UNIVERSITY OF CALIFORNIA, SAN DIEGO

Inference for Partially Identified Economic Models

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Hiroaki Kaido

Committee in charge:

Professor Halbert White, Chair
Professor Ivana Komunjer
Professor Dimitris Politis
Professor Andres Santos
Professor Allan Timmermann
Professor Rossen Valkanov

2010

The dissertation of Hiroaki Kaido is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2010

DEDICATION

To my wife Hitomi.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, I am deeply grateful to my advisor Halbert White for his guidance and constant encouragement throughout my graduate studies at the University of California, San Diego. I consider myself very fortunate to be advised by and to work with one of the greatest thinkers in econometrics. His insights and enthusiasm for academic research have always impressed me. I also thank him for letting me learn his rigorous approach to econometrics, which I would like to emulate throughout my academic career.

I would also like to acknowledge other dissertation committee members. Ivana Komunjer introduced me to the field of partial identification when I was in my second year, which helped me develop some of the key ideas in this dissertation. I also thank Allan Timmermann and Rossen Valkanov who provided valuable guidance on both theoretical and applied aspects of Chapter 1. For the development of Chapter 3, I have benefitted greatly from discussing various aspects of this chapter with Andres Santos. I would like to thank him for his comments and his patience with me. Dimitris Politis also provided helpful comments on this chapter.

Chapter 1, in full, is a reprint of the material as it appears in the *Journal of Financial Econometrics* 2009. Kaido, Hiroaki; White, Halbert, Oxford University Press, 2009. I was the primary investigator and author of this paper. I would like to thank anonymous referees and the editors of the journal: Eric Renault and René Garcia.

Chapter 2 is currently being prepared for submission for publication of the material. Kaido, Hiroaki; White, Halbert. I was the primary investigator and author of this material.

I am indebted to my former mentors Makoto Saito, Yoshiro Tsutsui, Tokuo Iwaisako, and Tatsuyoshi Saijo. I thank them for advising me at earlier stages of my studies and encouraging me to pursue my graduate studies in the U.S.

My graduate studies at UCSD have been supported by the Nakajima Foundation. Thanks to them, I could devote my time and energy into my research. I hereby gratefully acknowledge their generous support for the past five years.

I went through the tough coursework together with my friends and have shared lots of fun times with them since then. I especially thank Danielken Molina and Alberto Rossi for their friendship. My thanks also go to Ayelen Banegas, Jess Diamond, Dalia Ghanem, Youjin Hahn, Chulyoung Kim, Daniel Lima, Aren Megerdichian, Philip Neary, Sarada, Lucas Siga and Hee-Seung Yang. I have been very fortunate because many friends have been sharing their academic interests with me. I thank Tom Daula, Ben Fissel, Ben Gillen, Min Seong Kim, Suyong Song, and Xun Lu for spending time with me to discuss various ideas. I would like to also thank Kazuhiro Ikeda, Munechika Katayama, Tatsuyoshi Okimoto, Ryo Sugihara and their families for their support.

I came to San Diego alone as a new graduate student in 2005. After five years, I am now completing this thesis with the support of my family. I would like to thank my wife Hitomi for her love, her continuous support, and her decision to come to the U.S. to start living together with me in 2008. That is why I dedicate this thesis to her. My daughter Misaki, born in February, 2010, has brought a lot of happiness and an exciting change to my life. I have also been encouraged by many people in Japan. I especially thank my brother Toshiaki and my parents in law Jiro and Noriko Ito. Finally, I would like to thank my parents Nobuchika and Nobuko Kaido for their continuous and unlimited support.

VITA AND PUBLICATIONS

1980	Born, Hyogo, Japan
2003	B. A. in Economics, Osaka University, Osaka
2005	M. A. in Economics, Hitotsubashi University, Tokyo
2010	Ph. D. in Economics, University of California, San Diego

Hiroaki Kaido, Halbert White, "Inference on Risk Neutral Measures for Incomplete Markets", *Journal of Financial Econometrics*, 7, 2009.

Hiroaki Kaido, "Re-evaluating Consumption CAPM: Evidence from Pseudo-Cohort Data on Household Asset Holdings in Japan", *Gendai Finance*, 21, 2007.

ABSTRACT OF THE DISSERTATION

Inference for Partially Identified Economic Models

by

Hiroaki Kaido

Doctor of Philosophy in Economics

University of California, San Diego, 2010

Professor Halbert White, Chair

When a sample of data does not fully reveal the "true" data generating structure (or parameter) but gives information that bounds the set of observationally equivalent structures, an economic model is said to be *partially identified*. This dissertation develops and applies estimation and inference methods for economic models whose population features are only partially identified.

In Chapter 1 (co-authored with Halbert White), I apply econometric techniques from the partial identification literature to address a fundamental problem in asset pricing theory. Namely, that the market price of risk is only identified as a set under incomplete markets. I construct a set estimator and confidence regions for the set of market risk prices. I further show that it is possible to test hypotheses of economic interest without fully identifying the market price of risk.

The econometric techniques used in Chapter 1 are developed by Chapter 2 (co-authored with Halbert White). When the dimension of the parameter space is large, this is a particular challenge for set-valued estimators, as high dimensionality can create computational difficulties and seriously hamper the interpretation of estimation results. We study how the use of a natural two-stage extension of the Chernozhukov, Hong, and Tamer's (2007) (CHT) framework can exploit *a priori* knowledge about the data generating process to mitigate the problems otherwise associated with set estimation in high-dimensional parameter spaces.

Chapter 3 unifies two general approaches recently proposed in the literature, the *criterion function approach* and *support function approach*. CHT develop a theory of set estimation and inference for the set Θ_I of parameter values that minimize a criterion function $Q(\theta)$. The support function approach provides an alternative characterization of CHT's level-set estimator by its supporting hyperplanes. This results in an estimation and inference method that has the wide applicability of the criterion function approach and the computational tractability of the support function approach. By establishing the asymptotic distribution of the properly normalized support function of the level set estimator, I provide Wald-type inference tools to conduct tests regarding the identified set Θ_I and a point θ_0 in the identified set.

Chapter 1

Inference on Risk-Neutral Measures for Incomplete Markets

1.1 Introduction

For a continuous time, continuous state model, Harrison and Kreps (1979) have shown the equivalence between the absence of arbitrage and the existence of Q, the risk neutral probability measure (henceforth RNP), which is absolutely continuous with respect to the data generating measure. Subsequently, Delbaen and Schachermayer (1994) have shown that a condition called "no free lunch with vanishing risk" is equivalent to the existence of an RNP, which is mutually absolutely continuous with respect to the data generating measure. This result is known as the first fundamental theorem of financial economics (1st FTFE). In general, in the absence of arbitrage, the price of a financial asset can be computed simply as the expected value of its payoffs under the risk neutral probability, discounted by the risk-free rate. By comparing the RNP Q to the actual data-generating probability measure (DGP) P, one can recover agents' attitude toward risk.

Another object closely related to the RNP is the stochastic discount factor (SDF), also known as the pricing kernel. Harrison and Kreps (1979) and Harrison and Pliska (1981) show that the existence of the SDF is also equivalent to the absence of arbitrage. Further, they show that the uniqueness of the RNP (equivalently SDF) is equivalent to market completeness. This is known as the second fundamental theorem of financial economics (2nd FTFE).

There is a rich literature on the estimation of the SDF. The SDF depends generally on the state variables driving asset prices. Financial economists and macroeconomists have shown that a specific functional form for the SDF can be derived from the equilibrium prices generated by rational economic agents for assets with given payoff streams. Well known examples are the CAPM studied by Sharpe (1964) and Lintner (1965) and the consumption CAPM studied by Lucas (1978) and Breeden (1979). The state variables determining the SDF in these examples are the tangent portfolio return and aggregate consumption. The standard approach is to estimate parameters associated with this function and test whether the estimated SDF can price assets correctly or not. One drawback of this approach is the requirement of observable state variables. If the state variables are measured only poorly, this directly affects the bias and precision of estimators and the level and power of tests. Further, the functional form implied by the economic model need not be correctly specified; misspecification has similar adverse effects.

Recent studies (e.g., Aït-Sahalia and Lo, 1998; Chernov and Ghysels, 2000; and Rosenberg and Engle, 2002) show that one can estimate the RNP using only asset prices. These are usually measured very precisely. Further, high frequency data are often available. These rich data sets make possible the use of nonparametric techniques that can avoid the potential misspecification problem. So far, the literature has focused on estimating on a single risk neutral probability measure (or SDF) projected on price information. Aït-Sahalia and Lo (1998) nonparametrically estimate the RNP density. Chernov and Ghysels (2000) propose a method to estimate parameters associated with the RNP and the actual DGP jointly, using a time series of asset returns and option prices. Rosenberg and Engle (2002) estimate a unique RNP projected on the asset returns.

When markets are incomplete, there exists a set $\mathcal{Q}_I$ of RNPs identified by the distribution of observed asset prices. $\mathcal{Q}_I$ is identified in the sense that any of its elements generates the same distribution of observed asset prices. That is, there are multiple observationally equivalent economic structures Q. In this case, the economic structure is only *partially identified* by the observed data. The study of partial identification was pioneered by Charles Manski; see, e.g., Manski (2003). In this paper, we contribute to the finance literature by applying the techniques of partial identification to develop methods of estimation and inference for the set of RNPs $\mathcal{Q}_I$ identified by a given vector of asset prices without assuming market completeness nor using projection methods.

Our specific focus here is on the vector of time t market prices of risk, λ_t, a key element of the Girsanov transformation. In the absence of arbitrage, λ_t exists but is not uniquely identified by the asset price process when markets are incomplete. Instead, λ_t belongs to an identified set $\Lambda_{I,t}$ associated with $\mathcal{Q}_I$. By further imposing a bound on $\|\lambda_t\| := (\lambda_t'\lambda_t)^{1/2}$, we obtain an identified set denoted $\Lambda_{I,t}^M$. We then show that $\Lambda_{I,t}^M$ can be represented in terms of a set of minimizers of a certain criterion function. This enables us to apply the extremum set-estimation approach of Chernozhukov, Hong, and Tamer (2007) (henceforth the

CHT framework) to construct a set estimator and a confidence region for $\Lambda_{I,t}^{M}$ and to conduct hypothesis tests. In this application, we first concentrate out diffusion parameters and apply a two-stage procedure introduced by Kaido and White (2008) that helps to reduce the dimension of the associated set-valued estimators. To the best of our knowledge, this is the first application of such a procedure.

For concreteness, we pay particular attention to the case in which a standard multivariate geometric Brownian motion determines the evolution of asset prices. In this case, $\lambda_t = \lambda_0$, a non-random and time-invariant vector. As our applications show, there are cases where unspanned Brownian motions represent relevant risks (e.g. international aggregate risk). When the researcher's interest attaches to how those risks could be potentially evaluated, the RNP (or SDF) projected on asset prices does not provide enough information. In such cases, we need a tool to conduct tests on hypothesis regarding unspanned risks. The multivariate Black-Scholes setting is a simple starting point for illustrating the idea of this new tool.

If the projected RNP (or SDF) does not provide enough information. There could be two ways to proceed. One could fully specify the agent's preference and the endowment process to uniquely determine the equilibrium RNP, but as we will show, this is not strongly necessary for conducting statistical inference on features of the RNPs. Our alternative approach has an advantage that we can still test hypotheses of economic interest without fully specifying structural details of the model and also without fully identifying the equilibrium RNP. For this, we make use of statistical inference framework and associated subsampling algorithm developed by Romano and Shaikh (2008).

The paper is organized as follows. Section 1.2 specifies the asset price data generating process. In Section 1.3, we discuss the identification of the market price of risk. Section 1.4 sets forth our econometric framework for estimation and inference. In Section 1.5 , we apply our results to study international risk sharing and risk premia associated with market capitalization range indexes. Section 1.6 discusses extensions of our framework to more general multivariate asset price processes. Section 1.7 concludes with a summary and a discussion of directions for future research.

1.2 The Asset Price Process

For given positive finite T, let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, P)$ be a complete filtered probability space. The filtration $\{\mathcal{F}_t\} = \{\mathcal{F}_t\}_{t\in[0,T]}$ is assumed to satisfy the usual properties (e.g., Protter, 2005). Unless otherwise noted, $t \in [0,T]$ throughout. As is common, we take $\mathcal{F} = \mathcal{F}_T$. 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

$$dS_t = \mu_{0t}dt + \sigma_{0t}dW_t, \quad t \in [0,T],$$

where $\{W_t\}$ is a vector of $n \in \mathbb{N}$ independent standard Brownian motions under P adapted to the filtration $\{\mathcal{F}_t\}$, $\{\mu_{0t}\}$ is an $\mathbb{R}^d-$ valued adapted drift process, and $\{\sigma_{0t}\}$ is an $\mathbb{R}^{d\times n}-$ valued adapted diffusion coefficient process. We assume without loss of generality that S_t^0 is the price of the risk-free bond with known rate of return r. Let the discounted asset prices be $S_t^{*i} = S_t^i/S_t^0, i = 1, ..., d$.

Given an $\mathbb{R}^n-$ valued adapted process $\{\lambda_t\}$ such that $\int_0^T \|\lambda_t\|^2 dt < \infty$, $a.s. - P$, the Girsanov transformation defines a new adapted process $\{\tilde{W}_t\}$ by adjusting the drift of the original Brownian motion:

$$\tilde{W}_t = W_t + \int_0^t \lambda_s ds.$$

The absence of arbitrage (equivalently, the existence of the risk neutral measure) holds only for λ_t such that

$$\sigma_{0t}\lambda_t = \mu_{0t} - rS_t, \quad t \in [0,T], \quad a.s. - P. \tag{1.2.1}$$

Such a vector λ_t is called a *market price of risk*. Without further assumptions, and specifically without assuming market completeness, the market prices of risk form a set

$$\Lambda_{I,t} := \{\lambda_t : \sigma_{0t}\lambda_t = \mu_{0t} - rS_t\}.$$

We let $\mathbf{\Lambda}_I$ denote the set-valued process $\{\Lambda_{I,t}, t \in [0,T]\}$.

For our purposes here, it suffices to define market completeness in terms of $\Lambda_{I,t}$. We say that markets are *complete at* t when $\Lambda_{I,t}$ has a unique element; otherwise, we say markets are *incomplete at* t.

Under a risk neutral measure Q, $\tilde{W}_t$ follows a standard Brownian motion. After the change of measure from P to Q, the discounted asset return process can be represented by linear combinations of Brownian motions under Q:

$$\frac{dS_t^{*i}}{S_t^{*i}} = \sigma_{0t}^i \cdot d\tilde{W}_t \quad i = 1, ..., d,$$

where σ_{0t}^i is the $1 \times n$ ith row of σ_{0t}. That is, under Q, any asset is expected to earn a return equal to the risk-free rate. Using this result, the prices of redundant securities can be computed by taking the expectation under Q. (See, for example, Duffie, 2001, and Williams, 2006.)

In order to study identification in a simple but important special case in what follows, we consider a running example in which $\mu_{0t}^i = \mu_0^i S_t^i$ and $\sigma_{0t}^i = \sigma_0^i S_t^i$ for $i = 1, ..., d$. We call this specification a *multivariate Black-Scholes economy.* We formalize this as follows.

Assumption 1.2.1 (Multivariate Black-Scholes): *Let* $\{W_t\}$ *be a vector of* $n \in \mathbb{N}$ *independent standard Brownian motions under* P *adapted to the filtration* $\{\mathcal{F}_t\}$. *Let* $\{S_t\}$ *be a vector of* $d \in \mathbb{N}$ *asset prices such that* $S_0^i = 1$ *and solving the stochastic differential equations*

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

where $\mu_0 \in \mathbb{R}^d$ *has elements* μ_0^i, $i = 1, ..., d$, *and* $\sigma_0 \in \mathbb{R}^{d \times n}$ *has* $1 \times n$ *rows* σ_0^i, $i = 1, ..., d$. *Further,* $\{S_t\}$ *does not admit arbitrage.*

For this process, the market prices of risk always lie in the non-random time-invariant set

$$\Lambda_{I,0} = \{\lambda : \sigma_0 \lambda = \mu_0 - r\iota\},$$

where ι is a d-dimensional vector of ones. Any process $\{\lambda_t\}$ such that $\lambda_t \in \Lambda_{I,0}$, $t \in [0, T]$, is an admissible market price of risk process in this economy. As we know that the true Black-Scholes market price of risk is a constant, say λ_0, we consider only non-random, time-invariant processes $\{\lambda_t\}$ such that for all $t \in [0, T]$, $\lambda_t = \lambda$ for some fixed $\lambda \in \Lambda_{I,0}$.

1.3 Identifying the Market Price of Risk

1.3.1 The market price of risk and the RNP

Under the change of measure from the objective measure P to the risk neutral measure Q, the risk adjustment is fully determined by the Radon-Nikodym derivative[1] dQ/dP. In the continuous-time setting, one can define a density process of Radon-Nikodym derivatives $\boldsymbol{\xi} := \{\xi_t\}$ where $\xi_t := E_t[dQ/dP]$, with $E_t(\cdot) := E(\cdot \, |\mathcal{F}_t)$. As dQ/dP is $\mathcal{F} = \mathcal{F}_T-$ measurable, we have $\xi_T = dQ/dP$. The history $\lambda^t := \{\lambda_\tau, \, \tau \in [0, t]\}$ uniquely indexes the density ξ_t and therefore characterizes the risk adjustment. In general, given an adapted process $\{\lambda_t\}$, the corresponding densities can be written

$$\xi_t = \exp\left(-\int_0^t \lambda_s \cdot dW_s - \frac{1}{2}\int_0^t \|\lambda_s\|^2 ds \right) \quad t \in [0, T]. \tag{1.3.1}$$

Accordingly, $\boldsymbol{\xi}$ is also known as the *stochastic exponential* of $\{-\lambda_t\}$. In the multivariate Black-Scholes economy, ξ_t simplifies to

$$\xi_t = \exp\left(-\lambda_0 \cdot W_t - \frac{1}{2}\|\lambda_0\|^2 t \right), \tag{1.3.2}$$

where λ_0 is the true market price of risk in the Black-Scholes economy.

In general, there are multiple processes $\{\lambda_t\}$ consistent with the no-arbitrage requirement. This implies that there are multiple ways to change the measure from P to Q. Therefore, even if P is identified by the observed data, Q cannot be uniquely identified under incomplete markets.

The role of the Radon-Nikodym derivative is best understood in terms of the pricing equation. Consider a "European-type" asset paying zero for $t < T$ and $\varphi(W_T)$ in period T, where φ is a Borel measurable real-valued function. Let

[1] The Radon-Nikodym derivative $\mathcal{D}$ is defined as an $\mathcal{F}$-measurable strictly positive random variable such that for any $A \in \mathcal{F}$, $Q(A) = \int_A \mathcal{D} dP$.

$\varphi_{\boldsymbol{\lambda}} : \mathbb{R}^n \to \mathbb{R}$ be a measurable function such that

$$\varphi_{\boldsymbol{\lambda}}(\tilde{W}_T) = \varphi_{\boldsymbol{\lambda}}\left(W_T + \int_0^T \lambda_s ds\right) = \varphi(W_T).$$

For example, a contingent claim that pays 1 monetary unit if W_T is in a measurable set A and zero otherwise has a payoff $\varphi(W_T) = 1_{\{W_T \in A\}}$, where $1_{\{\cdot\}}$ is the indicator function taking the value 1 if the condition in brackets $\{\cdot\}$ is true and 0 otherwise. Then the payoff function in terms of $\tilde{W}_T$ is $\varphi_{\boldsymbol{\lambda}}(\tilde{W}_T) = 1_{\{\tilde{W}_T \in A_\lambda\}}$, where A_λ is a translation of A by $\int_0^T \lambda_s ds$.

Generally, there are two equivalent ways to compute the asset price p_0 at $t = 0$ for such an asset[2]. We have

$$p_0 = E^P\left[m_T\ \varphi(W_T)\right] = e^{-rT} E^Q[\varphi_{\boldsymbol{\lambda}}(\tilde{W}_T)].$$

The first equality uses the DGP P and the $\mathcal{F}_T$-measurable *stochastic discount factor* (SDF) m_T. The second equality uses the RNP Q and the risk-free rate to price the payoff $\varphi_{\boldsymbol{\lambda}}(\tilde{W}_T)$. To represent the SDF, we write

$$\begin{aligned} E^P\left[m_T\ \varphi(W_T)\right] &= e^{-rT} E^Q[\varphi_{\boldsymbol{\lambda}}(\tilde{W}_T)] = e^{-rT} \int \varphi_{\boldsymbol{\lambda}}(\tilde{W}_T) dQ \\ &= \int e^{-rT} \varphi(W_T) \frac{dQ}{dP} dP = \int e^{-rT} \xi_T\ \varphi(W_T) dP \\ &= E^P\left[e^{-rT} \xi_T\ \varphi(W_T)\right]. \end{aligned}$$

Thus, $m_T = e^{-rT}\xi_T$, the discounted Radon-Nikodym derivative. The SDF discounts the future payoff by e^{-rT} and adjusts its risk by ξ_T. If $\lambda_t = 0$ for all t, then $\xi_T = 1$ and no risk adjustment takes place. This is the case of risk neutrality. For the multivariate Black-Scholes economy, ξ_T is a log-normal random variable with mean 1 and variance $e^{\|\lambda_0\|^2 T} - 1$; in this case, risk neutrality is equivalent to $\lambda_0 = 0$.

The density process ξ of Radon-Nikodym derivatives is a stochastic process defined by the stochastic integral in (1.3.1). For what follows, we will take the variance of ξ_T to be finite. This condition is known as the L^2-*reducibility of*

[2] See Aït-Sahalia and Lo (2000) or chapter 6 of Duffie (2001), for example.

$\{\lambda_t\}$ (see, e.g., Duffie, 2001). Further, this finiteness assumption has a portfolio interpretation and a link to the option pricing bound studied in Cochrane and Saá-Requejo (2000). To ensure that ξ_T has finite variance in the Black-Scholes economy, we simply bound[3] λ_0 :

Assumption 1.3.1 (Bounded Risk Price): *For the Black-Scholes economy, there exists* $0 < M < \infty$ *such that* $\|\lambda_0\| \leq M$.

For the Black-Scholes economy, the identified set for the market price of risk is thus

$$\Lambda_{I,0}^M := \{\lambda : \sigma_0 \lambda = \mu_0 - r\iota, \quad \|\lambda\| \leq M\}.$$

An illustration of the identified set $\Lambda_{I,0}^M$ with $d = 1$ and $n = 2$ is given by Figure 1.3.1. In this example, the risk exposure of the single traded asset is determined by a vector $\sigma_0 \in \mathbb{R}^2$ such that both elements of σ_0 are non-zero; and $\lambda_0 \in \mathbb{R}^2$ is the true market price of risk. As there are two fundamental sources of risk in this economy, the traded security does not reveal λ_0. Instead it reveals all λ's that lie on the iso-risk premium line perpendicular to σ_0, as λ_0 is observationally equivalent to any other λ on the iso-risk premium line. To see this, fix σ_0, and let $N(\sigma_0)$ be the null space of σ_0. Consider $\lambda := \lambda_0 + b\eta$, where $b \in \mathbb{R}$ and $\eta \in N(\sigma_0)$. Then, λ_0 and λ give the same value of the risk premium by construction. As the joint distribution of the $d = 1$ discounted asset prices is fully characterized by the drift (risk premium) and the variance-covariance structure, λ_0 and λ are observationally equivalent. Thus, one cannot identify λ_0 by simply examining the distribution of asset prices. Instead, this distribution only reveals the iso-risk premium line. Together with L^2-boundedness, the identified set becomes a finite line segment. In more general cases, the identified set is a finite subset of an affine subspace orthogonal to the row space of σ_0.

[3]This boundedness also implies that the Radon-Nikodym derivative satisfies the Novikov condition. See Duffie (2001) and Williams (2006). We discuss a more general condition in Section 6.

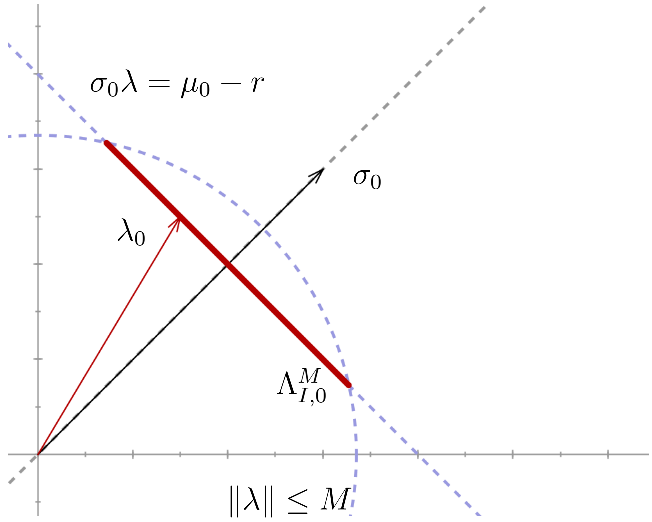

Figure 1.1: The identified set ($d = 1$ and $n = 2$)

1.3.2 A common factor structure

When σ_0 satisfies m a priori restrictions $\rho(\sigma_0) = 0$, these can facilitate estimation of the market price of risk and may even suffice to identify σ_0. A leading case of such restrictions is that of common factors. For the Black-Scholes case, we impose this formally as follows.

Assumption 1.3.2 (Common Factors): *Each asset return depends on its unique idiosyncratic risk and $n - d$ common factors.*

Assumption 1.3.2 implies that the returns are correlated with each other only through common risk factors. The $n - d$ common risk factors are represented by $n - d$ independent Brownian motions. An example with $d = 3$ and $n = 4$ is

$$\sigma_0 = \begin{bmatrix} \sigma_{11} & 0 & 0 & \sigma_{14} \\ 0 & \sigma_{22} & 0 & \sigma_{24} \\ 0 & 0 & \sigma_{33} & \sigma_{34} \end{bmatrix}.$$

With common factors, each asset return depends only on $(n - d + 1)$ Brownian motions. This reduces the number of nonzero parameters in σ_0 from nd in the general case to $(n - d + 1)d$.

In general, the $d \times d$ asset returns covariance matrix $\Sigma_0 := \sigma_0 \sigma_0'$ provides $d(d+1)/2$ restrictions on σ_0. The zero restrictions of Assumption 1.3.2 impose $(d-1)d$ further restrictions. This reduces the dimension of the unidentified aspects of σ_0. Specifically, when $n \leq (3d-1)/2$, Assumption 1.3.2 ensures that one can fully identify σ_0 from elements of Σ_0. Our examples in Section 1.5 take $d = 3$ and $n = 4$, a case in which σ_0 is fully identified.

We emphasize that this is inherently a *structural* restriction; that is, the data are generated by a process obeying this condition. Although alternative representations of the asset price process may exist that do not obey this restriction, these have only a stochastic and not a structural interpretation. Assumption 1.3.2 thus specifies an economic interpretation for the vector of Brownian motions. We interpret the first d elements as idiosyncratic risks and the last $n - d$ elements as common risks. As Section 1.5 illustrates, the meaning of W_t may vary depending on the given application.

Above, we represented the restrictions on σ_0 as $\rho(\sigma_0) = 0$, an $m \times 1$ zero vector. The common factors assumption has a simple representation of this form. For these cases, we have

$$\rho(\sigma_0) = \rho_0 vec(\sigma_0),$$

where $vec(\sigma_0)$ stacks the columns to yield an $nd \times 1$ column vector, and ρ_0 is an $m \times nd$ matrix. The matrix ρ_0 has rows whose elements are zero, except for a one in the position that identifies an element of σ_0 that is to take the value zero. For the example above with $n = 4$ and $d = 3$, $m = 6$. Further, the third row of the 6×12 matrix ρ_0 contains a one in the fourth position (corresponding to σ_{12}, which is set to zero), with the remaining row elements zero.

1.4 Econometric Framework

In this section, we propose estimation and hypothesis testing procedures for the market price of risk in the Black-Scholes economy following the set-estimation and hypothesis testing frameworks of Chernozhukov, Hong, and Tamer (2007). These authors study extremum estimators where the criterion functions do not have a unique minimizer. For estimation, the basic idea is to use lower contour sets of the sample criterion function as set-valued estimators or confidence regions. For hypothesis testing, Romano and Shaikh (2006) propose a subsampling procedure. In this section, we exploit these methods by showing that the identified set for the Black-Scholes economy risk price, $\Lambda_{I,0}^{M}$, can be characterized as a set of minimizers of a specific criterion function.

1.4.1 Applying the CHT framework

In the multivariate Black-Scholes economy, the vector of returns of d securities over the time interval $[s,t]$ obeys a multivariate normal distribution with mean $(t-s)(\mu_0 - (\|\sigma_0^1\|^2, \cdots, \|\sigma_0^d\|^2)'/2)$ and covariance matrix $(t-s)\Sigma_0$. Eq. (1.2.1) implies that the drift μ_0 is determined, once we specify (σ_0, λ_0) and r. Therefore, for any given constant r, the joint density of asset returns depends only on σ_0 and λ_0.

Consider a partition $\pi := \{0 =: t_0, t_1, ..., t_{N-1}, t_N := T\}$ of the interval $[0,T]$. Suppose we observe a series of asset prices $\{S_{t_j}\}_{j=0}^{N}$ over this partition. Let R_{t_j} be the $d \times 1$ vector of asset returns from period t_j to t_{j+1}: i.e., the ith element of R_{t_j} is $R_{t_j}^i := \ln S_{t_j}^i - \ln S_{t_{j-1}}^i$, $i = 1, ..., d$. Let $f(R_{t_1}, ..., R_{t_N}; \theta)$ denote the likelihood of a sample of asset returns at $\theta := (\sigma, \lambda) \in \Theta := \mathbb{S} \times \Lambda \subseteq \mathbb{R}^{d \times n} \times \mathbb{R}^n$, where $\mathbb{S}$ is a non-empty subset of $\mathbb{R}^{d \times n}$. In the multivariate Black-Scholes economy, returns are independent over time, so that

$$f(R_{t_1}, \cdots, R_{t_N}; \theta) = \prod_{j=1}^{N} f(R_{t_j}; \theta),$$

where $f(R_{t_j}; \theta)$ defines the likelihood for asset returns in t_j; this is a $d-$variate normal likelihood.

The coefficients $\theta_0 := (\sigma_0, \lambda_0) \in \Theta$ index the true DGP measure P. Let the *criterion function* $\bar{Q}_N : \Theta \to \bar{\mathbb{R}}_+$ be the shifted expected negative average log likelihood defined by

$$\bar{Q}_N(\theta) := E^P\left[-N^{-1}\sum_{j=1}^{N}\ln f(R_{t_j};\theta)\right] - q_{0,N}, \tag{1.4.1}$$

where

$$q_{0,N} := E^P\left[-N^{-1}\sum_{j=1}^{N}\ln f(R_{t_j};\theta_0)\right].$$

The criterion function thus has minimum value 0 at θ_0. This minimum is not unique; letting $\Theta_0^M := \{\sigma_0\} \times \Lambda_{I,0}^M$, we also have

$$\bar{Q}_N(\Theta_0^M) = 0.$$

Further, the asset return covariances only reveal $\Sigma_0 = \sigma_0\sigma_0'$, so σ_0 cannot be identified from observations of d asset returns without further restrictions. Specifically, let

$$\Theta_{I,0}^M := \{(\sigma,\lambda) \in \Theta : \sigma\sigma' = \Sigma_0, \quad \sigma\lambda = \mu_0 - r\iota, \quad \|\lambda\| \leq M\}.$$

Then $\Theta_0^M \subset \Theta_{I,0}^M$, and $\Theta_{I,0}^M$ contains all the minimizers of $\bar{Q}_N$. That is,

$$\bar{Q}_N(\Theta_{I,0}^M) = 0 \quad \text{and} \quad \bar{Q}_N(\theta) > 0 \quad \text{for } \theta \notin \Theta_{I,0}^M.$$

Working with $\bar{Q}_N$ and $\Theta_{I,0}^M$ enables us to apply the CHT framework to our problem.

Accordingly, let $Q_N : \Omega \times \Theta \to \bar{\mathbb{R}}_+$ be the sample criterion function defined by

$$Q_N(\theta) = -N^{-1}\sum_{j=1}^{N}\ln f(R_{t_j};\theta) - q_N, \tag{1.4.2}$$

where $q_N = \inf_\Theta -N^{-1}\sum_{j=1}^{N}\ln f(R_{t_j};\theta)$. Following Chernozhukov, Hong, and Tamer (2007), we define an ϵ-level set of the sample criterion function by

$$\hat{\Theta}_N(\epsilon) := \{\theta \in \Theta : N \cdot Q_N(\theta) \leq \epsilon\}.$$

When we choose ϵ properly, the random set $\hat{\Theta}_N(\epsilon)$ is a consistent set estimator or a confidence region for the identified set.

There are, however, several challenges to directly applying the CHT framework to our problem. First, the identified set $\Theta^M_{I,0}$ has a high dimension, $nd + n$. This leads to computational difficulties and can also hamper the interpretation of results. Further, for any fixed value of σ such that $\sigma\sigma' = \Sigma_0$, every element of the set $\Lambda^M_{I,0}(\sigma) := \{\lambda \in \Lambda : \sigma\lambda = \mu_0 - r\iota, \ \ \|\lambda\| \leq M\}$ minimizes $\bar{Q}_N$. This suggests that, as one changes the value of σ, the set $\Lambda^M_{I,0}(\sigma)$ rotates. Consequently, $\Theta^M_{I,0}$ may cover quite a large subset of Θ. Finally, $\Theta^M_{I,0}$ need not be convex. This may cause additional technical difficulties.

These difficulties can be mitigated or avoided by applying a two-stage procedure proposed by Kaido and White (2008), described next.

1.4.2 A two-stage procedure

In this section, we describe a two-stage procedure proposed by Kaido and White (2008) that reduces the dimension of the set estimator and the associated confidence region. With sufficient restrictions, some elements of σ_0 can even be fully identified. In such cases, we can replace identified elements of σ_0 in the sample criterion function with their consistent estimators. Even if this is not possible, restrictions on elements of σ_0 can still simplify estimation substantially.

We summarize Kaido and White's (2008) measurability and consistency results for the two-stage set estimator as follows. Let $m \in \mathbb{N}$ be the number of restrictions on σ_0 and let $\rho : \mathbb{R}^{d\times n} \to \mathbb{R}^m$ embody these restrictions as

$$\rho(\sigma_0) = 0.$$

The identified (σ, λ) values that satisfy all our restrictions are the elements of

$$\Theta^M_{I,0,\rho} := \{(\sigma, \lambda) \in \Theta : \sigma\sigma' = \Sigma_0, \quad \rho(\sigma) = 0, \quad \sigma\lambda = \mu_0 - r\iota, \quad \|\lambda\| \leq M\}.$$

Let $\hat{\Sigma}_N$ be a bounded consistent estimator of Σ_0, and let $K(\mathbb{S})$ be a collection of closed subsets of $\mathbb{S}$. Define a first-stage restricted set-estimator $\hat{S}_N : \Omega \to K(\mathbb{S})$

of σ_0 by

$$\hat{S}_N(\omega) = \{\sigma \in \mathbb{S} : \sigma\sigma' = \hat{\Sigma}_N(\omega), \quad \rho(\sigma) = 0\}. \tag{1.4.3}$$

This is a random set of diffusion coefficients that are consistent with the sample covariance of the returns and that satisfy the restriction $\rho(\sigma) = 0$.

Using this first-stage set estimator, let the second-stage set-estimator for $\Theta^M_{I,0,\rho}$ be defined by

$$\hat{\Theta}_N(\omega) := \{(\sigma, \lambda) \in \Theta : NQ_N(\omega, \sigma, \lambda) \leq \hat{\epsilon}(\omega), \quad \sigma \in \hat{S}_N(\omega)\}, \tag{1.4.4}$$

where $\hat{\epsilon}$ is now permitted to be random.

An important special case occurs when the restrictions suffice to identify σ_0. When $\hat{\sigma}_N$ is a consistent estimator of σ_0, the first-stage set estimator becomes a singleton, i.e., $\hat{S}_N = \{\hat{\sigma}_N\}$. The second-stage set estimator is then $\hat{\Theta}_N = \{\hat{\sigma}_N\} \times \hat{\Lambda}_N$, where

$$\hat{\Lambda}_N(\omega) := \{\lambda \in \Lambda : NQ_N(\omega, \hat{\sigma}_N(\omega)), \lambda) \leq \hat{\epsilon}(\omega)\}. \tag{1.4.5}$$

1.4.3 Effros-measurability

The first step in analyzing the two-stage set estimator is to establish its measurability. A useful measurability concept for set-valued functions is *Effros*-measurability. 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. See Molchanov (2005) for details.

Definition 1.4.1 (Effros-Measurability): *Let $(\Omega, \mathcal{F})$ be a measurable space. Let $l \in \mathbb{N}$, and let $\mathcal{G}$ be a topology on $\mathbb{R}^l$. Let $K(\mathbb{R}^l)$ be a collection of closed subsets of $\mathbb{R}^l$. A map $X : \Omega \to K(\mathbb{R}^l)$ is* Effros-measurable *with respect to $\mathcal{F}$ if, for each open set $G \in \mathcal{G}$,*

$$X^-(G) := \{\omega : X(\omega) \cap G \neq \emptyset\} \in \mathcal{F}.$$

The next result establishes Effros-measurability for general two-stage estimators.

THEOREM 1.4.1: *Let* $(\Omega, \mathcal{F}, P)$ *be a complete probability space, and let* $\Theta = \mathbb{S} \times \Lambda$, *where* $\mathbb{S}$ *and* Λ *are compact subsets of finite dimensional Euclidian spaces.*

Let $Q : \Omega \times \Theta \to \bar{\mathbb{R}}_+$ *be such that for each* $\theta \in \Theta$, $Q(\,\cdot\,, \theta)$ *is measurable-*$\mathcal{F}$ *and for* $F \in \mathcal{F}$ *with* $P(F) = 1$, $Q(\omega, \,\cdot\,)$ *is continuous on* Θ *for each* $\omega \in F$.

Let $\hat{S} : \Omega \to K(\mathbb{S})$ *be Effros-measurable with respect to* $\mathcal{F}$.

Then for any measurable $\hat{\epsilon} : \Omega \to \mathbb{R}^+$, *the* $\hat{\epsilon}$*- level set* $\hat{\Theta}_{\hat{\epsilon}} : \Omega \to K(\mathbb{S} \times \Lambda)$, *defined by*

$$\hat{\Theta}(\omega, \hat{\epsilon}(\omega)) = \{(\sigma, \lambda) \in \Theta : Q(\omega, \sigma, \lambda) \leq \hat{\epsilon}(\omega), \quad \sigma \in \hat{S}(\omega)\},$$

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

The proof of this and other formal results can be found in Kaido and White (2008).

For the special case where $\hat{S}$ is a singleton, e.g., when the diffusion coefficient is point identified, we have the following result.

COROLLARY 1.4.1: *Let the conditions of Theorem 1.4.1 hold, and suppose* $\hat{S}$ *is a singleton such that* $\hat{S} = \{\hat{\sigma}\}$, *where* $\hat{\sigma} : \Omega \to \mathbb{S}$ *is measurable-*$\mathcal{F}$. *Then*

(i) For each $\lambda \in \Lambda$, $\tilde{Q}(\,\cdot\,, \lambda) := Q(\,\cdot\,, \hat{\sigma}(\cdot), \lambda)$ *is a measurable function on* Ω *and for* $\tilde{F} \in \mathcal{F}$ *with* $P(\tilde{F}) = 1$, $\tilde{Q}(\omega, \,\cdot\,)$ *is continuous on* Λ *for each* $\omega \in \tilde{F}$.

(ii) For any measurable $\hat{\epsilon} : \Omega \to \mathbb{R}_+$, *the* $\hat{\epsilon}$*- level set* $\hat{\Lambda}_{\hat{\epsilon}} : \Omega \to K(\Lambda)$, *defined by*

$$\hat{\Lambda}(\omega, \hat{\epsilon}(\omega)) = \{\lambda \in \Lambda : \tilde{Q}(\omega, \lambda) \leq \hat{\epsilon}(\omega)\},$$

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

The following proposition establishes the Effros-measurability of our constrained first-stage estimator, enabling us to apply the above results.

PROPOSITION 1.4.1: *Let* $(\Omega, \mathcal{F})$ *be a measurable space, and let* $\mathbb{S}$ *be a compact subset of* $\mathbb{R}^{d \times n}$, *where* d *and* n *are finite positive integers.*

Let Ψ *be a set of bounded symmetric positive semi-definite matrices. Let* $\hat{\Sigma} : \Omega \to \Psi$ *be measurable-*$\mathcal{F}$, *and let* $\rho : \mathbb{S} \to \mathbb{R}^m$ *be continuous, where* m *is a finite positive integer.*

Let $\hat{S} : \Omega \to K(\mathbb{S})$ be defined by

$$\hat{S}(\omega) = \{\sigma \in \mathbb{S} : \sigma\sigma' = \hat{\Sigma}(\omega), \quad \rho(\sigma) = 0\}.$$

Then $\hat{S}$ is Effros-measurable with respect to $\mathcal{F}$.

1.4.4 Consistent set estimation

Next we provide results ensuring the consistency of the two-stage set estimator. Consistency is expressed in terms of the Hausdorff metric on the space of closed sets.

DEFINITION 1.4.2: *Let Θ be a compact subset of a finite dimensional Euclidean space. For any two closed subsets A and B of Θ, the Hausdorff metric is*

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

where $\|\cdot\|$ is the Euclidean norm, and $d_H(A, B) := \infty$ if either A or B is empty.

The Effros-measurability of the two-stage set estimator implies the measurability of the Hausdorff distance between the set estimator and the identified set[4]. This makes it possible to discuss the consistency of this set estimator. Our first result provides conditions under which the general two-stage set estimator is consistent.

THEOREM 1.4.2: *Let $(\Omega, \mathcal{F}, P)$ and $\Theta = \mathbb{S} \times \Lambda$ satisfy the conditions of Theorem 1.4.1, and suppose that for $N = 1, 2, ...$, Q_N and $\hat{S}_N$ satisfy the conditions on Q and $\hat{S}$ imposed in Theorem 1.4.1.*

Suppose there exists $\bar{Q}_N : \Theta \to \bar{\mathbb{R}}_+$ such that $\sup_{\theta \in \Theta} |Q_N(\,\cdot\,, \theta) - \bar{Q}_N(\theta)| = o_p(1)$. Let $S \in K(\mathbb{S})$ and define

$$\Theta_I := \operatorname*{arg\,min}_{(\sigma, \lambda) \in S \times \Lambda} \bar{Q}_N(\theta),$$

[4]This follows from Theorem 2.25 (vi) p.37 in Molchanov (2005).

such that $\bar{Q}_N(\Theta_I) = 0$ *for all* N *sufficiently large.*

Let $\hat{\epsilon}_N$ *be* $\mathcal{F}$*-measurable such that* $\hat{\epsilon}_N/N = o_p(1)$ *and*

$$\lim_{N\to\infty} P\left[\omega : \sup_{\theta\in\Theta_I} Q_N(\omega,\theta) \leq \hat{\epsilon}_N(\omega)/N\right] = 1.$$

Suppose further that $d_H(\hat{S}_N, S) = o_p(1)$*, and let*

$$\hat{\Theta}_N(\omega) := \{(\sigma,\lambda) \in \Theta : Q_N(\omega,\sigma,\lambda) \leq \hat{\epsilon}_N(\omega)/N, \quad \sigma \in \hat{S}_N(\omega)\}.$$

Then $\hat{\Theta}_N$ *is Effros-measurable with respect to* $\mathcal{F}$*, and* $d_H(\hat{\Theta}_N, \Theta_I) = o_p(1)$*.*

The next result treats the important special case in which S is fully identified (i.e., S is a singleton). This shows that the natural second-stage set estimator $\hat{\Lambda}_N$ is a consistent estimator for the identified set Λ_I.

