A., I. Canay, and P. Guggenberger (2012): "Distortions of Asymptotic Confidence Size in Locally Misspecified Moment Inequality Models," *Econometrica*, 80(4), 1741–1768.

Chen, X., E. Tamer, and A. Torgovitsky (2011): "Sensitivity Analysis in Semiparametric Likelihood Models," Cowles Foundation Discussion Paper 1836.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791–1828.

Cochrane, J., and J. Saá-Requejo (2000): "Beyond arbitrage: Good-deal asset price bounds in incomplete markets," *Journal of Political Economy*, 108(1), 79–119.

Debreu, G. (1967): "Integration of Correspondences," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of Calififornia Press, II-1 pp.351-372.

Dempster, A., N. Laird, D. Rubin, et al. (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–38.

Kaido, H., and H. White (2009): "Inference on Risk Neutral Measures for Incomplete Markets," *Journal of Financial Econometrics*, 7(3), 199–246.

Liu, X., and Y. Shao (2003): "Asymptotics for Likelihood Ratio Tests under Loss of Identifiability," *The Annals of Statistics*, 31(3), 807–832.

Lukacs, E. (1975): *Stochastic Convergence*. New York: Academic Press.

Manski, C., and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546.

Manski, C. F. (2003): *Partial Identification of Probability Distributions (Springer Series in Statistics)*. Springer, 1st edn.

Molchanov, I. (2005): *Theory of Random Sets*. Berlin: Springer.

Ponomareva, M., and E. Tamer (2010): "Misspecification in Moment Inequality Models: Back to Moment Equalities?," *Econometrics Journal*, 10, 1–21.

Rockafellar, R. T., and R. J.-B. Wets (2005): *Variational Analysis*. Berlin: Springer, 2nd edn.

Romano, J. P., and A. Shaikh (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78(1), 169–211.

Santos, A. (2012): "Inference in nonparametric instrumental variables with partial identification," *Econometrica*, 80(1), 213–275.

Stinchcombe, M., and H. White (1992): "Some Measurability Results for Extrema of Random Functions Over Random Sets," *The Review of Economic Studies*, 59(3), 495–514.

van der Vaart, A., and J. Wellner (2000): *Weak Convergence and Empirical Processes: With Applications to Statistics*. Berlin: Springer.

White, H. (1994): *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.