Corollary 1.4.2: *Let the conditions of Theorem 1.4.2 hold, and suppose that* S *is a singleton,* $S = \{\sigma_0\}$*. Let*

$$\Lambda_I := \arg\min_{\lambda\in\Lambda} \bar{Q}_N(\sigma_0,\lambda).$$

Let $\hat{\sigma}_N : \Omega \to \mathbb{S}$ *be measurable-*$\mathcal{F}$ *such that* $\hat{\sigma}_N = \sigma_0 + o_p(1)$*, and let*

$$\hat{\Lambda}_N(\omega) := \{\lambda \in \Lambda : Q_N(\omega, \hat{\sigma}_N(\omega), \lambda) \leq \hat{\epsilon}_N(\omega)/N\}.$$

Then $\hat{\Lambda}_N$ *is Effros-measurable with respect to* $\mathcal{F}$*, and* $d_H(\hat{\Lambda}_N, \Lambda_I) = o_p(1)$*.*

Figure 1.4.4 illustrates. As the sample size N increases, the set estimator $\hat{\Lambda}_N$ represented by the shaded region in figure 1.4.4 shrinks down to the identified set Λ_I, which is a line segment here.

Note that CHT give an additional condition (condition C.3.) for achieving consistency and polynomial convergence rate of the Hausdorff metric without setting $\hat{\epsilon}_N \geq \sup_{\Theta_I} NQ_n(\theta)$. Due to the first-stage estimation of σ, this condition may not hold for our applications. Therefore, we need to choose a sequence $\{\hat{\epsilon}_N\}$ that satisfies conditions in Theorem 1.4.2. Specifically, we choose a sequence

$\hat{\epsilon}_N = q_N + \kappa_N$, where $q_N = \inf_{\lambda \in \Lambda} Q_N(\hat{\sigma}_N, \lambda)$ and $\kappa_N \propto \ln N$.

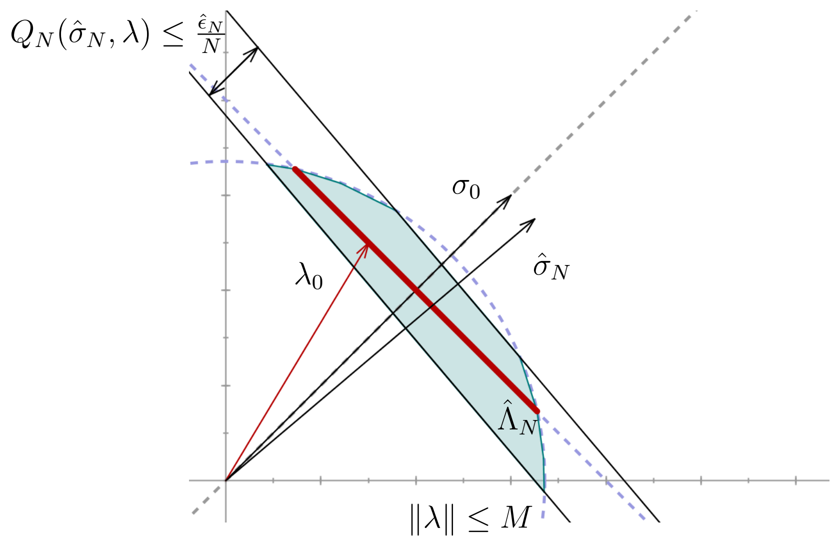

Figure 1.2: Set Estimator $\hat{\Lambda}_N$ ($d = 1$ and $n = 2$)

1.4.5 Hypothesis testing

Set estimation is useful when interest attaches to the characteristics of the identified set. If instead one wishes to test hypotheses regarding the identified set, it is not necessary to estimate the identified set. Specifically, let R be a closed subset of Θ (or Λ), where R is a set of parameters that satisfy the restrictions of interest. For example, R may represent a set of market prices of risk that are consistent with risk-neutrality or international risk sharing.

As the true coefficient value θ_0 is in the identified set, if θ_0 also satisfies the restrictions, the identified set Θ_I has a non-empty intersection with R. We thus consider hypotheses

$$H_o^{\Theta} : \Theta_I \cap R \neq \emptyset \quad vs. \quad H_A^{\Theta} : \Theta_I \cap R = \emptyset.$$

The null states that there is at least one element in the identified set satisfying the restrictions. Rejection means that none of the parameters in the identified set satisfies the restrictions, implying that θ_0 does not satisfy the restrictions.

In the Black-Scholes example, where interest attaches to λ_0, we consider hypotheses

$$H_o^{\Lambda} : \Lambda_I \cap R \neq \emptyset \quad vs. \quad H_A^{\Lambda} : \Lambda_I \cap R = \emptyset.$$

Because R is a closed subset of the compact parameter space, the hypotheses above are equivalent to

$$H_o^{\Theta} : \inf_{\theta \in \Theta \cap R} \bar{Q}_N(\theta) = 0 \qquad \text{or} \qquad H_o^{\Lambda} : \inf_{\lambda \in R} \bar{Q}_N(\sigma_0, \lambda) = 0.$$

Such hypotheses are considered in the partially identified case by Romano and Shaikh (2008) for parametric inference and by Santos (2007) for nonparametric inference.

To test these hypotheses in our two-stage framework, we replace $\bar{Q}_N$ and Θ with their sample analogs Q_N and $\hat{S}_N \times \Lambda$, which leads to the test statistics

$$\hat{T}_N(\Theta, R) = \inf_{\theta \in (\hat{S}_N \times \Lambda) \cap R} a_N Q_N(\theta) \quad \text{and} \quad \hat{T}_N(\Lambda, R) = \inf_{\lambda \in R} a_N Q_N(\hat{\sigma}_N, \lambda),$$

where a_N is a normalizing constant such that $\sup_{\theta \in \Theta_I} Q_N(\theta) = O_p(1/a_N)$ or $\sup_{\lambda \in \Lambda_I} Q_N(\sigma_0, \lambda) = O_p(1/a_N)$. In our problem, $a_N = N$, so the test statistics can be written

$$\begin{aligned} \hat{T}_N(\Theta, R) &= \sup_{\theta \in \hat{S}_N \times \Lambda} \sum_{j=1}^{N} \ln f(R_{t_j}; \theta) - \sup_{\theta \in (\hat{S}_N \times \Lambda) \cap R} \sum_{j=1}^{N} \ln f(R_{t_j}; \theta). \\ \hat{T}_N(\Lambda, R) &= \sup_{\lambda \in \Lambda} \sum_{j=1}^{N} \ln f(R_{t_j}; \hat{\sigma}_N, \lambda) - \sup_{\lambda \in R} \sum_{j=1}^{N} \ln f(R_{t_j}; \hat{\sigma}_N, \lambda). \end{aligned}$$

These can be viewed as log-likelihood ratio statistics for partially identified models.

To maintain a tight focus for the discussion to follow, we now restrict attention to the Black-Scholes case that will be the subject of our empirical examples. This is the case where Assumptions 1.2.1 and 1.3.1 hold, and the common factor structure of Assumption 1.3.2 ensures that σ_0 is point-identified. Thus, we restrict attention to $\hat{T}_N(\Lambda, R)$, where we take $\Lambda = \Lambda^M$. (We leave the notation $\hat{T}_N(\Lambda, R)$ unchanged for simplicity.) To test H_o^{Λ}, we require asymptotic critical values for

$\hat{T}_N(\Lambda, R)$.

Obtaining these critical values presents interesting challenges. Space precludes a rigorous derivation here, as handling all the necessary formalities is fairly involved. Nevertheless, the intuition behind our approach is straightforward, so we offer the following heuristic discussion.

We start by noting that the presence of $\hat{\sigma}_N$ in $\hat{T}_N(\Lambda, R)$ may have an impact on its limiting distribution. To accommodate this, we can proceed in a manner analogous to the fully identified case. There, one can often exploit a two-term mean value or Taylor-like expansion. The following straightforward high-level result applies when θ_0 is interior to Θ and the likelihood function is sufficiently smooth. Analogous but more elaborate results hold even when θ_0 is not interior to Θ.

PROPOSITION 1.4.2: *Let $\{a_N\}$ be a sequence of real numbers and for $p \in \mathbb{N}$, suppose that $\theta_0 \in \mathbb{R}^p$ and that $\{\hat{Q}_N : \Omega \to \mathbb{R}\}$, $\{Q_N : \Omega \to \mathbb{R}\}$, $\{\hat{\theta}_N : \Omega \to \mathbb{R}^p\}$, $\{g_N : \Omega \to \mathbb{R}^p\}$, and $\{H_N : \Omega \to \mathbb{R}^{p\times p}\}$ are sequences of measurable functions such that*

$$a_N\hat{Q}_N = a_N Q_N + a_N g_N' \left(\hat{\theta}_N - \theta_0\right) + a_N \left(\hat{\theta}_N - \theta_0\right)' H_N \left(\hat{\theta}_N - \theta_0\right)/2 + o_P(1),$$

where, for random matrices Z_0, Z_1, Z_2, Z_3 of suitable dimension,

$$(a_N Q_N,\ a_N^{1/2}\left(\hat{\theta}_N - \theta_0\right)',\ a_N^{1/2} g_N',\ (vec(H_N))') \xrightarrow{d} (Z_0, Z_1', Z_2', (vec(Z_3))').$$

Then

$$a_N\hat{Q}_N \xrightarrow{d} Z_0 + Z_2'Z_1 + Z_1'Z_3Z_1/2.$$

In our application, $a_N = N$, $p = dn$, $\theta_0 = vec(\sigma_0)$, $a_N\hat{Q}_N = \hat{T}_N(\Lambda, R)$, $a_N Q_N = T_N(\Lambda, R; \sigma_0)$, where

$$T_N(\Lambda, R; \sigma) := \sup_{\lambda\in\Lambda} \sum_{j=1}^{N} \ln f(R_{t_j}; \sigma, \lambda) - \sup_{\lambda\in R} \sum_{j=1}^{N} \ln f(R_{t_j}; \sigma, \lambda),$$

$\hat{\theta}_N = vec(\hat{\sigma}_N)$, $g_N = N^{-1}(\partial/\partial\sigma)T_N(\Lambda, R; \sigma_0)$, and $H_N = N^{-1}(\partial^2/\partial\sigma\partial\sigma')T_N(\Lambda, R; \sigma_0)$. Under our assumptions, $N^{1/2}(\hat{\sigma}_N - \sigma_0)$ and $N^{-1/2}(\partial/\partial\sigma)T_N(\Lambda, R; \sigma_0)$ will generally jointly obey a central limit theorem, and $N^{-1}(\partial^2/\partial\sigma\partial\sigma')T_N(\Lambda, R; \sigma_0)$ converges in probability to a constant matrix. The desired limiting distribution follows provided $T_N(\Lambda, R; \sigma_0)$ also converges in distribution (jointly with the other random variables).

For this, we can apply results of Liu and Shao (2003), whose theorem 3.1 provides general regularity conditions for the non-identified case ensuring that

$$\lim_{N\to\infty} 2T_N(\Lambda, R_0; \sigma_0) = \sup_{S\in\mathcal{F}_\Lambda} \max(W_S, 0)^2,$$

where $R_0 := \{\lambda_0\}$, W_S defines a centered Gaussian process $\{W_S : S \in \mathcal{F}_\Lambda\}$ with uniformly continuous sample paths and covariance kernel

$$E(W_{S_1}W_{S_2}) = E(S_1S_2),$$

and $\mathcal{F}_\Lambda$ is a specific Donsker class of functions, a set of limits of generalized score functions S (see Liu and Shao, 2003, eq.(3.1)).

In our application, interest attaches to

$$T_N(\Lambda, R; \sigma_0) = T_N(\Lambda, R_0; \sigma_0) - T_N(R, R_0; \sigma_0),$$

so Liu and Shao's theorem 3.1 implies that under H_o^Λ,

$$\lim_{N\to\infty} 2T_N(\Lambda, R; \sigma_0) = \sup_{S\in\mathcal{F}_\Lambda} \max(W_S, 0)^2 - \sup_{S\in\mathcal{F}_R} \max(W_S, 0)^2.$$

Although this gives the asymptotic distribution only for $T_N(\Lambda, R; \sigma_0)$, the extension to the required joint convergence appears straightforward.

Proposition 1.4.2 then delivers the asymptotic distribution of $\hat{T}_N(\Lambda, R)$. As this appears to be a complicated distribution, we seek computationally simple methods for obtaining the desired critical values. One particularly appealing approach is to use the method of subsampling, as the existence of the limiting distribution for $\hat{T}_N(\Lambda, R)$ generally suffices to ensure that subsampling can generate valid asymptotic critical values.

Specifically, we obtain valid asymptotic critical values for $\hat{T}_N(\Lambda, R)$ by applying a subsampling algorithm proposed by Romano and Shaikh (2008). Let $b := b_N < N$ be a sequence of integers such that $b \to \infty$ and $b/N \to 0$. Let B_N be the number of randomly chosen subsamples of size b from a sample of size N, and let $\hat{T}_{N,b,k}$ be the subsampled test statistic for the kth subsample of size b, specifically

$$\hat{T}_{N,b,k} := \inf_{\lambda \in R} -\sum_{j \in \mathcal{J}_k} \ln f(R_{t_j}; \hat{\sigma}_{N,b,k}, \lambda) - \inf_{\lambda \in \Lambda^M} \left(-\sum_{j \in \mathcal{J}_k} \ln f(R_{t_j}; \hat{\sigma}_{N,b,k}, \lambda) \right),$$

where $\mathcal{J}_k$ is the $b-$element set of indexes for the kth subsample. Note that for each k, $\hat{T}_{N,b,k}$ is evaluated using a first stage estimate $\hat{\sigma}_{N,b,k}$, computed for that subsample.

Next, let $\alpha \in (0,1)$ be a prespecified significance level for the test, and define

$$d_{N,1-\alpha} = \inf \left\{ x : B_N^{-1} \sum_{1 \le k \le B_N} 1_{\{\hat{T}_{N,b,k} \le x\}} \ge 1 - \alpha \right\}.$$

This is a subsampling estimator for the asymptotic $1-\alpha$ quantile of $\hat{T}_N(\Lambda, R)$. Theorem 3.4 (i) of Romano and Shaikh (2008) then ensures that when H_o^Λ holds,

$$\liminf_{N \to \infty} P\left(\hat{T}_N(\Lambda, R) \le d_{N,1-\alpha}\right) \ge 1 - \alpha,$$

so that $d_{N,1-\alpha}$ provides a valid asymptotic critical value for testing H_o^Λ. When the alternative H_A^Λ holds, $d_{N,1-\alpha}$ diverges, but at a sufficiently slow rate that the test based on $d_{N,1-\alpha}$ is nevertheless consistent, a consequence of $b/N \to 0$.

1.4.6 Confidence regions

Confidence regions can be constructed using a subsampling procedure proposed by Chernozhukov, Hong, and Tamer (2007) (CHT). For this it suffices that

$$\sup_{\lambda \in \Lambda_{I,0}^M} N\, Q_N(\hat{\sigma}_N, \lambda) \xrightarrow{d} Z,$$

where Z is a random variable.

Care is required in verifying this condition due to the presence of $\hat{\sigma}_N$. One might consider using Proposition 8 to establish this. The natural choices for this are $a_N = N$, $p = dn$, $\theta_0 = vec(\sigma_0)$, $a_N \hat{Q}_N = \sup_{\lambda \in \Lambda^M_{I,0}} N\ Q_N(\hat{\sigma}_N, \lambda)$, $a_N Q_N = \sup_{\lambda \in \Lambda^M_{I,0}} N\ Q_N(\sigma_0, \lambda)$, $\hat{\theta}_N = vec(\hat{\sigma}_N)$, $g_N = (\partial/\partial\sigma) \sup_{\lambda \in \Lambda^M_{I,0}} Q_N(\sigma_0, \lambda)$, and $H_N = (\partial^2/\partial\sigma\partial\sigma') \sup_{\lambda \in \Lambda^M_{I,0}} Q_N(\sigma_0, \lambda)$. It then suffices to verify that $\sup_{\lambda \in \Lambda^M_{I,0}} N\ Q_N(\sigma_0, \lambda)$ converges in distribution jointly with $N^{1/2}(\hat{\sigma}_N - \sigma_0)$ and $N^{1/2}(\partial/\partial\sigma)$ $\sup_{\lambda \in \Lambda^M_{I,0}} Q_N(\sigma_0, \lambda)$, and that $(\partial^2/\partial\sigma\partial\sigma') \sup_{\lambda \in \Lambda^M_{I,0}} Q_N(\sigma_0, \lambda)$ converges in probability to a constant matrix. The first condition corresponds to the key primitive condition assumed by CHT (the "CHT condition"), and under mild conditions a central limit theorem holds for $N^{1/2}(\hat{\sigma}_N - \sigma_0)$ and $(\partial^2/\partial\sigma\partial\sigma') \sup_{\lambda \in \Lambda^M_{I,0}} Q_N(\sigma_0, \lambda)$ converges as required.

Nevertheless, it is not clear whether $N^{1/2}(\partial/\partial\sigma) \sup_{\lambda \in \Lambda^M_{I,0}} Q_N(\sigma_0, \lambda)$ converges in distribution; in particular, nothing appears to ensure that this quantity has (limiting) mean zero, so the central limit theorem need not hold. Accordingly, we seek an alternative approach.

A promising way to proceed is to recast our two-stage estimator as a single-stage estimator; as we show, this yields a straightforward formulation of the CHT condition. If this recasting is indeed possible, one might ask why this approach is not used from the outset. The main reason is that our likelihood-based approach is more robustly applicable to identifying the set of market risk prices of interest than the method of moments-based single-stage approach described next. In general settings, moment-based methods may introduce spurious zeros into the single-stage objective function, thereby possibly altering the apparent identified set in undesired ways. We describe how this can happen below. Using a two-stage approach permits us to ensure that the identified set is that associated with the risk prices of interest. The single-stage recasting can then be used with this identified set to deliver conditions justifying the CHT subsampling procedure.

To recast our two-stage estimator as a single-stage method of moments estimator, let $\beta := (\mu', vech(\Sigma)', \lambda')' \in \mathbb{B}$, say, and define the functions

$$\begin{aligned} m_0(R_{t_j};\beta) &= R_{t_j} - (t_j - t_{j-1})(\mu - (\Sigma_{11}, ..., \Sigma_{dd})'/2) \\ m_1(R_{t_j};\beta) &= vech[m_0(R_{t_j};\beta)m_0(R_{t_j};\beta)' - (t_j - t_{j-1})\Sigma] \\ m_2(R_{t_j};\beta) &= (\partial/\partial\lambda) \ln f(R_{t_j};\varsigma(\Sigma),\lambda), \end{aligned}$$

where Σ is a $d \times d$ symmetric positive semi-definite matrix with diagonal elements Σ_{ii}, $i = 1, ..., d$, and $\varsigma(\Sigma)$ is such that $\varsigma(\Sigma)\varsigma(\Sigma)' = \Sigma$. The functions m_0 and m_1 yield moment equations for estimating μ_0 and Σ_0. The function m_2 is the log-likelihood score with respect to λ.

Let $m := (m_1', m_2', m_3')'$, and define $\hat{m}_N(\beta) := (\hat{m}_{0,N}(\beta)', \hat{m}_{1,N}(\beta)', \hat{m}_{2,N}(\beta)')'$, where

$$\hat{m}_{i,N}(\beta) := N^{-1}\sum_{j=1}^{N} m_i(R_{t_j};\beta), \quad i = 0, 1, 2.$$

Then there generally exists a unique solution ($\hat{\mu}_N$, $\hat{\Sigma}_N$) to the first two moment equations satisfying

$$\hat{m}_{0,N}(\hat{\mu}_N, vech(\hat{\Sigma}_N), \lambda) = 0 \quad \text{and} \quad \hat{m}_{1,N}(\hat{\mu}_N, vech(\hat{\Sigma}_N), \lambda) = 0$$

for all λ, as m_0 and m_1 do not depend on λ. This delivers a first-stage estimator $\hat{\Sigma}_N$ that is generally $N^{1/2}-$consistent for Σ_0. From this we construct $\hat{\sigma}_N = \varsigma(\hat{\Sigma}_N)$. Further, for all $\lambda \in \tilde{\Lambda}_N$, say, where $\tilde{\Lambda}_N$ is the subset of $\hat{\Lambda}_N$ containing the zeros of $Q_N(\hat{\sigma}_N, \lambda)$, we have

$$\hat{m}_{2,N}(\hat{\mu}_N, vech(\hat{\Sigma}_N), \lambda) = 0.$$

That is, the zeros of $\hat{m}_{2,N}$ correspond to (a subset of) our second stage set estimator.

Collecting these facts, we have that for all $\beta \in \{\hat{\mu}_N\} \times \{vech(\hat{\Sigma}_N)\} \times \tilde{\Lambda}_N$,

$$\hat{m}_N(\beta) = 0.$$

Although it is not a typical feature of the likelihood for the Black-Scholes economy, in more general settings, there may be other zeros of $\hat{m}_N(\beta)$, as the likelihood scores may have zeros corresponding to local minima, maxima, or inflection points of the likelihood function. These are the "spurious" zeros referred to above. Nevertheless, because we will not rely on $\hat{m}_N$ to define the identified set of interest, this will not create difficulties.

By making two more identifications, we can state a version of the CHT condition justifying subsampling in the present context. First, we define the single-stage sample objective function

$$\tilde{Q}_N(\beta) := \hat{m}_N(\beta)'\hat{m}_N(\beta).$$

Certain minimizers of this function correspond to our two-stage estimators. Note that this is a standard method of moments objective function; because there are no over-identifying moment conditions, this is also the generalized method of moments objective function (Hansen, 1982). We have $\tilde{Q}_N(\beta) \geq 0$, with the minimum attained at zero because of the lack of over-identification. Finally, define the identified set

$$\mathbb{B}^M_{I,0} := \{\mu_0\} \times \{vech(\Sigma_0)\} \times \Lambda^M_{I,0}.$$

The CHT condition justifying subsampling can now be stated as

$$\sup_{\beta \in \mathbb{B}^M_{I,0}} N\ \tilde{Q}_N(\beta) \xrightarrow{d} Z.$$

To implement the CHT method, we first construct a "preliminary" consistent set estimator, say $\hat{\Lambda}_{N,0}$, and let $l = 1$. Next, we randomly choose B_N subsets of size b, and compute $\hat{\epsilon}_l$ as the $1 - \alpha$ quantile of the statistics

$$\mathcal{Z}_{N,b,k} := \sup_{\lambda \in \hat{\Lambda}_{N,l-1}} bQ_{N,b,k}(\hat{\sigma}_{N,b,k}, \lambda), \quad k = 1, ..., B_N,$$

where $Q_{N,b,k}(\hat{\sigma}_{N,b,k}, \lambda)$ is the criterion function evaluated for the kth $b-$element subset drawn from the full sample of N observations. We then use $\hat{\epsilon}_l$ to get a new set estimator $\hat{\Lambda}_{N,l} = \{\lambda \in \Lambda : N\ Q_N(\hat{\sigma}_N, \lambda) \leq \hat{\epsilon}_l\}$.

We may repeat this process for $l = 2, ..., L$. The final set estimator

$$\hat{\Lambda}_N = \{\lambda \in \Lambda : NQ_N(\hat{\sigma}_N, \lambda) \leq \hat{\epsilon}\}$$

is a $1-\alpha$ confidence set for $\Lambda^M_{I,0}$, taking $\hat{\epsilon} = \hat{\epsilon}_L$. That is, $\liminf_{N\to\infty} P(\Lambda^M_{I,0} \subseteq \hat{\Lambda}_N) \geq 1-\alpha$.

Moreover, $\hat{\Lambda}_N$ is a consistent set estimator when we choose $\hat{\epsilon} = \min(\hat{\epsilon}_L, q_N + \kappa_N)$ for any $\kappa_N \propto \ln N$, where $q_N := \inf_{\lambda\in\Lambda} Q_N(\hat{\sigma}_N, \lambda)$.

1.5 Applications

In this section, we illustrate set estimation and hypothesis testing with two examples. The first studies international risk sharing. The second studies risk premia for market capitalization range index returns.

1.5.1 International risk sharing

A three-country asset price process

Consider three portfolios with prices S^1_t, S^2_t, and S^3_t, each of which is traded in the domestic market of each country $i = 1, 2, 3$. We assume that investors can potentially participate in all three markets. In addition, we assume there is an international risk free asset with a known rate of return r. Let $S_t = (S^1_t, S^2_t, S^3_t)'$. Suppose $\{S_t\}$ is generated by a multivariate Black-Scholes process with $d = 3$ and $n = 4$. Suppose further that Assumptions 1.3.1 and 1.3.2 hold. The identifying restriction on the diffusion coefficient, therefore, is $\rho(\sigma_0) = \rho_0 vec(\sigma_0) = 0$, as described in Section 1.3.2. We thus interpret the first three elements of dW_t as country-specific risks and the fourth element as international risk.

The true market price of risk λ_0 is a 4×1 vector that satisfies $\sigma_0\lambda_0 = \mu_0 - r\iota$ and the bound $\|\lambda_0\| \leq M$. The first three elements of λ_0 represent risk premia on the country specific risks, and the fourth element represents a risk premium on the international risk. Because $d = 3$ and $n = 4$, λ_0 is not point identified. The identified set for the market price of risk, therefore, is $\Lambda^M_{I,0} = \{\lambda : \sigma_0\lambda - \mu_0 -$

$r\iota, \|\lambda\| \leq M\}$. Using set estimation, we can estimate the set of market prices of risk (and therefore risk neutral measures) that are compatible with the behavior of portfolio returns.

In this example, Assumption 1.3.2 fully identifies σ_0, so the identified set for the diffusion coefficients is a singleton, $S = \{\sigma_0\}$. Let $\hat{\Sigma}_N$ be the standard sample covariance estimator. This is a $\sqrt{N}-$consistent estimator of Σ_0 under Assumption 1.2.1. Using the relationship $\sigma_0\sigma_0' = \Sigma_0$, we define a first stage estimator $\hat{\sigma}_N$ to be the (unique) estimator such that $\hat{\sigma}_N\hat{\sigma}_N' = \hat{\Sigma}_N$. This estimator $\hat{\sigma}_N$ is then a $\sqrt{N}-$consistent estimator of σ_0. Given the first-stage estimator $\hat{\sigma}_N$, we estimate the identified set $\Lambda_{I,0}^M$ using eq. (1.4.5). This gives a set of market prices of risk compatible with the observed domestic portfolio returns across the three countries. Assumptions 1.2.1, 1.3.1, and 1.3.2 ensure that the regularity conditions of Corollary 1.4.2 hold, so this is a consistent set estimator of $\Lambda_{I,0}^M$.

We turn now to hypothesis testing. If there is an integrated international financial market, country specific risks should be diversified away. This implies a simple hypothesis that risk premia for the country-specific risks are zero, whereas those who accept the international aggregate risk receive a nonzero risk premium as a reward. According to Lewis (1995), complete markets and optimal risk-sharing imply that the stochastic discount factor varies only with the common international component and is independent of any country specific disturbances. She tests this hypothesis by regressing consumption growth on a constant (the common international component) and domestic output growth (a proxy for country-specific risk), using cross-country data.

In our framework, the international risk sharing hypothesis can be tested using panel data on portfolio returns rather than consumption data[5]. If asset prices are determined in general equilibrium for the integrated world market as described by Lewis (1995), the stochastic discount factor $m(W_t)$ depends only on the international risk. This implies that the elasticities of the pricing kernel with respect to the country-specific risks are zero: i.e., $-\partial \ln m(W_t)/\partial W_t^i = 0$ for $i = 1, 2, 3$. Recall that $m(W_t) = e^{-rt}\xi_t$ and that ξ_t is given by (1.3.2) in the

[5]An approach similar to ours was taken by Campbell and Hamao (1992). They used U.S. and Japanese stock returns to investigate capital market integration based on a single factor model.

multivariate Black-Scholes economy. Thus, the vector of elasticities equals λ. If the country-specific risks are fully diversified away, the first three elements of λ should be 0. We thus let R be the subset of Λ such that

$$R := \{\lambda \in \Lambda : \lambda_1 = \lambda_2 = \lambda_3 = 0\}.$$

The null hypothesis is that there is at least one element λ in the identified set that is consistent with full international risk sharing. This is equivalent to $\Lambda^M_{I,0} \cap R \neq \emptyset$. We can test this hypothesis using the statistic $\hat{T}_N(\Lambda, R) = \inf_{\lambda \in R} NQ_N(\hat{\sigma}_N, \lambda)$ and the subsampling procedure described in Section 1.4.5.

Empirical results

For our empirical study, we consider the financial markets of the U.S., Japan, and Europe. For these three regions, we use standard publicly available data obtained from the Global Financial Database. For the U.S., we use the S&P 500 composite price index, a value weighted index that represents about 75% of the market capitalization of the New York Stock Exchange. To ensure comparability across countries, we use the Tokyo Stock Exchange Price Index (TOPIX) for Japan and the Morgan Stanley Capital International (MSCI) Europe Price Index for Europe. The TOPIX is a value weighted index of all securities traded on the first section of the Tokyo Stock Exchange. The MSCI Europe Price Index is a free-float-adjusted market capitalization-weighted index constructed from indices in 16 developed markets: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.

We use the monthly 1-month T-Bill yield taken from the Center for Research on Security Prices (CRSP) to construct the short term risk-free rate. Specifically, we take the average yield over the whole sample period as our constant risk-free rate r, which is 5.6232% per annum.

The first and last months for which we are able to obtain complete data for all three portfolios and the T-Bill yield are January 1970 and December 2007, for a total of 456 observations. We remove the top and bottom 2.5% of returns from our sample to ensure that the results are not influenced by large outliers. This reduces the sample size to 405. Panel A in Table 1.1 shows summary statistics for the four variables over the full sample period. Panel B reports their variance and correlation coefficients.

Table 1.2 reports the first stage estimate $\hat{\sigma}_N$ for σ_0 with standard errors in parentheses, computed by the delta method. The estimates for the full sample (70:1-07:12) used in the second stage estimation appear in the last column. To assess the stability of the sample, we also report estimates for four sub-periods (70:1-79:12, 80:1-89:12, 90:1-99:12, and 00:1-07:12) in columns 2-5. The estimated coefficients are stable across sub-periods in most cases, although the diffusion coefficient of the MSCI index on its idiosyncratic risk is poorly estimated, especially during the 1980's and 1990's.

To consistently estimate the identified set $\Lambda_{I,0}^M$, we must choose $\hat{\epsilon}$ satisfying the conditions of Theorem 1.4.2. Any $\hat{\epsilon}$ that grows slower than N ensures the consistency of our estimator. We thus choose $\hat{\epsilon}_0 = q_N + \kappa_N$, where $q_N = \inf_{\lambda\in\Lambda} Q_N(\hat{\sigma}_N, \lambda)$ and $\kappa_N \propto \ln N$, and we form the consistent set estimator $\hat{\Lambda}_{N,0} = \{\lambda : NQ_N(\hat{\sigma}_N, \lambda) \leq \hat{\epsilon}_0\}$. In this example, $\Lambda_{I,0}^M$ is a line segment in four-dimensional Euclidean space. The set estimator, therefore, is a four-dimensional cylinder that shrinks down to this line with probability approaching 1.

We project this cylinder to lower dimensional Euclidean spaces to understand its shape. Figure 1.B.1 shows the convex hulls of boundary points of the four-dimensional cylinder projected onto three-dimensional spaces. Note that the surface of the original set is smooth, but the set is approximated by a polygon because of the discretization of the grid. Figure 1.B.1 shows our second stage set estimator projected onto two-dimensional subspaces.

We can construct a confidence region or another consistent set-estimator using the CHT subsampling procedure described above, using $\hat{\Lambda}_{N,0}$ as our preliminary estimator. The 95% confidence region is smaller than the preliminary set estimator and contains the origin, as depicted in Figures 1.B.1 and 1.B.1.

Finally, we formally test the international risk sharing hypothesis. The statistic $\hat{T}_N(\Lambda, R)$ is 9.12 in our sample. We estimate the critical value for $\hat{T}_N(\Lambda, R)$ by subsampling. Table 1.3 provides critical values for different choices of b and B_N. For all of these critical values, we reject the null hypothesis of international risk sharing. Figure 1.B.1 shows the corresponding subsampling distribution with $b = 40$ and $B_N = 5,000$.

As an experiment, we also computed subsampled critical values (not reported here) always using the full sample first-stage estimator in the subsampling exercise. The critical values for the different choices of b and B_N are largely similar to those in Table 1.3, suggesting that the first-stage estimation is not having much impact on the asymptotic distribution of our test statistic.

The fact that this test rejects, whereas the 95% confidence interval contains the origin provides mixed evidence for the risk-sharing hypothesis. Possibly, the direct hypothesis testing approach is more powerful; but without further investigation, we cannot rule out the possibility that this mixed result is due to variations associated with the method of subsampling that would be mitigated in larger samples.

1.5.2 Risk Premia on Cap Range Index Returns

An asset price process for three cap range indexes

Our second illustration concerns risk premia for market capitalization range ("cap range") index returns. Since the seminal work of Fama and French (1993, 1996), many empirical studies have shown that there are sources of priced risk beyond just that associated with movements in the market portfolio. One of these risk factors is known to be related to firm size. Using cap range index returns, we study risk premia on both size-specific risk factors and the market factor without assuming the uniqueness of the risk price (or, equivalently, the risk neutral measure).

For this, suppose Assumption 1.2.1 holds and that there are three portfolios ("large cap," "mid cap," and "small cap") whose returns are generated by a multivariate Black-Scholes economy with $d = 3$ and $n = 4$. As above, we impose Assumptions 1.3.1 and 1.3.2, so that the index return for each cap range is driven only by its idiosyncratic factor and the market factor. Because this structure is exactly the same as in the previous example, we can use the same set-estimation methods.

Previous studies have found that the small cap ($j = 3$) risk and the market risk are priced in the market (see, e.g., Fama and French, 1993, 1996; and Liew and Vassalou, 2000). If the other risks are diversified away, we expect $\lambda_1 = \lambda_2 = 0$. We thus let R be the subset of Λ^M such that

$$R := \{\lambda \in \Lambda^M : \lambda_1 = \lambda_2 = 0\}.$$

Therefore, we consider the null hypothesis that there is at least one parameter value in the identified set that is compatible with the irrelevance of the large cap and mid cap risks: $H_o : \Lambda^M_{I,0} \cap R \neq \emptyset$. Again, we can test this hypothesis using the framework in section 1.4.5.

Empirical results

For our empirical study, we consider three subclasses of firm sizes using the S&P/Citigroup Global Cap Range Index Returns. There, stocks are classified on the basis of their float-adjusted market capitalization. We examine daily returns for the following three indexes: *large cap* ($>$ \$5 billion), *mid cap* (\$1-\$5 billion), and *small cap* ($<$ \$1 billion). The first and last days for which we are able to obtain complete data for all three index returns are August 1, 1989 and December 31, 2007, for a total of 4,805 observations. After removing the top and bottom 2.5% of returns, we obtain 4,420 observations. Once again, we use the monthly 1-month T-Bill yield from CRSP to construct the short term risk-free rate for the same sample period. Our constant risk-free rate r is the average of these rates, 4.0633% per annum.

Table 1.4 reports summary statistics. Table 1.5 reports the first stage estimate $\hat{\sigma}_N$, with standard errors computed using the delta method. The estimated coefficients are stable across sub-periods.

The second-stage set estimator with $M = 20$ is depicted in Figures 1.B.2

and 1.B.2. The 95% confidence region is depicted in Figures 1.B.2 and 1.B.2 We observe that a non-zero premium on the market risk and a zero premium on the small cap risk is plausibly compatible with the returns distribution. For the premia on the large cap and mid cap indexes, the upper left panels of Figures 1.B.2 and 1.B.2 show that the origin is in the set estimator and the confidence region, implying that the irrelevance of the large cap and the mid cap risks is compatible with the returns distribution.

Next, we formally test the irrelevance of large cap and mid cap risks. The test statistic $\hat{T}_N(\Lambda, R)$ is 0.12 in our sample. As before, we estimate the critical value for $\hat{T}_N(\Lambda, R)$ by subsampling. Table 1.6 provides critical values for different choices of b and B_N. For example, with $b = 80$ and $B_N = 1,000$ the critical value is 1.70. Figure 1.B.2 shows this subsampling distribution. For none of the tabulated critical values do we reject the null hypothesis. Therefore, the irrelevance of the large cap and mid cap risks is statistically compatible with the observed returns distribution.

1.6 Modeling More General Asset Price Processes

Our discussion so far has focused mainly on the Black-Scholes case for clarity and conciseness. To the extent that this case is overly simplistic, our empirical results constitute only an illustrative first step in the study of risk pricing in incomplete markets. Nevertheless, much of our analysis and discussion extends to more general processes, providing the foundation for more sophisticated empirical studies. In this section we discuss some aspects of this extension.

A more general data generating process whose special cases are often used in applications is the following geometric process:

Assumption 1.6.1 (Multivariate Geometric Process): *Let $\{W_t\}$ be a vector of $n \in \mathbb{N}$ independent standard Brownian motions under P adapted to the filtration $\{\mathcal{F}_t\}$. Let $\{S_t\}$ be a vector of $d \in \mathbb{N}$ assets such that $S_0^i = 1$ and solving the* *stochastic differential equations*

$$dS_t^i = \mu_{0t}^i S_t^i dt + \sigma_{0t}^i S_t^i dW_t, \quad t \in [0,T], \quad i = 1, ..., d,$$

where μ_{0t} *has elements* $\mu_{0t}^i : \Omega \to \mathbb{R}$ *and* σ_{0t} *has* $1 \times n$ *rows* $\sigma_{0t}^i : \Omega \to \mathbb{R}^n$, *adapted to* $\mathcal{F}_t$, $i = 1, ..., d$. *Further,* $\{S_t\}$ *does not admit arbitrage.*

Under this assumption, $\{\mu_{0t}\}$ and $\{\sigma_{0t}\}$ are general adapted processes. For example, one may posit that a version of Assumption 1.3.2 holds, such that

$$\sigma_{0t} = \begin{bmatrix} \theta_{0\sigma}^1 (S_t^1)^{v-1} & 0 & 0 & \cdots & 0 & \theta_{0\sigma}^{d+1} (S_t^1)^{w-1} \\ 0 & \theta_{0\sigma}^2 (S_t^2)^{v-1} & 0 & \cdots & 0 & \theta_{0\sigma}^{d+2} (S_t^2)^{w-1} \\ \vdots & & \ddots & \cdots & & \vdots \\ \vdots & & & \ddots & \cdots & \vdots \\ 0 & \cdots & \cdots & 0 & \theta_{0\sigma}^d (S_t^d)^{v-1} & \theta_{0\sigma}^{2d} (S_t^d)^{w-1} \end{bmatrix},$$

where $v, w \in [0,1]$, and $(S_t^i)^{v-1}$ denotes the price of the ith security at t raised to the power $v-1$. Letting $\{\lambda_{0t}\}$ denote the true risk price process, suppose λ_{0t} has elements $\lambda_{0t}^i = \theta_{0\lambda}^i (S_t^i)^{1-v}$ for $i = 1, ..., d$ and $\lambda_{0t}^n = \theta_{0\lambda}^n$ for $n = d+1$. Then, by the no arbitrage condition, the drift is determined by $\mu_{0t}^i = \theta_{0\sigma}^{d+i} \theta_{0\lambda}^n (S_t^i)^{w-1} + (\theta_{0\sigma}^i \theta_{0\lambda}^i + r)$ for $i = 1, ..., d$. The process is indexed by the coefficient vector $\theta_0 = (\theta_{0\sigma}^1, \cdots, \theta_{0\sigma}^{2d}, \theta_{0\lambda}^1, \cdots, \theta_{0\lambda}^n) \in \mathbb{R}^{2d+n}$.

For the Black-Scholes economy, $v = w = 1$. The general case in which $v = w$ corresponds to a multivariate version of the constant elasticity of variance (CEV) process often used to model stock prices, short rates, forward rates, and stochastic volatilities. Various other cases of interest arise by varying v and w. For instance, choosing $v = 1/2$ and $w = 0$ gives a process whose idiosyncratic component follows a square root process, as in Cox, Ingersoll, and Ross (1985), and whose aggregate component follows a Brownian motion.

Further, because the $\sigma-$field $\mathcal{G}_t := \sigma(W_\tau, \tau \in [0,t])$ generated by the $t-$history of the multivariate Brownian motion $\{W_t\}$ may be a proper subset of $\mathcal{F}_t$, this assumption also covers certain more general stochastic volatility processes.

The analog of Assumption 1.3.2 becomes

ASSUMPTION 1.6.2 (Envelope Process): *There exists an adapted process* $\{M_t\}$ *such that* $0 < M_t < \infty$ *for* $t \in [0,T]$, $||\lambda_{0t}|| \le M_t$, *and* $E^P\left[\exp\left(\int_0^T M_t^2 dt\right)\right] < \infty$.

This condition ensures that $var(\xi_T) < \infty$, where

$$\xi_T = \exp\left(-\int_0^T \lambda_{0s} \cdot dW_s - \frac{1}{2}\int_0^T \|\lambda_{0s}\|^2 ds\right).$$

Under these assumptions, the market prices of risk at time t belong to the random set

$$\Lambda_{I,t}^M := \Lambda_{I,t} \cap \Lambda^{M_t},$$

where

$$\Lambda_{I,t} := \{\lambda : \sigma_{0t}\lambda = \mu_{0t} - r\iota\} \quad \text{and} \quad \Lambda^{M_t} := \{\lambda : ||\lambda|| \le M_t\}.$$

To apply maximum likelihood methods, we parameterize λ_{0t} and σ_{0t} as follows:

ASSUMPTION 1.6.3 (Parametric Specification): *Let* Θ *be a compact subset of* $\mathbb{R}^p$, $p \in \mathbb{N}$. *(i) For* $t \in [0,T]$, *the functions* $\ell_t : \Omega\times\Theta \to \mathbb{R}^n$ *and* $s_t : \Omega\times\Theta \to \mathbb{R}^{d\times n}$ *are such that for each* $\theta \in \Theta$, $\ell_t(\cdot,\theta)$ *and* $s_t(\cdot,\theta)$ *are measurable*$-\mathcal{F}_t$, *and for each* $\omega \in \Omega$, $\ell_t(\omega,\cdot)$ *and* $s_t(\omega,\cdot)$ *are continuous on* Θ*; (ii) for each* $\theta \in \Theta$ *and* $t \in [0,T]$, $||\ell_t(\cdot,\theta)|| \le M_t$*; (iii) there exists* $\theta_0 \in \Theta$ *such that for* $t \in [0,T]$, $\sigma_{0t} = s_t(\cdot,\theta_0)$ *and* $\lambda_{0t} = \ell_t(\cdot,\theta_0)$.

Similar to our discussion above, Θ is the parameter space; for convenience, we assume that it implicitly embodies any prior restrictions known to hold for θ_0, such as $\rho(\theta_0) = 0$. The first part of this assumption specifies the parametric functions ℓ_t and s_t. In the second part, we require that the bound of Assumption 1.6.2 holds for all θ in Θ. The third part ensures that this specification is correct, in that there is a parameter value θ_0 in Θ corresponding to the true arbitrage-free process generating asset returns.

This assumption implies a parameterization for μ_{0t} of the form $m_t(\cdot,\theta) = s_t(\cdot,\theta)\ell_t(\cdot,\theta) + r\iota$. Alternatively, one may directly parameterize μ_{0t} instead of λ_{0t};

the no arbitrage condition then implies a parameterization for λ_{0t}. For brevity, we leave aside this possibility here.

Successive conditioning yields a likelihood function for returns $R_{t_j} := \ln S_{t_j} - \ln S_{t_{j-1}}$ defined by

$$f_N(R_{t_1}, \cdots, R_{t_N}; \theta) = \prod_{j=1}^{N} f_{t_j}(R_{t_j}; \theta \mid \mathcal{H}_{t_{j-1}}),$$

where $f_{t_j}(R_{t_j}; \theta \mid \mathcal{H}_{t_{j-1}})$ defines the likelihood for returns in t_j given the information $\mathcal{H}_{t_{j-1}}$, where $\sigma(R_{t_1}, \cdots, R_{t_{j-1}}) \subseteq \mathcal{H}_{t_{j-1}}$. This likelihood function does not necessarily have a closed form expression. In such cases, we may rely on an approximation of the likelihood function. See Aït-Sahalia (2002, 2008) and Kristensen (2008), for example.

Analogous to the Black-Scholes case, the criterion function $\bar{Q}_N : \Theta \to \bar{\mathbb{R}}_+$ is the shifted expected negative average log-likelihood defined by

$$\bar{Q}_N(\theta) := E^P \left[-N^{-1} \sum_{j=1}^{N} \ln f_{t_j}(R_{t_j}; \theta \mid \mathcal{H}_{t_{j-1}}) \right] - q_{0,N},$$

where

$$q_{0,N} := E^P \left[-N^{-1} \sum_{j=1}^{N} \ln f_{t_j}(R_{t_j}; \theta_0 \mid \mathcal{H}_{t_{j-1}}) \right].$$

The identified set is again the set of zeros of $\bar{Q}_N$,

$$\Theta_I := \{\theta \in \Theta : \bar{Q}_N(\theta) = 0\}.$$

The identified market prices of risk at time t are then given by the Effros-measurable set

$$\lambda_t(\Theta_I) := \{\lambda : \lambda = \ell_t(\theta), \quad \theta \in \Theta_I\} \subset \Lambda^M_{I,t}.$$

It is not immediately obvious that $\lambda_t(\Theta_I) = \Lambda^M_{I,t}$. Ensuring that this holds may require further conditions. Nevertheless, the correct specification assumption ensures that $\lambda_{0t} \in \lambda_t(\Theta_I)$.

The sample criterion function is given by $Q_N : \Omega \times \Theta \to \bar{\mathbb{R}}_+$, defined by

$$Q_N(\theta) = -N^{-1} \sum_{j=1}^{N} \ln f_{t_j}(R_{t_j}; \theta \mid \mathcal{H}_{t_{j-1}}) - q_N,$$

where $q_N = \inf_\Theta -N^{-1} \sum_{j=1}^{N} \ln f_{t_j}(R_{t_j}; \theta \mid \mathcal{H}_{t_{j-1}})$.

As when we apply the CHT approach above, we define an ϵ-level set for the sample criterion function by

$$\hat{\Theta}_N(\epsilon) := \{\theta \in \Theta : N \cdot Q_N(\theta) \leq \epsilon\}.$$

When we choose ϵ properly, the random set $\hat{\Theta}_N(\epsilon)$ is Effros measurable and is a consistent set estimator or confidence region for Θ_I. The estimated prices of risk at time t are given by $\lambda_t(\hat{\Theta}_N(\epsilon))$.

Hypothesis tests for the risk price process $\{\lambda_t(\Theta_I)\}$ can be conducted by inverting the confidence interval process $\{\lambda_t(\hat{\Theta}_N(\epsilon))\}$ or using a likelihood ratio test for $\theta \in \Theta_R$, where $\Theta_R \subset \Theta$ expresses the restrictions specified by a null hypothesis of interest, e.g.,

$$H_o^\Lambda : \lambda_{0|\mathcal{T}} \in \Lambda_R, \quad t \in \mathcal{T},$$

where $\mathcal{T}$ is a given subset of $[0, T]$, $\lambda_{0|\mathcal{T}}$ denotes the process $\{\lambda_{0t}\}$ restricted to $\mathcal{T}$, and Λ_R is a given subset of $\Lambda^{M_\mathcal{T}} := \{\{\Lambda^{M_t}\}, t \in \mathcal{T}\}$.

As before, two-stage estimation can help in mitigating the challenges arising in estimating θ_0. Specifically, let $\theta_0 := (\theta_{01}, \theta_{02}) \in \Theta_1 \times \Theta_2 =: \Theta$, and let $\hat{\Theta}_{1N}$ be a first-stage set estimator for θ_{01}.

Using this, let the second-stage set estimator for the identified set be defined by

$$\hat{\Theta}_N(\omega) := \{(\theta_1, \theta_2) \in \Theta : NQ_N(\omega, \theta_1, \theta_2) \leq \hat{\epsilon}(\omega), \quad \theta_1 \in \hat{\Theta}_{1N}(\omega)\},$$

where $\hat{\epsilon}$ may be random, as before.

When the available restrictions suffice to fully identify θ_{01}, we have $\hat{\Theta}_{1N} = \{\hat{\theta}_{1N}\}$, say. The second-stage set estimator is then $\hat{\Theta}_N = \{\hat{\theta}_{1N}\} \times \hat{\Theta}_{2N}$, where

$$\hat{\Theta}_{2N}(\omega) := \{\theta_2 \in \Theta_2 : NQ_N(\omega, \hat{\theta}_{1N}(\omega), \theta_2) \leq \hat{\epsilon}(\omega)\}.$$

Subsampling remains an appealing method for constructing confidence regions here. To provide conditions ensuring its validity, in particular the CHT condition, recasting a two-stage procedure as a single-stage procedure may again prove convenient. Depending on the particular circumstances, it may be possible to use a method of moments approach analogous to that discussed in Section 1.4.6. In other cases, it may be helpful to exploit an exponentially tilted likelihood, along the lines proposed by Kitamura and Stutzer (1997).

Formally ensuring consistency and convergence in distribution of these estimators will require careful specification of further regularity conditions appropriate to the specific context of interest. Nevertheless, the framework sketched here should prove helpful in pursuing these results.

1.7 Concluding remarks

In this paper, we study an econometric framework useful for estimating and testing hypotheses about the price of risk in the absence of complete markets. We state results ensuring the Effros-measurability and consistency of set estimators for the vector of market risk prices, and we discuss the construction of hypothesis tests and confidence sets using subsampling.

Our results build on the seminal work of Chernozhukhov, Hong, and Tamer (2007) for estimation and testing in partially identified models. To handle the challenges associated with jointly estimating all parameters of the model, we apply a two-stage method introduced by Kaido and White (2008). For the present application, we estimate covariance parameters in the first stage and risk prices in the second stage. For hypothesis testing, we make use of a subsampling procedure proposed by Romano and Shaikh (2006, 2008). To illustrate, we apply our methods to estimate market risk prices and test hypotheses concerning international risk sharing and market capitalization range indexes.

By providing new methods for inference on risk neutral measures in incomplete markets, our work thus complements that of Aït-Sahalia and Lo (1998), Chernov and Ghysels (2000), Clement, Gourieroux, and Monfort (2000), and Abadir and Rockinger (2003), among others.

An interesting direction for further research is to study investor risk preferences in the absence of the identification of market risk prices. This may create an opportunity to extend the work of Aït-Sahalia and Lo (2000), Jackwerth (2000), and Rosenberg and Engle (2002).

One of the key assumptions in our framework is a bound, M_t, on the market price of risk. Not only does this bound sharpen our set estimators, but it also plays a key role when using the estimated risk neutral measure to price non-redundant securities. Cochrane and Saá-Requejo (2000) show that this type of L^2 bound on the Radon-Nikodym derivative (or SDF) delivers sharper upper and lower bounds on the price of the non-redundant security. In related work, Bernardo and Ledoit (2000) consider a L^∞ bound on the Radon-Nikodym derivative. Further investigation of the choice of M_t, particularly the use of empirical evidence to choose M_t, is an interesting topic for further research.

Yet another interesting topic is the development of tests for market completeness *per se*. Such tests will require careful specification of the nature of the alternative complete and incomplete market structures, together with a theory of estimation and inference for parameters partially identified only under the alternative, possibly on the boundary of the parameter space. This will require extension of work of Davies (1977, 1987) and Andrews (1999, 2001) to the context of partial identification.

To maintain a sharp focus for our results, we have considered in detail the multivariate Black-Scholes economy. Nevertheless, our framework applies more broadly, and we sketch some features of its application to more general geometric processes. Extension to asset prices generated by Levy processes or subordinated processes are other interesting possibilities deserving attention in future work. Methods of estimation and inference for such potentially more realistic asset-price generating processes will then make possible increasingly refined empirical studies of risk pricing in incomplete markets.

1.A Tables

1.A.1 Tables for International Risk Sharing

Table 1.1: Summary Statistics for Stock Index Returns and the T-Bill Yield

A:	S&P500	TOPIX	MSCI	T-Bill
Mean	0.0061	0.0067	0.0069	0.0047
Std. Dev.	0.0343	0.0504	0.0378	0.0023
Min	-0.0893	-0.1141	-0.1027	0.0007
Max	0.0829	0.1382	0.0926	0.0135
Skewness	-0.0425	-0.0986	-0.0167	0.2800
Kurtosis	-0.1081	-0.0484	0.1247	0.3700
Obs	405	405	405	405

B:	S&P500	TOPIX	MSCI
S&P500	0.00118		
TOPIX	0.23588	0.00254	
MSCI	0.50724	0.38051	0.00143

The sample period is 1970:1-2007:12 with 456 monthly observations. We remove observations corresponding to the top and bottom 2.5% of returns for each series. This reduces the sample size to 405. We compute returns from the S&P500, TOPIX, and MSCI Europe Price Indexes obtained from the Global Financial Database for stock index returns. For the risk-free rate, we average monthly 1-month T-Bill yields taken from the CRSP database. We report robust measures of skewness and kurtosis. Skewness is computed as $SK = (Q_3 + Q_1 - 2Q_2)/(Q_3 - Q_1)$, where Q_i is the ith quartile of the return. Kurtosis is computed as $KR = (E_7 - E_5 + E_3 - E_1)/(E_6 - E_2) - 1.23$, where E_i is the ith octile. See Kim and White (2004) for details. Panel B reports variance (diagonal) and correlation (off-diagonal) coefficients for the returns.

Table 1.2: First-Stage Diffusion Coefficient Estimates

	2	3	4	5	6
	70:1-79:12	80:1-89:12	90:1-99:12	00:1-07:12	70:1-07:12
σ_{11}	0.1106	0.1101	0.0925	0.0520	0.0982
	(0.0292)	(0.0324)	(0.0336)	(0.0616)	(0.0337)
σ_{22}	0.1327	0.1546	0.1833	0.1518	0.1580
	(0.0531)	(0.0410)	(0.0428)	(0.0450)	(0.0441)
σ_{33}	0.0915	0.0000	0.0460	0.0676	0.0572
	(0.0514)	(0.2288)	(0.1348)	(0.0618)	(0.1188)
σ_{14}	0.0554	0.0534	0.0748	0.1038	0.0698
	(0.0472)	(0.0433)	(0.0458)	(0.0401)	(0.0447)
σ_{24}	0.1029	0.0690	0.0731	0.0648	0.0769
	(0.0733)	(0.0602)	(0.0595)	(0.0494)	(0.0607)
σ_{34}	0.0945	0.1606	0.1176	0.1112	0.1198
	(0.0575)	(0.1057)	(0.0593)	(0.0472)	(0.0632)
Number of Obs.	111	100	105	89	405

Columns 2-5 report the estimated diffusion coefficients (with standard errors in parentheses) for the following sample periods: 1970:1-1979:12, 1980:1-1989:12, 1990:1-1999:12, and 2000:1-2007:12. The last column reports the estimation results for the full sample.

Table 1.3: Critical Values for Various Choices of b and B_N

b	B_N	$d^{\Lambda}_{405,0.95}$
80	5000	6.36
80	1000	6.03
80	500	6.36
40	5000	4.63
40	1000	4.70
40	500	5.13
20	5000	3.81
20	1000	4.06
20	500	4.12

The third column reports 95% critical values of the test statistic. For each subsample, we estimate the diffusion coefficient $\hat{\sigma}_{N,b,k}$ and compute the statistic $\hat{T}_{N,b,k}$.

1.A.2 Tables for Market Cap Index Returns

Table 1.4: Summary Statistics for Cap Range Index Returns and the T-Bill Yield

A:	Large Cap	Mid Cap	Small Cap	T-Bill
Mean	0.00042	0.00051	0.00052	0.00016
Std. Dev.	0.00732	0.00706	0.00761	0.00007
Min	-0.02043	-0.02019	-0.02214	0.00003
Max	0.02021	0.01942	0.02117	0.00031
Skewness	0.03916	0.03459	-0.02127	-0.41912
Kurtosis	0.25411	0.22744	0.26137	0.04876
Obs	4,420	4,420	4,420	221

B:				
Large Cap	0.000054			
Mid Cap	0.860242	0.000050		
Small Cap	0.731907	0.904943	0.000058	

The sample period is 08/01/1989-12/31/2007 with 4,805 daily observations for the cap range index returns. We remove observations corresponding to the top and bottom 2.5% of returns for each series. This reduces the sample size to 4,420. We compute returns from the S&P/Citigroup Global Cap Range Index data for: *large cap* ($>$ \$5 billion), *mid cap* (\$1-\$5 billion), and *small cap* ($<$ \$1 billion). For the risk-free rate, we average monthly 1-month T-Bill yields taken from the CRSP database for the same sample period (221 monthly observations). We report robust measures of skewness and kurtosis. Skewness is computed as $SK = (Q_3 + Q_1 - 2Q_2)/(Q_3 - Q_1)$, where Q_i is the ith quartile of the return. Kurtosis is computed as $KR = (E_7 - E_5 + E_3 - E_1)/(E_6 - E_2) - 1.23$, where E_i is the ith octile. See Kim and White (2004) for details. Panel B reports variance (diagonal) and correlation (off-diagonal) coefficients for the returns.

Table 1.5: First-Stage Diffusion Coefficient Estimates

	2	3	4
	1989-1999	2000-2007	1989-2007
σ_{11}	0.0662	0.0622	0.0654
	(0.0044)	(0.0043)	(0.0044)
σ_{22}	0.0000	0.0000	0.0000
	(0.0055)	(0.0078)	(0.0059)
σ_{33}	0.0515	0.0618	0.0591
	(0.0038)	(0.0049)	(0.0043)
σ_{14}	0.0961	0.1044	0.0991
	(0.0069)	(0.0067)	(0.0068)
σ_{24}	0.1079	0.1302	0.1181
	(0.0058)	(0.0065)	(0.0062)
σ_{34}	0.0848	0.1383	0.1084
	(0.0061)	(0.0076)	(0.0071)
Number of Obs.	2,603	1,817	4,420

Standard errors in parentheses.

Table 1.6: Critical Values for Various Choices of b and B_N

b	B_N	$d_{4,420,0.95}$
320	1000	1.68
320	500	1.55
160	1000	1.54
160	500	1.74
80	1000	1.70
80	500	1.59
40	1000	1.66
40	500	1.62

The third column reports 95% critical values of the test statistic. For each subsample, we estimate the diffusion coefficient $\hat{\sigma}_{N,b,k}$ and compute the statistic $\hat{T}_{N,b,k}$.

1.B Figures

1.B.1 Figures for International Risk Sharing

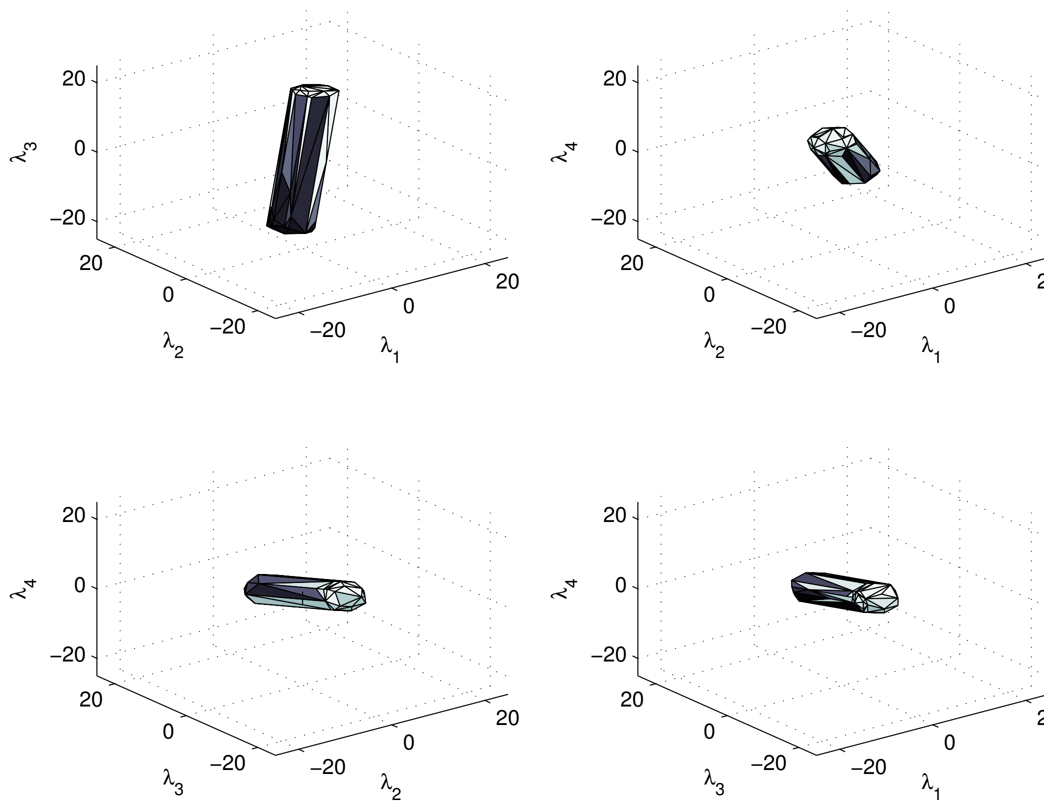

Figure 1.3: Second-Stage Set Estimator (Convex Hulls of Projections to Three-Dimensional Subspaces)

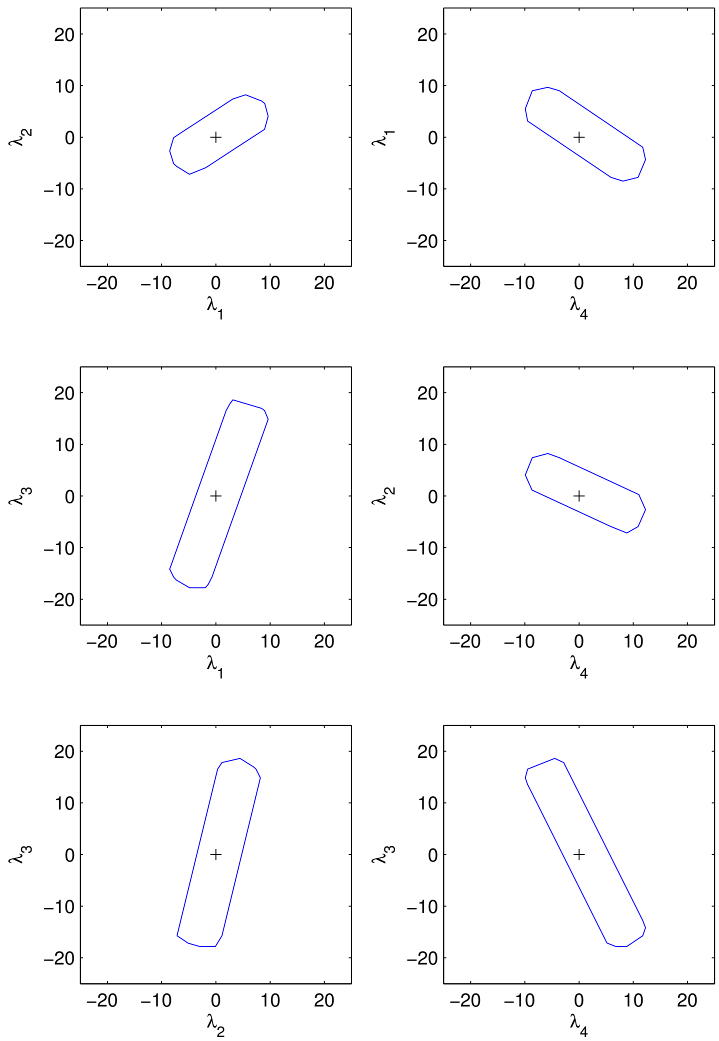

Figure 1.4: Second-Stage Set Estimator (Projections of Boundary Points to Two-Dimensional Subspaces)

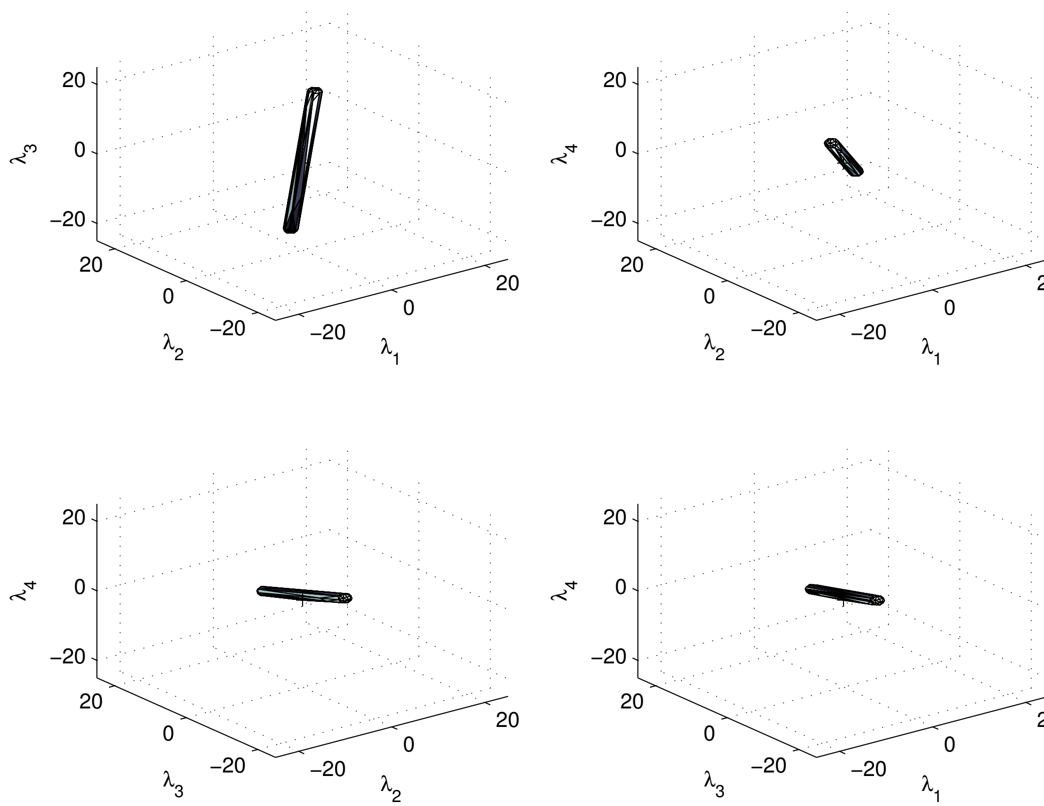

Figure 1.5: 95% Confidence Region (Convex Hulls of Projections to Three-Dimensional Subspaces)

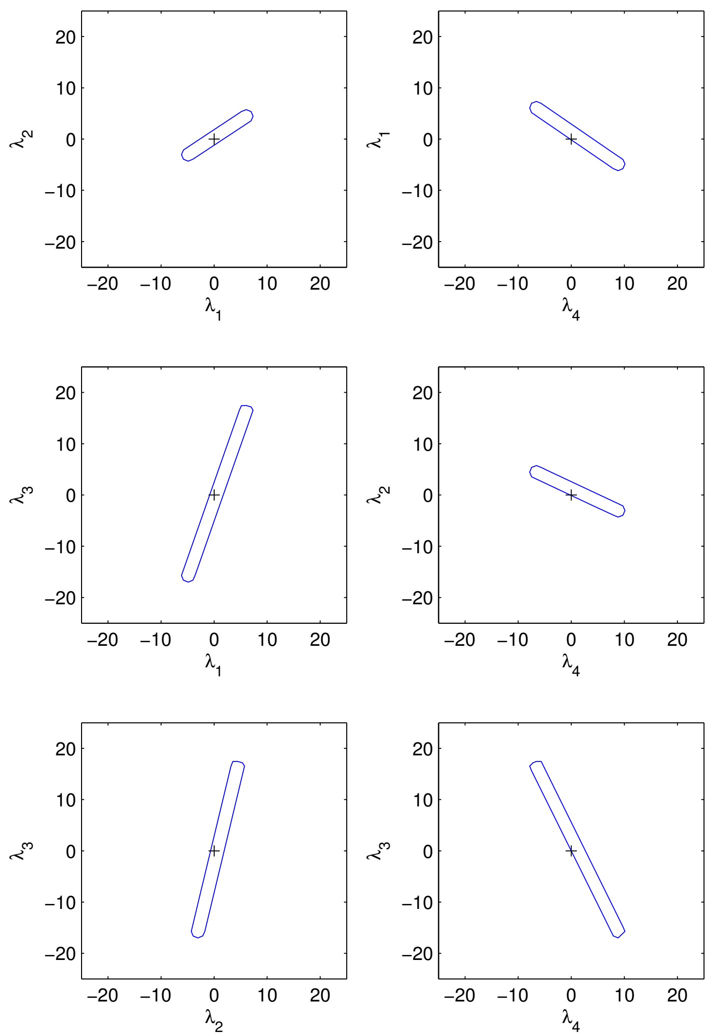

Figure 1.6: 95% Confidence Region (Projections of Boundary Points to Two-Dimensional Subspaces)

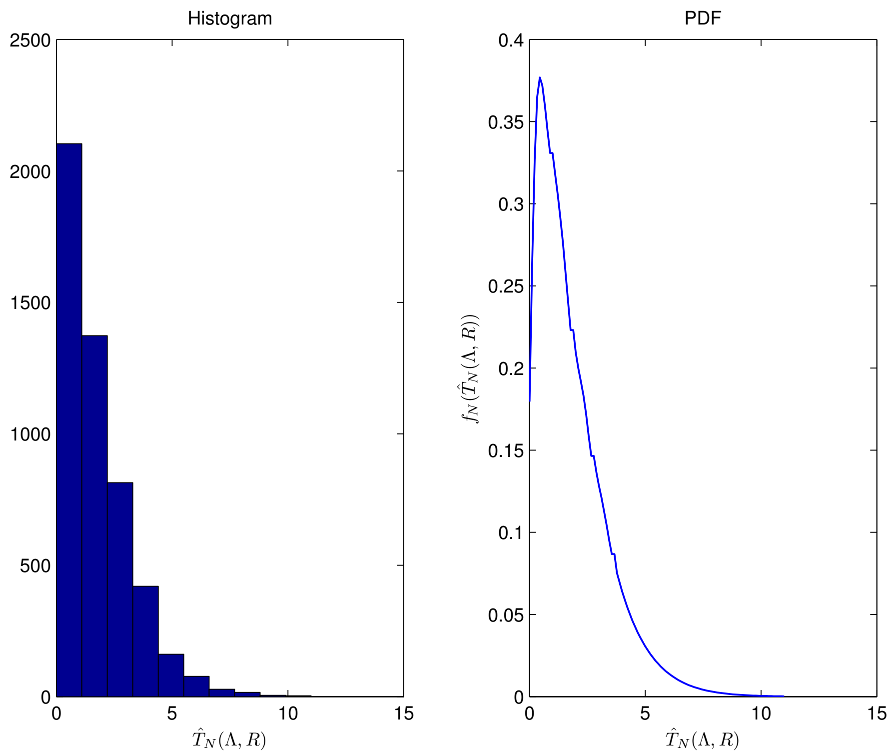

Figure 1.7: Subsampling Distribution of the LR Test Statistic $(N = 405, b = 40, B_N = 5,000)$

1.B.2 Figures for Returns on Market Cap Index Returns

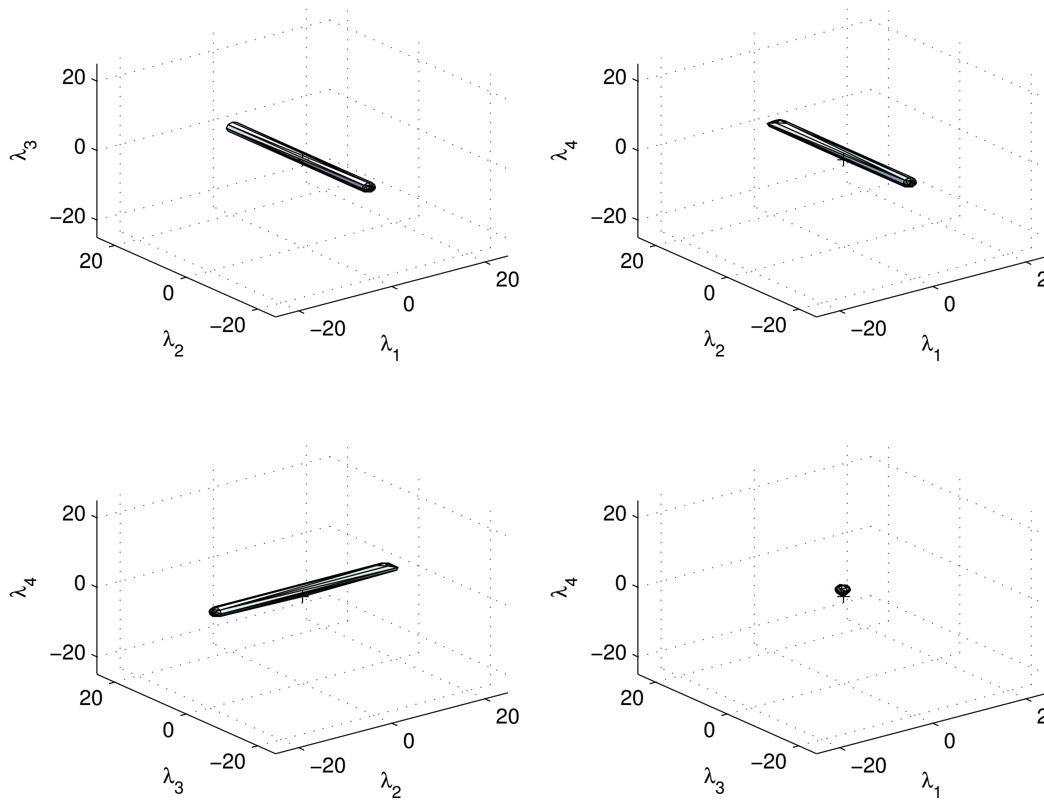

Figure 1.8: Second-Stage Set Estimator (Convex Hulls of Projections to Three-Dimensional Subspaces)

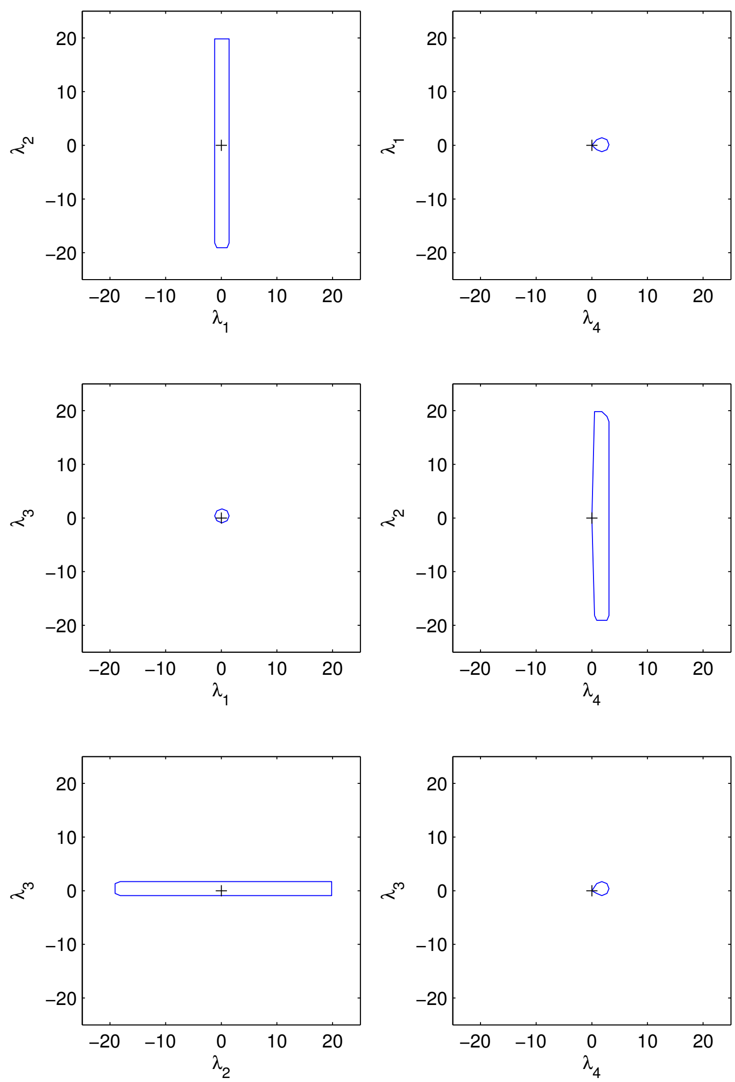

Figure 1.9: Second-Stage Set Estimator (Projections of Boundary Points to Two-Dimensional Subspaces)

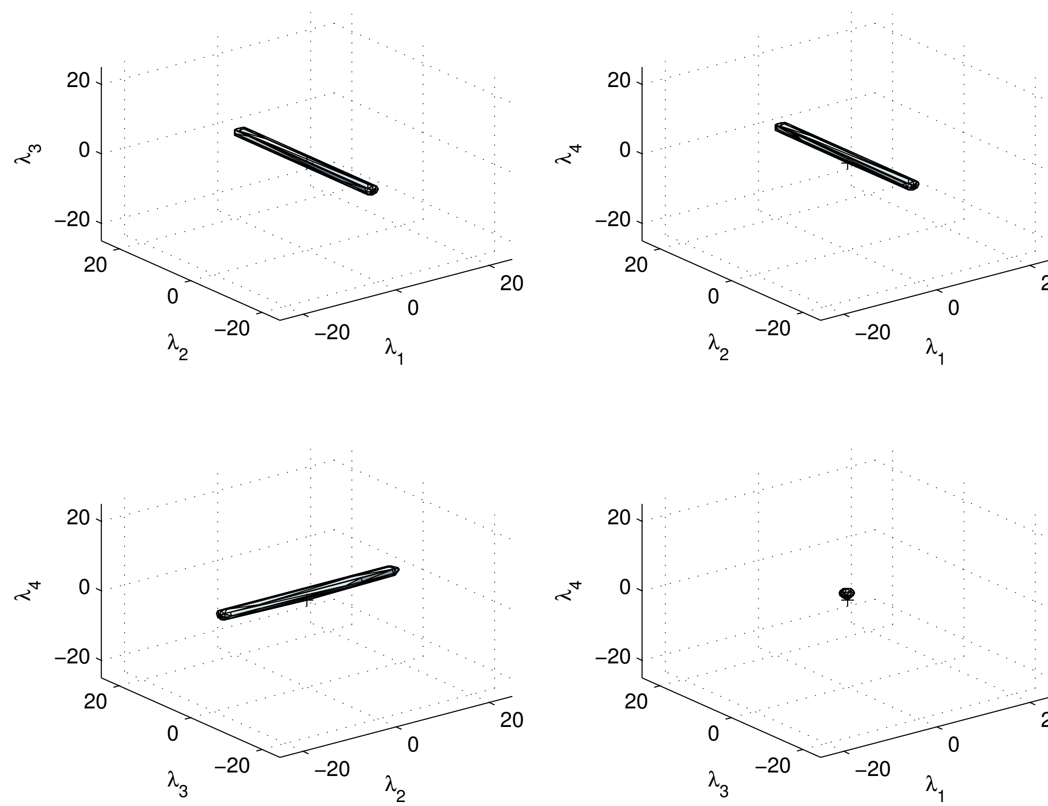

Figure 1.10: Confidence Region (Convex Hulls of Projections to Three-Dimensional Subspaces)

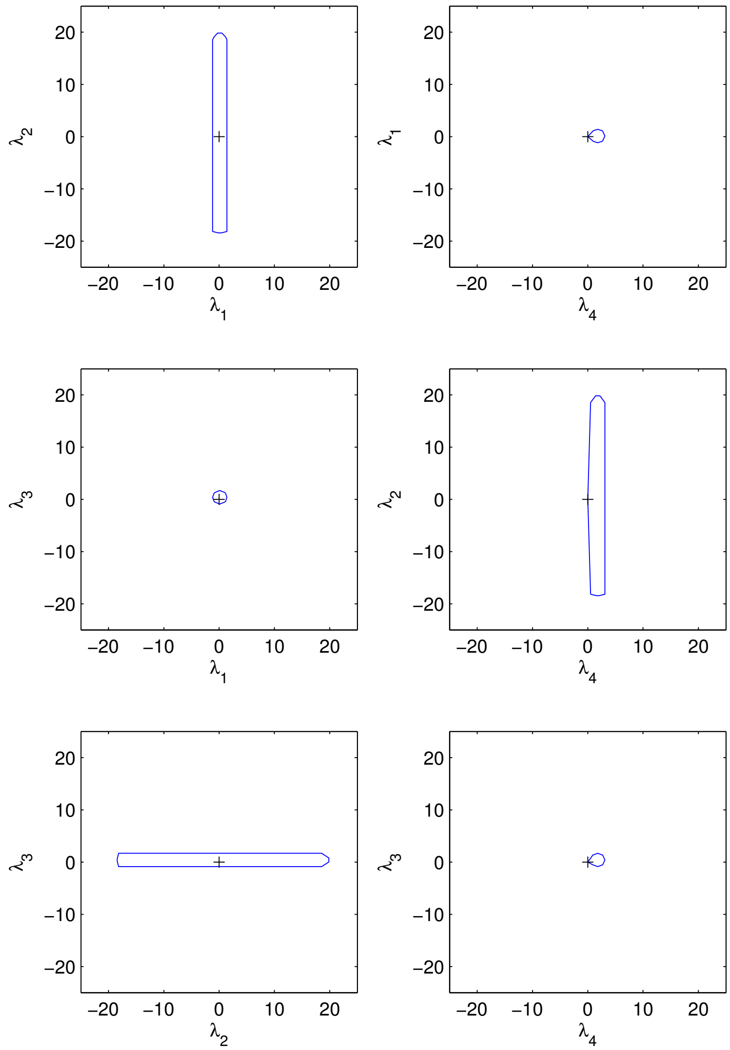

Figure 1.11: Confidence Region (Projections of Boundary Points to Two-Dimensional Subspaces)

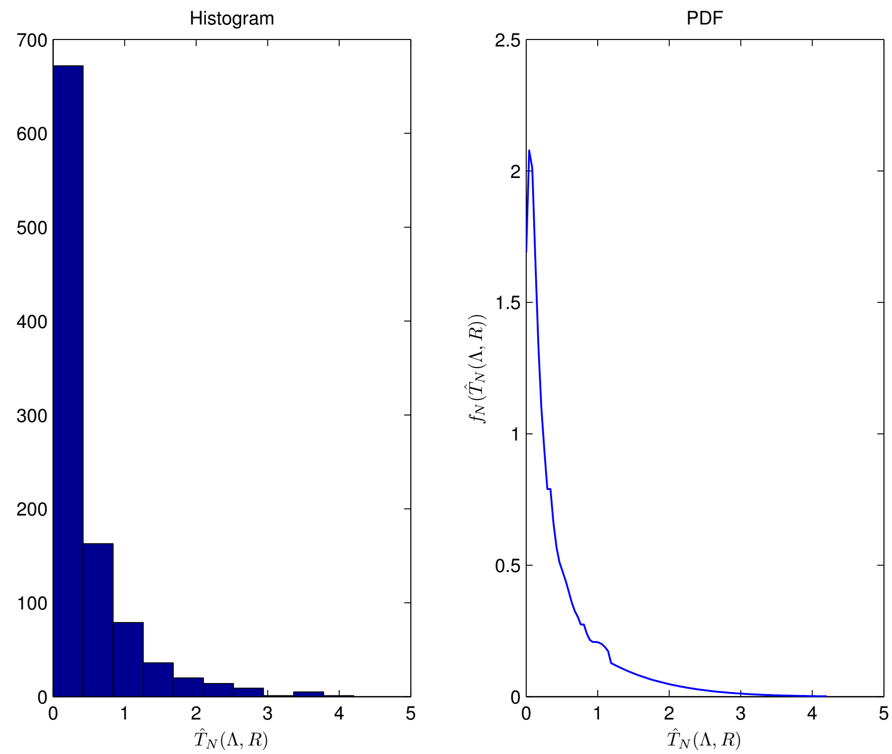

Figure 1.12: Subsampling Distribution of the LR Test Statistic ($N = 4,420, b = 320, B_N = 1,000$)

Bibliography

Abadir, K., and M. Rockinger (2003): “Density Functionals, with an Option-Pricing Application,” *Econometric Theory*, 19(05), 778–811.

Aït-Sahalia, Y. (2002): “Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach,” *Econometrica*, 70(1), 223–262.

——— (2008): “Closed-form Likelihood Expansions for Multivariate Diffusions,” *Annals of Statistics*, 36(2), 906–937.

Aït-Sahalia, Y., and A. W. Lo (1998): “Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices,” *The Journal of Finance*, 53(2), 499–547.

——— (2000): “Nonparametric Risk Management and Implied Risk Aversion,” *Journal of Econometrics*, 94(1-2), 9–51.

Andrews, D. W. K. (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.

Bernardo, A., and O. Ledoit (2000): “Gain, Loss and Asset Pricing,” *Journal of Political Economy*, 108(1), 144.

Breeden, D. T. (1979): “An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities,” *Journal of Financial Economics*, 7(3), 265–296.

Campbell, J. Y., and Y. Hamao (1992): “Predictable Stock Returns in the United States and Japan: A Study of Long-Term Capital Market Integration,” *The Journal of Finance*, 47(1), 43–69.

Chernov, M., and E. Ghysels (2000): “A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation,” *Journal of Financial Economics*, 56(3), 407–458.

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

Clement, E., C. Gourieroux, and A. Monfort (2000): "Econometric Specification of the Risk Neutral Valuation Model," *Journal of Econometrics*, 94(1-2), 117–143.

Cochrane, J., and J. Saá-Requejo (2000): "Beyond Arbitrage: Good-Deal Asset Price Bounds in Incomplete Markets," *Journal of Political Economy*, 108(1), 79.

Cox, J. C., J. Ingersoll, Jonathan E., and S. A. Ross (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53(2), 385–408.

Davies, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 64(2), 247–254.

Davies, R. B. (1987): "Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 74(1), 33–43.

Delbaen, F., and W. Schachermayer (1994): "A General Version of the Fundamental Theorem of Asset Pricing," *Mathematische Annalen*, V300(1), 463–520.

Duffie, D. (2001): *Dynamic Asset Pricing Theory.* 3rd ed. Princeton: Princeton University Press.

Fama, E. F., and K. R. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33(1), 3–56.

——— (1996): "Multifactor Explanations of Asset Pricing Anomalies," *The Journal of Finance*, 51(1), 55–84.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–1054.

Harrison, J. M., and D. M. Kreps (1979): "Martingales and Arbitrage in Multiperiod Securities Markets," *Journal of Economic Theory*, 20(3), 381–408.

Harrison, J. M., and S. R. Pliska (1981): "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Processes and their Applications*, 11(3), 215–260.

Jackwerth, J. (2000): "Recovering Risk Aversion from Option Prices And Realized Returns," *Review of Financial Studies*, 13(2), 433–451.

KAIDO, H., AND H. WHITE (2008): "A Two-Stage Procedure for Partially Identified Models," Discussion Paper, University of California, San Diego.

KIM, T.-H., AND H. WHITE (2004): "On More Robust Estimation of Skewness and Kurtosis," *Finance Research Letters*, 1(1), 56–73.

KITAMURA, Y., AND M. STUTZER (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65(4), 861–874.

KRISTENSEN, D. (2008): "Pseudo-Maximum Likelihood Estimation in Two Classes of Semiparametric Diffusion Models," Working Paper, Columbia University.

LEWIS, K. K. (1995): "Puzzles in International Financial Markets," in G. M. Grossman, and K. Rogoff (eds.) *Handbook of International Economics*, vol. 3, New York: Elsevier, pp. 1913–1971.

LIEW, J., AND M. VASSALOU (2000): "Can Book-to-Market, Size, and Momentum be Risk Factors that Predict Economic Growth?" *Journal of Financial Economics*, 57(2), 221–245.

LINTNER, J. (1965): "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *The Review of Economics and Statistics*, 47(1), 13–37.

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

LUCAS, R. E. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46(6), 1429–1445.

MANSKI, C. F. (2003): *Partial Identification of Probability Distributions*. Berlin: Springer.

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

PROTTER, P. E. (2005): *Stochastic Integration and Differential Equations*, 3rd ed. Berlin: Springer.

ROMANO, J. P., AND A. SHAIKH (2006): "Inference for the Identified Set in Partially Identified Econometric Models," Working Paper, University of Chicago.

——— (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, Special Issue in Honor of T. W. Anderson, forthcoming.

ROSENBERG, J. V., AND R. F. ENGLE (2002): "Empirical Pricing Kernels," *Journal of Financial Economics*, 64(3), 341–372.

Santos, A. (2007): "Inference in Nonparametric Instrumental Variables with Partial Identification," Working Paper, University of California, San Diego.

Sharpe, W. F. (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *The Journal of Finance*, 19(3), 425–442.

Williams, R. J. (2006): *Introduction to the Mathematics of Finance*. New York: American Mathematical Society.

This chapter, in full, is a reprint of the material as it appears in the *Journal of Financial Econometrics* 2009. Kaido, Hiroaki; White, Halbert, Oxford University Press, 2009. I was the primary investigator and author of this paper.

Chapter 2

A Two-Stage Procedure for Partially Identified Models

2.1 Introduction

Statistical inference for partially identified economic models, pioneered by Manski (see Manski, 2003, and the references there), 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 challenge for any estimator arises when the dimension of the parameter space is large. This is a particular challenge for set-valued estimators, as high dimensionality can create computational difficulties and seriously hamper the interpretation of estimation results. If some *a priori* knowledge about the data generating process is available, however, we can exploit this knowledge to reduce the dimension and/or the volume of the set estimator. Indeed, we often encounter cases in practice where some of the parameters can be identified and estimated separately from the rest. Economic theory, e.g., in the form of optimization or equilibrium conditions, may impose additional parameter restrictions. This paper studies how a natural two-stage extension of the CHT framework can exploit such knowledge to mitigate the problems otherwise associated with set estimation in high-dimensional parameter spaces.

Specifically, we consider cases satisfying the following conditions: (i) the parameter vector consists of two sub-vectors: a "first-stage" parameter vector and a "second-stage" parameter vector; (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 a criterion function. These conditions are often satisfied by models studied in the literature. For example, Bajari, Benkard, and Levin (2007)'s 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.

An important feature here is that we do not necessarily require the model to be correctly specified. The study of partially identified misspecified models has begun very recently. Panomareva and Tamer (2009) and Bugni, Canay, and Guggenberger (2010) study misspecified moment inequality models. Kline and Santos (2010) study misspecified quantile models with missing data. In contrast we study general extremum estimation with possible misspecification.

The paper is organized as follows. Section 2.2 summarizes CHT's econometric framework and formalizes the two-stage structure just described, without requiring correct specification. Section 2.3 gives several illustrative examples. Section 2.4 studies measurability and consistency of the two-stage set estimator. Section 2.5 studies inference based on a quasi-likelihood ratio statistic, extending results of Liu and Shao (2004), and Section 2.6 concludes. An appendix contains formal proofs.

2.2 The Data Generating Process and the Model

2.2.1 CHT Framework and Two-Stage Structure

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

Assumption 2.2.1: *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 sub-vectors, $\theta_1 \in \Theta_1$ 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.

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

$$Q_n(\omega, \theta) = n^{-1} \sum_{i=1}^{n} q(X_i(\omega), \theta) - \inf_{\theta \in \Theta} \; n^{-1} \sum_{i=1}^{n} q(X_i(\omega), \theta), \tag{2.2.1}$$

where $\{X_i : \Omega \to \mathcal{X}\}$ 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

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

Following CHT, we define the *identified set* as the set of minimizers of $\bar{Q}_n$:

Definition 2.2.1: *The* unrestricted identified set $\Theta_{I,n}^u$ *is defined as*

$$\Theta_{I,n}^u := \{\theta \in \Theta : \bar{Q}_n(\theta) = 0\}. \tag{2.2.2}$$

Examples of studies in which the identified set is defined in this way are those of Romano and Shaikh (2006), Ciliberto and Tamer (2006), Chernozhukov, Hong, and Tamer (2007), Bugni (2008), Bajari, Benkard, and Levin (2008), and Kaido and White (2009).

Observe that $\Theta_{I,n}^u$ has an n index, due to the n index of $\bar{Q}_n$. With stationary data, the n index is unneccesary; 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_I^u.

Under regularity conditions, Q_n eventually reveals Θ_I^u. CHT's approach is to use the level sets of Q_n to construct confidence sets and a set estimator for Θ_I^u.

A practical challenge occurs when the unrestricted 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 "two-stage" structures, defined as follows.

DEFINITION 2.2.2: *$\bar{Q}$ has* two-stage structure *if there exists $\Theta_{I,1}^u$ such that*

$$\Theta_I^u = \{\theta \in \Theta : \bar{Q}(\theta_1, \theta_2) = 0, \qquad \theta_1 \in \Theta_{I,1}^u\}. \tag{2.2.3}$$

The literature contains many examples with this structure. Often, the first stage is fully identified, so $\Theta_{I,1}^u$ is simply $\{\theta_1^0\}$, say. We provide details for three examples in the next section.

From now on, we mainly consider two-stage structures. Formally, we impose

ASSUMPTION 2.2.2: *$\bar{Q}$ has two-stage structure.*

The challenge of large dimension can be further addressed by exploiting prior knowledge about the underlying parameters. To describe this, we require a notion of the "true" or "pseudo-true" parameter vector for which this prior knowledge holds. To define these notions, we introduce the following assumption:

ASSUMPTION 2.2.3: *There exist a set $\mathcal{P}$ of complete probability measures on $(\Omega, \mathfrak{F})$ such that $P^0 \in \mathcal{P}$ and a given surjective mapping $\mathcal{T} := (\mathcal{T}_1', \mathcal{T}_2')' : \mathcal{P} \to \Theta$ such that $\theta^0 := \mathcal{T}(P^0) \in \Theta_I^u$.*

With this assumption, the "(pseudo-) true" parameter is $\theta^0 := \mathcal{T}(P^0)$, with $\theta_1^0 := \mathcal{T}_1(P^0)$ and $\theta_2^0 := \mathcal{T}_2(P^0)$. Consistent with White (1994), we call $\mathcal{P}$ the *model*. This is the set of all probability measures that we view as possibly having generated the data. Because of surjectivity, there is (at least) one element of the model for each θ in Θ.

In the "correctly specified" case[1], we view θ^0 as the true parameter, with $\mathcal{T}$ implementing the mapping from probability measures in $\mathcal{P}$ to the corresponding elements of Θ. A typical case occurs with $\mathcal{P} := \{P_\theta : \theta \in \Theta\}$, where each P_θ is a complete probability measure on $(\Omega, \mathfrak{F})$. Then $\mathcal{T}$ is the projection mapping such that $\theta = \mathcal{T}(P_\theta)$. In this case, $P^0 = P_{\theta^0}$. Requirement (ii) automatically holds in the correctly specified case.

When the model is misspecified, we view θ^0 as the pseudo-true parameter that would be identified from $\bar{Q}$ in the presence of the information required to resolve partial identification to full identification. In this case, $\mathcal{T}$ depends on $\bar{Q}$, as it implements the mapping from any P in $\mathcal{P}$ to the corresponding optimizer of $\bar{Q}$. (See White (1994, ch.3) for details.) Proper choice of $\mathcal{T}$ ensures that surjectivity holds in the misspecified case.

When θ_1^0 satisfies a priori restrictions, these can restrict the first-stage identified set. Such restrictions may take several forms. Our first example specifies explicit restrictions.

Restriction 2.2.1: Let $m_1 \in \mathbb{N}$, and let $\rho : \Theta_1 \to \mathbb{R}^{m_1}$ be a given measurable function. θ_1^0 satisfies $\rho(\theta_1^0) = 0$.

We also consider implicit restrictions.

Restriction 2.2.2: 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.

Restrictions 2.2.1 and 2.2.2 are often useful for simplifying first-stage estimation. They define a set Θ_1^r of parameter values satisfying the restrictions. For example, if both explicit and implicit restrictions hold, we have

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

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.*

[1]This case requires $\bar{Q}$ to be suitably well behaved.

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 2.2.3: *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 2.2.4: *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 $\Theta_I = \{\theta_1^0\} \times \Theta_{I,2}$.

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 several ways to construct the first-stage estimator in section 2.4.3. 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 2.2.4 (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 2.4.

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, extending the CHT framework. To motivate this, recall that the population criterion function $\bar{Q}$ (eventually) achieves its minimum (zero) value on Θ_I^u and gives a strictly positive value outside this set. Generally, the sample criterion function Q_n well approximates $\bar{Q}$ as n tends to infinity. As CHT show, a level set of Q_n with a level decreasing to 0 at a proper rate is a good estimator for Θ_I^u. 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 2.2.5 (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*

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

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

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

2.3 Examples

In this section, we present three examples exhibiting the two-stage structure described in the previous section. The first example is a linear regression model with an interval-valued outcome variable subject to sample selection.

EXAMPLE 2.3.1: *[Interval Censored Outcome with Sample Selection] Suppose a vector of explanatory variables* $(X_i, Z_i, D_i) \in \mathbb{R}^{d_1 \times d_2} \times \{0,1\}$ *is observable for* $i = 1, \cdots, n$. *Suppose that for samples with* $D_i = 1$, *we observe a pair* (Y_{1i}, Y_{2i}) *that satisfies the conditional moment inequalities:*

$$E[Y_{1i}|X_i, Z_i, D_i = 1] \leq E[Y_i^*|X_i, Z_i, D_i = 1] \leq E[Y_{2i}|X_i, Z_i, D_i = 1] \quad a.s.,$$

where Y_i^* *is a latent variable determined by* $Y_i^* = D_i Y_i^{**}$, *and the conditional mean* *of Y_i^{**} is a linear function of X_i:*

$$Y_i^{**} = X_i'\beta_0 + U_i, \quad E[U_i|X_i] = 0, \quad i = 1, \cdots, n.$$

The coefficients β_0 belong to a nonempty compact subset of $\mathbb{R}^{d_1}$. Suppose further that the value of D_i is determined by

$$D_i = 1\left\{Z_i'\gamma_0 + V_i \geq 0\right\}, \quad i = 1, \cdots, n,$$

where V_i is an unobservable random variable.

This is a more general version of the interval outcome regression models studied in Manski and Tamer (2002), Romano and Shaikh (2006), Chernozhukov, Hong, and Tamer (2007), and Beresteanu and Molinari (2008), extended to permit sample selection on the dependent variable. When (U_i, V_i) are jointly normal, it is well known that

$$E[Y_i^*|X_i, Z_i, D_i = 1] = X_i'\beta_0 + \alpha_0\lambda(Z_i'\gamma_0)$$

for some $\alpha_0 \in \mathbb{R}$, where λ is the inverse Mills ratio (Heckman, 1979)[2]. In this example, the first-stage parameter and the second-stage parameter are γ and (α, β) respectively. The true first-stage parameter γ_0 can be identified up to scale by observing D_i and can be fully identified if we normalize the standard deviation of V_i. Let V_i have standard deviation 1. Then the first-stage identified set is $\Theta_{I,1} = \{\gamma_0\}$.

Let $W_i \in \mathbb{R}^J$ be a vector of positive-valued transformations of (X_i, Z_i, D_i). For example, let $W_i = \{1\{(X_i, Z_i, D_i) \in \mathcal{X}_j\}, j = 1, \cdots, J\}$ for a suitable collection of sets $\mathcal{X}_j$. The moment inequalities can be written

$$E[Y_{1i}W_i] \leq \beta_0' E[X_i W_i] + \alpha_0 E[\lambda(Z_i'\gamma_0)W_i] \leq E[Y_{2i}W_i], \quad i = 1, \cdots, n.$$

The coefficients α_0 and β_0 are not identified because of the outcome variable in-

[2]It is possible to relax the joint normality assumption. In general, the conditional expectation takes the form $E[Y_i^*|X_i, Z_i, D_i = 1] = X_i'\beta_0 + h(Z_i'\gamma_0)$, where $h(Z_i'\gamma_0) = E[U_i|V_i \geq -Z_i'\gamma_0]$.

terval censoring. Following CHT's notation, we let $m_i(\alpha,\beta,\gamma) := ((Y_{1i} - \beta' X_i - \alpha\lambda(Z_i'\gamma))W_i, -(Y_{2i} - \beta' X_i - \alpha\lambda(Z_i'\gamma))W_i)'$. Then the moment inequalities are

$$E[m_i(\alpha_0, \beta_0, \gamma_0)] \leq 0.$$

Given a symmetric positive-definite matrix $S \in \mathbb{R}^{2J\times 2J}$, the population criterion function is

$$\bar{Q}(\alpha,\beta,\gamma) := \left\| E[m_i(\alpha,\beta,\gamma)]' S^{1/2} \right\|_+^2,$$

where $\|x\|_+ = \|\max(x,0)\|$. Given γ_0, the second-stage identified set is $\Theta_{I,2} := \{(\alpha,\beta) : \bar{Q}_n(\alpha,\beta,\gamma_0) = 0\}$. Now define the sample criterion function by replacing P with an empirical measure P_n:

$$Q_n(\omega,\alpha,\beta,\gamma) := \left\| \hat{E}_n[m_i(\alpha,\beta,\gamma)]' S_n^{1/2} \right\|_+^2,$$

where $\hat{E}_n[m_i(\alpha,\beta,\gamma)] = n^{-1}\sum_{i=1}^n m_i(\alpha,\beta,\gamma)$ and S_n is a symmetric positive definite matrix for each n. In this example, a consistent point estimate $\hat{\gamma}_n$ of γ_0 can be obtained from a first-stage probit estimation. Then, for each n, the second-stage set estimator is $\hat{\Theta}_{2n} = \{(\alpha,\beta) : nQ_n(\omega,\alpha,\beta,\hat{\gamma}_n) \leq \hat{\epsilon}_n\}$ for some $\hat{\epsilon}_n$ properly chosen.

The second example is Bajari, Benkard, and Levin's (2007; BBL) analysis of dynamic models of imperfect competition.

Example 2.3.2: *[Dynamic Models of Imperfect Competition] Let $n, L, M \in \mathbb{N}$. 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 $\pi_i : A \times S \times \mathcal{V}_i \to \mathbb{R}$, the pure-strategy Markov perfect equilibria (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 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[3]. Their population criterion function is defined by

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

When α_0 is fully identified, the first-stage identified set is $\Theta_{I,1} = \{\alpha_0\}$, and the second-stage identified set is $\Theta_{I,2} = \{\gamma : \bar{Q}(\alpha_0,\gamma) = 0\}$. The identified set for (α,γ) is therefore $\{\alpha_0\} \times \Theta_{I,2}$.

[3]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.

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[4]. 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

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

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 (2006). The consequences of the first-stage estimation for second-stage inference, however, were not explicitly considered. Our analysis will provide a rigorous way to explicitly account for these consequences. Further, 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.

The third example is Kaido and White's (2009) study of the market price of risk in incomplete markets. We present one of their main cases.

Example 2.3.3: *[Market Price of Risk in Incomplete Markets] 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*

[4] BBL propose drawing the x-values independently from H.

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[5]. The first assumption ensures the existence of a risk-neutral measure and an associated market price of risk λ satisfing $\sigma_0\lambda = \mu_0 - r\iota$, where ι is a vector of ones. The second ensures that the true market price of risk λ_0 has a finite upper bound M on its norm. Thus, they define the parameter space Θ_2 for λ to be $\Theta_2 = \{\lambda : \|\lambda\| \le M\}$.

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 implicit restrictions described in Restriction 2.2.2. The restricted first-stage identified set is

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

where $s(x, A) = A - xx'$ for $x \in \mathbb{R}^{d\times N}$ and $A \in \mathbb{R}^{d\times d}$. A natural estimator for $\Theta_{I,1}$ is $\hat{\Theta}_{1n} := \{\sigma : s(\sigma, \hat{\Sigma}_n) = 0\}$, where $\hat{\Sigma}_n$ is the sample covariance matrix of asset returns. 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 $\bar{q}_n := \inf_{\sigma,\lambda} -E[n^{-1}\sum_{i=1}^{n} \ln f(R_{t_i};$

[5] See Cochrane and Saá-Requejo (2000) for example.

$\sigma, \lambda)]$. The identified set for (σ, λ) is defined by $\Theta_I := \{(\sigma, \lambda) : \bar{Q}_n(\sigma, \lambda) = 0, a.a.n, \text{and } \sigma \in \Theta_{I,1}\}$. 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 $\hat{\Theta}_n := \{(\sigma, \lambda) : n\bar{Q}_n(\sigma, \lambda) \leq \hat{\epsilon}_n, \sigma \in \hat{\Theta}_{1n}\}$.

Under additional assumptions, the diffusion coefficient can be fully identified. Kaido and White (2009) consider examples that satisfy additional explicit restrictions of the type considered in Restriction 2.2.1. 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 $\rho : \mathbb{R}^{d\times N} \to \mathbb{R}^m$ such that $\rho(\sigma_0) = 0$ for some $m \in \mathbb{N}$. In this case, the first-stage identified set is

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

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 $\hat{\Theta}_{2n} = \{\lambda : nQ_n(\hat{\sigma}_n, \lambda) \leq \hat{\epsilon}_n\}$.

2.4 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 the set estimator is a set-valued random element that is measurable in an appropriate sense. The consistency result is a straightforward extension of CHT.

2.4.1 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 $\mathcal{G}$ be a topology on $\mathcal{A}$. 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 2.4.1 (Effros-measurability): *A map* $X : \Omega \to \mathcal{F}(\mathcal{A})$ *is* Effros-measurable *with respect to* $\mathfrak{F}$ *if for each open set* $G \in \mathcal{G}$

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

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[6]. To show the measurability of the set estimators defined in Definition 2.2.5, 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 $\zeta : \mathbb{R}^d \to \bar{\mathbb{R}}$ is *lower semicontinuous* (*lsc*) if $\liminf_{x \to \bar{x}} \zeta(x) \geq \zeta(\bar{x})$ for every $\bar{x} \in \mathbb{R}^d$.

Definition 2.4.2 (Epigraph and Proper Normal Integrand): *If* $\zeta : \mathbb{R}^d \to \bar{\mathbb{R}}$ *is lsc, then*

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

is called the epigraph *of* ζ.

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 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*

[6]Effros-measurability is defined for maps with a more general domain $\mathcal{F}(\mathbb{E})$, where $\mathbb{E}$ is a Polish space, but for our purposes, it suffices to take $\mathbb{E} = \mathcal{A} \subset \mathbb{R}^d$. Accordingly, all the definitions and propositions taken from Molchanov (2005) are restated for this case. Details on the measurability of random sets are discussed in chapter 1 of Li, Ogura, and Kreinovich (2002) and chapter 1 of Molchanov (2005).

identically equal to $+\infty$.

The following are useful facts about normal integrands.

FACT 2.4.1 (By M, Proposition 3.6, p.340): *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 2.4.2 (By M, Proposition 3.10 (i), p.342): *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 the following results from Stinchcombe and White (1992; SW).

LEMMA 2.4.1 (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 2.4.2 (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[7]*. Then the function* $\zeta^* : \Omega \to \bar{\mathbb{R}}$ *defined by*

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

is analytic.

REMARK 2.4.1: The result of Lemma 2.4.2 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.

We can now state a general result for Effros-measurability of two-stage set estimators:

THEOREM 2.4.1: *(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*;*

[7] 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).

(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

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

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

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

Corollary 2.4.1: *Suppose Assumptions 2.2.1 and 2.2.4 (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 2.2.5 are Effros-measurable.*

2.4.2 Consistency

In this section, we show that the two-stage set estimators of Definition 2.5 converge in probability to the identified set. The consistency is in terms of Hausdorff metric on $\mathcal{F}(\Theta_2)$. 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.

Theorem 2.4.2: *(i) Let $(\Omega, \mathfrak{F}, P)$ and $\Theta = \Theta_1 \times \Theta_2$ satisfy the conditions of Theorem 2.4.1, 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 2.4.1; (ii) Suppose there exists $\bar{Q} : \Theta \to \bar{\mathbb{R}}_+$ such that $\sup_\Theta |Q_n(\theta) - \bar{Q}(\theta)| = o_p(1)$. Let $\Theta_{I,1} \in \mathcal{F}(\Theta_1)$ and define*

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

(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 $d_H(\hat{\Theta}_{1n}, \Theta_{I,1}) = o_p(1)$, 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 for $\bar{Q}$. Nevertheless, $\bar{\Theta}_I = \Theta_I$ when $\bar{Q}$ does have two-stage structure. In the absence of two-stage structure, the result above is useful for partial identification analysis of profile estimators, like the EM algorithm (Dempster, Laird, and Rubin, 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 2.4.2: *(i) Let the conditions of Theorem 2.4.2 hold, and suppose that $\Theta_{I,1}$ is a singleton, $\Theta_{I,1} = \{\theta_1^0\}$; (ii) Let*

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

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

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

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 uniform convergence assumption.

Assumption 2.4.1: *$\sup_\Theta |Q_n(\theta) - \bar{Q}(\theta)| = o_p(1)$.*

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 2.4.3: *Suppose Assumptions 2.2.1 (i) and 2.2.2 hold. Then* $\bar{\Theta}_I = \Theta_I$. *If Assumptions 2.2.1 (ii), 2.2.4, and 2.4.1 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 2.2.5 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 measurable-*$\mathfrak{F}$ *such that* $\hat{\theta}_{1n} - \theta_1^0 = o_p(1)$, *then the second-stage estimator* $\hat{\Theta}_{2n}$ *of Definition 2.2.5 is consistent:* $d_H(\hat{\Theta}_{2n}, \Theta_{2,I}) = o_p(1)$.

2.4.3 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 those of Chernozhukov, Hong, and Tamer (2007), Kaido (2009), or Bereseteanu and Molinari (2008). It remains, however, to study set estimation with the *a priori* restrictions considered in Section 2.2. We therefore take an estimator $\hat{\Theta}_{1n}^u$ of $\Theta_{I,1}^u$ as given and study how the restrictions affect the set estimation in the first stage.

Case 1: Restriction 2.2.1

We first consider the simplest case, where Restriction 2.2.1 holds. We have the following result.

THEOREM 2.4.3: *Suppose that (i)* $\rho : \Theta_1 \to \mathbb{R}^{m_1}$ *is a continuous function; (ii)* $\Theta_1^r = \{\theta_1 \in \Theta_1 : \rho(\theta_1) = 0\}$; *(iii)* $\hat{\Theta}_{1n}^u$ *is Effros measurable. Then*

(i) $\hat{\Theta}_{1n} := \hat{\Theta}_{1n}^u \cap \Theta_1^r$ *is Effros measurable, i.e. Assumption 2.2.4 holds.*

(ii) If Assumption 2.2.3 and Restriction 2.2.1 also hold, and $d_H(\hat{\Theta}_{1n}^u, \Theta_{I,1}^u) = o_p(1)$, *then* $d_H(\hat{\Theta}_{1n}, \Theta_{I,1}) = o_p(1)$.

Case 2: Restrictions 2.2.1 and 2.2.2

Next, we consider cases where we impose both Restrictions 2.2.1 and 2.2.2. A special case is that only condition 2.2.2 holds. Since condition 2.2.2 involves another parameter that is fully identified, we need a more general treatment. The next theorem establishes the Effros-measurability and consistency of the first-stage set estimator.

THEOREM 2.4.4: *Suppose that (i)* $\rho : \Theta_1 \to \mathbb{R}^{m_1}$ *is a continuous function; (ii)* $s : \Theta_1 \times \Psi \to \mathbb{R}^{m_2}$ *is a continuous function; (iii)* $\Theta_1^r = \{\theta_1 \in \Theta_1 : s(\theta_1, \psi_0) = 0, \rho(\theta_1) = 0\}$*; (iv) there is a point estimator* $\hat{\psi}_n : \Omega \to \Psi$ *that is* $\mathfrak{F}$*-measurable and* $\hat{\psi}_n - \psi_0 = o_p(1)$*; (v)* $\hat{\Theta}_{1n}^r = \{\theta_1 \in \Theta_1 : s(\theta_1, \hat{\psi}_n) = 0, \rho(\theta_1) = 0\}$ *(vi)* $\hat{\Theta}_{1n}^u$ *is Effros measurable. Then*

(i) $\hat{\Theta}_{1n} := \hat{\Theta}_{1n}^u \cap \hat{\Theta}_{1n}^r$ *is Effros measurable, i.e. Assumption 2.2.4 holds.*

(ii) If Assumption 2.2.3 and Restrictions 2.2.1 and 2.2.2 also hold, and $d_H(\hat{\Theta}_{1n}^u, \Theta_{I,1}^u) = o_p(1)$*, then* $d_H(\hat{\Theta}_{1n}, \Theta_{I,1}) = o_p(1)$*.*

2.5 Two-Stage Inference using the 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. Let R be a closed subset of Θ (or Θ_2), where R is a set of parameter values that satisfy the restrictions of interest.

As the (pseudo-) true parameter value θ^0 is in the identified set, if θ^0 satisfies the given restrictions, the identified set has nonempty intersection with R. We can thus consider the hypotheses

$$\begin{aligned} H_0^{\Theta} &: \quad \Theta_I \cap R \neq \varnothing \text{ versus } \Theta_I \cap R = \varnothing \\ H_0^{\Theta_2} &: \quad \Theta_{I,2} \cap R \neq \varnothing \text{ versus } \Theta_{I,2} \cap R = \varnothing. \end{aligned}$$

Because R is a closed subset of the compact parameter space, these null hypotheses are equivalent to

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

Such hypotheses are considered in the partially identified case in the single stage context by Romano and Shaikh (2008) for parametric inference and by Santos (2007) for nonparametric inference.

To test these hypotheses in our two-stage framework, we replace $\bar{Q}_n$ and Θ with their sample analogs Q_n and $\hat{\Theta}_{1n} \times \Theta_2$, which leads to the test statistics

$$\hat{T}_n(\Theta, R) = \inf_{\theta \in (\hat{\Theta}_{1n} \times \Theta_2) \cap R} a_n Q_n(\theta) \quad \text{and} \quad \hat{T}_n(\Theta_2, R) = \inf_{\theta_2 \in R} a_n Q_n(\hat{\theta}_{1n}, \theta_2)$$

Below, we focus on the cases where Q_n takes the quasi-maximum likelihood form in Eq. (2.2.1). Let $a_n = n$, a typical case. Then the statistics can be written as

$$\begin{aligned} \hat{T}_n(\Theta, R) &= \inf_{\theta \in (\hat{\Theta}_{1n} \times \Theta_2) \cap R} \sum_{i=1}^{n} q(X_i(\omega), \theta) - \inf_{\theta \in (\hat{\Theta}_{1n} \times \Theta_2)} \sum_{i=1}^{n} q(X_i(\omega), \theta) \\ \hat{T}_n(\Theta_2, R) &= \inf_{\theta_2 \in R} \sum_{i=1}^{n} q(X_i(\omega), \hat{\theta}_{1n}, \theta_2) - \inf_{\theta_2 \in \Theta_2} \sum_{i=1}^{n} q(X_i(\omega), \hat{\theta}_{1n}, \theta_2) \end{aligned} \tag{2.5.1}$$

When $q(x, \theta) = -\ln f(x, \theta)$, where $f(\cdot, \theta)$ is a probability density function for each θ, these statistics can be viewed as quasi-likelihood ratio statistics.

In general, the first stage estimation impacts the distribution of the statistics. This impact is quite involved for $\hat{T}_n(\Theta, R)$. Thus, to gain insight and give useful results, in what follows, we focus on the statistic $\hat{T}_n(\Theta_2, R)$, where the first stage parameter is fully identified. This includes Examples 2.3.1 and 2.3.2 and important special cases of Example 2.3.3. We leave analysis of $\hat{T}_n(\Theta, R)$ to future work, as that analysis requires much more space than available here.

When the first stage parameter is point identified, we can exploit a two-term mean-value expansion. The following straightforward high-level result applies when θ_1^0 is interior to Θ_1 and the function $q(x, \cdot, \theta_2)$ is sufficiently smooth. Analogous but more elaborate results hold even when θ_1^0 is not interior to Θ_1.

PROPOSITION 2.5.1: *Let $\{a_n\}$ be a sequence of real numbers and for $p \in \mathbb{N}$, suppose that $\theta_0 \in \mathbb{R}^p$ and that $\{\hat{\mathcal{Q}}_n : \Omega \to \mathbb{R}\}$, $\{\mathcal{Q}_n : \Omega \to \mathbb{R}\}$, $\{\hat{\theta}_n : \Omega \to \mathbb{R}^p\}$, $\{g_n : \Omega \to \mathbb{R}^p\}$, and $\{H_n : \Omega \to \mathbb{R}^{p\times p}\}$ are sequences of measurable functions such that*

$$a_n\hat{\mathcal{Q}}_n = a_n\mathcal{Q}_n + a_n g_n'\left(\hat{\theta}_n - \theta_0\right) + a_n\left(\hat{\theta}_n - \theta_0\right)' H_n\left(\hat{\theta}_n - \theta_0\right)/2 + o_p(1), \quad (2.5.2)$$

where, for random matrices Z_0, Z_1, Z_2, Z_3 of suitable dimension,

$$(a_n\mathcal{Q}_n,\ a_n^{1/2}\left(\hat{\theta}_n - \theta_0\right)',\ a_n^{1/2}g_n',\ (vec(H_n))') \xrightarrow{d} (Z_0, Z_1', Z_2', (vec(Z_3))').$$

Then

$$a_n\hat{\mathcal{Q}}_n \xrightarrow{d} Z_0 + Z_2'Z_1 + Z_1'Z_3Z_1/2.$$

In our application, $a_n = n$, $p = d_1$, $\theta_0 = \theta_0^1$, $a_n\hat{\mathcal{Q}}_n = \hat{T}_n(\Theta_2, R)$, $a_n\mathcal{Q}_n = T_n(\Theta_2, R; \theta_0^1)$, where

$$T_n(\Theta_2, R; \theta_1^0) := \inf_{\theta_2\in R}\sum_{i=1}^n q(X_i, \theta_1^0, \theta_2) - \inf_{\theta_2\in\Theta_2}\ \sum_{i=1}^n q(X_i, \theta_1^0, \theta_2),$$

$\hat{\theta}_n = \hat{\theta}_{1n}$, $g_n = n^{-1}(\partial/\partial\theta_1)T_n(\Theta_2, R; \theta_1^0)$, and $H_n = n^{-1}(\partial^2/\partial\theta_1\partial\theta_1')T_n(\Theta_2, R; \theta_1^0)$. Therefore, to establish weak convergence of the statistic, it suffices to establish joint convergence of $(T_n(\Theta_2, R; \theta_1^0), \sqrt{n}(\hat{\theta}_{1n} - \theta_1^0), n^{-1/2}\partial/\partial\theta_1 T_n(\Theta_2, R; \theta_1^0), n^{-1}\partial^2/\partial\theta_1\partial\theta_1'$ $T_n(\Theta_2, R; \theta_1^0))$. Assumption 2.5.1 below summarizes a high-level condition sufficient for the desired result.

Before stating the assumption, we introduce additional notation. For any sequence $\{Z_i\}$ of random vectors, let $\mathbb{G}_n(Z_i) := \sqrt{n}(\hat{E}_n(Z_i) - E(Z_i))$. The space of bounded functions on a set $\mathcal{D}$ will be denoted $l^\infty(\mathcal{D})$. We focus on a set $\mathcal{D} = \{\phi : \mathcal{X} \to \mathbb{R}\}$, which is a class of measurable functions such that

$$\sup_{\phi\in\mathcal{D}} |\phi(x) - E(\phi(X))| < \infty, \qquad \text{for all } x. \quad (2.5.3)$$

We call $\mathcal{W}_n : \Omega \to l^\infty(\mathcal{D})$ an *empirical process* on $\mathcal{D}$ if for every $\phi \in \mathcal{D}$,

$$\pi_\phi \mathcal{W}_n = \mathbb{G}_n(\phi(X_i)) = \sqrt{n}(\hat{E}_n(\phi(X_i)) - E(\phi(X_i))),$$

where $\pi_\phi : l^\infty(\mathcal{D}) \to \mathbb{R}$ is the projection of the argument process at ϕ. For simplicity, in what follows, we write $\mathcal{W}_n(\phi) = \pi_\phi \mathcal{W}_n$; there will be no risk of confusion. Following van der Vaart and Wellner (2000), we define the P^0-Donsker property[8].

Definition 2.5.1 (P^0-Donsker class): *A class of measurable functions $\mathcal{D} = \{\phi : \mathcal{X} \to \mathbb{R}\}$ satisfying* (2.5.3) *is said to be P^0-Donsker if the empirical process $\mathcal{W}_n$ on $\mathcal{D}$ converges weakly to a tight measurable element $\mathcal{W} : \Omega \to l^\infty(\mathcal{D})$.*

Assumption 2.5.1: *(i) There exist a P^0-Donsker class $\mathcal{S} := \{\phi : \mathcal{X} \to \mathbb{R} : \sup_{\phi\in\mathcal{S}} |\phi(x) - E(\phi(X))| < \infty, \forall x \in \mathcal{X}\}$ and a continuous function $g : l^\infty(\mathcal{S}) \to \mathbb{R}$ such that $T_n(\Theta_2, R; \theta_1^0) = g(\mathcal{W}_n) + o_p(1)$; (ii) There exist measurable functions $\psi_j : \mathcal{X} \to \mathbb{R}^{d_1}, j = 1, 2, 3$ and $\psi_k : \mathcal{X} \to \mathbb{R}^{d_1^2}, k = 4, 5$ such that*

$$\sqrt{n}(\hat{\theta}_{1n} - \theta_1^0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_1(X_i) + o_p(1) \tag{2.5.4}$$

$$\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta_1} T_n(\Theta_2, R; \theta_1^0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_2(X_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_3(X_i) + o_p(1) \tag{2.5.5}$$

$$\frac{1}{n} vec\left(\frac{\partial^2}{\partial \theta_1 \partial \theta_1'} T_n(\Theta_2, R; \theta_1^0)\right) = \frac{1}{n} \sum_{i=1}^n \psi_4(X_i) - \frac{1}{n} \sum_{i=1}^n \psi_5(X_i) + o_p(1), \tag{2.5.6}$$

with $E(\psi_j(X_i)) = 0$, $j = 1, 2, 3$, $E(\psi_j(X_i)) < \infty$, $j = 4, 5$, and $E(\psi_j(X_i)'\psi_j(X_i)) < \infty$, $j = 1, 2, 3$; (iii) For any finite m, let $S_l \in \mathcal{S}, l = 1, \cdots, m$, and suppose the sequence of $m + 5$-vectors

$$(\mathcal{W}_n(S_1), \cdots, \mathcal{W}_n(S_m), \mathbb{G}_n(\psi_1(X_i)), \mathbb{G}_n(\psi_2(X_i)), \mathbb{G}_n(\psi_3(X_i)), \hat{E}_n(\psi_4(X_i)), \hat{E}_n(\psi_5(X_i)))'$$

jointly converges in distribution to $(\mathcal{W}(S_1), \cdots, \mathcal{W}(S_m), W_1, \cdots, W_3, H_4, H_5)'$,

[8] For details on weak convergence and tightness, see van der Vaart and Wellner (2000).

where $(\mathcal{W}(S_1), \cdots, \mathcal{W}(S_m), W_1, W_2, W_3)$ *follows a mean zero multivariate Normal distribution and* H_4 *and* H_5 *are constant vectors.*

Assumption 2.5.1 (i) requires that $T_n(\Theta_2, R; \theta_1^0)$ converges to a continuous function of a Gaussian process $\mathcal{W}$ defined on some set $\mathcal{S}$. For correctly specified models, Liu and Shao (2003) give primitive conditions that ensure this requirement, which we will elaborate below, together with more primitive conditions ensuring Assumption 2.5.1. Note that we need weak convergence of $\mathcal{Y}_n := (\mathcal{W}_n, \mathbb{G}_n(\psi_1(X_i)), \cdots, \mathbb{G}_n(\psi_3(X_i)), \hat{E}_n(\psi_4(X_i)), \hat{E}_n(\psi_5(X_i)))'$ in the product space $l^\infty(\mathcal{S}) \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_1^2} \times \mathbb{R}^{d_1^2}$. For this, it is sufficient to have weak convergence of $\mathcal{W}_n$, of $\mathbb{G}_n(\psi_j(X_i)), j = 1, \cdots, 3$, and of $\hat{E}_n(\psi_k(X_i)), k = 4, 5$ (i.e. convergence in probability), and finite dimensional convergence of $\mathcal{W}_n$, because joint tightness follows from marginal tightness[9]. Generally, a central limit theorem and weak law of large numbers ensure the finite-dimensional convergences. Assumption 2.5.1 (ii) requires that $\hat{\theta}_n$ has an asymptotic linear representation with influence function ψ_1. Many estimators, including the maximum likelihood estimator (MLE), satisfy this requirement. Similarly, an envelope theorem ensures the existence of $\psi_j, j = 2, \cdots, 5$, when $q(x, \cdot, \theta_2)$ satisfies appropriate differentiability conditions. Assumption 2.5.1 (iii) ensures that $\frac{1}{n}\frac{\partial^2}{\partial\theta_1\partial\theta_1'}T_n(\Theta_2, R; \theta_1^0)$ converges to a constant matrix. We denote this matrix by H.

Now we can state the desired result.

Theorem 2.5.1: *Suppose Assumption 2.2.1 holds with* (2.2.1) *defining* Q_n, *Assumption 2.2.2 and 2.2.4 hold with* $\Theta_{I,1} = \{\theta_1^0\}$, *and that Assumption 2.5.1 holds. Suppose* $\hat{T}_n(\Theta_2, R)$ *admits the expansion in Eq.* (2.5.2) *with* $a_n = n$, $p = d_1$, $\theta_0 = \theta_0^1$, $a_n\hat{\mathcal{Q}}_n = \hat{T}_n(\Theta_2, R)$, $a_n\mathcal{Q}_n = T_n(\Theta_2, R; \theta_0^1)$, $\hat{\theta}_n = \hat{\theta}_{1n}$, $g_n = n^{-1}(\partial/\partial\theta_1)T_n(\Theta_2, R; \theta_1^0)$, *and* $H_n = n^{-1}(\partial^2/\partial\theta_1\partial\theta_1')T_n(\Theta_2, R; \theta_1^0)$. *Then*

$$\hat{T}_n(\Theta_2, R) \xrightarrow{d} g(\mathcal{W}) + (W_2 - W_3)'W_1 + W_1'HW_1,$$

where $\mathcal{W}$ *is a tight Gaussian process on* $\mathcal{S}$.

[9]See Hall and Loynes (1977) or Billingsley (1999), page 65, Problem 5.9.

We now provide more primitive conditions that ensure Assumption 2.5.1. These permit us to apply the framework of Liu and Shao (2003). For this, we will impose correct model specification for the first time. We add the following structure to the DGP and model:

Assumption 2.5.2 (Additional Assumptions 1): *(i) Let* $\{X_i : \Omega \to \mathcal{X}, i = 1, 2, \cdots\}$ *be an IID sequence of random k-vectors on* $(\Omega, \mathfrak{F}, \mathbb{P}^0)$. *(ii) For each* $i = 1, 2, \cdots$, P^0 *(induced on* $\mathcal{X}$*) is absolutely continuous with respect to a* σ*-finite measure* μ. *Let* $f^0 := dP^0/d\mu$ *be the (Radon-Nikodym) density. (iii) Let* $d \in \mathbb{N}$ *and* $\Theta \subset \mathbb{R}^d$, *and for each* $\theta \in \Theta$ *let* P_θ *be absolutely continuous with respect to* μ *with* $f(\cdot, \theta) := dP_\theta/d\mu$. *Define* $\mathcal{P}_\Theta := \{P_\theta : \theta \in \Theta\}$ *and suppose* $P^0 \in \mathcal{P}_\Theta$. *(iv) For each* $\theta \in \Theta$, *let*

$$Q_n(\theta) := n^{-1} \sum_{i=1}^{n} - \ln f(X_i; \theta) - \inf_{\Theta} n^{-1} \sum_{i=1}^{n} - \ln f(X_i; \theta).$$

Assumption 2.5.2 (i) imposes IID structure. This is for convenience only. It can be substantially relaxed. Assumption 2.5.2 (ii) ensures the existence of a density function, and Assumption 2.5.2 (iii) enforces correct model specification. Assumption 2.5.2 (iv) further specifies the functional form of the criterion function. The statistic $\hat{T}_n(\Theta, R)$, therefore, takes the form in Eq. (2.5.1) with $q(x, \theta) = - \ln f(x; \theta)$. For each $(\theta_1, \theta_2) \in \Theta$ and $x \in \mathcal{X}$, let $l_{\theta_1, \theta_2}(x) := f(x; \theta_1, \theta_2)/f^0(x)$. We rewrite the statistic as follows:

$$\begin{aligned} T_n(\Theta_2, R; \theta_1^0) &= \inf_{\theta_2 \in R} \sum_{i=1}^{n} - \ln f(X_i; \theta_1^0, \theta_2) - \inf_{\theta_2 \in \Theta_2} \sum_{i=1}^{n} - \ln f(X_i; \theta_1^0, \theta_2), \\ &= \sup_{\theta_2 \in \Theta_2} \sum_{i=1}^{n} \ln l_{\theta_1^0, \theta_2}(X_i) - \sup_{\theta_2 \in R} \sum_{i=1}^{n} \ln l_{\theta_1^0, \theta_2}(X_i). \end{aligned} \qquad (2.5.7)$$

Our goal here is to show that $\sum_{i=1}^{n} \ln l_{\theta_1^0, \theta_2}(X_i)$ converges weakly to a well-defined limit. To establish this, we follow Liu and Shao (2003) and use a (truncated) quadratic expansion of the statistic with respect to the Hellinger metric between

$P_{\theta_1^0,\theta_2}$ and P^0:

$$H(\theta_2) := \left(E\left(\sqrt{l_{\theta_1^0,\theta_2}(X_i)} - 1 \right)^2 /2 \right)^{1/2}.$$

Note that $H(\theta_2) = 0$ for all $\theta_2 \in \Theta_{2,I}$ when we define the population criterion function by $\bar{Q}(\theta) := E[-\ln f(X_i;\theta)] - \inf_{\theta\in\Theta} E[-\ln f(X_i;\theta)]$. We use the Hellinger metric to define neighborhoods of $\Theta_{2,I}$. For each $\epsilon > 0$, let $\Theta_{2I}^{\epsilon} := \{\theta_2 \in \Theta_2 : 0 < H(\theta_2) \leq \epsilon\}$ and $\Theta_{2R}^{\epsilon} := \{\theta_2 \in R \cap \Theta_2 : 0 < H(\theta_2) \leq \epsilon\}$. We call these ϵ-Hellinger neighborhoods.

Let $L_2(P^0)$ be the set of all functions $\phi : \mathcal{X} \to \mathbb{R}$ such that $E(\phi^2(X_i)) < \infty$. We equip $L_2(P^0)$ with the L_2-norm $\|\phi\| := E(\phi^2(X_i))^{1/2}$. For likelihood ratios that are square integrable, let $D(\theta_2) := (E(l_{\theta_1^0,\theta_2}(X_i) - 1)^2)^{1/2}$ be the L_2-metric between $P_{\theta_1^0,\theta_2}$ and P^0.

Now for each $\theta_2 \in \Theta_2$, define $S_{\theta_2} := (l_{\theta_1^0,\theta_2} - 1)/D(\theta_2)$. We call S_{θ_2} the *generalized score function*. This function has the properties that $E(S_{\theta_2}(X_i)) = 0$ for all $\theta_2 \in \Theta_{2,I}$ and that $E(S_{\theta_2}^2(X_i)) = 1$ for all $\theta_2 \in \Theta_2$ if the second moment of $l_{\theta_1^0,\theta_2}(X_i)$ exists. Liu and Shao's (2003) main idea is to approximate the likelihood-ratio statistic by the supremum of a quadratic function of the generalized score on an ϵ-Hellinger neighborhood. The following subsets of $L_2(P^0)$ play a central role when we derive the asymptotic distribution of $T_n(\Theta, R; \theta_1^0)$. For each $\epsilon > 0$, let

$$\begin{aligned}
\mathcal{S}_I^{\epsilon} &:= \{S_{\theta_2} \in L_2(P^0) : S_{\theta_2} = (l_{\theta_1^0,\theta_2} - 1)/D(\theta_2), \theta_2 \in \Theta_{2I}^{\epsilon}\} \\
\mathcal{S}_R^{\epsilon} &:= \{S_{\theta_2} \in L_2(P^0) : S_{\theta_2} = (l_{\theta_1^0,\theta_2} - 1)/D(\theta_2), \theta_2 \in \Theta_{2R}^{\epsilon}\} \\
\bar{\mathcal{S}}_I^{\epsilon} &:= \Big\{ S \in L_2(P^0) : \exists \{\theta_2^m\} \subset \Theta_{2I}^{\epsilon} \text{ such that,} \\
&\qquad \lim_{m\to\infty} \|(l_{\theta_1^0,\theta_2^m}(X_i) - 1)/D(\theta_2^m) - S\| = 0, \text{ as } D(\theta_2^m) \to 0. \Big\} \\
\bar{\mathcal{S}}_R^{\epsilon} &:= \Big\{ S \in L_2(P^0) : \exists \{\theta_2^m\} \subset \Theta_{2R}^{\epsilon} \text{ such that,} \\
&\qquad \lim_{m\to\infty} \|(l_{\theta_1^0,\theta_2^m}(X_i) - 1)/D(\theta_2^m) - S\| = 0, \text{ as } D(\theta_2^m) \to 0. \Big\}
\end{aligned}$$

The set $\mathcal{S}_I^{\epsilon}$ is the collection of generalized score functions whose indexes θ_2 belong to the ϵ-Hellinger neighborhood Θ^{ϵ}_{2I} of $\Theta_{2,I}$. Similarly, $\mathcal{S}^{\epsilon}_{R}$ collects generalized score functions whose indexess belong to the restricted ϵ-Hellinger neighborhood Θ^{ϵ}_{2R}. $\bar{\mathcal{S}}^{\epsilon}_{I}$ and $\bar{\mathcal{S}}^{\epsilon}_{R}$ are the L^2-closures of $\mathcal{S}^{\epsilon}_{I}$ and $\mathcal{S}^{\epsilon}_{R}$ respectively.

We introduce two additional regularity conditions, Assumptions 2.5.3 and 2.5.4. To state these, we require some further definitions.

Definition 2.5.2: *For given $\epsilon > 0$, a set $\mathcal{S}$ of square integrable real-valued functions is **(ϵ-) complete** if for any sequence $\{\theta^m_2\} \subset \Theta^{\epsilon}_2$ with $H(\theta^m_2) \to 0$, there exists a subsequence $\{\theta^{m_k}_2\}$ of $\{\theta^m_2\}$ such that $S_{\theta^{m_k}_2}$ converges to some $S \in \mathcal{S}$ in $L_2(P^0)$. $\mathcal{S}$ **admits (ϵ-) continuous paths** if for all $S \in \mathcal{S}$, there exists a path $\{\theta_2(t,S) : 0 < t \leq \epsilon\} \subset \Theta^{\epsilon}_2$ such that $\theta_2(t,S)$ is continuous in t, $H(\theta_2(t,S)) = t$, and $\lim_{t\to 0} S_{\theta_2(t,S)} = S$ in $L_2(P^0)$. An* envelope function *$\Phi : \mathcal{X} \to \mathbb{R}$ for a class $\mathcal{D}$ is a function such that $\sup_{\phi\in\mathcal{D}} |\phi(x)| \leq \Phi(x)$ for all $x \in \mathcal{X}$.*

The following assumption, together with Assumption 2.5.2, ensures that we can apply Liu and Shao (2003)'s Theorem 3.1, which ensures Assumption 2.5.1 (i).

Assumption 2.5.3: *(i) For any $\theta_2 \in \Theta_2$, $l_{\theta^0_1,\theta_2} \in L_2(P^0)$. (ii) For some $\epsilon > 0$, $\mathcal{S}^{\epsilon}_{I}$ is a P^0-Donsker class with a square integrable envelope function. $\bar{\mathcal{S}}^{\epsilon}_{I}$ and $\bar{\mathcal{S}}^{\epsilon}_{R}$ are complete and admit continuous paths.*

Note that $\mathcal{S}^{\epsilon}_{R}$ is a subset of $\mathcal{S}^{\epsilon}_{I}$. Therefore, under Assumption 2.5.3 (ii), $\mathcal{S}^{\epsilon}_{R}$ is also a P^0-Donsker class. Given Assumption 2.5.3, we can show that $T_n(\Theta_2, R; \theta^0_1) = g(\mathcal{W}_n)$, where $\mathcal{W}_n$ is an empirical process defined for each $S_\theta \in \mathcal{S}^{\epsilon}_{I}$ by

$$\begin{aligned} \mathcal{W}_n(S_\theta) &= \mathbb{G}_n(S_\theta(X_i)) \\ &= \sqrt{n}\hat{E}_n((l_{\theta^0_1,\theta_2}(X_i) - 1)/D(\theta_2)), \end{aligned} \tag{2.5.8}$$

and the continuous function g is given by $g(\mathcal{W}_n) = \sup_{S\in\bar{\mathcal{S}}^{\epsilon}_{I}} \max\{\mathcal{W}_n(S), 0\}^2 - \sup_{S\in\bar{\mathcal{S}}^{\epsilon}_{R}} \max\{\mathcal{W}_n(S), 0\}^2$.

Our remaining task is to ensure Assumption 2.5.1 (ii) and (iii). As many estimators satisfy Eq. (2.5.4), we maintain the assumption that the first-stage estimator is asymptotically linear with influence function ψ_1. Now we consider more primitive conditions that ensure the existence of $\psi_j, j = 2, \cdots, 5$. Specifically, we first give conditions under which $\psi_2(\cdot) = -\partial/\partial\theta_1 \ln(\cdot;\theta_1^0,\theta_2^R)$ and $\psi_3(\cdot) = -\partial/\partial\theta_1 \ln f(\cdot;\theta_1^0,\theta_2^0)$, where $\theta_2^R \in \Theta_{2,I} \cap R$. Additional smoothness requirements on f further ensure the existence of ψ_4 and ψ_5.

Note that if R and Θ_2 are defined by convex inequality constraints, a classic envelope theorem[10] ensures the differentiability of the value functions $\inf_{\theta_2\in R}\sum_{i=1}^n - \ln f(x;\cdot,\theta_2)$ and $\inf_{\theta_2\in\Theta_2}\sum_{i=1}^n - \ln f(x;\cdot,\theta_2)$ on Θ_1. This therefore ensures the existence of ψ_2 and ψ_3, but the requirement that the sets are defined by convex inequality constraints may be too restrictive. For example, the set R could be given by a nonconvex restriction $R := \{\theta_2 : g(\theta_2) = 0\}$ such as $g(x) = x_1x_2 - 1$. Instead of adding a convex structure to R and Θ_2, we use results of Milgrom and Segal (2002) who establish a more general envelope theorem applicable to cases where R and Θ_2 are arbitrary sets. Following them, we mildly extend their smoothness concept[11].

Definition 2.5.3 (Equidifferentiality): *Let Θ be an arbitrary set, and let $B \subseteq \mathbb{R}$ be an open set. The family of functions $\{\phi(\theta,\cdot) : B \to \mathbb{R} | \theta \in \Theta\}$ is equidifferentiable at $y \in B$ if $(\phi(\theta,y') - \phi(\theta,y))/(y'-y)$ converges uniformly over Θ as $y' \to y$.*

When A is an infinite set, equidifferentiability is stronger than the usual differentiability. A simple sufficient condition for equidifferentiability is the continuous differentiability of $\phi(\theta,\cdot)$ on B for each θ and the compactness of Θ.

For each $x \in \mathcal{X}$, consider the family $\{f(x;\cdot,\theta_2) : \theta_2 \in \Theta_2\}$. To state our equidifferentiability condition, for any $\theta_1 \in \Theta_1$, let $\theta_{l,1}$ be its $l-$th component. Consider the vector $(\theta_{1,1},\cdots,\theta_{l-1,1},\theta_{l+1,1},\cdots,\theta_{d_1,1})'$, obtained by removing $\theta_{l,1}$ from θ_1. We denote this vector $\theta_{-l,1}$ and let $\Theta_{-l,1} := \{\theta_{-l,1} : \exists\theta_{l,1}$ such that $(\theta_{1,1},\cdots,\theta_{l-1,1},\theta_{l,1},\theta_{l+1,1},\cdots,\theta_{d_1,1})' \in \Theta_1\}$. For each $x \in \mathcal{X}$, $\theta_{-l,1} \in \Theta_{-l,1}$, and $\theta_2 \in \Theta_2$, let $\tilde{f}_l(x;\theta_{-l,1},\theta_2,\theta_{l,1})$ denote the map $\theta_{l,1} \mapsto f(x,(\theta_{1,1},\cdots,\theta_{l-1,1},\theta_{l,1},\theta_{l+1,1},\cdots,\theta_{d_1,1})',\theta_2)$. As we formally state below, our smoothness requirement for the existence of ψ_2 and ψ_3 is that for each x, the family $\{\ln\tilde{f}_l(x;\theta_{-l,1}^0,\theta_2,\cdot) : \theta_2 \in \Theta_2\}$

[10] See for example, Mas-Colell, Whinston, and Green (1995).

[11] Milgrom and Segal (2002) define equidifferentiability for a scalar $y \in (0,1)$. Here, we extend their definition to a vector taking values in a Euclidean space.

is equidifferentiable at $\theta^0_{l,1}$ for $l = 1, \cdots, d_1$.

Similarly, for the existence of ψ_4 and ψ_5, we require an additional smoothness condition. Let $\theta_{-l,-m,1}$ be the vector obtained by removing $\theta_{l,1}$ and $\theta_{m,1}$ from θ_1 and let

$$\begin{aligned}\Theta_{-l,-m,1} := \{\theta_{-l,-m,1} : \exists(\theta_{l,1}, \theta_{m,1}) \text{ such that} \\ (\theta_{1,1}, \cdots, \theta_{l-1,1}, \theta_{l,1}, \theta_{l+1,1}, \cdots, \theta_{m-1,1}, \theta_{m,1}, \theta_{m+1,1}, \cdots, \theta_{d_1,1})' \in \Theta_1\}.\end{aligned}$$

For each $x \in \mathcal{X}$, $\theta_{-l,-m,1} \in \Theta_{-l,-m,1}$, and $\theta_2 \in \Theta_2$, let $\tilde{f}_{l,m}(x; \theta_{-l,-m,1}, \theta_2, \theta_{l,1}, \theta_{m,1})$ denote the map $(\theta_{l,1}, \theta_{m,1}) \mapsto f(x, (\theta_{1,1}, \cdots, \theta_{l-1,1}, \theta_{l,1}, \theta_{l+1,1}, \cdots, \theta_{m-1,1}, \theta_{m,1}, \theta_{m+1,1}, \cdots, \theta_{d_1,1})', \theta_2)$. Our smoothness requirement for the existence of ψ_4 and ψ_5 is that for each x, the family

$$\left\{\partial/\partial\theta_{l,1} \ln \tilde{f}_{l,m}(x; \theta^0_{-l,-m,1}, \theta_2, \theta^0_{l,1}, \cdot) : \theta_2 \in \Theta_2\right\}$$

is equidifferentiable at $\theta^0_{m,1}$ for all $l, m = 1, \cdots, d_1$.

In order to apply Milgrom and Segal's (2002) results, we must also ensure that the (possibly set-valued) solutions of $\inf_{\theta_2 \in R} \sum_{i=1}^n - \ln f(x_i; \theta_1, \theta_2)$ and $\inf_{\theta_2 \in \Theta_2} \sum_{i=1}^n - \ln f(x_i; \theta_1, \theta_2)$ are well defined. For each $x^n = (x_1, \cdots, x_n)' \in \mathcal{X}^n$ and $\theta_1 \in \Theta_1$, let

$$\begin{aligned}R^*(x^n, \theta_1) &:= \arg\min_{\theta_2 \in R} \sum_{i=1}^n - \ln f(x_i; \theta_1, \theta_2) \\ \Theta^*_2(x^n, \theta_1) &:= \arg\min_{\theta_2 \in \Theta_2} \sum_{i=1}^n - \ln f(x_i; \theta_1, \theta_2).\end{aligned}$$

We define *selections* of these solutions as follows.

Definition 2.5.4 (Selection): *Let $p \in \mathbb{N}$ and $B \subseteq \mathbb{R}^p$. Let $(A, \mathcal{A}, \lambda)$ be a measure space. A* selection *$u : A \to B$ of an Effros-measurable set-valued map $U : A \to \mathcal{F}(B)$ is a measurable function such that $u(a) \in U(a), \lambda - a.e.$*

The following assumption suffices to apply Milgrom and Segal (2002)'s Theorem 3, which we use to establish the existence of $\psi_j, j = 2, \cdots, 5$.

Assumption 2.5.4: *(i) For any* $x^n \in \mathcal{X}^n$ *and any* $\theta_1 \in \Theta_1$, $R^*(x^n, \theta_1) \neq \varnothing$ *and* $\Theta_2^*(x^n, \theta_1) \neq \varnothing$*;*

(ii.a) For any $x \in \mathcal{X}$*, the family* $\{\ln \tilde{f}_l(x; \theta^0_{-l,1}, \theta_2, \cdot) : \theta_2 \in \Theta_2\}$ *is equidifferentiable at* $\theta^0_{l,1}$ *for all* $l = 1, \cdots, d_1$*;*

(ii.b) For any $x \in \mathcal{X}$, $\sup_{\theta_2 \in \Theta_2} |d/d\theta_{l,1} \ln \tilde{f}_l(x; \theta^0_{-l,1}, \theta_2, \theta^0_{l,1})| < \infty$ *for all* $l = 1, \cdots, d_1$*;*

(ii.c) For any $x^n \in \mathcal{X}^n$*, selections* $\theta^*_{2,R}(x^n, \cdot) \in R^*(x^n, \cdot)$, $\theta^*_{2,\Theta_2}(x^n, \cdot) \in \Theta^*_2(x^n, \cdot)$*, and* $x \in \mathcal{X}$*, the maps* $\partial/\partial\theta_1 \ln f(x; \theta^0_1, \theta^*_{2,R}(x^n, \cdot))$ *and* $\partial/\partial\theta_1 \ln f(x; \theta^0_1, \theta^*_{2,\Theta_2}(x^n, \cdot))$ *are continuous at* θ^0_1*;*

(iii.a) For any $x \in \mathcal{X}$*, the family* $\{\partial/\partial\theta_{l,1} \ln \tilde{f}_{l,m}(x; \theta^0_{-l,-m,1}, \theta_2, \theta^0_{l,1}, \cdot) : \theta_2 \in \Theta_2\}$ *is equidifferentiable at* $\theta^0_{m,1}$ *for all* $l, m = 1, \cdots, d_1$*;*

(iii.b) For any $x \in \mathcal{X}$, $\sup_{\theta_2 \in \Theta_2} |\partial^2/\partial\theta_{l,1}\partial\theta_{m,1} \ln \tilde{f}_{l,m}(x; \theta^0_{-l,-m,1}, \theta_2, \theta^0_{l,1}, \theta^0_{m,1})| < \infty$ *for all* $l, m = 1, \cdots, d_1$*;*

(iii.c) For any $x^n \in \mathcal{X}^n$*, selections* $\theta^*_{2,R}(x^n, \cdot) \in R^*(x^n, \cdot)$, $\theta^*_{2,\Theta_2}(x^n, \cdot) \in \Theta^*_2(x^n, \cdot)$*, and* $x \in \mathcal{X}$ *the maps* $\partial^2/\partial\theta_1\partial\theta_1' \ln f(x; \theta^0_1, \theta^*_{2,R}(x^n, \cdot))$ *and* $\partial^2/\partial\theta_1\partial\theta_1' \ln f(x; \theta^0_1, \theta^*_{2,\Theta_2}(x^n, \cdot))$ *are continuous at* θ^0_1*;*

It is straightforward to verify that given Assumption 2.5.4 (i) and (ii), the conditions of Milgrom and Segal's (2003) Theorem 3 are satisfied. For each $X^n = (X_1, \cdots, X_n)$ and $x \in \mathcal{X}$, define

$$\psi_{n2}(x) := -\frac{\partial}{\partial\theta_1} \ln f(x; \theta^0_1, \theta^*_{2,R}(X^n, \theta^0_1))$$
$$\psi_{n3}(x) := -\frac{\partial}{\partial\theta_1} \ln f(x; \theta^0_1, \theta^*_{2,\Theta_2}(X^n, \theta^0_1)).$$

Notice that $\psi_{n2}(X_i)$ and $\psi_{n3}(X_i)$ depend on the sample X^n through the selections $\theta^*_{2,R}(X^n, \theta^0_1)$ and $\theta^*_{2,\Theta_2}(X^n, \theta^0_1)$; but with probability approaching 1, Hausdorff consistency ensures that these selections are restricted to $\Theta_{2,I} \cap R$ and $\Theta_{2,I}$ respectively.

For each $x \in \mathcal{X}$, define

$$\psi_2(x) := -\frac{\partial}{\partial\theta_1} \ln f(x; \theta_1^0, \theta_2^R) \tag{2.5.9}$$
$$\psi_3(x) := -\frac{\partial}{\partial\theta_1} \ln f(x; \theta_1^0, \theta_2^0), \tag{2.5.10}$$

where $\theta_2^R \in \Theta_{2,I} \cap R$. If $\hat{E}_n\psi_{nj}(X_i) - \hat{E}_n\psi_j(X_i) = o_p(n^{-1/2}), j = 2, 3$, then the representation in Eq. (2.5.5) holds with ψ_2 and ψ_3 defined above.

Similarly, for each $x \in \mathcal{X}$, define

$$\psi_{n4}(x) = vec\left(-\frac{\partial^2}{\partial\theta_1\partial\theta_1'} \ln f(x; \theta_1^0, \theta_{2,R}^*(X^n, \theta_1^0))\right)$$
$$\psi_{n5}(x) = vec\left(-\frac{\partial^2}{\partial\theta_1\partial\theta_1'} \ln f(x; \theta_1^0, \theta_{2,\Theta_2}^*(X^n, \theta_1^0))\right).$$

For each $x \in \mathcal{X}$, define

$$\psi_4(x) := vec\left(-\frac{\partial^2}{\partial\theta_1\partial\theta_1'} \ln f(x; \theta_1^0, \theta_2^R)\right) \tag{2.5.11}$$
$$\psi_5(x) := vec\left(-\frac{\partial^2}{\partial\theta_1\partial\theta_1'} \ln f(x; \theta_1^0, \theta_2^0)\right). \tag{2.5.12}$$

It is straightforward to see that the representation in Eq. (2.5.6) holds if for $j = 4, 5$, $\hat{E}_n(\psi_{nj}(X_i)) - \hat{E}_n(\psi_{nj}(X_i)) = o_p(1)$.

The following theorem verifies Assumption 2.5.1.

THEOREM 2.5.2: *(i) Suppose that Assumptions 2.5.2 and 2.5.3 hold. Then, Assumption 2.5.1 (i) holds with* $\mathcal{S} = \bar{\mathcal{S}}_I^\epsilon$ *and* $g(w) := \sup_{S\in\bar{\mathcal{S}}_I^\epsilon} \max\{w(S), 0\}^2 - \sup_{S\in\bar{\mathcal{S}}_R^\epsilon} \max\{w(S), 0\}^2$.

(ii) Suppose further that the first-stage estimator $\hat{\theta}_{1n}$ *is asymptotically linear with influence function* ψ_1*, with* $E(\psi_1(X_i)) = 0$*, and* $E|\psi_{1,h}(X_i)|^2 < \infty$ *for all* $h = 1, \cdots, d_1$*. Suppose that Assumption 2.5.4 holds. Suppose also that*

$$E\left(\frac{\partial}{\partial\theta_1} \ln f(X_i; \theta_1^0, \theta_2^0)\right) = 0, \quad E\left(\frac{\partial}{\partial\theta_1} \ln f(X_i; \theta_1^0, \theta_2^R)\right) = 0,$$

and for any $\theta_2 \in \Theta_2$*,*

$$E\left(vec\left(-\frac{\partial^2}{\partial\theta_1\partial\theta_1'}\ln f(X_i;\theta_1^0,\theta_2)\right)\right) < \infty,$$
$$E\left(\frac{\partial}{\partial\theta_1}\ln f(X_i;\theta_1^0,\theta_2)'\frac{\partial}{\partial\theta_1}\ln f(X_i;\theta_1^0,\theta_2)\right) < \infty.$$

Suppose further that $\hat{E}_n(\psi_{nj}(X_i)) - \hat{E}_n(\psi_j(X_i)) = o_p(n^{-1/2}), j = 2,3$ *and that* $\hat{E}_n(\psi_{nj}(X_i)) - \hat{E}_n(\psi_j(X_i)) = o_p(1)$*,* $j = 4,5$*. Then Assumption 2.5.1 (ii) holds with* $\psi_j, j = 2\cdots,5$ *given in Eqs.* (2.5.10) *and* (2.5.12)*, and Assumption 2.5.1 (iii) holds.*

An immediate corollary of this theorem is the following.

Corollary 2.5.1: *Suppose Assumption 2.2.1 holds with* (2.2.1) *defining* Q_n*, and Assumption 2.2.2 and 2.2.4 hold with* $\Theta_{I,1} = \{\theta_1^0\}$*. Suppose* $\hat{T}_n(\Theta_2,R)$ *admits the expansion in Eq.* (2.5.2) *with* $a_n = n$*,* $p = d_1$*,* $\theta_0 = \theta_0^1$*,* $a_n\hat{\mathcal{Q}}_n = \hat{T}_n(\Theta_2,R)$*,* $a_n\mathcal{Q}_n = T_n(\Theta_2,R;\theta_0^1)$*,* $\hat{\theta}_n = \hat{\theta}_{1n}$*,* $g_n = n^{-1}(\partial/\partial\theta_1)T_n(\Theta_2,R;\theta_1^0)$*, and* $H_n = n^{-1}(\partial^2/\partial\theta_1\partial\theta_1')T_n(\Theta_2,R;\theta_1^0)$*. Suppose further that the conditions of Theorem 2.5.2 hold. Then*

$$\hat{T}_n(\Theta_2,R) \xrightarrow{d} g(\mathcal{W}) + (W_2 - W_3)'W_1 + W_1'HW_1,$$

where $\mathcal{W}$ *is a zero mean tight Gaussian process on* $\bar{\mathcal{S}}_I^\epsilon$ *with covariance kernel* $E[\mathcal{W}(S_1)\mathcal{W}(S_2)] = E[S_1S_2]$ *for all* $S_1, S_2 \in \bar{\mathcal{S}}^\epsilon$ *and* $g(w) := \sup_{S\in\bar{\mathcal{S}}_I^\epsilon}\max\{w(S),0\}^2 - \sup_{S\in\bar{\mathcal{S}}_R^\epsilon}\max\{w(S),0\}^2$*;* $(\mathcal{W}(S_1),\cdots,\mathcal{W}(S_m),W_1,W_2,W_3)$ *follows a mean zero multivariate normal distribution with covariances*

$$E[\mathcal{W}(S_{m_j})\mathcal{W}(S_{m_k})] = E[S_{m_j}S_{m_k}], \quad \forall m_j, m_k = 1,\cdots,m$$
$$E[\mathcal{W}(S_{m_j})W_k] = E[S_{m_j}\psi_k(X_i)], \quad \forall m_j = 1,\cdots,m, \quad k = 1,2,3$$
$$E[W_jW_k] = E[\psi_j(X_i)\psi_k(X_i)], \quad \forall \quad k = 1,2,3;$$

and H *is such that* $vec(H) = E[\psi_4(X_i)] - E[\psi_5(X_i)]$*.*

2.6 Concluding Remarks

This paper studies an estimation and inference procedure for a parameter that has a two-stage structure. This structure enable us to estimate a subvector of the parameter or its identified set separately from the rest. As we illustrate, various applied studies use this structure. Our procedure constructs a two-stage set estimator by taking an appropriate level 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 in Bajari, Benkard, and Levin (2007), but its measurability and its applicability to hypothesis testing have not been previously studied. We give conditions for the measurability of the two-stage set estimator and establish consistency of the two-stage estimator, extending results of Chernozhukov, Hong, and Tamer (2007). For testing hypothesis about the second-stage parameter, we propose a test based on a quasi-likelihood ratio type statistic and study its asymptotic distribution. We give primitive conditions for an important special case based on results of Liu and Shao (2004). This test is especially useful when the researcher is interested in testing a hypothesis that involves only a subset of the whole parameter vector. A future task is to extend this testing method to the general case where the first-stage parameter is also partially identified.

2.A Mathematical Appendix

Proof of Theorem 2.4.1. For any $E \subseteq \Theta$, let $\hat{\Theta}^{-}(E) := \{\omega : \hat{\Theta}(\omega, \hat{\epsilon}(\omega)) \cap E \neq \varnothing\}$. We first note the following fact (Theorem 1.2.3 in Molchanov, 2005). As Θ is a subset of a complete separable metric space, $\hat{\Theta}(\omega, \hat{\epsilon}(\omega))$ is Effros-measurabile if and only if

$$\hat{\Theta}^{-}(F) \in \mathfrak{F}, \quad \forall F \in \mathcal{F}(\Theta). \tag{2.A.1}$$

For our purposes, it is more convenient to work with closed sets. We therefore establish the Effros-measurability by showing (2.A.1).

Let $F \in \mathcal{F}(\Theta)$. If $F = \varnothing$, then $\hat{\Theta}(\omega, \hat{\epsilon}(\omega))^{-}(F) = \varnothing \in \mathfrak{F}$. Below, we consider cases where F is nonempty. For any $\omega \in \Omega$,

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

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

$$\{\omega : \hat{\Theta}(\omega, \hat{\epsilon}(\omega)) \cap F \neq \varnothing\} = \{\omega : R(\omega) \cap F \neq \varnothing\} \cap \left\{\omega : \inf_{R_F(\omega)} \zeta(\omega, \theta_1, \theta_2) \leq \hat{\epsilon}(\omega)\right\}.$$

Therefore, it suffices to show that the random set R is Effros-measurable and that the infimum of the random function ζ over the random closed set R_F is also measurable in usual sense. For the Effros-measurability of R, observe that

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

where $F_1 := \{\theta_1 \in \Theta_1 : (\theta_1, \theta_2) \in F \text{ for some } \theta_2 \in \Theta_2\}$. It is obvious that R_F is also Effros-measurable.

For the measurability of $\inf_{R_F(\omega)} \zeta(\omega, \theta_1, \theta_2)$, we apply Lemma 2.4.2. Conditions (i) and (ii) of Lemma 2.4.2 are satisfied by our hypothesis. Assumption (ii) of Theorem 2.4.1 ensures that ζ is jointly measurable by Lemma 2.4.1, which ensures condition (iii) of Lemma 2.4.2. Condition (iv) of Lemma 2.4.2 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 2.4.2 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 (1996). Given this result, we can apply Lemma 2.4.1 to establish the jointly measurability of $\tilde{\zeta}$. By Fact 2.4.1, $\tilde{\zeta}$ is a normal integrand. This ensures the measurability of its level sets. □

Proof of Theorem 2.4.2. To show $d_H(\hat{\Theta}_n, \bar{\Theta}_I) = o_p(1)$, we need both (i) $\sup_{\theta \in \bar{\Theta}_I} d(\theta, \hat{\Theta}_n) = o_p(1)$ and (ii) $\sup_{\theta \in \hat{\Theta}_n} d(\theta, \bar{\Theta}_I) = o_p(1)$.

We first show (i). By condition (iv), for any $\delta > 0$, there exists $N_\delta \geq 0$ such that $P(E_n) \geq 1 - \delta$ for all $n \geq N_\delta$, where $E_n := \left\{\omega : \sup_{\theta \in \bar{\Theta}_I} a_n Q_n(\theta) \leq \hat{\epsilon}_n\right\}$. Below, we take $\omega \in E_n$. For any $\epsilon > 0$, let $\hat{\Theta}_n^\epsilon(\omega) := \{\theta \in \Theta : d(\theta, \hat{\Theta}_n(\omega)) \leq \epsilon\}$ be the ϵ-expansion of $\hat{\Theta}_n$. Then, $\sup_{\theta \in \hat{\Theta}_n^\epsilon(\omega)} Q_n(\omega, \theta) > \hat{\epsilon}_n$ as $\hat{\Theta}_n$ is the $\hat{\epsilon}_n$-level set and that $Q_n(\omega, \cdot)$ is continuous. Therefore, we have

$$\sup_{\theta \in \bar{\Theta}_I} a_n Q_n(\omega, \theta) < \sup_{\theta \in \hat{\Theta}_n^\epsilon(\omega)} a_n Q_n(\omega, \theta)$$

This implies $\bar{\Theta}_I \subset \hat{\Theta}_n^\epsilon(\omega)$. Therefore, $\sup_{\theta \in \bar{\Theta}_I} d(\theta, \hat{\Theta}_n(\omega)) \leq \epsilon$. Since ϵ was arbitrary, $\sup_{\theta \in \bar{\Theta}_I} d(\theta, \hat{\Theta}_n) = o_p(1)$.

For (ii), we need to show that for any $\epsilon > 0$, $\sup_{\theta_1 \in \hat{\Theta}_n} d(\theta_1, \bar{\Theta}_I) \leq \epsilon$ 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 $\bar{\Theta}_I = \arg\min_{\theta\in\Theta} \bar{\zeta}$. Let

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

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

By the triangle inequality,

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

where the second equality holds since $\Theta_{I,1}$ and $\hat{\Theta}_{1n}$ are closed under our assumption 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

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

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

Combining steps (i) and (ii), we conclude that $d(\hat{\Theta}_n, \bar{\Theta}_I) = o_p(1)$. □

Proof of Theorem 2.4.3. Since Θ_r is a fixed closed set defined as a preimage of a continuous function ρ, Effros-measurability and consistency trivially follow from those of $\hat{\Theta}_{ur,1n}$. □

Proof of Theorem 2.4.4. Let $g : \Theta_1 \times \Psi \to \mathbb{R}^{m_1+m_2}$ be a mapping such that

$$g(\theta_1, \psi) = \begin{bmatrix} s(\theta_1, \psi) \\ \rho(\theta_1) \end{bmatrix}.$$

For (i), we again consider intersections with closed sets. Let $F \in \mathcal{F}(\Theta_1)$. Consider the set

$$R_F(\omega) = \{\theta_1 \in \Theta_1 : g(\theta_1, \hat{\psi}_n(\omega)) = 0\} \cap F.$$

Let $\tilde{g} : \Omega \times \Theta_1$ be a measurable map $(\omega, \theta_1) \mapsto g(\theta_1, \hat{\psi}_n(\omega))$. By conditions (a), (b), (d), and Lemma 2.4.1, $\tilde{g}$ is jointly meansurable, following the proof of Example 3.1 in Stinchcombe and White (1992), we can show $grR_F = \tilde{g}^{-1}(\{0\}) \cap (\Omega \times F)$. Therefore, $grR_F \in \mathfrak{F} \otimes \mathcal{B}_{\Theta_1}$ for any $F \in \mathcal{F}(\Theta_1)$. This is equivalent to the Effros measurability of $\hat{\Theta}_{r,1n}$ by fundamental measurability theorem. This implies the Effros measurability of $\hat{\Theta}_{1n}$.

For (ii), we use the fact that the convergence in Hausdorff metric is equivalent to the general notion of set convergence called Painlevé-Kuratowski convergence (PK-converngece) when the parameter space is bounded. See section 4.C in Rockafellar and Wets (2005). For completeness we give the definition of PK-convergence below.

Definition 2.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

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

We write $F_n \overset{PK}{\to} F$ *or* $PK - \lim F_n = F$[12].

By assumption, $\hat{\Theta}^u_{1n}$ converges to $\Theta^u_{I,1}$ in Hausdorff metric in probability. Therefore, it has a subsequence $\hat{\Theta}^u_{1n_k}$ that converges in Painlevé-Kuratowski sense with probability one. We use the following lemma.

Lemma 2.A.1 (Hit or miss criteria: Theorem 4.5 in RW): *Let* $\mathbb{E}$ *be a locally compact Hausdorff second countable space (LCHS). 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* $F \cap G \neq \emptyset$*, one has* $F_n \cap G \neq \emptyset$ *for all sufficiently large* n*.*
2. $\limsup_{n\to\infty} F_n \subseteq F$ *iff for every compact set* K *with* $F \cap K = \emptyset$*, one has* $F_n \cap K = \emptyset$ *for all sufficiently large* n*.*

Now let G be an open set such that $(\Theta^u_{I,1} \cap \Theta^r_1) \cap G \neq \emptyset$. This implies $\Theta^u_{I,1} \cap G \neq \emptyset$ and $\Theta^r_1 \cap G \neq \emptyset$. Similarly, let K be a compact set such that $(\Theta^u_{I,1} \cap \Theta^r_1) \cap K = \emptyset$. Since $\Theta^u_{I,1} \cap \Theta^r_1 \neq \emptyset$, we must have $\Theta^u_{I,1} \cap K = \emptyset$ or $\Theta^r_1 \cap K = \emptyset$.

We first show that every subsequence of $\hat{\Theta}_{1n_k}$ has a further subsequence that satisfies $\hat{\Theta}_{1n_{k_j}} \cap G \neq \emptyset$. By the hypothesis and Lemma 2.A.1 1, $\hat{\Theta}^u_{1n_k} \cap G \neq \emptyset$ for sufficiently large n almost surely. Now let $\theta^*_1 \in \Theta^r_{I,1} \cap G$. Since G is open and $s(\theta^*_1, \cdot)$ is continuous, for any $\epsilon > 0$, there exists an open ball $B(\theta^*_1, \delta)$ with some radius $\delta > 0$ such that $|s(\theta_1, \psi_0)| < \epsilon$ and $|\rho(\theta_1)| < \epsilon$ for all $\theta_1 \in B(\theta^*_1, \delta)$. Note that $\hat{\Theta}^r_{1n}$ converges in Hausdorff metric in probability to Θ^r_1. Therefore, it has a

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

subsequence such that for all n_l large and for all $\theta_{1,n_l} \in \hat{\Theta}^r_{n_l}$, $|s(\theta_{1,n_l}, \psi_0)| < \epsilon$ and $\rho(\theta_{1,n_l}) < \epsilon$ almost surely. Take $\{n_{k_j}\} := \{n_k\} \cap \{n_l\}$. Then, $\hat{\Theta}_{1n_{k_j}} \cap G \neq \emptyset$.

We now show that every subsequence of $\hat{\Theta}_{1n}$ has a further subsequence that satisfies $\hat{\Theta}_{1n_{k_j}} \cap K = \emptyset$. By the hypothesis and Lemma 2.A.1 2, if $\Theta^u_{I,1} \cap K = \emptyset$, $\hat{\Theta}^u_{1n_k} \cap K \neq \emptyset$ for sufficiently large n_k almost surely. Therefore, further subsequences also satisfy this condition. Now, if $\Theta^r_{I,1} \cap K = \emptyset$, take $\theta^*_1 \in K$. Then for some $\delta > 0$, $s(\theta^*, \psi^0) \geq \delta$ and $\rho(\theta^*_1) \geq \delta$, but the Hausdorff consistency of $\hat{\Theta}^r_{1n}$ again implies it has a subsequence such that for all n_l large and for all $\theta_{1,n_l} \in \hat{\Theta}^r_{n_l}$, $|s(\theta_{1,n_l}, \psi_0)| < \delta$ and $\rho(\theta_{1,n_l}) < \delta$ almost surely. Therefore $\hat{\Theta}^r_{1n_l} \cap K = \emptyset$.

Now we have shown that every subsequence of $\hat{\Theta}_{1n}$ has a further subsequence that satisfies the hit-or-miss criteria almost surely. Therefore $d_H(\hat{\Theta}_{1n_{k_j}}, \Theta^r_{I,1}) = o_{as}(1)$, but this also implies $d_H(\hat{\Theta}_{1n_{k_j}}, \Theta^r_{I,1}) = o_p(1)$. Therefore, the original sequence converges in probability, which is the conclusion of the theorem. □

Proof of Theorem 2.5.1. As Assumption 2.2.1 holds with Q_n given in (2.2.1), we may write

$$T_n(\Theta_2, R; \theta^0_1) = \inf_{\theta_2 \in R} \sum_{i=1}^n q(X_i, \theta^0_1, \theta_2) - \inf_{\theta_2 \in \Theta_2} \sum_{i=1}^n q(X_i, \theta^0_1, \theta_2).$$

Assumption 2.2.2 ensures the two-stage structure, and Assumption 2.2.4 ensures the existence of a first-stage estimator $\hat{\theta}_{1n}$. By Assumption 2.5.1 (i), $\mathcal{W}_n \Rightarrow \mathcal{W}$, where $\mathcal{W}$ is tight. Assumption (ii) and (iii) imply that

$$(\mathbb{G}_n(\psi_1(X_i)), \mathbb{G}_n(\psi_2(X_i)), \mathbb{G}_n(\psi_3(X_i)), \hat{E}_n(\psi_4(X_i)), \hat{E}_n(\psi_5(X_i)))' \xrightarrow{d} (W_1, W_2, W_3, E(\psi_4(X_i)), E(\psi_5(X_i)))'$$

and $\mathcal{Y}_n \xrightarrow{f.d.} \mathcal{Y}$, where

$$\mathcal{Y} := (\mathcal{W}, W_1, W_2, W_3, E(\psi_4(X_i)), E(\psi_5(X_i)))',$$

As the joint tightness is implied by the marginal tightness, this ensures that $\mathcal{Y}_n \Rightarrow$

$\mathcal{Y}$. By Assumption 2.5.1 and the continuous mapping theorem,

$$\left(T_n(\Theta_2, R; \theta_1^0), \sqrt{n}(\hat{\theta}_{1n} - \theta_1^0), \frac{1}{\sqrt{n}}\frac{\partial}{\partial \theta_1}T_n(\Theta_2, R; \theta_1^0), \frac{1}{n}vec\left(\frac{\partial^2}{\partial\theta_1\partial\theta_1'}T_n(\Theta_2, R; \theta_1^0)\right)\right)' \xrightarrow{d} (g(\mathcal{W}), W_1, W_2 - W_3, H)'.$$

Now apply Proposition 2.5.1 with $a_n = n$, $p = d_1$, $\theta_0 = \theta_0^1$, $a_n\hat{\mathcal{Q}}_n = \hat{T}_n(\Theta_2, R)$, $a_n\mathcal{Q}_n = T_n(\Theta_2, R; \theta_0^1)$, $\hat{\theta}_n = \hat{\theta}_{1n}$, $g_n = n^{-1}(\partial/\partial\theta_1)T_n(\Theta_2, R; \theta_1^0)$, and $H_n = n^{-1}$ $(\partial^2/\partial\theta_1\partial\theta_1')T_n(\Theta_2, R; \theta_1^0)$. Then, the conclusion follows. □

Proof of Theorem 2.5.2 . (i): First, Assumption 2.5.3 gives the required Donsker-class. Given Assumption 2.5.2, we may write

$$T_n(\Theta_2, R; \theta_1^0) = \sup_{\theta_2\in\Theta_2} \sum_{i=1}^n \ln l_{\theta_1^0,\theta_2}(X_i) - \sup_{\theta_2\in R}\sum_{i=1}^n \ln l_{\theta_1^0,\theta_2}(X_i).$$

Given Assumptions 2.5.2 and 2.5.3, Theorem 3.1 in Liu and Shao (2003) ensures that for each $\theta_2 \in \Theta_2$, there exist an $\epsilon > 0$ and $S_{\theta_2} \in \bar{\mathcal{S}}_I^\epsilon$ such that

$$\max\{\sum_{i=1}^n \ln l_{\theta_1^0,\theta_2}(X_i), 0\} = \max\{\sqrt{n}D(\theta_2)\mathbb{G}_n(S_{\theta_2}(X_i)) - n/2D^2(\theta_2), 0\} + o_p(1).$$

Given this representation,

$$\begin{aligned}\sup_{\theta_2\in\Theta_2}\sum_{i=1}^n \ln l_{\theta_1^0,\theta_2}(X_i) &= \sup_{\theta\in\Theta_2, \sum_{i=1}^n \ln l_{\theta_1^0,\theta_2}(X_i)>0} \max\{\mathbb{G}_n S_{\theta_2}(X_i), 0\}^2 + o_p(1)\\ &= \sup_{S_{\theta_2}\in\bar{\mathcal{S}}_I^\epsilon} \max\{\mathbb{G}_n S_{\theta_2}(X_i), 0\}^2 + o_p(1)\end{aligned}$$

following the proof of Theorem 2.3 in Liu and Shao (2003). Similarly for $\sup_{\theta_2\in R}$ $\sum_{i=1}^n \ln l_{\theta_1^0,\theta_2}(X_i)$, we have

$$\sup_{\theta_2\in R}\sum_{i=1}^n \ln l_{\theta_1^0,\theta_2}(X_i) = \sup_{S_{\theta_2}\in\bar{\mathcal{S}}_R^\epsilon} \max\{\mathbb{G}_n S_{\theta_2}, 0\}^2 + o_p(1)$$

Combining these results gives the conclusion of (i).

(ii): The asymptotic linearity of the first-stage estimator is given by the hypothesis. Given Assumption 2.5.4 (i) and (ii), the conditions of Theorem 3 in Milgrom and Segal hold for $\inf_{\theta_2\in R}\sum_{i=1}^n -\ln f(x;\cdot,\theta_2)$ and $\inf_{\theta_2\in\Theta_2}\sum_{i=1}^n -\ln f(x;\cdot,\theta_2)$. Then, we obtain

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta_1}T_n(\Theta_2,R;\theta_1^0)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_{n2}(X_i)-\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_{n3}(X_i).$$

As $\hat{E}_n\psi_{nj}(X_i)-\hat{E}_n\psi_j(X_i)=o_p(n^{-1/2}), j=2,3$, we may write

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta_1}T_n(\Theta_2,R;\theta_1^0)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_2(X_i)-\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_3(X_i)+o_p(1),$$

where ψ_2 and ψ_3 are of the form (2.5.10). This ensures the representation in (2.5.5). For the representation (2.5.6), we apply Theorem 3 in Milgrom and Segal one more time. This is possible under Assumption 2.5.4 (iii). Then, we obtain

$$\frac{1}{n}vec\left(\frac{\partial^2}{\partial\theta_1\partial\theta_1'}T_n(\Theta_2,R;\theta_1^0)\right)=\frac{1}{n}\sum_{i=1}^n\psi_{n4}(X_i)-\frac{1}{n}\sum_{i=1}^n\psi_{n5}(X_i).$$

As $\hat{E}_n(\psi_{nj}(X_i))-\hat{E}_n(\psi_j(X_i))=o_p(1), j=4,5$, we may write

$$\frac{1}{n}vec\left(\frac{\partial^2}{\partial\theta_1\partial\theta_1'}T_n(\Theta_2,R;\theta_1^0)\right)=\frac{1}{n}\sum_{i=1}^n\psi_4(X_i)-\frac{1}{n}\sum_{i=1}^n\psi_5(X_i)+o_p(1),$$

where ψ_4 and ψ_5 are of the form in (2.5.12).

The hypothesis of the theorem implies that $E(\psi_j(X_i))=0$, $j=1,2,3$ and $E(\psi_j(X_i)'\psi_j(X_i))<\infty$, $j=1,2,3$. Furthermore, for any $S\in\bar{\mathcal{S}}_I^\epsilon$, $E(S^2(X_i))<\infty$. Therefore,

$$(\mathcal{W}_n(S_1),\cdots,\mathcal{W}_n(S_m),\mathbb{G}_n(\psi_1(X_i)),\mathbb{G}_n(\psi_2(X_i)),\mathbb{G}_n(\psi_3(X_i)))'$$

jointly obeys the multivariate central limit theorem for IID random variables. Note further that $E(\psi_j(X_i))<\infty$, $j=4,5$ by our hypothesis. The weak law of large numbers then implies $E(\psi_j(X_i))\overset{p}{\to}E(\psi_j(X_i))=:H_j$, $j=4,5$. By Slutsky's lemma, Assumption 2.5.1 (iii) follows.

□

Proof of Corollary 2.5.1. The conclusion is immediate, invoking Theorem 2.5.1.

□

Bibliography

BAJARI, P., C. L. BENKARD, AND J. LEVIN (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75(5), 1331–1370.

BERESTEANU, A., AND F. MOLINARI (2006): "Asymptotic Properties for a Class of Partially Identified Models," Cemmap Working Paper CWP 10/06.

BILLINGSLEY, P. (1995): *Probability and Measure*. Wiley-Interscience, 3 edn.

——— (1999): *Convergence of Probability Measures (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 2 edn.

BUGNI, F. A., I. CANAY, AND P. GUGGENBERGER (2010): "Asymptotic Distortions in Locally Misspecified Moment Inequality Models," Discussion Paper, University of California San Diego.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models1," *Econometrica*, 75(5), 1243–1284.

DEBREU, G. (1967): "Integration of correspondences.," Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, University of Calififornia Press, II-1 351-372.

HALL, W., AND R. LOYNES (1977): "Weak convergence of processes related to likelihood ratios," *Ann. Statist.*, 5(2), 330–341.

HECKMAN, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–161.

KAIDO, H. (2009): "A Dual Approach to Inference for Partially Identified Econometric Models," Working Paper, University of California, San Diego.

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

KLINE, P., AND A. SANTOS (2010): "Interval Estimation of Potentially Misspecified Quantile Models in the Presence of Missing Data," Discussion Paper, University of California San Diego.

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

Ponomareva, M., and E. Tamer (2009): "Misspecification in Moment Inequality Models: Back to Moment Equalities?," Discussion Paper, Northwestern University.

Romano, J. P., and A. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, Special Issue in Honor of T. W. Anderson, forthcoming.

Romano, J. P., and A. M. Shaikh (2006): "Inference for the Identified Set in Partially Identified Econometric Models," Working Paper, University of Chicago.

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.

White, H. (1996): *Estimation, Inference and Specification Analysis (Econometric Society Monographs)*. Cambridge University Press.

This chapter is currently being prepared for submission for publication of the material. Kaido, Hiroaki; White, Halbert. I was the primary investigator and author of this material.

Chapter 3

A Dual Approach to Inference for Partially Identified Econometric Models

3.1 Introduction

Statistical inference for partially identified economic models is a growing field in econometrics. The field was pioneered by Charles Manski in the 1990's (see Manski, 2003, and the references there), and there have since been substantial theoretical extensions and applications. In this literature, the economic structures of interest are characterized by an *identified set* Θ_I, rather than by a single point in the parameter space $\Theta \subset \mathbb{R}^d, d \in \mathbb{N}$. Elements of the identified set lead to observationally equivalent data generating processes. A sample of data generated by any of the parameter values in the identified set, therefore, gives us information about the identified set, but not about the underlying "true" parameter value generating the observed data.

Chernozhukov, Hong, and Tamer (2007) (CHT) study estimation and statistical inference on Θ_I within a general extremum estimation framework. These authors have shown that a level-set estimator based on a properly chosen sequence of levels for the criterion function consistently estimates the identified set, defined as a set of minimizers. They use a quasi-likelihood ratio (QLR) statistic to construct a confidence set that asymptotically covers the identified set with at least a prespecified probability. This criterion function approach is applicable to a broad class of problems.

Another popular approach is to estimate the boundary of Θ_I directly. This estimate can then be used to conduct inference for Θ_I. This is an attractive alternative if the boundary of the identified set is easily estimable. Much of the literature has studied the case where Θ_I is a closed interval (e.g. Horowitz and Manski, 1998, 2000, Manski, 2003, and Imbens and Manski, 2004). Recent studies extend this approach to the case where Θ_I is a multi-dimensional compact convex set (Beresteanu and Molinari, 2008 (BM) and Bontemps, Magnac, and Maurin, 2008). When Θ_I is compact and convex, its support function provides a tractable representation by summarizing the location of the supporting hyperplanes of Θ_I.

So far, the criterion function approach and the support function approach have been viewed as distinct. Each has its advantages and challenges. The criterion function approach is widely applicable, but constructing the level set can be computationally demanding. The support function approach, on the other hand, is more direct and computationally tractable for some problems, but it has been applied to a limited class of models when parameters are multi-dimensional. A main contribution of this paper is to unify these approaches within a general framework. We do this by studying an inference method that exploits the wide applicability of the criterion function approach and the tractability of the support function approach. To the best of our knowledge, this is the first such attempt.

In this paper, we focus on econometric models with compact convex identified sets, which enables us to characterize the identified set by its support function[1]. This class includes many econometric models studied recently, e.g., regression with interval data (Manski and Tamer, 2004, Magnac and Maurin, 2008) and an asset pricing model in incomplete markets (Kaido and White, 2009). Following CHT, our estimator of Θ_I is the level set $\hat{\Theta}_n = \{\theta : Q_n(\theta) \leq t_n\}$ of a criterion function $Q_n(\cdot)$ for some sequence of levels $\{t_n\}$. Collecting all the parameter values at which $Q_n(\theta)$ does not exceed the specified level can be computationally demanding. Our alternative method stores the values $\max_{Q_n(\theta)\leq t_n}\langle p, \theta\rangle$ for different unit vectors p. This yields the support function $s(\cdot, \hat{\Theta}_n)$ of the set estimator. The required computation is straightforward, and one can fully recover the set estimator from its support function. This can result in computational savings that range from modest to dramatic.

Another significant contribution here is a new automated step-up algorithm for selecting the tuning parameter t_n. As explained above, the criterion function approach requires the researcher to choose the level t_n of the criterion function to construct the set estimator (CHT; Bugni, 2009). Our iterative algorithm removes the arbitrariness in the choice of t_n. We relate this to a multiple testing problem. Our algorithm can be interpreted as the reduced form of a step-up procedure that controls the familywise error rate (FWER) of hypotheses that are indexed by compact convex sets. This understanding provides a link to Romano and Shaikh's (2009) recent work on a step-down procedure.

Our approach is particularly well suited to conducting hypothesis tests and

[1]Our analysis applies to the convex hull of the identified set if it is nonconvex.

constructing confidence collections and confidence sets. For this, we first show that the asymptotic distribution of the properly normalized (centered and scaled) support function is that of a specific stochastic process on the unit sphere. The normalized support function lets us measure the distance between sets using the Hausdorff metric common in the literature. This enables us to test the hypothesis that the identified set coincides with a given set, i.e., $H_0 : \Theta_I = \Theta_0$. The test can be inverted to construct a confidence collection that contains the identified set as an element, with some prescribed confidence level. Inference methods for this type of hypothesis are as yet unavailable within CHT's framework.

The normalized support function also lets us test whether the identified set includes a specific set or point. That is, for a given set Θ_0 or point θ_0, we can test $H_0 : \Theta_0 \subseteq \Theta_I$ or $H_0 : \theta_0 \in \Theta_I$. The former test can be inverted to construct another confidence collection, containing each subset of the identified set as an element, with at least some prescribed confidence level. Further, taking the union of the elements of this collection yields a confidence set that covers the identified set. This confidence set is comparable to CHT's confidence sets, constructed by inverting their QLR statistic. Similarly, the test for $\theta_0 \in \Theta_I$ can be inverted to construct a confidence set for each point in the identified set. This set is comparable to those of Imbens and Manski (2004), CHT, Romano and Shaikh (2008), and Andrews and Guggenberger (2009).

The construction of confidence collections and confidence sets by inverting the normalized support function was first proposed by BM for the case where Θ_I is a linear transformation of the Aumann expectation of set-valued random variables. Bontemps, Magnac, and Maurin (2007) consider a confidence set for a point in the identified set, when Θ_I is characterized by incomplete linear moment restrictions. Our analysis further contributes by extending these results to the general case where Θ_I is the set of minimizers of a criterion function.

Closely related to our work here is that of BM, who develop an estimation and inference framework based on their set-average estimator, a (Minkowski) average of independent and identically distributed (IID) set-valued random variables. One of BM's key ideas is to embed the space of compact convex sets into a subset of the space of continuous functions (Hörmander, 1955; Beer, 1993). In this paper, we follow a similar approach to study the asymptotic behavior of our set estimator. But instead of using a set-averaging approach, we analyze a version of the sample criterion function using *weak epiconvergence* to derive the asymptotic distribution of the normalized support function of the level-set estimator. Weak epiconvergence is a relatively new concept that characterizes the limit of the infimum of stochastic processes over compact sets and has proven useful for studying the asymptotic behavior of extremum estimators with point identification (Knight, 1999; Chernozhukov and Hong, 2004; and Han and Phillips, 2006). Our analysis shows that weak epiconvergence is ideally suited to study extremum estimators of partially identified models[2].

We apply our theory to econometric models characterized by finitely many moment inequalities. This class has been extensively studied. Recent research in this area includes Andrews, Berry, and Jia (2004), Pakes, Porter, Ho, and Ishii (2005), CHT, Fan and Park (2007), Galichon and Henry (2007), BM, Guggenberger, Hahn, and Kim (2008), Rosen (2008), Andrews and Guggenberger (2009), Andrews and Soares (2009), Bugni (2009), Canay (2009), Galichon and Henry (2009), Hahn and Ridder (2009), Moon and Schorfheide (2009), and Yıldız (2009). We contribute to this literature by establishing a new equivalence result within this class. Our Wald-type statistic (squared directed Hausdorff distance) and CHT's QLR statistic converge in distribution to the same limit under some regularity conditions. As a result, the Wald confidence set, i.e., the union of all elements in the confidence collection constructed from the Wald statistic, is asymptotically equivalent to CHT's confidence set, a level set whose level is a specific quantile of the QLR statistic.

A special case of this result is the equivalence result previously given by BM. They show that the Wald statistic based on their set-average estimator is asymptotically equivalent to CHT's QLR statistic within the class of (one-dimensional)

[2]To the best of our knowledge, Chernozhukov, Hong, and Tamer (2007) is the first article that adapted the idea of weak epiconvergence to partially identified models. They used a modified version, which is called "weak sup-convergence," to study the asymptotic distribution of their QLR statistic. Here we work directly with weak epiconvergence.

interval-identified models. Our results show that this can be attributed to: (i) the asymptotic equivalence of the Wald statistic and the QLR-statistic within a more general class; and (ii) the fact that the set-average estimator coincides with the level-set estimator when Θ_I is a closed interval.

The paper is organized as follows. In section 3.2, we summarize CHT's econometric framework and introduce some useful background. We establish the asymptotic distribution of the normalized support function and develop our inference methods in section 3.3. Section 3.4 studies moment inequality models and presents the equivalence result. We present Monte Carlo simulation results in section 3.5 and conclude in section 3.6. We collect together our mathematical proofs in the mathematical appendix.

3.2 The CHT Framework and Some Useful Background

In this section, we briefly summarize the framework of CHT and introduce basic notions in the theory of variational analysis and random sets.

3.2.1 Criterion Function Approach

Our first assumption describes the data generation process and the sample and population criterion functions. For this we require the following definition, where we let $\mathbb{R}_+ := [0, \infty)$ and $\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$.

Definition 3.2.1: *Let $\mathcal{S} \subset \mathbb{R}^d$, $d \in \mathbb{N}$. The function $f : \mathcal{S} \to \bar{\mathbb{R}}_+$ is* proper *on $\mathcal{S}$ if $f(x) < \infty$ for at least one $x \in \mathcal{S}$. If f is proper on $\mathcal{S} = \mathbb{R}^d$, we say f is* proper.

Assumption 3.2.1: *Let $d \in \mathbb{N}$ and $Q : \mathbb{R}^d \to \bar{\mathbb{R}}_+$ be a Borel measurable function. Let $\Theta \subset \mathbb{R}^d$ be compact and convex, with a nonempty interior. Let $(\Omega, \mathfrak{F}, P)$ be a complete probability space. For $n = 1, 2, ...,$ let $Q_n : \Omega \times \mathbb{R}^d \to \bar{\mathbb{R}}_+$ be jointly measurable such that $Q_n(\omega, \cdot)$ is proper on Θ for all $\omega \in F \in \mathfrak{F}$, $P(F) = 1$,* *and for all* $\omega \in \Omega$ *and* $\theta \notin \Theta$, $Q_n(\omega, \theta) = \infty$.

The set Θ is the parameter space, which we take here to be of finite dimension. Compactness is a standard assumption on Θ for extremum estimation. Convexity and nonempty interior help us to avoid the "parameters on the boundary problem" for partially identified models[3]. The probability measure P governs the stochastic properties of the data generating process (e.g., independence or dependence, stationarity or heterogeneity). When, as is assumed here, $Q_n(\omega, \cdot)$ is proper on Θ for all $\omega \in F \in \mathfrak{F}$, $P(F) = 1$, we say "Q_n is proper on Θ *a.s.*" For convenience in what follows, we define $Q_n(\omega, \cdot)$ outside of Θ to take the value ∞.

The function Q_n acts as our sample criterion function, for example,

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

where $\{X_i : \Omega \to \mathbb{R}\}$ is a sequence of random variables and q is a suitable function, e.g., $q(x, \theta) = (x - \theta)^2$ for scalar x and θ. Observe that the second term ensures that we always have $Q_n(\omega, \theta) \geq 0$. As is common, we may write $Q_n(\theta)$ as a shorthand for $Q_n(\,\cdot\,, \theta)$.

Another common choice for Q_n is that associated with generalized method of moments (GMM) estimation,

$$\begin{aligned} Q_n(\omega, \theta) = {} & [n^{-1} \sum_{i=1}^{n} m(X_i(\omega), \theta)]' \; \hat{V}_n^{-1}(\omega) \; [n^{-1} \sum_{i=1}^{n} m(X_i(\omega), \theta)] \\ & - \inf_{\theta \in \Theta} \; [n^{-1} \sum_{i=1}^{n} m(X_i(\omega), \theta)]' \; \hat{V}_n^{-1}(\omega) \; [n^{-1} \sum_{i=1}^{n} m(X_i(\omega), \theta)], \end{aligned}$$

where m is a suitable vector-valued function such that $E[m(X_i, \theta)] = 0$ for one or more values of θ, and $\hat{V}_n$ is an estimator of V, a suitably chosen covariance matrix.

The function Q is the population criterion function. Under assumptions given below, Q_n converges to Q in a suitable sense. The population analog Q will

[3] This point is already mentioned by CHT, which we do not pursue in this paper. They provided sufficient conditions to ensure the parameters in the interior case. Our assumption is based on Lemma 4.1 of CHT.

thus inherit certain properties (e.g., properness) from the sample criterion function Q_n. Without loss of generality, we normalize the minimum value of Q to 0, i.e. $\inf_\Theta Q(\theta) = 0$. For example, when $\{X_i\}$ is stationary and the expectations exist, the population analog of the first example above is

$$Q(\theta) = E[q(X_i(\cdot), \theta)] - \inf_{\theta \in \Theta} E[q(X_i(\cdot), \theta)].$$

Following Chernozhukov, Hong, and Tamer (2007), we define the identified set as the set of minimizers of Q:

Definition 3.2.2 (Identified set): *The* identified set Θ_I *satisfies*

$$\Theta_I := \{\theta \in \Theta : Q(\theta) = 0\}. \tag{3.2.1}$$

There are numerous examples where the identified set can be written as in (3.2.1). See Manski and Tamer (2002), Bajari, Benkard, and Levin (2007), Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008, 2009), Ciliberto and Tamer (2009), and Kaido and White (2009). Leading examples are the cases where Θ_I is a closed interval in $\mathbb{R}$ or an ellipsoid in $\mathbb{R}^2$. Θ_I is a primary object of interest here. In particular, we are concerned with estimation and inference for Θ_I.

We ensure next that Θ_I is a compact convex set contained in the interior of Θ, $\Theta^o := \text{int}(\Theta)$.

Assumption 3.2.2: *(i)* Θ_I *is nonempty, closed, and convex; (ii)* $\Theta_I \subset \Theta^o$.

The compactness of Θ and Assumption 3.2.2 (i) imply the compactness of Θ_I. Assumption 3.2.2 (ii) removes the trivial case $\Theta_I = \Theta$ and the "parameters on the boundary" case. The latter case is definitely of interest, but to keep a tight focus here, we leave this for analysis elsewhere.

Let $\{a_n\}$ be a sequence of positive constants, and define a stochastic process ζ_n on $\mathbb{R}^d$ by

$$\zeta_n(\theta) := a_n Q_n(\theta), \qquad \theta \in \mathbb{R}^d.$$

The constants a_n normalize the criterion function so that ζ_n converges in distribution to a limit process in an appropriate mode, as we discuss further below. We now define the set estimator of interest here as a level set of ζ_n:

DEFINITION 3.2.3 (Set estimator): *For sequences* $\{t_n \in \mathbb{R}_+\}$ *and* $\{a_n \in \mathbb{R}_+\}$*, the* set estimator *is*

$$\hat{\Theta}_n(t_n) := \{\theta \in \Theta : \zeta_n(\theta) \leq t_n\} = \{\theta \in \Theta : a_n Q_n(\theta) \leq t_n\}.$$

To discuss convergence of $\hat{\Theta}_n(t_n)$ to Θ_I, we require suitable distance measures. For this (here and throughout), let $\mathcal{K}$ be a collection of closed subsets in $\mathbb{R}^d$, and let $\|\cdot\|$ denote the Euclidean norm on $\mathbb{R}^d$. We measure the distance between sets in $\mathcal{K}$ using the following Hausdorff distances.

DEFINITION 3.2.4 (Directed Hausdorff distance and Hausdorff metric): *For any* $A, B \in \mathcal{K}$*, the* directed Hausdorff distance *is defined as*

$$\vec{d}_H(A, B) := \sup_{a \in A} d(a, B),$$

where $d(a, B) := \inf_{b \in B} \|b - a\|$ *and* $\vec{d}_H(A, B) := \infty$ *if either* A *or* B *is empty. The* Hausdorff metric *is defined as*

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

The directed Hausdorff distance takes the value 0 when $A \subseteq B$ and a positive value otherwise[4]. This is useful in checking the coverage of the set estimator. For convenience, we refer to either of these as "Hausdorff distance measures."

CHT give a set of conditions (C.1 and C.2 in their paper) sufficient for the consistency of $\hat{\Theta}_n(t_n)$ for Θ_I in the Hausdorff metric and for deriving its convergence rate. Those conditions are general enough to be satisfied by many examples involving moment inequalities and equalities. Following CHT's conditions C.1 and

[4]The directed Hausdorff distance is formally the *lower Hausdorff hemimetric*. A hemimetric d defined on a set $\mathbb{E}$ is a mapping $\mathbb{E} \times \mathbb{E} \to \mathbb{R}$ such that for any $x, y, z \in \mathbb{E}$, (*i*) $d(x, y) \geq 0$, (*ii*) $d(x, z) \leq d(x, y) + d(y, z)$, and (*iii*) $d(x, x) = 0$. In other words, a hemimetric satisfies some properties of a metric, but fails to satisfy symmetry ($d(x, y) = d(y, x)$) and identity ($d(x, y) = 0$ if and only if $x = y$). There is also an upper Hausdorff hemimetric, which corresponds to $\vec{d}_H(B, A)$.

C.2, we assume the following.

Assumption 3.2.3: *(i)* $\sup_{\theta\in\Theta}\{Q(\theta)-Q_n(\theta)\}_+ = o_p(1)$. *(ii)* $\sup_{\theta\in\Theta_I} Q_n(\theta) = O_p(1/a_n)$. *(iii) There exist positive constants* (δ,κ,γ) *such that for any* $\epsilon\in(0,1)$, *there are* $(\kappa_\epsilon, n_\epsilon)$ *such that for all* $n\ge n_\epsilon$

$$Q_n(\theta)\ge \kappa\min\{d(\theta,\Theta_I),\delta\}^\gamma,$$

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

Under this assumption, the level-set estimator $\hat{\Theta}_n(t_n)$ is consistent in the Hausdorff metric and has a convergence rate $r_n = (a_n/\max\{1,\kappa_n\})^{1/\gamma}$, when t_n satisfies $t_n \ge \sup_{\Theta_I} a_n Q_n(\theta)$ with probability tending to 1. Such a sequence $\{t_n\}$ of levels can be constructed by setting $t_n = t\kappa_n$, where $t>0$ and κ_n is a slowly diverging sequence, e.g., $\kappa_n = \log\log n$. Theorem 3.B.1 in the Appendix summarizes CHT's consistency and rate of convergence results for interested readers.

The following condition, CHT's degeneracy condition (C.3), often holds for econometric models that involve finitely many moment inequalities.

Assumption 3.2.4 (Degeneracy): *(i) There is a sequence of subsets* Θ_n *of* Θ, *which could be data dependent (i.e., Effros-measurable functions on* Ω*), such that* Q_n *vanishes on these subsets, that is,* $Q_n(\theta)=0$ *for each* $\theta\in\Theta_n$, *for each* n, *and these sets can approximate the identified set arbitrarily well in the Hausdorff metric, that is,* $d_H(\Theta_n,\Theta_I)\le\epsilon_n$ *for some* $\epsilon_n = o_p(1)$. *(ii)* $\epsilon_n = O_p(1/a_n^{1/\gamma})$.

Under this additional condition, CHT show that it is possible to achieve consistency and an exact polynomial rate of convergence by choosing a constant level $t_n = t\in\mathbb{R}_+$. For later use, we summarize the results below.

Theorem 3.2.1: *Suppose Assumptions 3.2.1, 3.2.2, 3.2.3 (i), (ii), and 3.2.4 (i) hold. Then,* $d_H(\hat{\Theta}_n(t),\Theta_I) = o_p(1)$. *Suppose, in addition, Assumption 3.2.3 (iii) and 3.2.4 (ii) hold. Then,* $a_n^{1/\gamma} d_H(\hat{\Theta}_n(t),\Theta_I) = O_p(1)$.

For models with finitely many moment inequalities, the sample criterion function often vanishes on the set $\Theta_n = \{\theta\in\Theta : n^{-1}\sum_{j=1}^n m_{j,\theta}\le 0\}$, i.e., the set of parameter values at which sample moment inequalities are satisfied. When Θ_I has nonempty interior, the set of points satisfying the moment inequalities approximate Θ_I at $\sqrt{n}$ rate. In this case, Assumption 3.2.4 holds with $a_n = n$, $\gamma = 2$, and $\kappa_n = 1$. Section 3.4 studies this class of econometric models.

To keep a tight focus on the goal of unifying the criterion function and support function approaches, we maintain Assumption 3.2.4 in the following sections.

3.2.2 Support Function Approach

We begin by defining notions useful for characterizing compact convex sets: *support function*, *supporting plane,* and *support set*. For this, let $\langle x, y\rangle$ denote the (Euclidean) inner product of two vectors $x, y \in \mathbb{R}^d$. We write $\|p\| = \langle p, p\rangle^{1/2}$.

Definition 3.2.5 (Support function, supporting plane, and support set): *Let $F \in \mathcal{K}$ and $\mathbb{S}^{d-1} := \{p \in \mathbb{R}^d : \|p\| = 1\}$ be the unit sphere in $\mathbb{R}^d$. The* support function *s of F at $p \in \mathbb{S}^{d-1}$ is defined by*

$$s(p, F) = \sup_{x \in F} \langle p, x\rangle.$$

The supporting (hyper)plane *$\mathbb{H}(p, F)$ of F at $p \in \mathbb{S}^{d-1}$ is*

$$\mathbb{H}(p, F) = \{x \in \mathbb{R}^d : \langle p, x\rangle = s(p, F)\}.$$

The support set *$H(p, F)$ of F at $p \in \mathbb{S}^{d-1}$ is*

$$H(p, F) = \mathbb{H}(p, F) \cap F.$$

The value of the support function $s(p, F)$ measures the signed distance from the origin of the supporting plane $\mathbb{H}(p, F)$ of the set F with a normal vector p. Figure 3.2.2 illustrates this. When the set is strictly convex, its support set $H(p, F)$ for each $p \in \mathbb{S}^{d-1}$ is a singleton.

A maximization problem associated with the support function can be utilized to compute the level-set estimator $\hat{\Theta}_n(t)$. Consider the following problem for

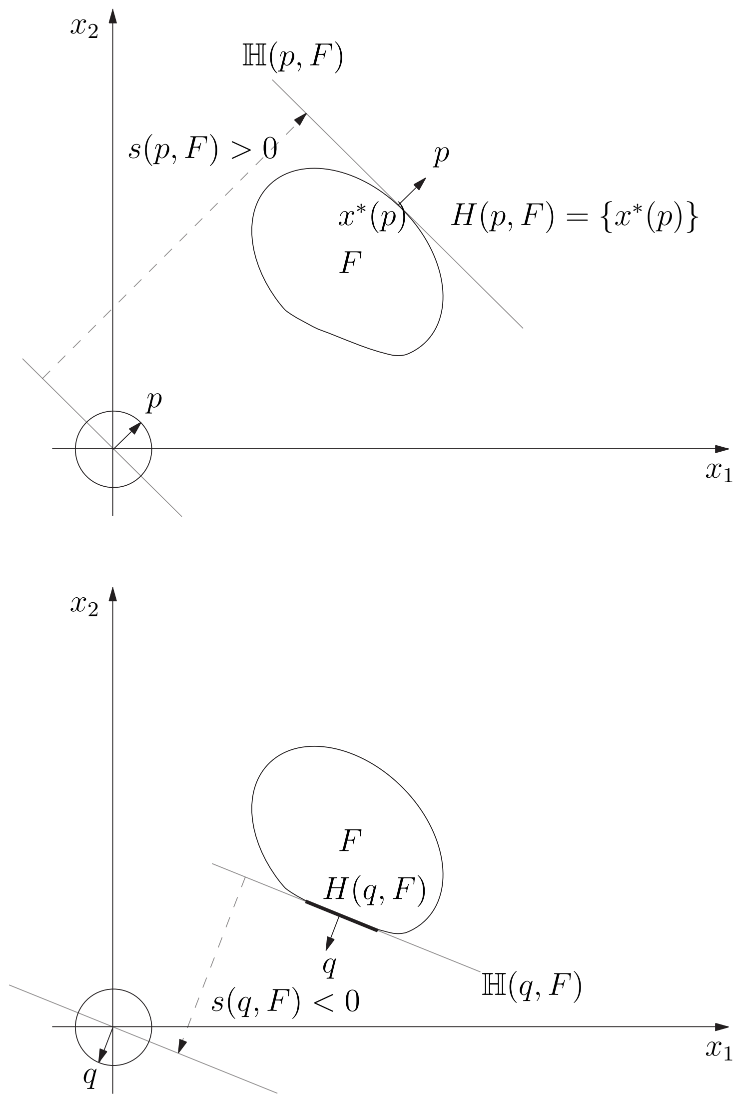

Figure 3.1: Support function, supporting plane, and support set

a given $p \in \mathbb{S}^{d-1}$ and $t \in \mathbb{R}_+$:

$$s(p, \hat{\Theta}_n(t)) := \sup_{\theta \in \Theta} \quad \langle p, \theta \rangle \qquad \text{s.t.} \quad a_n Q_n(\theta) \leq t. \tag{3.2.2}$$

When Q_n is convex in a neighborhood of Θ_I, this is a convex programming problem, which is straightforward to solve numerically using standard algorithms. Often, such algorithms find a point $\hat{\theta}_n(p,t)$ in the support set $H(p, \hat{\Theta}_n(t))$ as a solution of the problem (3.2.2). Therefore, a straightforward algorithm to compute the level set estimator is the following.

ALGORITHM 3.2.1: *Choose $L \in \mathbb{N}$.*

Step 1 : *Generate a grid of points $\{p_1, \cdots, p_L\}$ on the unit sphere $\mathbb{S}^{d-1}$.*

Step 2 : *Solve the problem* (3.2.2) *for $p = p_l, l = 1, \cdots, L$. Store the solutions $\{\hat{\theta}_n(p_l, t), l = 1, \cdots, L\}$.*

When the grid is fine enough, the solutions $\{\hat{\theta}_n(p_l, t), l = 1, \cdots, L\}$ provide a good approximation to the boundary of $\hat{\Theta}_n(t)$.

In addition to providing a straightforward algorithm to compute the level set estimator, the support function itself contains useful information. Let $\mathcal{K}_c$ be a collection of compact convex subsets of $\mathbb{R}^d$. Every nonempty compact convex set is the intersection of its supporting half spaces. Thus, each element of $\mathcal{K}_c$ is uniquely determined by its support function. This suggests that properties of the metric space $(\mathcal{K}_c, d_H)$ may translate nicely to properties of a space of functions. Let $\mathcal{C}(\mathbb{S}^{d-1})$ be the space of bounded continuous functions on $\mathbb{S}^{d-1}$. Let $\|\cdot\|_{\mathcal{C}(\mathbb{S}^{d-1})}$ be the uniform norm on $\mathbb{S}^{d-1}$: i.e., $\|f\|_{\mathcal{C}(\mathbb{S}^{d-1})} = \sup_{x \in \mathbb{S}^{d-1}} |f(x)|$. Let $d_{\mathcal{C}(\mathbb{S}^{d-1})}$ be the metric induced by this norm. Let $\oplus$ denote the *Minkowski addition* operator, such that $F_1 \oplus F_2 = cl\{f_1 + f_2 : f_1 \in F_1, f_2 \in F_2\}$. The Hörmander embedding theorem is

THEOREM 3.2.2 (Hörmander's isometric embedding theorem): *The mapping $F \mapsto s(\cdot, F)$ is an isometric embedding of $(\mathcal{K}_c, d_H)$ into a closed convex cone* *in* $(\mathcal{C}(\mathbb{S}^{d-1}), d_{\mathcal{C}(\mathbb{S}^{d-1})})$ *that preserves Minkowski addition and non-negative multiplication; i.e. for any* F_1 *and* $F_2 \in \mathcal{K}_c$,

$$d_H(F_1, F_2) = \|s(\cdot, F_1) - s(\cdot, F_2)\|_{\mathcal{C}(\mathbb{S}^{d-1})} = \sup_{p \in \mathbb{S}^{d-1}} |s(p, F_1) - s(p, F_2)|$$

$$s(p, F_1 \oplus F_2) = s(p, F_1) + s(p, F_2),$$

and for any $\lambda \in \mathbb{R}_+$,

$$s(p, \lambda F_1) = \lambda s(p, F_1).$$

Details for this theorem are in Beer (1993) and Li, Ogura, and Kreinovich (2002)[5].

For our purposes, the fact that the mapping defined by the support function is an isometry is important. Consider the process:

$$\mathcal{Z}_n(p, t) := a_n^{1/\gamma} \left(s(p, \hat{\Theta}_n(t)) - s(p, \Theta_I) \right).$$

This process is useful for conducting inference. Theorem 3.2.2 ensures that when $\hat{\Theta}_n(t), \Theta_I \in \mathcal{K}_c$, the distance $a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I)$ equals $\sup_{p \in \mathbb{S}^{d-1}} |\mathcal{Z}_n(p, t)|$, a functional of $\mathcal{Z}_n(\cdot, t)$.

For the directed Hausdorff distance, we have the following result.

THEOREM 3.2.3: *Given any two compact convex sets* $F_1, F_2 \in \mathcal{K}_c$, *the directed Hausdorff distance satisfies*

$$\vec{d}_H(F_1, F_2) = \sup_{p \in \mathbb{B}^d} \{s(p, F_1) - s(p, F_2)\} = \sup_{p \in \mathbb{S}^{d-1}} \{s(p, F_1) - s(p, F_2)\}_+,$$

where $\mathbb{B}^d := \{p \in \mathbb{R}^d : \|p\| \leq 1\}$.

For the proof, see BM Lemma A.1.

From this result, together with Assumptions 3.2.1 and 3.2.2 (i), we have

[5] Hörmander's embedding theorem holds in a more general environment. If the underlying space $\mathbb{E}$ is separable, then we can isometrically embed $(\mathcal{K}_c(\mathbb{E}), d_H)$ into a closed convex cone in $\mathcal{C}(\mathbb{S}^*)$, where $\mathbb{S}^*$ is the unit sphere in the dual space $\mathbb{E}^*$. We can use metrics that metrize either the strong norm topology or the weak* topology. This permits extending our framework to handle nonparametric estimation, one of our future tasks.

that for given $t \in \mathbb{R}_+$

$$\begin{aligned} a_n^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) &= \sup_{p \in \mathbb{S}^{d-1}} \{-\mathcal{Z}_n(p,t)\}_+ \\ a_n^{1/\gamma} \vec{d}_H(\hat{\Theta}_n(t), \Theta_I) &= \sup_{p \in \mathbb{S}^{d-1}} \{\mathcal{Z}_n(p,t)\}_+. \end{aligned}$$

If for given t we can find a stochastic process $\mathcal{Z}(\cdot,t)$ such that $\mathcal{Z}_n(\cdot,t)$ converges suitably in distribution to $\mathcal{Z}(\cdot,t)$, then the desired limiting distributions of our Hausdorff distance measures follow from the continuous mapping theorem, as these distance measures are continuous functions of $\mathcal{Z}_n(\cdot,t)$. Thus, we focus on deriving the asymptotic distribution of $\mathcal{Z}_n(\cdot,t)$.

As we show, this distribution is a stochastic process on $\mathbb{S}^{d-1}$. In leading cases, this is a Gaussian process. Moreover, its dependence on t is typically straightforward. Specifically, t often affects only the mean of the limiting process and in a manner known a priori. Thus, there exists a known function μ such that for all $t \in \mathbb{R}_+$, $\mathcal{Z}^*(\cdot) := \mathcal{Z}(\cdot,t) - \mu(t)$ is a mean zero process on $\mathbb{S}^{d-1}$, where $\mathcal{Z}(\cdot,t)$ is the desired weak limit of $\mathcal{Z}_n(\cdot,t)$.

3.2.3 Convergence Concepts

To define the required convergence concepts, consider a sequence of stochastic processes $\{\xi_n\}$ defined on a complete separable metric space $(\mathbb{E}, d)$, so that for $n = 1, 2, ..., \xi_n : \Omega \times \mathbb{E} \to \bar{\mathbb{R}}$ is jointly measurable, where $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$. For simplicity, we often suppress the dependence of ξ_n on $\omega \in \Omega$, but this should be understood implicitly. In specific contexts, we also view ξ_n as a mapping from the sample space Ω to a space of functions on $\mathbb{E}$.

The simplest convergence in distribution concept for stochastic processes is weak convergence in finite dimensions, defined next. We use the notation $\xrightarrow{d}$ to denote the usual convergence in distribution (weak convergence) for a vector of finite dimension (as in, e.g., White, 2001, p.65).

Definition 3.2.6 (Finite dimensional weak convergence): *Let $(\mathbb{E}, d)$ be a complete separable metric space. A sequence of stochastic processes $\{\xi_n, n \geq 1\}$ on*

$\mathbb{E}$ *is said to* weakly converge in finite dimension *to a limit* ξ*, denoted* $\xi_n \overset{f.d.}{\to} \xi$*, if for any finite m-tuple* $(x_1, ..., x_m)$*, where* $x_j \in \mathbb{E}$ *for each* $j = 1, ..., m$*,*

$$(\xi_n(x_1), ..., \xi_n(x_m)) \overset{d}{\to} (\xi(x_1), ..., \xi(x_m)).$$

It is well known that the finite dimensional weak convergence is equivalent to weak convergence in the uniform metric when the sequence $\{\xi_n\}$ is *tight* in $l^\infty(\mathbb{E})$, where $l^\infty(\mathbb{E})$ is the space of uniformly bounded functions on $\mathbb{E}$; see, e.g., van der Vaart and Wellner (2000). We denote $\xi_n \overset{u.d.}{\to} \xi$ when ξ_n weakly converges to a stochastic process ξ in the uniform metric.

Here, a main goal is to find $\mathcal{Z}(\cdot, t)$ such that $\mathcal{Z}_n(\cdot, t) \overset{u.d.}{\to} \mathcal{Z}(\cdot, t)$ for $\mathcal{Z}_n(\cdot, t)$ defined above. In order to achieve this goal, we make use of the notion of weak epiconvergence given next.

Definition 3.2.7 (Weak epiconvergence): *A sequence of stochastic processes* $\{\xi_n, n \geq 1\}$ *on* $\mathbb{E}$ *is said to* weakly epiconverge *to a limit* ξ*, denoted* $\xi_n \overset{e.d.}{\to} \xi$*, if for any compact subsets*[6] $R_1, ..., R_k$ *of* $\mathbb{E}$ *with open interiors* $R_1^o, ..., R_m^o$ *and any finite m-tuple of real numbers* $\tau_1, ..., \tau_m$*,*

$$\begin{aligned} & P\left(\inf_{x \in R_1} \xi(x) > \tau_1, ..., \inf_{x \in R_m} \xi(x) > \tau_m\right) \\ \leq & \liminf_{n \to \infty} P\left(\inf_{x \in R_1} \xi_n(x) > \tau_1, ..., \inf_{x \in R_m} \xi_n(x) > \tau_m\right) \quad (3.2.3) \\ \leq & \limsup_{n \to \infty} P\left(\inf_{x \in R_1^o} \xi_n(x) \geq \tau_1, ..., \inf_{x \in R_m^o} \xi_n(x) \geq \tau_m\right) \\ \leq & P\left(\inf_{x \in R_1^o} \xi(x) \geq \tau_1, ..., \inf_{x \in R_m^o} \xi(x) \geq \tau_m\right). \quad (3.2.4) \end{aligned}$$

We call the condition given by (3.2.3) the *lower epilimit condition*. Similarly, we call that given by (3.2.4) the *upper epilimit condition*[7].

[6] In this definition, the sets $R_1, ..., R_k$ can instead be taken from a class of relatively compact sets $\mathcal{V}$ such that (i) $\mathcal{V}$ is closed under finite union and intersection; (ii) each compact set K in $\mathbb{E}$ is representable as the intersection of a decreasing sequence in $\mathcal{V}$; and (iii) each open set G in $\mathbb{E}$ is representable as the union of an increasing sequence in $\mathcal{V}$. A typical example for such a $\mathcal{V}$ is a class of closed rectangles. See Pflug (1992) for details.

[7] These names are motivated by Proposition 7.29 in Rockafellar and Wets (2005).

Weak epiconvergence is generally useful for studying the limiting distribution of extremum estimators, especially when the criterion function assumes the value infinity, which often occurs in constrained optimization problems[8]. This concept is weaker than weak convergence (on compact sets) in the uniform metric (Pflug, 1995, Proposition 1) and is equivalent to finite dimensional weak convergence when the sequence $\{\xi_n\}$ satisfies a condition called "stochastic equi-lower-semicontinuity" (Knight, 1999, Theorem 2).

For our purposes, weak epiconvergence of a version of the criterion function ζ_n helps ensure the finite dimensional weak convergence of $\mathcal{Z}_n(\cdot,t)$. The desired results then follow by establishing tightness of $\{\mathcal{Z}_n(\cdot,t)\}$.

3.3 Inference Using the Normalized Support Function

In this section, we present our first main results. We begin by establishing the duality that relates the finite dimensional distribution of the normalized support function $\mathcal{Z}_n(\cdot,t)$ to that of the infimum of a localized criterion function $\tilde{\zeta}_n = a_n Q_n(\theta + \lambda/a_n^{1/\gamma})$ over a class of compact sets. We further show that $\mathcal{Z}_n(\cdot,t)$ converges weakly in the uniform metric to a stochastic process on $\mathbb{S}^{d-1}$ under appropriate regularity conditions on $\tilde{\zeta}_n$. We then present our inference methods using functionals of $\mathcal{Z}_n(\cdot,t)$.

3.3.1 Asymptotic Distribution of the Normalized Support Function

We first add a mild regularity condition on the criterion function. For this, we use the following definition.

[8]Details on weak epiconvergence can be found in Pflug (1992), Geyer (1994), Pflug (1995), Knight (1999), Geyer (2003), and Molchanov (2005), among others. Recent applications of weak epiconvergence in econometrics include Chernozhukov and Hong (2004), Chernozhukov (2005), and Han and Philips (2006).

Definition 3.3.1 (Lower semicontinuity): *The 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* $\bar{x} \in \mathbb{R}^d$.

If a function $f : \Omega \times \mathbb{R}^d \to \bar{\mathbb{R}}$ is such that $f(\omega, \cdot)$ is lsc for all $\omega \in F \in \mathfrak{F}$, $P(F) = 1$, then we say f is lower semicontinuous almost surely (lsc $a.s.$). A subtle problem for inference here is that $\hat{\Theta}_n(t)$ may be empty with positive probability in finite samples. To handle this, we use the following convention. We set $s(p, \hat{\Theta}_n(t)) = s(p, \hat{\Theta}_n(\underline{t}_n))$ if $\hat{\Theta}_n(t) = \varnothing$, where $\underline{t}_n := \inf_\Theta a_n Q_n(\theta)$. This convention ensures that $\mathcal{Z}_n(\cdot, t) \in \mathcal{C}(\mathbb{S}^{d-1})$ $a.s.$ Note that $P(\hat{\Theta}_n(t) = \varnothing) \to 0$ under the conditions of Theorem 3.2.1, so this adjustment becomes less and less likely as $n \to \infty$.

The following lemma establishes the duality between the minimization of the criterion function and the maximization of the corresponding inner product. This lemma provides a way to relate the stochastic behavior of the support function $s(\cdot, \hat{\Theta}_n(t))$ to that of the original criterion function $\zeta_n(\cdot) = a_n Q_n(\cdot)$.

Lemma 3.3.1 (Duality 1): *Suppose that Assumption 3.2.1 holds. Let* $n \in \mathbb{N}$ *and* $t \in \mathbb{R}_+$ *be given. Suppose* ζ_n *is lsc a.s. Then, for any* $u \in \mathbb{R}$ *and* $p \in \mathbb{S}^{d-1}$

$$s(p, \hat{\Theta}_n(t)) < u \quad \Leftrightarrow \quad \inf_{\theta \in K_{u,p} \cap \Theta} \zeta_n(\theta) > t,$$

with probability 1, where $K_{u,p}$ *is the half space*

$$K_{u,p} := \{\theta \in \mathbb{R}^d : \langle p, \theta \rangle \geq u\}.$$

By this lemma, we can relate the support function of the level set estimator to the criterion function[9]. Our goal is then to relate the normalized support function $\mathcal{Z}_n(p, t)$ to a localized version of the criterion function.

We define a process $\tilde{\zeta}_n$ whose behavior captures that of ζ_n for local deviations from the boundary points of Θ_I. For this, let $\partial\Theta_I$ be the boundary of Θ_I; this coincides with the collection of support points of Θ_I: i.e., $\partial\Theta_I := \{\theta : \theta \in H(p, \Theta_I),$

[9]Note that if $\hat{\Theta}_n(t) = \varnothing$, we take $s(p, \hat{\Theta}_n(t)) = \sup_{\theta \in \varnothing} \langle p, \theta \rangle = -\infty$.

$p \in \mathbb{S}^{d-1}\}$. Define a stochastic process $\tilde{\zeta}_n$ on $\partial\Theta_I \times \mathbb{R}^d$ by

$$\tilde{\zeta}_n(\theta, \lambda) := \zeta_n(\theta + \lambda / a_n^{1/\gamma}), \qquad \theta \in \partial\Theta_I, \lambda \in \mathbb{R}^d.$$

The quantity $\theta + \lambda/a_n^{1/\gamma}$ represents a deviation of order $a_n^{-1/\gamma}$ in the direction λ from θ.

To apply the previous lemma, we first note that

$$\{\omega : \mathcal{Z}_n(p,t) < u\} = \left\{\omega : s(p, \hat{\Theta}_n(t)) < s(p, \Theta_I) + u/a_n^{1/\gamma}\right\}.$$

Applying Lemma 3.3.1 to $s(p, \hat{\Theta}_n(t)) < s(p, \Theta_I) + u/a_n^{1/\gamma}$ yields

$$\{\omega : \mathcal{Z}_n(p,t) < u\} = \left\{\omega : \inf_{\tilde{\theta} \in K_{s(p,\Theta_I)+u/a_n^{1/\gamma},p} \cap \Theta} \zeta_n(\tilde{\theta}) > t\right\}. \tag{3.3.1}$$

Next, for each $\tilde{\theta} \in K_{s(p,\Theta_I)+u/a_n^{1/\gamma},p} \cap \Theta$, we decompose $\tilde{\theta}$ by letting $\tilde{\theta} = \theta + (\tilde{\theta} - \theta)$ where $\theta \in H(p, \Theta_I)$[10]. We define $\lambda := a_n^{1/\gamma}(\tilde{\theta} - \theta)$. We can therefore write

$$\tilde{\theta} = \theta + \lambda/a_n^{1/\gamma}.$$

The motivation for rescaling λ by $1/a_n^{1/\gamma}$ is that $u/a_n^{1/\gamma}$ appears in the subscript of K in eq. (3.3.1). By this decomposition, θ represents the part of $\tilde{\theta}$ that gives the inner product value $s(p, \Theta_I)$, and $\lambda/a_n^{1/\gamma}$ represents the part of $\tilde{\theta}$ that gives an inner product value greater than or equal to $u/a_n^{1/\gamma}$. This decomposition is illustrated in figure 3.2.

It is easy to show that λ satisfies $\langle p, \lambda \rangle \geq u$, so $\lambda \in K_{u,p}$. In addition, since $\lambda = a_n^{1/\gamma}(\tilde{\theta} - \theta)$ with $\tilde{\theta} \in \Theta$, λ belongs to a shifted and rescaled space $a_n^{1/\gamma}(\Theta - \theta)$ $:= \{\lambda \in \mathbb{R}^d : \lambda = a_n^{1/\gamma}(\tilde{\theta} - \theta), \tilde{\theta} \in \Theta\}$. Thus, $\lambda \in K_{u,p} \cap a_n^{1/\gamma}(\Theta - \theta)$. Using this

[10]Note that θ here is not necessarily unique; however, this has no impact on the arguments to follow.

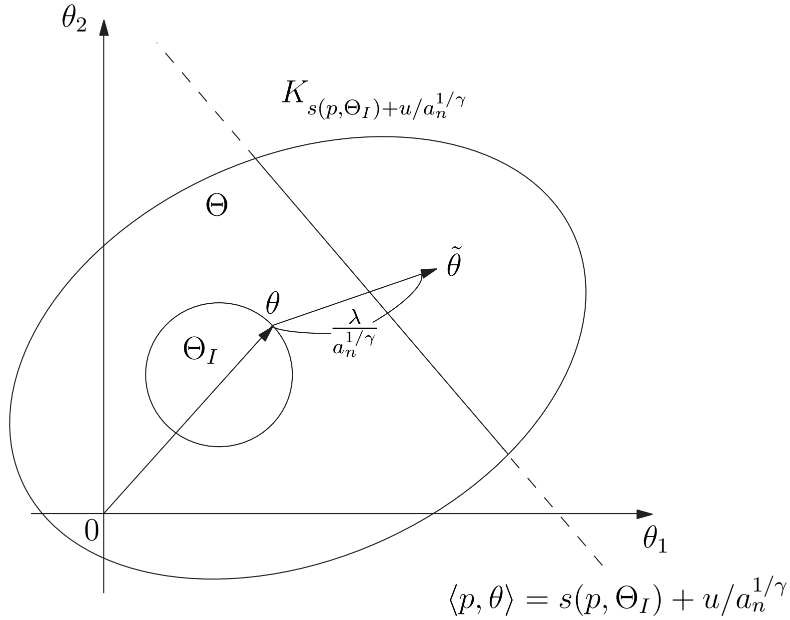

Figure 3.2: Decomposition of $\tilde{\theta}$

decomposition, we can rewrite the event in eq. (3.3.1) as

$$\{\omega : \mathcal{Z}_n(p,t) < u\} = \left\{\omega : \inf_{\theta \in H(p,\Theta_I)} \inf_{\lambda \in K_{u,p} \cap [a_n^{1/\gamma}(\Theta - \theta)]} \tilde{\zeta}_n(\theta, \lambda) > t\right\}.$$

Let $r_{n,u,p}$ be the correspondence defined on $H(p, \Theta_I)$ by

$$r_{n,u,p}(\theta) := K_{u,p} \cap [a_n^{1/\gamma}(\Theta - \theta)], \quad n = 1, 2, \dots .$$

For each $\theta \in H(p, \Theta_I)$, the set $K_{u,p} \cap [a_n^{1/\gamma}(\Theta - \theta)]$ is an image of $r_{n,u,p}(\theta)$. Figures 3 and 4 in Appendix B illustrate how this image changes when θ moves along $H(p, \Theta_I)$ for a fixed n[11]. The graph of this correspondence is

$$R_{n,u,p} := \{(\theta, \lambda) : \lambda \in r_{n,u,p}(\theta), \theta \in H(p, \Theta_I)\},$$

[11] Appendix B is available from `http://econ.ucsd.edu/~hkaido/pdf/supmat.pdf`.

which is illustrated in figure 5. Thus,

$$\{\omega : \mathcal{Z}_n(p,t) < u\} = \left\{\omega : \inf_{(\theta,\lambda)\in R_{n,u,p}} \tilde{\zeta}_n(\theta,\lambda) > t\right\}.$$

To analyze this event, it is important to understand the behavior of $R_{n,u,p}$ as n increases.

For this, let $\partial\Theta = \Theta \setminus \Theta^o$ denote the boundary of Θ. We call elements of $\Theta_I \cap \partial\Theta$ *identified parameters on the boundary* (*of* Θ). The remaining elements of Θ_I are *identified parameters in the interior* (*of* Θ). How $R_{n,u,p}$ behaves in the limit depends on whether or not there is an identified parameter on the boundary of Θ.

Specifically, if, as Assumptions 3.2.1 and 3.2.2 (*ii*) ensure, there are no identified parameters on the boundary, then $a_n^{1/\gamma}(\Theta-\theta)$ converges to $\mathbb{R}^d$ in the sense of Painlevé-Kuratowski (PK)[12]. Thus, for any (u,p), we have the PK convergences

$$K_{u,p} \cap a_n^{1/\gamma}(\Theta - \theta) \to K_{u,p} \quad \text{and}$$
$$R_{n,u,p} \to R_{u,p} := H(p,\Theta_I) \times K_{u,p}.$$

This case is depicted in figure 6.

On the other hand, if there is an identified parameter on the boundary of Θ, the limit of the sequence of graphs $\{R_{n,u,p}\}_{n=1}^{\infty}$ has a form that depends on the structure of Θ. In this case, the local parameter space may be approximated by a cone, following the ideas of Geyer (1994) and Andrews (1999). This case is definitely of interest, but in order to keep a tight focus here, we leave this for analysis elsewhere.

Our next result provides conditions ensuring that $R_{n,u,p}$ behaves in such a way that the infimum of the stochastic process $\tilde{\zeta}_n$ over $R_{n,u,p}$ is close to the infimum over $R_{u,p}$ in a stochastic sense when n is sufficiently large.

[12]For a sequence $\{C_n\}_{n\in\mathbb{N}}$ of subsets of $\mathbb{R}^d$, the *inner limit* is the set $\liminf_{n\to\infty} C_n := \{x : \exists\{x_n\}_{n\in\mathbb{N}}$ such that $x_n \to x$ and $x_n \in C_n, \forall n\}$ while the *outer limit* is the set $\limsup_{n\to\infty} C_n := \{x : \exists\{x_{n_k}\}_{k\in\mathbb{N}}$ such that $x_{n_k} \to x$ and $x_{n_k} \in C_{n_k}, \forall k\}$. The *limit* of the sequence exists if inner and outer limit sets are equal: $\lim_{n\to\infty} C_n = \liminf_{n\to\infty} C_n = \limsup_{n\to\infty} C_n$. When $\lim_{n\to\infty} C_n$ exists and equal to a set C, the sequence $\{C_n\}_{n\in\mathbb{N}}$ is said to converge to C in the *Painlevé-Kuratowski* sense. See Rockafellar and Wets (2005, ch.4) for details.

LEMMA 3.3.2: *Suppose Assumptions 3.2.1 and 3.2.2 hold. Suppose that* $\tilde{\zeta}_n$ *is lsc a.s. and that there exists* $\bar{\epsilon} > 0$ *such that for any* $0 < \epsilon < \bar{\epsilon}$,

$$\liminf_{n\to\infty} \left\{ (\theta, \lambda) \in R_{u,p} : \tilde{\zeta}_n(\theta, \lambda) < \inf_{R_{u,p}} \tilde{\zeta}_n(\theta, \lambda) + \epsilon \right\} \neq \varnothing, \tag{3.3.2}$$

almost surely. Then for any $0 < \epsilon < \bar{\epsilon}$ *there exists a finite integer* N_ϵ *such that for all* $(u, p) \in \mathbb{R} \times \mathbb{S}^{d-1}$

$$P\left(\left| \inf_{(\theta,\lambda)\in R_{n,u,p}} \tilde{\zeta}_n(\theta, \lambda) - \inf_{(\theta,\lambda)\in R_{u,p}} \tilde{\zeta}_n(\theta, \lambda) \right| \geq \epsilon \right) \leq \epsilon, \quad \forall n \geq N_\epsilon.$$

Since the function $\zeta_n = a_n Q_n$ is defined for all $\theta \in \mathbb{R}^d$, the infima above are well defined. When condition (3.3.2) holds, we say that $\{\tilde{\zeta}_n\}$ obeys the *nonempty limit* ϵ*-argmin condition*. This requires that the sequence $\{\tilde{\zeta}_n\}$ stabilizes in such a way that its ϵ-argmin set does not keep moving around. Properness ensures that the difference of the infima in the conclusion is not of the form $\infty - \infty$. This conclusion is an analog of Condition S.1 assumed by CHT, motivated by results of Chernoff (1954) and Andrews (1999).

In order to apply weak epiconvergence to $\tilde{\zeta}_n$, we need to control the limiting behavior of the finite-dimensional distributions of the infima of $\tilde{\zeta}_n$ over a family of compact sets. As $R_{u,p}$ is a closed but unbounded set, we need to replace it with a compact set. As Salinetti and Wets (1986) and Molchanov (2005) show, this can be done under a regularity condition known as *equi-inf-compactness*, defined as follows.

DEFINITION 3.3.2 (Equi-inf-compactness): *The sequence of stochastic processes* $\{\xi_n\}$ *is* equi-inf-compact *if for every* $\alpha \in \mathbb{R}$ *there exists a compact set* L_α *such that* $\{x : \xi_n(x) \leq \alpha\} \subset L_\alpha$ *a.s. for all* $n \geq 1$.

If this condition holds for $\{\tilde{\zeta}_n\}$, we can approximate the limit of the infima of $\{\tilde{\zeta}_n\}$ over the closed unbounded set $R_{u,p}$ by the infimum over a compact set $\tilde{R}_{u,p} := R_{u,p} \cap L_{u,p}$ with $L_{u,p}$ properly chosen. Then we can apply weak epiconvergence by checking the limiting behavior of the infima of $\tilde{\zeta}_n$ over compact sets $\{\tilde{R}_{u_j,p_j}, j = 1, 2, ..., m\}$.

We now state a second duality result, relating $\mathcal{Z}_n$ and $\tilde{\zeta}_n$.

LEMMA 3.3.3 (Duality 2): *Suppose that Assumptions 3.2.1 and 3.2.2 hold. Let $t \in \mathbb{R}_+$ be given. Suppose that $\{\tilde{\zeta}_n\}$ obeys the nonempty limit ϵ-argmin condition, that $\{\tilde{\zeta}_n\}$ is equi-inf-compact, and that $\tilde{\zeta}_n$ is lsc a.s. for all n sufficiently large. Then, for any finite m-tuple $\{(u_j, p_j) \in \mathbb{R} \times \mathbb{S}^{d-1}\}_{j=1}^m$, there exist compact sets $L_{u_j,p_j}, j = 1, ..., m$, such that, with $\tilde{R}_{u,p} := R_{u,p} \cap L_{u,p}$,*

$$\liminf_{n\to\infty} P(\mathcal{Z}_n(p_1, t) < u_1, ..., \mathcal{Z}_n(p_m, t) < u_m)$$
$$\geq \liminf_{n\to\infty} P\left(\inf_{(\theta,\lambda)\in\tilde{R}_{u_1,p_1}} \tilde{\zeta}_n(\theta, \lambda) > t, ..., \inf_{(\theta,\lambda)\in\tilde{R}_{u_m,p_m}} \tilde{\zeta}_n(\theta, \lambda) > t\right).$$

This lemma ensures that, to study the (finite-dimensional) asymptotic behavior of $\mathcal{Z}_n(\cdot, t)$, it suffices to study the asymptotic behavior of the infima of $\tilde{\zeta}_n$ over compact sets. The right hand side of this inequality can be controlled if $\tilde{\zeta}_n$ weakly epiconverges to a known limiting process $\tilde{\zeta}$. If so, we can seek a process $\mathcal{Z}$ such that

$$P(\mathcal{Z}(p_1, t) < u_1, ..., \mathcal{Z}(p_m, t) < u_m)$$
$$= P\left(\inf_{(\theta,\lambda)\in\tilde{R}_{u_1,p_1}} \tilde{\zeta}(\theta, \lambda) > t, ..., \inf_{(\theta,\lambda)\in\tilde{R}_{u_m,p_m}} \tilde{\zeta}(\theta, \lambda) > t\right).$$

The portmanteu theorem then implies $\mathcal{Z}_n(\cdot, t) \stackrel{f.d.}{\to} \mathcal{Z}(\cdot, t)$.

The next theorem establishes this; it further gives the asymptotic distributions of the Hausdorff distances. For this, we formally impose sufficient regularity on $\{\tilde{\zeta}_n\}$.

ASSUMPTION 3.3.1 (Local Process Regularity): *(i) For all n sufficiently large, $\tilde{\zeta}_n$ is, almost surely, lsc, and Q_n is convex in a neighborhood of Θ_I. (ii) The sequence $\{\tilde{\zeta}_n\}$ obeys the nonempty limit ϵ-argmin condition, is equi-inf-compact, and weakly epiconverges to a stochastic process $\tilde{\zeta}$.*

Assumption 3.3.1 (ii) is stronger than strictly necessary. It appears that weak epiconvergence can be replaced by the lower epilimit condition without affecting our conclusions.

THEOREM 3.3.1: *Suppose that Assumptions 3.2.1, 3.2.2, 3.2.3, 3.2.4, and 3.3.1 hold. For each $t \in \mathbb{R}_+$ and $\theta \in \partial\Theta_I$, let $\hat{\Lambda}(t,\theta)$ be a random level set of the map $\lambda \mapsto \tilde{\zeta}(\theta,\lambda)$ defined by*

$$\hat{\Lambda}(t,\theta) = \{\lambda : \tilde{\zeta}(\theta,\lambda) \leq t\}.$$

Suppose that the limiting process $\tilde{\zeta}$ is such that $\hat{\Lambda}(t,\theta)$ is nonempty a.s. for each $t \in \mathbb{R}_+$ and $\theta \in \partial\Theta_I$.

Then for each $t \in \mathbb{R}_+$,

(i) $\mathcal{Z}_n(\cdot,t) \overset{f.d.}{\to} \mathcal{Z}(\cdot,t)$, where $\mathcal{Z}(\cdot,t)$ is a stochastic process on $\mathbb{S}^{d-1}$, which has the representation

$$\mathcal{Z}(p,t) = \sup_{\theta \in H(p,\Theta_I)} s\left(p, \hat{\Lambda}(t,\theta)\right); \tag{3.3.3}$$

(ii) letting m be a finite integer and $\{(u_j,p_j) \in \mathbb{R} \times \mathbb{S}^{d-1}\}_{j=1}^m$ an m-tuple, the finite dimensional distributions of $\mathcal{Z}(\cdot,t)$ satisfy

$$\begin{aligned} &P(\mathcal{Z}(p_1,t) < u_1, ..., \mathcal{Z}(p_m,t) < u_m) \\ &\qquad = P\left(\inf_{(\theta,\lambda)\in\tilde{R}_{u_1,p_1}} \tilde{\zeta}(\theta,\lambda) > t, ..., \inf_{(\theta,\lambda)\in\tilde{R}_{u_m,p_m}} \tilde{\zeta}(\theta,\lambda) > t\right); \end{aligned}$$

(iii) $\mathcal{Z}_n(\cdot,t) \overset{u.d.}{\to} \mathcal{Z}(\cdot,t)$, ensuring that

$$\begin{aligned} &a_n^{1/\gamma} d_H(\hat{\Theta}_n(t),\Theta_I) \overset{d}{\to} \|\mathcal{Z}(\cdot,t)\|_{\mathcal{C}(\mathbb{S}^{d-1})} \\ &a_n^{1/\gamma} \vec{d}_H(\Theta_I,\hat{\Theta}_n(t)) \overset{d}{\to} \sup_{p\in\mathbb{S}^{d-1}} \{-\mathcal{Z}(p,t)\}_+ \quad \text{and} \\ &a_n^{1/\gamma} \vec{d}_H(\hat{\Theta}_n(t),\Theta_I) \overset{d}{\to} \sup_{p\in\mathbb{S}^{d-1}} \{\mathcal{Z}(p,t)\}_+ . \end{aligned}$$

3.3.2 Inference for the Identified Set

Using the asymptotic distribution results of the previous section, we now study hypothesis testing on Θ_I, together with confidence collections and confidence sets constructed by inverting the Wald statistics. To the best of our knowledge, Beresteanu and Molinari (2008) is the first article discussing statistical inference and confidence collections based on the Hausdorff distance measures. Our results below are mostly parallel to the results presented in sections 2.2 and 2.3 of their paper.

We first study hypothesis testing, confidence collections, and confidence sets based on the Hausdorff metric. Let $\Theta_0 \in \mathcal{K}_c$ be a given compact convex set, and consider testing

$$H_0 : \Theta_I = \Theta_0, \quad vs. \quad H_1 : \Theta_I \neq \Theta_0. \tag{3.3.4}$$

Recall that $d_H(\Theta_I, \Theta_0) = 0$ if and only if $\Theta_I = \Theta_0$. A natural statistic for the test, therefore, is the scaled Hausdorff metric $T_n(t) = a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_0)$. Under the null hypothesis, the statistic has the limiting distribution $\|\mathcal{Z}(\cdot, t)\|_{\mathcal{C}(\mathbb{S}^{d-1})}$ by Theorem 3.3.1; under the alternative, it diverges to ∞ with probability approaching one because the statistic is not properly centered. Let $\alpha \in (0, 1)$ be a significance level. We obtain a test of asymptotic level α by rejecting the null hypothesis when $T_n(t)$ exceeds the asymptotic critical value

$$c_{1-\alpha}(t) := \inf \left\{ x : P\left(\|\mathcal{Z}(\cdot, t)\|_{\mathcal{C}(\mathbb{S}^{d-1})} \leq x \right) \geq 1 - \alpha \right\},$$

where $\mathcal{Z}(\cdot, t)$ is the stochastic process given in Theorem 3.3.1.

It is often difficult to compute this critical value directly, as the required asymptotic distribution differs from case to case. Specifically, the properties of $\mathcal{Z}(\cdot, t)$ depend on the weak epilimit $\tilde{\zeta}$ and therefore on the functional form of the criterion function. Also, the distribution of $\mathcal{Z}(\cdot, t)$ depends on the characteristics of the true identified set Θ_I. For some special cases, it might be possible to simulate the asymptotic distribution of the relevant process to obtain the critical value, but this approach is not generally applicable.

As a practical alternative, we now propose a straightforward subsampling procedure that yields generally valid asymptotic critical values under the high-level assumptions provided above and mild regularity conditions on the rate at which the subsample size grows. For concreteness, we present a procedure for the important class of cases in which the sample criterion function Q_n is constructed from a sample $\{X_i : \Omega \to \mathbb{R}^k\}_{i=1}^n$ of IID random vectors.

ASSUMPTION 3.3.2: *Let Assumption 3.2.1 hold with* $Q_n(\omega, \theta) = \tilde{Q}_n(X_1(\omega), \cdots, X_n(\omega), \theta)$ *where* $\tilde{Q}_n : \prod_{i=1}^n \mathbb{R}^k \times \mathbb{R}^d \to \bar{\mathbb{R}}_+$ *is jointly measurable,* $n = 1, 2, \cdots$, *and* $\{X_i\}$ *is an IID sequence of random k-vectors,* $k \in \mathbb{N}$.

It is straightforward to extend our results to a sample of stationary and strong mixing time series. See Politis, Romano, and Wolf (1999, Ch 3) for details.

ALGORITHM 3.3.1 (Subsampling for level-set estimators): *Let* $t > 0$ *and* $0 < \alpha < 1$ *be given. 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.

Step 1. *For* $k = 1, \cdots, N_{n,b}$, *construct* $\hat{\Theta}_{n,b,k}(t)$, *the set estimator for the k-th subsample, computed as a t-level set of the criterion function* $\zeta_{n,b,k}(X_{k_1}, \cdots, X_{k_b}, \theta) = a_b \tilde{Q}_{n,b,k}(X_{k_1}, \cdots, X_{k_b}, \theta)$, *with the obvious notation.*

Step 2. *For* $k = 1, \cdots, N_{n,b}$, *compute*

$$\hat{T}_{n,b,k}(t) = a_b^{1/\gamma} d_H \left(\hat{\Theta}_{n,b,k}(t), \hat{\Theta}_n(t) \right).$$

Step 3. *Compute the 100×(1−α)% quantile of the subsampling distribution, given by*

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

where

$$\hat{F}_{n,b}(x,t) := N_{n,b}^{-1} \sum_{1 \leq k \leq N_{n,b}} 1_{\{\hat{T}_{n,b,k}(t) \leq x\}}.$$

For any t, let $F(x,t) := P[\|\mathcal{Z}(\cdot,t)\|_{\mathcal{C}(\mathbb{S}^{d-1})} \leq x]$ define the cumulative distribution function (CDF) of $\|\mathcal{Z}(\cdot,t)\|_{\mathcal{C}(\mathbb{S}^{d-1})}$. The next theorem is a basic result for subsampling the Hausdorff metric of level set estimators.

THEOREM 3.3.2: *Suppose the conditions of Theorem 3.3.1 and Assumption 3.3.2 hold. Suppose that* $b \to \infty$ *and* $b/n \to 0$ *as* $n \to \infty$. *Let* $\hat{F}_{n,b}(\cdot, t)$ *and*

$\hat{c}_{n,b,1-\alpha}(t)$ be computed by Algorithm 3.3.1.

(i) If x is a continuity point of $F(\cdot,t)$, then $\hat{F}_{n,b}(x,t) \to F(x,t)$ in probability;

(ii) If $F(\cdot,t)$ is continuous, then $\sup_x |\hat{F}_{n,b}(x,t) - F(x,t)| \to 0$ in probability;

(iii) If $F(\cdot,t)$ is continuous at $c_{1-\alpha}(t)$, then

$$\lim_{n\to\infty} P\left(a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \leq \hat{c}_{n,b,1-\alpha}(t)\right) = 1 - \alpha.$$

As a corollary of this result, we can show that the test has the correct level and is consistent against any fixed alternative hypothesis.

Corollary 3.3.1: *Suppose the conditions of Theorem 3.3.2 hold. Let Θ_0 be a nonempty compact convex subset of Θ^o.*

(i) If $\Theta_0 = \Theta_I$ and $F(\cdot,t)$ is continuous and strictly increasing at $c_{1-\alpha}(t)$, then $\hat{c}_{n,b,1-\alpha}(t) = c_{1-\alpha}(t) + o_p(1)$, and the test has asymptotic rejection probability α:

$$\lim_{n\to\infty} P\left(T_n(t) > \hat{c}_{n,b,1-\alpha}(t)\right) = \alpha.$$

(ii) If $\Theta_0 \neq \Theta_I$, then the test is consistent:

$$\lim_{n\to\infty} P\left(T_n(t) > \hat{c}_{n,b,1-\alpha}(t)\right) = 1.$$

When $N_{n,b}$ is large, we can instead employ a stochastic approximation to $\hat{F}_{n,b}(\cdot,t)$ by randomly drawing subsamples, with or without replacement. See Politis, Romano, and Wolf (1999, Sec. 2.4) for details.

The confidence collection is the collection of compact convex sets such that our test does not reject the null hypothesis in (3.3.4) when any set in the collection is taken to be Θ_0. This collection can be obtained by inverting our test statistic:

$$\hat{\mathcal{X}}_{n,b,1-\alpha}(t) = \left\{\Psi \in \mathcal{K}_c : a_n^{1/\gamma} d_H\left(\hat{\Theta}_n(t), \Psi\right) \leq \hat{c}_{n,b,1-\alpha}(t)\right\}.$$

We show next that, by the duality of the test, this collection includes Θ_I with probability $1-\alpha$ asymptotically.

To construct a confidence set, consider the union of the elements of $\hat{\mathcal{X}}_{n,b,1-\alpha}(t)$,

$$\hat{\Psi}_{n,b,1-\alpha}(t) := \bigcup \{\Psi : \Psi \in \hat{\mathcal{X}}_{n,b,1-\alpha}(t)\}.$$

We show that this can be easily computed by expanding our level set estimator by an amount $\hat{\epsilon}_{n,b,1-\alpha}(t) := \hat{c}_{n,b,1-\alpha}(t)/a_n^{1/\gamma}$. This yields a confidence set whose precise level is difficult to determine but which is bounded below by $1-\alpha$ and is thus conservative.

THEOREM 3.3.3: *Suppose the conditions of Theorem 3.3.2 hold. Suppose $F(\cdot,t)$ is continuous at $c_{1-\alpha}(t)$. Then*

(i) $\lim_{n\to\infty} P\left(\Theta_I \in \hat{\mathcal{X}}_{n,b,1-\alpha}(t)\right) = 1-\alpha$*;*

(ii) $\hat{\Psi}_{n,b,1-\alpha}(t) = \hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}}(t)$*, where* $\hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}}(t)$ *is a closed* $\hat{\epsilon}_{n,b,1-\alpha}(t)$*-envelope of* $\hat{\Theta}_n(t)$ *given by*

$$\hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}}(t) := \{\theta : d_H(\theta, \hat{\Theta}_n(t)) \leq \hat{\epsilon}_{n,b,1-\alpha}(t)\};$$

(iii) Further,

$$\lim_{n\to\infty} P\left(\Theta_I \subseteq \hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}(t)}(t)\right) \geq 1-\alpha.$$

Next, we consider hypothesis testing, confidence collections, and confidence sets based on the directed Hausdorff distance. Again, let $\Theta_0 \in \mathcal{K}_c$ be a given compact convex set. We consider testing

$$H_0 : \Theta_0 \subseteq \Theta_I \quad vs. \quad H_1 : \Theta_0 \nsubseteq \Theta_I. \tag{3.3.5}$$

Recall that $\vec{d}_H(\Theta_0, \Theta_I) = 0$ if and only if $\Theta_0 \subseteq \Theta_I$. We therefore test this hypothesis using the scaled directed Hausdorff distance $T_n^{\rightarrow}(t) := a_n^{1/\gamma} \vec{d}_H(\Theta_0, \hat{\Theta}_n(t))$. By focusing on the inclusion relationship above, we can directly compare our results to those of BM. We can also test the reverse inclusion using $a_n^{1/\gamma} \vec{d}_H(\hat{\Theta}_n(t), \Theta_0)$ (the scaled upper Hausdorff hemimetric) as described below.

Following BM, we use the triangle inequality

$$\vec{d}_H(\Theta_0, \hat{\Theta}_n(t)) \leq \vec{d}_H(\Theta_0, \Theta_I) + \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)). \tag{3.3.6}$$

The first term on the right vanishes under the null. Under the assumptions of Theorem 3.1, the second term has a well-defined limiting distribution when scaled by $a_n^{1/\gamma}$. We now specify a subsampling algorithm similar to Algorithm 3.3.1 to approximate this distribution.

ALGORITHM 3.3.2: *Implement Algorithm 3.3.1 but with $\vec{d}_H(\hat{\Theta}_n(t), \hat{\Theta}_{n,b,k}(t))$ replacing $d_H(\hat{\Theta}_{n,b,k}(t),\ \hat{\Theta}_n(t))$.*

The analogs of $\hat{T}_{n,b,k}(t)$, $F_n(x,t)$, $\hat{c}_{n,b,1-\alpha}(t)$, and $\hat{F}_{n,b}(t)$ are denoted $\hat{T}^{\rightarrow}_{n,b,k}(t)$, $F^{\rightarrow}_n(x,t)$, $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)$, and $\hat{F}^{\rightarrow}_{n,b}(t)$. Similarly, let $F^{\rightarrow}(x,t) := P(\sup_{p\in\mathbb{S}^{d-1}} \{-\mathcal{Z}(p,t)\}_+ \leq x)$ define the CDF of $\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+$. Also, let $c^{\rightarrow}_{1-\alpha}(t) := \inf\{x : P(\ \sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+ \leq x) \geq 1-\alpha\}$ define the asymptotic critical value for level $\alpha \in (0,1)$. The next theorem establishes the validity of subsampling for the directed Hausdorff distance.

THEOREM 3.3.4: *Suppose the conditions of Theorem 3.3.1 and Assumption 3.3.2 hold. Suppose that $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. Let $\hat{F}^{\rightarrow}_{n,b}(\cdot,t)$ and $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)$ be computed by Algorithm 3.3.2.*

(i) If x is a continuity point of $F^{\rightarrow}(\cdot,t)$, then $\hat{F}^{\rightarrow}_{n,b}(x,t) \to F^{\rightarrow}(x,t)$ in probability;

(ii) If $F^{\rightarrow}(\cdot,t)$ is continuous except at $x=0$, then for any $\epsilon > 0$, $\sup_{|x|\geq\epsilon} |\hat{F}^{\rightarrow}_{n,b}(x,t) - F^{\rightarrow}(x,t)| \to 0$ in probability;

(iii) If $F^{\rightarrow}(\cdot,t)$ is continuous at $c^{\rightarrow}_{1-\alpha}(t)$, then

$$\lim_{n\to\infty} P\left(a_n^{1/\gamma}\vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) \leq \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right) = 1-\alpha.$$

The directed Hausdorff distance has a discontinuity at $x = 0$. The consistency result (ii) is, therefore, weaker than the uniform convergence of subsampling CDFs over the whole real line. It establishes the uniform convergence of the subsampling CDF on compact sets excluding 0. As Bugni (2008) discusses, this weaker consistency result is sufficient for the purpose of hypothesis testing and constructing confidence sets as we are often interested in approximating the $90, 95$, and 99 percentiles. As we will discuss in Section 3.3.3, it is possible to choose t so that the discontinuity does not occur at the quantiles of interest.

Results for the rejection probability and the consistency against fixed alternatives now follow as before.

COROLLARY 3.3.2: *Suppose the conditions of Theorem 3.3.4 hold. Let* Θ_0 *be a nonempty compact convex subset of* Θ^o.

(i) If $\Theta_0 \subseteq \Theta_I$ *and* $F^{\rightarrow}(\cdot, t)$ *is continuous and strictly increasing at* $c^{\rightarrow}_{1-\alpha}(t)$, *then* $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t) = c^{\rightarrow}_{1-\alpha}(t) + o_p(1)$, *and the test has asymptotic rejection probability bounded above by* α*:*

$$\lim_{n\to\infty} P\left(T^{\rightarrow}_n(t) > \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right) \leq \alpha;$$

(ii) If $\Theta_0 \nsubseteq \Theta_I$, *then the test is consistent:*

$$\lim_{n\to\infty} P\left(T^{\rightarrow}_n(t) > \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right) = 1.$$

To construct a confidence collection, invert the test statistic to obtain

$$\hat{\mathcal{X}}^{\rightarrow}_{n,b,1-\alpha}(t) = \left\{\Psi \in \mathcal{K}_c : a_n^{1/\gamma} \vec{d}_H\left(\Psi, \hat{\Theta}_n(t)\right) \leq \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right\}.$$

As we show, this gives a conservative confidence collection. To construct a confidence set, consider the union of the elements of $\hat{\mathcal{X}}^{\rightarrow}_{n,b,1-\alpha}(t)$,

$$\hat{\Psi}^{\rightarrow}_{n,b,1-\alpha}(t) := \bigcup\{\Psi : \Psi \in \hat{\mathcal{X}}^{\rightarrow}_{n,b,1-\alpha}(t)\}.$$

Analogous to our previous result, we show that this can be easily computed by expanding our level set estimator by an amount $\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t) := \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)/a_n^{1/\gamma}$. In contrast to our previous result, this confidence set has asymptotic level $1-\alpha$. This is a Wald-type confidence set that is directly comparable to the QLR-type confidence set studied by CHT.

THEOREM 3.3.5: *Suppose the conditions of Theorem 3.3.4 hold. Suppose*

$F^{\rightarrow}(\cdot, t)$ is continuous at $c^{\rightarrow}_{1-\alpha}(t)$. Then

(i) For each $\Theta_0 \subseteq \Theta_I$, $\lim_{n\to\infty} P\left(\Theta_0 \in \hat{\mathcal{X}}^{\rightarrow}_{n,b,1-\alpha}(t)\right) \geq 1-\alpha$ with equality when $\Theta_0 = \Theta_I$;

(ii) $\hat{\Psi}^{\rightarrow}_{n,b,1-\alpha}(t) = \hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}}(t)$, where $\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}}(t)$ is a closed $\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)$-envelope of $\hat{\Theta}_n(t)$ given by

$$\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}}(t) := \left\{\theta : d(\theta, \hat{\Theta}_n(t)) \leq \hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)\right\};$$

(iii) Furthermore, for t small enough,

$$\lim_{n\to\infty} P\left(\Theta_I \subseteq \hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}}(t)\right) = 1-\alpha.$$

Note that for the confidence set to achieve the coverage probability $1-\alpha$, we must set t small enough. We will discuss how to choose t in the next subsection.

Results for testing the reverse inclusion

$$H_0 : \Theta_I \subseteq \Theta_0 \quad vs. \quad H_1 : \Theta_I \nsubseteq \Theta_0. \tag{3.3.7}$$

follow similarly, based on the statistic $T_n^{\leftarrow}(t) := a_n^{1/\gamma} \vec{d}_H(\hat{\Theta}_n(t), \Theta_0)$. A subsampling algorithm that can be used to approximate the relevant limiting distribution is

Algorithm 3.3.3: *Implement Algorithm 3.3.1 but with $\vec{d}_H(\hat{\Theta}_{n,b,k}(t), \hat{\Theta}_n(t),)$ replacing $d_H(\hat{\Theta}_{n,b,k}(t), \hat{\Theta}_n(t))$.*

The analogs of $\hat{T}_{n,b,k}(t), F_n(x,t), \hat{c}_{n,b,1-\alpha}(t)$, and $\hat{F}_{n,b}(t)$ are denoted $\hat{T}^{\leftarrow}_{n,b,k}(t)$, $F_n^{\leftarrow}(x,t)$, $\hat{c}^{\leftarrow}_{n,b,1-\alpha}(t)$, and $\hat{F}^{\leftarrow}_{n,b}(t)$. Similarly, let $F^{\leftarrow}(x,t) := P(\sup_{p\in\mathbb{S}^{d-1}}\{\mathcal{Z}(p,t)\}_+ \leq x)$ define the CDF of $\sup_{p\in\mathbb{S}^{d-1}}\{\mathcal{Z}(p,t)\}_+$; and let $c^{\leftarrow}_{1-\alpha}(t) := \inf\{x : P(\sup_{p\in\mathbb{S}^{d-1}} \{\mathcal{Z}(p,t)\}_+ \leq x) \geq 1-\alpha\}$ define the asymptotic critical value for level $\alpha \in (0,1)$. Results analogous to Theorem 3.3.4, Corollary 3.3.2, and Theorem 3.3.5 now follow analogously.

3.3.3 Choice of Level

As we will see in section 3.4, we can often properly weight the criterion function so that the level t only affects the mean of the limiting process $\mathcal{Z}(p,t)$. In this case, we can re-center the process $\mathcal{Z}_n(p,t)$ by a known function $\mu(t)$ or a consistent estimator $\hat{\mu}_n(t)$, so that the choice of level becomes asymptotically irrelevant for inference.

Even if we do not have a known form for $\mu(t)$ nor a consistent estimator, it is possible to remove the arbitrariness in the choice of t. In this section, we show that, at least asymptotically, the choice of t does not matter for constructing confidence sets for Θ_I. The construction of the confidence set is based on Theorem 3.3.5.

For each $\alpha \in (0,1)$, let $t^*_{1-\alpha}$ be the smallest t such that $c^{\rightarrow}_{1-\alpha}(t) = 0$. That is,

$$t^*_{1-\alpha} := \inf\{t \in \mathbb{R}_+ : c^{\rightarrow}_{1-\alpha}(t) = 0\}.$$

We will show that, for any $0 \leq t < t^*_{1-\alpha}$, confidence sets constructed in the manner of Theorem 3.3.5 are asymptotically equivalent to each other, in the sense that their difference (in the Hausdorff metric) is of stochastic order smaller than $a_n^{1/\gamma}$. In this sense, the initial choice of t does not matter for constructing the confidence set, given $t < t^*_{1-\alpha}$.

We start with the following lemma that shows $c^{\rightarrow}_{1-\alpha}$ is non-increasing on $[0, t^*_{1-\alpha}]$.

Lemma 3.3.4: *Suppose the conditions of Theorem 3.3.5 are satisfied. Then, for any* $0 \leq t < t' \leq t^*_{1-\alpha}$,

$$0 = c^{\rightarrow}_{1-\alpha}(t^*_{1-\alpha}) \leq c^{\rightarrow}_{1-\alpha}(t') \leq c^{\rightarrow}_{1-\alpha}(t) \leq c^{\rightarrow}_{1-\alpha}(0).$$

Recall that a confidence set $\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)}(t)$ is an expansion of the level set $\hat{\Theta}_n(t)$ by the amount $\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t) = \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)/a_n^{1/\gamma}$. Lemma 3.3.4 suggests that if we start with a large t, the amount we need to expand will be smaller, and at $t = t^*_{1-\alpha}$, we do not need to expand the set at all. The following theorem shows that, when the limiting process takes the form $\mathcal{Z}(p,t) = \mu(t) + \mathcal{Z}^*(p)$, this change in the amount of expansion makes all the confidence sets asymptotically equivalent, so that the initial choice of t is not essential as long as $t < t^*_{1-\alpha}$[13].

Theorem 3.3.6: *Suppose the conditions of Theorem 3.3.5 hold. Suppose that the limiting process takes the form $\mathcal{Z}(p,t) = \mu(t) + \mathcal{Z}^*(p)$ for each $(p,t) \in \mathbb{S}^{d-1} \times \mathbb{R}_+$ and $\mathcal{Z}_n(p,t) - \mathcal{Z}_n(p,t') = \mu(t) - \mu(t') + o_p(1)$ uniformly, where $\mu : \mathbb{R}_+ \to \mathbb{R}$ is an unknown function. Then for each $\alpha \in (0,1)$ and $0 \le t < t^*_{1-\alpha}$,*

$$d_H\left(\hat{\Theta}_n^{\overrightarrow{\hat{\epsilon}_{n,b,1-\alpha}}(t)}(t), \hat{\Theta}_n(t^*_{1-\alpha})\right) = o_p(a_n^{-1/\gamma}).$$

An immediate corollary is the following.

Corollary 3.3.3: *Suppose that the conditions of Theorem 3.3.6 hold. Then for each $\alpha \in (0,1)$ and for any $0 \le t \le t'^*_{1-\alpha}$,*

$$d_H\left(\hat{\Theta}_n^{\overrightarrow{\hat{\epsilon}_{n,b,1-\alpha}}(t)}(t), \hat{\Theta}_n^{\overrightarrow{\hat{\epsilon}_{n,b,1-\alpha}}(t')}(t'))\right) = o_p(a_n^{-1/\gamma}).$$

Theorem 3.3.6 raises an interesting research question. CHT construct a confidence set $\hat{\Theta}_n(\hat{\tau}_{n,b,1-\alpha})$ such that $\lim_{n\to\infty} P(\Theta_I \subseteq \hat{\Theta}_n(\hat{\tau}_{n,b,1-\alpha})) = 1-\alpha$, where $\hat{\tau}_{n,b,1-\alpha}$ is a subsampling estimate of the $1-\alpha$ quantile $\tau^*_{1-\alpha}$ of the limiting distribution of their QLR-statistic $\sup_{\Theta_I} a_n Q_n(\theta)$. If $t^*_{1-\alpha} = \tau^*_{1-\alpha}$ holds, the confidence sets based on the QLR-approach and our approach are asymptotically equivalent. The question is under what conditions the asymptotic equivalence holds. In section 3.4, we give a partial answer to this question. For models that involve finitely many moment inequalities, we will provide conditions on the criterion function and weighting matrix that ensure $t^*_{1-\alpha} = \tau^*_{1-\alpha}$.

Based on these results, we propose a generic algorithm to construct the confidence set.

Algorithm 3.3.4: *(Iterative Algorithm) Set $\kappa > 0$ small. Initialize $l = 1$,*

[13] The reason we cannot allow the equality $t = t^*_{1-\alpha}$ is because the subsampling fails to estimate the quantile at which the distribution is discontinuous.

and choose t_l small enough.

Step 1. *Construct the set estimator $\hat{\Theta}_n(t_l)$. Estimate the asymptotic $1-\alpha$ quantile $c^{\rightarrow}_{1-\alpha}(t_l)$ of the scaled directed Hausdorff distance $a_n^{1/\gamma}\vec{d}_H(\Theta_I, \hat{\Theta}_n(t_l))$ by Algorithm 3.3.2, obtainning $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t_1)$. Using $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t_l)$, expand $\hat{\Theta}_n(t_l)$ by $\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t_l) = \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t_l)/a_n^{1/\gamma}$ to obtain $\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t_l)}(t_l)$.*

Step 2. *Update the level by setting $t_{l+1} := \sup_{\theta\in\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t_l)}(t_l)} a_n Q_n(\theta)$.*

Step 3. *Repeat steps 1-2 until $|t_{l+1} - t_l| < \kappa$.*

The iterative algorithm can be proved to yield an increasing sequence $\{t_l, l = 1, 2, \cdots\}$ that tends to $t^*_{1-\alpha}$. As Theorem 3.3.6 shows, if the limiting process takes the form $\mathcal{Z}(p,t) = \mu(t) + \mathcal{Z}^*(p)$, one can stop at Step 1, as the iteration does not provide any first-order asymptotic improvement, although it may provide higher order refinements.

The iterative algorithm can be related to a multiple testing problem[14]. Romano and Shaikh (2009) is the first article that considered a step-wise procedure to construct a confidence set for Θ_I using CHT's QLR statistic. As we will show below, Algorithm 3.3.4 can be interpreted as a reduced form of a step-up procedure that controls the family-wise error rate of a multiple testing problem.

Consider the following family of hypotheses:

$$H_{\Theta_0} : \Theta_0 \subseteq \Theta_I, \quad \Theta_0 \in \mathcal{K}_c. \tag{3.3.8}$$

Each hypothesis in the family is indexed by a compact convex set Θ_0. For this multiple testing problem, we aim to control the following family-wise error rate (FWER):

$$FWER := P\left(\text{reject at least 1 hypothesis } H_{\Theta_0} \text{ s.t. } \Theta_0 \subseteq \Theta_I\right).$$

Below, we consider subcollections of $\mathcal{K}_c$. It is convenient to match each subcollection with a real number t. For this, we introduce a mapping L_n. For any $t \geq 0$,

[14] Details on the multiple testing problems can be found, for example, in Westfall and Young (1993) and Lehmann and Romano (2005, Ch.8).

consider a mapping L_n that assigns a subcollection, $L_n(t) = \{\Theta_0 : \Theta_0 \subseteq \hat{\Theta}_n(t)\}$. Given a subcollection $S \subseteq \mathcal{K}_c$, it is also possible to define a pseudo inverse mapping M_n by $t = M_n(S) := \inf\{t' : \Theta_0 \subseteq \hat{\Theta}_n(t'), \forall \Theta_0 \in S\}$[15].

We consider the following step-up procedure. The procedure starts with an initial subcollection $S_1 \subseteq \mathcal{K}_c$ of hypotheses. In the first step, we look at hypotheses in S_1^c, the collection of sets that are not in S_1. We then find the hypothesis that gives the least significant statistic value and compare this value with a common critical value. If the least significant statistic's value exceeds the common critical value, we reject all hypotheses in S_1^c and stop. Otherwise, we accept the hypotheses that give values below the common critical value and add them to S_1. We call this new collection S_2. In the next step, we test all hypotheses in S_2^c, which are not accepted in the first step. If the hypothesis with the least significant statistic is rejected, then we reject all the hypotheses in S_2^c and stop. Otherwise, we proceed to test the hypotheses that are not accepted in the first and second step. We repeat this until we stop.

Formally, the step-up procedure can be summarized as follows.

ALGORITHM 3.3.5: *(Step-up Procedure) Initialize $l = 1$ and t_l. Set $S_l = L_n(t_l)$.*

Step 1. *Construct the set estimator $\hat{\Theta}_n(t_l)$. If for all $\Theta_0 \in S_l^c$, $a_n^{1/\gamma}\vec{d}_H(\Theta_0, \hat{\Theta}_n(t_l)) > \hat{c}_{n,b,1-\alpha}^{\rightarrow}(t_l)$, then reject all hypotheses in S_l^c and stop.*

Step 2. *Otherwise, set $S_{l+1} = \{\Theta_0 \in \mathcal{K}_c : a_n^{1/\gamma}\vec{d}_H(\Theta_0, \hat{\Theta}_n(t_l)) \leq \hat{c}_{n,b,1-\alpha}^{\rightarrow}(t_l)\}$, set $t_{l+1} = M_n(S_{l+1})$, and proceed.*

Step 3. *Repeat Steps 1-2 until the procedure stops.*

For each l, S_l^c represents the family of hypotheses that are not previously accepted. In each step, the procedure compares the least significant test statistic $\inf_{\Theta_0 \in S_l^c} a_n^{1/\gamma}\vec{d}_H(\Theta_0, \hat{\Theta}_n(t_l))$ and a common critical value $\hat{c}_{n,b,1-\alpha}^{\rightarrow}(t_l)$. In each step, if the procedure does not reject all remaining hypotheses, it creates a random collection of compact convex sets S_l. The procedure continues until it rejects all

[15]Note that $M_n(L_n(t)) = t$ if Q_n is continuous, but in general $L_n(M_n(S)) \supseteq S$.

hypotheses that have not been previously accepted. If the procedure stops after the L-th iteration, we can use S_L as a confidence collection. Further, the union of all sets in this collection is a confidence set for Θ_I by Theorem 3.3.5 (ii).

If $Q_n(\theta)$ is continuous, the updating rule in Step 2 of Algorithm 3.3.5 can be written as

$$\begin{aligned}
t_{l+1} &= M_n(S_{l+1}) \\
&= \inf\{t : \Theta_0 \subseteq \hat{\Theta}_n(t), \quad \forall \Theta_0 \in S_{l+1}\} \\
&= \inf\left\{t : \sup_{\bigcup_{\Theta_0 \in S_{l+1}} \Theta_0} a_n Q_n(\theta) \leq t\right\} \\
&= \sup_{\hat{\Theta}_n^{\hat{c}_{n,b,1-\alpha}(t_l)/a_n^{1/\gamma}}(t_l)} a_n Q_n(\theta).
\end{aligned}$$

Thus, the step-up procedure yields the same sequence $\{t_l, l = 1, 2, \cdots\}$ as Algorithm 3.3.4. We can view the updating rule of Algorithm 3.3.4 as a reduced form of the step-up procedure.

We now establish that this procedure asymptotically controls the FWER.

Theorem 3.3.7 (Control of FWER): *Suppose the conditions of Theorem 3.3.6 hold. Then the step-up procedure asymptotically controls the FWER in the strong sense, i.e.,*

$$\lim_{n\to\infty} FWER \leq \alpha,$$

for all possible constellations of true and false hypotheses.

Our treatment of the multiple testing problem defined in eq. (3.3.8) enables us to use random collections of sets as building blocks. Alternatively, one may consider the family $\{H_{\theta_0} : \theta_0 \in \Theta_I\}$ indexed by $\theta_0 \in \Theta$. In this case, one may construct an analogous step-up procedure that starts with a initial set $\tilde{S}_1 \subset \Theta$ and steps up using the scaled directed Hausdorff distance statistic $a_n^{1/\gamma} \vec{d}_H(\{\theta_0\}, \hat{\Theta}_n(t_l))$. It can be shown that for any $l = 1, 2, \cdots$, $\tilde{S}_l$ equals $\bigcup_{S_l} \Theta_0$; this alternative procedure thus yields the same confidence set as before. This alternative approach is more analogous to Romano and Shaikh (2009)'s step-down procedure, as the hypotheses are indexed by $\theta_0 \in \Theta$. Both the step-up and step-down procedures control the family-wise error rate. Comparing the power of the two procedures is an interesting topic for future work.

3.3.4 Inference for Points in the Identified Set

As Imbens and Manski (2004) discuss, it is often of interest to test hypotheses regarding the true parameter value that generates the data[16]. As the true parameter value cannot be distinguished from any other element of Θ_I, a relevant question would be whether or not a given parameter value θ_0 is observationally equivalent to the data generating parameter value, i.e. $\theta_0 \in \Theta_I$. The scaled directed Hausdorff distance can be used to test this hypothesis. The test can then be inverted to yield a confidence set that asymptotically covers each point in Θ_I with at least a prespecified probability.

Bontemps, Magnac, and Maurin (2007) extend Imbens and Manskis' (2004) results to set-identified linear models. Our results in this section extend them further to the class of problems that can be studied in the extremum estimation framework.

Let $\theta_0 \in \Theta$, and consider testing

$$H_0 : \theta_0 \in \Theta_I \quad vs. \quad H_1 : \theta_0 \notin \Theta_I. \tag{3.3.9}$$

This can be equivalently stated as $H_0 : \langle p, \theta_0 \rangle \leq s(p, \Theta_I), \forall p \in \mathbb{S}^{d-1}$ $vs.$ $H_1 : \langle p, \theta_0 \rangle > s(p, \Theta_I), \exists p \in \mathbb{S}^{d-1}$. Suppose for the moment that $\theta_0 \in \partial\Theta_I$. In this case, there exists $p_0 \in \mathbb{S}^{d-1}$ such that $\langle p_0, \theta_0 \rangle = s(p_0, \Theta_I)$. Below, we restrict our attention to the cases where p_0 is the unique maximizer of $\langle p, \theta_0 \rangle - s(p, \Theta_I)$.

For each $n \in \mathbb{N}$, let $\hat{p}_n \in \mathbb{S}^{d-1}$ be a vector that maximizes $\langle p, \theta_0 \rangle - s(p, \hat{\Theta}_n(t))$. We use the directed Hausdorff distance statistic to test the hypothesis. Given θ_0,

[16] See also Woutersen (2006), Fan and Park (2007), and Stoye (2009) for extensions of Imbens and Manski's (2004) analysis.

we define the statistic

$$\begin{aligned}
T^{\rightarrow}_{n,\theta_0}(t) &:= a_n^{1/\gamma}\vec{d}_H(\{\theta_0\},\hat{\Theta}_n(t)) \\
&= \sup_{p\in\mathbb{S}^{d-1}} a_n^{1/\gamma}\left\{\langle p,\theta_0\rangle - s(p,\hat{\Theta}_n(t))\right\}_+ \\
&= a_n^{1/\gamma}\{\langle \hat{p}_n,\theta_0\rangle - s(\hat{p}_n,\hat{\Theta}_n(t))\}_+.
\end{aligned}$$

Lemmas in the Mathematical Appendix show $T^{\rightarrow}_{n,\theta_0}(t) \xrightarrow{d} \{-\mathcal{Z}(p_0,t)\}$ under appropriate regularity conditions.

For each $\alpha \in (0,1)$ and $t \in \mathbb{R}_+$, let

$$c^{\rightarrow}_{1-\alpha}(p,t) := \inf\left\{x : P\left(\{-\mathcal{Z}(p,t)\}_+ \le x\right) \ge 1-\alpha\right\}.$$

Similar to the inference for Θ_I, we estimate $c^{\rightarrow}_{1-\alpha}(p,t)$ by subsampling. An aspect specific to pointwise inference is that we use the quantile $c^{\rightarrow}_{1-\alpha}(p_0,t)$ evaluated at p_0.

Algorithm 3.3.6: *Let $t > 0$ and $0 < \alpha < 1$ be given. 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.*

Step 1. *For $k = 1,\cdots,N_{n,b}$, construct $\hat{\Theta}_{n,b,k}(t)$, the set estimator for the k-th subsample, computed as a t-level set of the criterion function $\zeta_{n,b,k}(X_{k_1},\cdots,X_{k_b},\theta) = a_b\tilde{Q}_{n,b,k}(X_{k_1},\cdots,X_{k_b},\theta)$, with the obvious notation.*

Step 2. *For $k = 1,\cdots,N_{n,b}$, and for each $p \in \mathbb{S}^{d-1}$, compute $\{-\mathcal{Z}_{n,b,k}(p,t)\}_+$, where*

$$\mathcal{Z}_{n,b,k}(p,t) = a_b^{1/\gamma}[s(p,\hat{\Theta}_{n,b,k}(t)) - s(p,\hat{\Theta}_n(t))].$$

Step 3. *For each $p \in \mathbb{S}^{d-1}$, compute the $100\times(1-\alpha)\%$ quantile of the subsampling distribution, given by*

$$\hat{c}^{\rightarrow}_{n,b,1-\alpha}(p,t) = \inf\left\{x : \hat{F}^{\rightarrow}_{n,b}(x,p,t) \ge 1-\alpha\right\},$$

where

$$\hat{F}^{\rightarrow}_{n,b}(x,p,t) := N_{n,b}^{-1} \sum_{1\le k\le N_{n,b}} 1_{\left\{\{-\mathcal{Z}_{n,b,k}(p,t)\}_+ \le x\right\}}.$$

For each $p \in \mathbb{S}^{d-1}$ and $t \in \mathbb{R}_+$, let $F^{\rightarrow}(x,p,t) := P\left(\{-\mathcal{Z}(p,t)\}_+ \le x\right)$ define the CDF of $\{-\mathcal{Z}(p,t)\}_+$. The next theorem establishes the validity of subsampling.

Theorem 3.3.8: *Suppose the conditions of Theorem 3.1 and Assumption 3.2 hold. Suppose that $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. Let $\hat{F}^{\rightarrow}_{n,b}(\cdot,p,t)$ and $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(p,t)$ be computed by Algorithm 3.3.6. Then for given $t \in \mathbb{R}_+$, $\alpha \in (0,1)$, and each $p \in \mathbb{S}^{d-1}$,*

(i) If x is a continuity point of $F^{\rightarrow}(\cdot,p,t)$, then $\hat{F}^{\rightarrow}_{n,b}(x,p,t) \to F^{\rightarrow}(x,p,t)$ in probability;

(ii) If $F^{\rightarrow}(\cdot,p,t)$ is continuous except at $x = 0$, then for any $\epsilon > 0$, $\sup_{|x|\ge\epsilon} |\hat{F}^{\rightarrow}_{n,b}(x,p,t) - F^{\rightarrow}(x,p,t)| \to 0$ in probability;

(iii) If $F^{\rightarrow}(\cdot,p,t)$ is continuous at $c^{\rightarrow}_{1-\alpha}(p,t)$, then

$$\lim_{n\to\infty} P\left(\{-\mathcal{Z}_n(p,t)\}_+ \le \hat{c}^{\rightarrow}_{n,b,1-\alpha}(p,t)\right) = 1-\alpha.$$

As a corollary of this result, we can show that the test has exact size for some $\theta_0 \in \Theta_I$ and is consistent against any fixed alternative hypothesis.

Corollary 3.3.4: *Suppose the conditions of Theorem 3.3.8 hold. Suppose the conditions of Lemma 3.B.7 also hold.*

(i) If $\theta_0 \in \Theta_I$ and if for given $t \in \mathbb{R}_+$ and $\alpha \in (0,1)$, $F^{\rightarrow}(\cdot,p_0,t)$ is continuous and strictly increasing at $c^{\rightarrow}_{1-\alpha}(p_0,t)$, then $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n,t) = c^{\rightarrow}_{1-\alpha}(p_0,t) + o_p(1)$, and the test has asymptotic rejection probability α :

$$\lim_{n\to\infty} \sup_{\theta_0\in\Theta_I} P\left(T^{\rightarrow}_{n,\theta_0}(t) > \hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n,t)\right) = \alpha.$$

(ii) If $\theta_0 \notin \Theta_I$, then for any $t \in \mathbb{R}_+$ and $\alpha \in (0,1)$, the test is consistent:

$$\lim_{n\to\infty} P\left(T^{\rightarrow}_{n,\theta_0}(t) > \hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n,t)\right) = 1.$$

The confidence set for θ_0 is obtained by inverting the test. Define

$$\check{\Theta}_{n,b,1-\alpha}(t) := \{\theta_0 \in \Theta : T^{\rightarrow}_{n,\theta_0}(t) \leq \hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t)\}.$$

Note that $\hat{c}^{\rightarrow}_{n,b,1-\alpha}$ depends on θ_0 through $\hat{p}_n$. This dependence of the critical value on θ_0 reflects how precisely each boundary point of Θ_I can be estimated. The following theorem shows that this confidence set has the correct coverage probability.

THEOREM 3.3.9: *Suppose the conditions of Theorem 3.3.8 hold. Then for a given* $\alpha \in (0, 1)$

$\lim_{n\to\infty} \inf_{\theta_0 \in \Theta_I} P\left(\theta_0 \in \check{\Theta}_{n,b,1-\alpha}(t)\right) = 1 - \alpha;$

Note the difference between this confidence set and that for the identified set. To construct $\check{\Theta}_{n,b,1-\alpha}(t)$, we use $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t)$ as a critical value, instead of $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)$. Intuitively, the former takes into account how precisely the support set of Θ_I for the specific normal vector p_0 is estimated. On the other hand, the latter takes into account how precisely the whole boundary of Θ_I is estimated.

Bontemps, Magnac, Maurin (2007) study a confidence set whose coverage is asymptotically valid uniformly over possible values of a nuisance parameter $\Delta = \sup_{p \in \mathbb{S}^{d-1}} s(p, \Theta_I) - s(-p, \Theta_I)$, the maximum length of the identified set. Developing an extension of Theorem 3.3.9 in this direction is an interesting topic for future work.

Power against Local Alternatives

So far, our discussion has been based on the fixed probability measure $P = P_{\theta_0,\Theta_I}$, where θ_0 is the parameter associated with the true DGP and Θ_I the identified set. In this section, we consider the power of the test against a sequence of alternatives $\{P_{\theta_n,\Theta_I}\}$ indexed by parameter values $\{\theta_n\}$ while fixing the identified set.

If θ_0 is in the interior of Θ_I, it can be shown that the test has no power against alternatives in the neighborhood of θ_0. In the following, we therefore study the case where $\theta_0 \in \partial\Theta_I$.

Let $\pi_{n,b,t} : \Theta \to [0,1]$ be the power function defined by

$$\pi_{n,b,t}(\theta) := P_{\theta,\Theta_I}\left(T^{\rightarrow}_{n,\theta}(t) > c^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t)\right).$$

Let $h > 0$. Consider the sequence of alternatives $\{\theta_n\}$ that satisfies $\theta_n := \theta_0 + \lambda/a_n^{1/\gamma}$, where $\theta_0 \in \partial\Theta_I$, and $\lambda \in \mathbb{R}^d$ satisfies $\langle p_0, \lambda\rangle = h$. In terms of the support function, the local parameter has the property that $\langle p_0, \theta_n\rangle = s(p_0, \Theta_I) + h/a_n^{1/\gamma}$. Therefore, for this sequence, the local deviation from the null hypothesis is measured by a distance in terms of the support function, and its magnitude is controlled by the parameter h.

The power of the test has the following properties.

Theorem 3.3.10: *Suppose the conditions of Theorem 3.3.8 hold. Then (i) The test is asymptotically locally unbiased:* $\liminf_{n\to\infty} \pi_{n,b,t}(\theta_n) \geq \alpha$ *for any* $h > 0$*;*

(ii) The limiting power function satisfies

$$\begin{aligned}\liminf_{n\to\infty} \pi_{n,b,t}(\theta_n) &\geq \liminf_{n\to\infty} P_{\theta_n,\Theta_I}\left(\inf_{\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0}} \tilde{\zeta}_n(\theta,\lambda) > t\right)\\ &\geq P_{\theta_0,\Theta_I}\left(\inf_{\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0}} \tilde{\zeta}(\theta,\lambda) > t\right),\end{aligned}$$

where

$$\begin{aligned}\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} &= H(p,\Theta_I) \times (K_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} \cap L)\\ &= \left\{(\theta,\lambda) : \langle p_0,\theta\rangle = s(p_0,\Theta_I), \langle p_0,\lambda\rangle \geq h - c^{\rightarrow}_{1-\alpha}(p_0,t), \lambda \in L\right\},\end{aligned}$$

for some compact set L.

The power depends on the event

$$\inf_{\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0}} \tilde{\zeta}_n(\theta,\lambda) > t. \tag{3.3.10}$$

This representation gives several insights. First, as a natural consequence of the construction of the local alternatives, the asymptotic power of the test is fully determined by p_0 and h, given t. Note that the infimum on the left hand side is essentially determined by the set

$$K_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} = \{\lambda : \langle p_0, \lambda\rangle \geq h - c^{\rightarrow}_{1-\alpha}(p_0,t)\}.$$

From this, one can see the test has the same power against two distinct alternatives θ_n and θ'_n that are on the same hyperplane with the normal vector p_0, which has the distance $h/a_n^{1/\gamma}$ to θ_0. This is because we reject the hypothesis if and only if the constraint set $K_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} \cap L$ is separated from the level-t set of $\tilde{\zeta}_n(\theta_0,\cdot)$. One can show that the probability of the event above is asymptotically α when $h = 0$. As h increases, the constraint set $K_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} \cap L$ escapes to the horizon, and the infimum over this set becomes arbitrarily large, which drives $\pi_{n,b,t}(\theta_n)$ to one as $h \to \infty$.

Second, for a given $t \in \mathbb{R}_+$ and $p_0 \in \mathbb{S}^{d-1}$, the slope of the power function as a function of h is determined by the shape of the criterion function. If the econometrician wishes to achieve a rapid increase in power against alternatives that are away from the null with a specific direction p_0, she should choose a criterion function Q_n to force $\tilde{\zeta}_n(\theta_0,\cdot)$ to grow more rapidly in the direction p_0. This makes it more likely for the left hand side of eq. (3.3.10) to exceed t.

Third, for a given $h \geq 0$ and $t \geq 0$, the power differs with $p_0 \in \mathbb{S}^{d-1}$ depending on the precision with which each support set is estimated. This explains why we use different critical values for different directions when we constructed $\check{\Theta}_{n,b,1-\alpha}(t)$.

An Extension of Pointwise Inference

A simple extension of pointwise inference yields a conservative test for a hypothesis that Θ_I has a nonempty intersection with a known set Θ_0. When Θ_0 is a set of parameter values that satisfy some restrictions, this test can be used to assess the validity of such restrictions. This type of hypothesis has been studied in Romano and Shaikh (2008) for parametric models and Santos (2007) for nonparametric models.

Now let $\Theta_0 \in \mathcal{K}$ be a nonempty closed subset of Θ. Consider testing

$$H_0 : \Theta_0 \cap \Theta_I \neq \varnothing \quad vs. \quad H_1 : \Theta_0 \cap \Theta_I = \varnothing. \tag{3.3.11}$$

Here, Θ_0 collects parameter values that satisfy the restrictions of interest. The null states that there is at least one element in the identified set satisfying the restrictions. Rejection means that none of the parameters in the identified set satisfy the restrictions, implying that the data generating parameter value does not satisfy the restrictions.

The null hypothesis can be equivalently stated as $H_0 : \exists \theta_0 \in \Theta_0$ such that $\{\theta_0\} \subseteq \Theta_I$. Note that $\inf_{\theta_0 \in \Theta_0} \vec{d}_H(\{\theta_0\}, \Theta_I) = 0$ under the null hypothesis, and $\inf_{\theta_0 \in \Theta_0} \vec{d}_H(\{\theta_0\}, \Theta_I) > 0$ under the alternative hypothesis. Therefore, a natural test statistic is $\inf_{\theta_0 \in \Theta_0} T^{\rightarrow}_{n,\theta_0}(t) = \inf_{\theta_0 \in \Theta_0} a_n^{1/\gamma} \vec{d}_H(\{\theta_0\}, \hat{\Theta}_n(t))$. The triangle inequality (3.3.6) implies

$$\inf_{\theta_0 \in \Theta_0} \vec{d}_H(\{\theta_0\}, \hat{\Theta}_n(t)) \leq \inf_{\theta_0 \in \Theta_0} \vec{d}_H(\{\theta_0\}, \Theta_I) + \vec{d}_H(\Theta_I, \hat{\Theta}_n(t)).$$

Under the null hypothesis, the first term on the right vanishes. Therefore, we can use the asymptotic conservative critical value $c^{\rightarrow}_{n,b,1-\alpha}(t)$ computed by Algorithm 3.3.2 to test the hypothesis. As a corollary to Theorem 3.3.4, results for the rejection probability and the consistency against fixed alternatives follow as before.

Corollary 3.3.5: *Suppose the conditions of Theorem 3.3.4 hold. Let Θ_0 be a nonempty closed subset of Θ^o.*

(i) If $\Theta_0 \cap \Theta_I \neq \varnothing$ and $F^{\rightarrow}(\cdot, t)$ is continuous and strictly increasing at $c^{\rightarrow}_{1-\alpha}(t)$, then $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t) = c^{\rightarrow}_{1-\alpha}(t) + o_p(1)$ and the test has asymptotic rejection probability bounded above by α :

$$\lim_{n\to\infty} P\left(\inf_{\theta_0 \in \Theta_0} T^{\rightarrow}_{n,\theta_0}(t) > \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right) \leq \alpha;$$

(ii) If $\Theta_0 \cap \Theta_I = \varnothing$, then the test is consistent:

$$\lim_{n\to\infty} P\left(\inf_{\theta_0 \in \Theta_0} T^{\rightarrow}_{n,\theta_0}(t) > \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)\right) = 1.$$

One may envisage other extensions of pointwise inference. One such possibility is inference on linear functionals of θ_0. This extension is straightforward in our framework, as any linear functional of θ_0 can be represented as $\langle p, \theta_0 \rangle$ for some $p \in \mathbb{R}^d$. This may also be extended to nonlinear functionals, but to keep a tight focus here, we leave that analysis to elsewhere.

3.4 Moment Inequality Models

In this section, we pay special attention to a class of economic models with an identified set defined by finitely many moment inequalities. This class has been extensively studied recently[17]. Leading examples in this class of models are a regression model with censored outcome variables (Manski and Tamer, 2002), entry game models (Ciliberto and Tamer, 2009), and dynamic game models (Bajari, Benkard, and Levin, 2007). We first show that this class can be studied within the framework developed above. We provide a set of conditions for this class that ensure the high level assumptions presented in sections 3.2 and 3.3. In section 3.4.2, we provide additional results that can be obtained by using CHT's quadratic criterion function. In particular, we establish the asymptotic equivalence of the squared directed Hausdorff distance statistic and CHT's QLR statistic.

3.4.1 General Results for Moment Inequality Models

In the following, we use E and $\hat{E}_n$ to denote the expectation operators with respect to the data generating probability measure and the empirical measure, respectively. We consider functions $m_j : \mathbb{R}^k \times \mathbb{R}^d \to \bar{\mathbb{R}}$, $j = 1, \cdots, J$, that define the following moment inequality restrictions.

$$E(m_j(X;\theta)) \leq 0, \quad j = 1, \cdots, J.$$

[17]Recent research in this area includes Andrews, Berry, and Jia (2004), Pakes, Porter, Ho, and Ishii (2005), Rosen (2008), CHT, Fan and Park (2007), Galichon and Henry (2007), BM, Andrews and Guggenberger (2009), Andrews and Soares (2009), Bugni (2009), Canay (2009), Galichon and Henry (2009), Hahn and Ridder (2009), Moon and Schorfheide (2009), and Yıldız (2009).

Interest attaches to the identified set, which comprises the values at which the moment inequalities are satisfied: i.e., $\Theta_I := \{\theta \in \Theta : E(m_j(X;\theta)) \le 0, j = 1, \cdots, J\}$.

Let m_θ be a $J \times 1$ vector whose j-th component is $m_{j,\theta} := m_j(X;\theta)$. Let $\mathcal{P}_J$ be the space of symmetric positive definite real-valued $J \times J$ matrices, and let $\bar{\mathcal{P}}_J$ be the space of symmetric positive definite extended real-valued $J \times J$ matrices. For any $\theta \in \mathbb{R}^d$, let $W(\theta) \in \bar{\mathcal{P}}_J$ be a weighting matrix, and let $\{\hat{W}_n : \Omega \times \mathbb{R}^d \to \bar{\mathcal{P}}_J\}$ be a sequence of (possibly random) positive definite weighting matrices. For brevity, we write $\hat{W}_n(\theta)$. We consider population and sample criterion functions of the form:

$$Q(\theta) = \varphi\left(E(m_\theta), W(\theta)\right)$$
$$Q_n(\theta) = \varphi\left(\hat{E}_n(m_\theta), \hat{W}_n(\theta)\right),$$

where $\varphi : \bar{\mathbb{R}}^J \times \bar{\mathcal{P}}_J \to \bar{\mathbb{R}}_+$ is a non-negative continuous function of the moment condition and the weighting matrix. For example, CHT and Romano and Shaikh (2008, 2009) consider the following functional form for Q_n:

$$Q_n(\theta) = \sum_{j=1}^{J} (\hat{W}_{jn}^{1/2}(\theta)\hat{E}_n(m_{j,\theta}))_+^2,$$

where $\hat{W}_{jn}(\theta)$ is the j-th diagonal element of $\hat{W}_n(\theta)$. Manski and Tamer (2002) and Rosen (2008) use the form:

$$Q_n(\theta) = \inf_{\mu \in \mathbb{R}_-^J} (\hat{E}_n(m_\theta) - \mu)' \hat{W}_n(\theta)(\hat{E}_n(m_\theta) - \mu),$$

where $\mathbb{R}_-^J = \{x \in \mathbb{R}^J : x_j \le 0, j = 1, \cdots, J\}$. We focus on a class of criterion functions that includes the examples above as special cases[18]. We assume the following regularity conditions on the parameter space, the moment conditions, and the "index function" φ.

[18] It would be interesting to extend our analysis here to a more general class within which we can also study moment equality models. Such a general class was considered in Andrews and Guggenberger (2009) and Andrews and Soares (2009).

Assumption 3.4.1: *Let* $J \in \mathbb{N}$. $\varphi : \bar{\mathbb{R}}^J \times \bar{\mathcal{P}}_J \to \bar{\mathbb{R}}_+$ *is a non-negative continuous function such that for any* $w \in \mathcal{P}_J$, $\varphi(y, w) = 0$ *if and only if* $y \le 0$, *i.e.* $y_j \le 0$ *for* $j = 1, \cdots, J$, *and* $\varphi(y, w) = \infty$ *if* y *or* w *contains an infinite element. Let* $\Theta \subset \mathbb{R}^d, d \in \mathbb{N}$, *be compact and convex with nonempty interior. Let* $W : \mathbb{R}^d \to \bar{\mathcal{P}}_J$ *be a measurable mapping, and suppose that* W *is finite and continous on* Θ *and that if* $\theta \notin \Theta$ *then* $\det(W(\theta)) = \infty$. *Let* $k \in \mathbb{N}$; *for each* $j = 1, \cdots, J$, $m_j : \mathbb{R}^k \times \mathbb{R}^d \to \bar{\mathbb{R}}$ *is jointly measurable, and for each* $x \in \mathbb{R}^k$, *if* $\theta \notin \Theta$ *then* $m_j(x, \theta) = \infty$. *Let* $(\Omega, \mathfrak{F}, P)$ *be a complete probability space. Let* $\{X_i : \Omega \to \mathbb{R}^k\}$ *be a sequence of identically distributed random vectors such that for each* $\theta \in \Theta$ *and* $j = 1, \cdots, J$, $E(m_j(X_i, \theta)) < \infty$. *Let* $\hat{W}_n : \Omega \times \mathbb{R}^d \to \bar{\mathcal{P}}_J$ *be jointly measurable, and suppose that for each* $\omega \in \Omega$, $\hat{W}_n(\omega, \cdot)$ *is finite and continuous on* Θ, *uniformly in* n, *and for each* $\omega \in \Omega$, *if* $\theta \notin \Theta$ *then* $\det \hat{W}_n(\omega, \theta) = \infty$. *Define* $Q(\theta) := \varphi\left(E(m_\theta), W(\theta)\right)$ *and* $Q_n(\theta) := \varphi(\hat{E}_n(m_\theta), \hat{W}_n(\theta))$.

Assumption 3.4.1 ensures that Assumption 3.2.1 holds for moment inequality models. The assumed continuity of W on Θ and its behavior outside of Θ ensures that its minimum eigenvalue is bounded from below by a positive constant over $\mathbb{R}^d$. The almost sure properness of the sample criterion function Q_n is ensured by the requirements that φ is a nonnegative function and that $E(m_{j,\theta})$ and $W(\theta)$ are finite on Θ. Using the criterion function Q, the identified set can be defined as $\Theta_I = \{\theta : Q(\theta) = 0\}$.

The following condition ensures Assumption 3.2.2.

Assumption 3.4.2: *(i) There exists* $\theta \in \Theta$ *such that* $E(m_{j,\theta}) \le 0$ *for* $j = 1, \cdots, J$. *The map* $\theta \mapsto \varphi(E(m_\theta), W(\theta))$ *is continuous and convex on* Θ; *(ii)* $\{\theta \in \Theta : \varphi(E(m_\theta), W(\theta)) = 0\} \subset \Theta^o$.

Assumption 3.4.2 (i) ensures nonemptiness, closedness, and convexity of the identified set. Assumption 3.4.2 (ii) ensures that the identified set is in the interior of Θ.

Conditions required for the consistency of the set estimator $\hat{\Theta}_n(t)$ are standard.[19] In particular, we must ensure the uniform convergence of Q_n. The rate of

[19]Strictly speaking, one needs to establish the measurability of $d_H(\hat{\Theta}_n(t), \Theta_I)$ to discuss con-

convergence depends on the choice of the index function φ. Here, we give primitive conditions on the moment conditions and the index function based on CHT's condition M.2. For this, we introduce the ϵ-contraction of Θ_I, which is defined by $\Theta_I^{-\epsilon} := \{\theta \in \Theta_I : d(\theta, \Theta \setminus \Theta_I) \geq \epsilon\}$ for $\epsilon > 0$.

Assumption 3.4.3: *(i-a) There exist* $0 < L_1 < \infty$ *and a continuous increasing function* $h_1 : \bar{\mathbb{R}}_+ \to \bar{\mathbb{R}}_+$ *such that for any* $w \in \mathcal{P}_J$ *and* $x, x^* \in \mathbb{R}^J$, $|\varphi(x, w) - \varphi(x^*, w)| \leq L_1 h_1(\|x - x^*\|)$*, and there exist* $0 < L_2 < \infty$ *and a continuous increasing function* $h_2 : \bar{\mathbb{R}}_+ \to \bar{\mathbb{R}}_+$ *such that for any* $x \in \mathbb{R}^J$ *and* $w, w^* \in \mathcal{P}_J$, $|\varphi(x, w) - \varphi(x, w^*)| \leq L_2 h_2(\max_{i,j} |w_{ij} - w_{ij}^*|)$*; (i-b)* $\{m_\theta : \theta \in \Theta\}$ *is a P-Donsker class, and* $\hat{W}_n(\theta) - W(\theta) = o_p(1)$ *uniformly over* Θ*; (ii-a)* $\sup_{\Theta_I} Q_n(\theta) = O_p(1/a_n)$*; (ii-b) There exist positive constants* (C_1, δ) *such that for any* $\theta \in \Theta$, $\|E(m_\theta)\|_+ \geq C_1 \min\{d(\theta, \Theta_I), \delta\}$*; (ii-c) There exist positive constants* (C_2, γ) *such that for any* $\omega \in \mathcal{P}_J$ *and* $x \in \mathbb{R}^J$, $\varphi(x, w) \geq C_2 \|x\|_+^\gamma$*; (ii-d) There exist positive constants* $(C_3, C_4, \bar{\epsilon})$ *such that for any* $0 \leq \epsilon \leq \bar{\epsilon}$ *and* $\theta \in \Theta_I^{-\epsilon}$, $\max_{1 \leq j \leq J} E(m_{j,\theta}) \leq -C_3 \epsilon$*, and* $d_H(\Theta_I^{-\epsilon}, \Theta_I) \leq C_4 \epsilon$.

Assumptions 3.4.3 (i-a,b) are sufficient for the uniform convergence of Q_n on Θ[20]. Assumption 3.4.3 (ii) collects conditions necessary for the convergence rate result. Condition (ii-a) requires the sample criterion function Q_n to vanish over the identified set at a rate of $1/a_n$. Assumption 3.4.3 (ii-b) requires the norm of $E(m_\theta)$ to be bounded from below by the distance from the identified set when θ is outside Θ_I. Together with Assumption 3.4.3 (ii-c), this ensures the existence of a polynomial minorant, which is required in Assumption 3.2.3 (ii). Assumption 3.4.3 (ii-d) requires the moment conditions to take strictly negative values on the contracted identified set. This enables us to approximate the identified set by its contraction $\Theta_I^{-\epsilon}$, on which the sample criterion function $a_n Q_n(\theta)$ vanishes. As CHT illustrate, Assumption 3.4.3 (ii-d) holds in many applications. This condition implies Assumption 3.2.4, which suffices to attain the exact rate of convergence

sistency. It is known that the measurability of $\hat{\Theta}_n(t)$ as a random closed set is sufficient for this purpose. For details about the measurability of level set estimators, see Kaido and White (2008).

[20]If we further assume that $\theta \mapsto \varphi(\hat{E}_n m_\theta, \hat{W}_n(\theta))$ is globally convex on Θ, a weaker assumption that $\hat{E}_n(\theta)$ and $W_n(\theta)$ converge in probability pointwise is sufficient as in Andersen and Gill (1982) Corollary II.2 and Newey and McFadden (1992) Theorem 2.7.

$a_n^{1/\gamma}$ without setting $t \geq \sup_\Theta a_n Q_n(\theta)$.

The next step is to show that $\tilde{\zeta}_n(\theta,\lambda) = nQ_n(\theta + \lambda/a_n^{1/\gamma})$ satisfies the local process regularity conditions given in Assumption 3.3.1. Most importantly, we need to ensure that $\tilde{\zeta}_n$ weakly epiconverges to a well-defined limit. To illustrate the key ideas, we take a slightly generalized version of CHT's criterion function as an example.

Let $x \circ y$ denote the entrywise (Hadamard) product of $x, y \in \mathbb{R}^J$. Let $s : \mathbb{R}^J \to \{1,0\}^J$ be a vector-valued mapping whose j-th component is $s_j = 1\{x_j > 0\}$. Let the index function be defined by $\varphi(x,w) := \left\|w^{1/2}x\right\|_+^2 := \|w^{1/2}(x \circ s)\|^2$. The sample criterion function is then $Q_n(\theta) = \|\hat{W}_n^{1/2}(\theta)\hat{E}_n(m_\theta)\|_+^2$. As the weighting matrix need not be diagonal, this is a slightly generalized version of the criterion function used by CHT.

With this choice of index function, we can take $a_n = n$ and $\gamma = 2$. That is, $nQ_n(\theta)$ has nondegenerate asymptotics, and $\sqrt{n}d_H(\hat{\Theta}_n(t), \Theta_I) = O_p(1)$, given Assumptions 3.4.1, 3.4.2, and 3.4.3. Suppose that m_θ allows a first-order expansion $m_{\theta^*} = m_\theta + \nabla' m_\theta(\theta^* - \theta) + o(|\theta^* - \theta|)$ on Θ^o, where ∇m_θ is a d-by-J matrix and $o(|\theta^* - \theta|)$ represents a small order term. Under these assumptions, we can write

$$
\begin{aligned}
\tilde{\zeta}_n(\theta,\lambda) &= \left\|\sqrt{n}\hat{E}_n(m_{\theta+\lambda/\sqrt{n}})\hat{W}_n^{1/2}(\theta+\lambda/\sqrt{n})\right\|_+^2 \\
&= \left\|[\sqrt{n}\hat{E}_n(m_\theta) + \hat{E}_n(\nabla' m_\theta)\lambda]\hat{W}_n^{1/2}(\theta+\lambda/\sqrt{n})\right\|_+^2 + o_p(1) \\
&= \left\|[\mathbb{G}_n m_\theta + \hat{E}_n(\nabla' m_\theta)\lambda + \sqrt{n}E m_\theta]\hat{W}_n^{1/2}(\theta+\lambda/\sqrt{n})\right\|_+^2 + o_p(1) \\
&= \left\|\mathcal{M}_n(\theta,\lambda)\hat{W}_n^{1/2}(\theta+\lambda/\sqrt{n})\right\|_+^2 + o_p(1),
\end{aligned}
$$

where we define $\mathbb{G}_n := \sqrt{n}(\hat{E}_n - E)$ and $\mathcal{M}_n(\theta,\lambda) = \mathbb{G}_n m_\theta + \hat{E}_n(\nabla' m_\theta)\lambda + \sqrt{n}E(m_\theta)$.

By the P-Donsker property of the moment functions, $\mathbb{G}_n m_\theta \overset{u.d.}{\to} \mathbb{G}(\theta)$ in $l^\infty(\Theta)$, where $\mathbb{G}$ is a $J \times 1$ zero-mean Gaussian process with almost surely continuous paths, and $Var(\mathbb{G}_j(\theta)) > 0$ for each $\theta \in \Theta$ and $j = 1, \cdots, J$. Together with the P-Donsker property, a set of general assumptions is often available to ensure that, for each $(\theta,\lambda) \in \Theta^o \times \mathbb{R}^d$, $\mathcal{M}_n(\theta,\lambda) \overset{f.d.}{\to} \mathcal{M}(\theta,\lambda) := \mathbb{G}(\theta) + \Pi(\theta)\lambda + \varsigma(\theta)$ and

$\hat{W}_n(\theta + \lambda/\sqrt{n}) \xrightarrow{p} W(\theta)$, where $\varsigma(\theta)$'s j-th component satisfies

$$\varsigma_j(\theta) = \begin{cases} -\infty & \text{if } E(m_{j,\theta}) < 0 \\ 0 & \text{if } E(m_{j,\theta}) = 0 \\ \infty & \text{if } E(m_{j,\theta}) > 0 \end{cases} \tag{3.4.1}$$

for $j = 1, 2, \cdots, J$. The components of $\varsigma(\theta)$ are unbounded if the corresponding population moment inequalities are not binding, but the truncation operator $(\cdot)_+$ makes the criterion function always bounded from below by 0, which ensures the properness of the limiting process.

Similarly, for general choice of φ, one can often show $\tilde{\zeta}_n(\theta, \lambda) = \varphi(\mathcal{M}_n(\theta, \lambda), \hat{W}_n(\theta + \lambda/a_n^{1/\gamma})) + o_p(1)$ for each θ and λ. Then the continuous mapping theorem implies

$$\tilde{\zeta}_n(\theta, \lambda) \xrightarrow{f.d.} \varphi\left(\mathcal{M}(\theta, \lambda), W(\theta)\right)$$

Recall that, to apply Theorem 3.3.1, we need to establish the weak epiconvergence of $\tilde{\zeta}_n$ instead of the weak finite-dimensional convergence. Provided that the weak finite-dimensional limit exists, Knight (1999) shows that the weak finite-dimensional limit is also the weak epi-limit if and only if the sequence $\tilde{\zeta}_n(\theta, \lambda)$ is equi-lower-semicontinuous[21]. A general sufficient condition that ensures the desired weak epiconvergence and other local process regularities is the following.

Assumption 3.4.4: *(i-a) For each $j = 1, \cdots, J$, and $x \in \mathbb{R}^k, m_j(x, \cdot)$ is continuously differentiable with respect to θ on Θ^o with a continuous gradient $\nabla m_\theta(x, \cdot) \in \mathbb{R}^{d \times J}$, and for some continuous mapping $\Pi : \Theta^o \mapsto \mathbb{R}^{J \times d}$ and each θ in Θ^o, $\hat{E}_n(\nabla' m_\theta) = \Pi(\theta) + o_p(1)$; (i-b) $\sqrt{n}E(m_\theta) = \varsigma(\theta) + o_p(1)$ for each θ in Θ, where ς is defined by Eq.* (3.4.1)*; (ii) The map $\theta \mapsto \varphi(\hat{E}_n(m_\theta), \hat{W}_n(\theta))$ is convex in a neighborhood of Θ_I; (iii) The map $(\theta, \lambda) \mapsto \varphi(\mathcal{M}_n(\theta, \lambda), \hat{W}_n(\theta + \lambda/a_n^{1/\gamma}))$ is equi-lower-semicontinuous on $\Theta^o \times \mathbb{R}^d$.*

The conditions in Assumption 3.4.4 are plausibly general. In addition to

[21] The mathematical appendix summarizes Knight (1999)'s results. When $\tilde{\zeta}_n(\theta, \lambda)$ is also globally convex, it suffices to check that the limiting function is finite on some open set. See Geyer (2003) for details.

the P-Donskerness of $\{m_\theta : \theta \in \Theta\}$, we only require the finite-dimensional pointwise convergence of other terms in $\mathcal{M}_n(\theta, \lambda)$. A standard LLN will ensure this requirement.

Next, the following theorem establishes Assumptions 3.2.1-3.2.4, and the local process regularity (Assumption 3.3.1), including weak epiconvergence.

THEOREM 3.4.1: *Suppose Assumptions 3.4.1, 3.4.2 3.4.3, and 3.4.4 hold. Then Assumptions 3.2.1, 3.2.2, 3.2.3, 3.2.4 and 3.3.1 are satisfied with weak epilimit* $\tilde{\zeta}(\theta, \lambda) := \varphi(\mathcal{M}(\theta, \lambda), W(\theta))$.

Theorem 3.3.1 now applies. An important corollary is the following.

COROLLARY 3.4.1: *Suppose Assumptions 3.4.1, 3.4.2 3.4.3, and 3.4.4 hold. Then* $\sqrt{n} d_H(\hat{\Theta}_n(t), \Theta_I) \xrightarrow{d} \|\mathcal{Z}(\cdot, t)\|_{\mathcal{C}(\mathbb{S}^{d-1})}$, *and* $\sqrt{n} \vec{d}_H(\hat{\Theta}_n(t), \Theta_I) \xrightarrow{d} \sup_{p \in \mathbb{S}^{d-1}} \{-\mathcal{Z}(\cdot, t)\}_+$, *where* $\mathcal{Z}(\cdot, t)$ *can be represented as*

$$\mathcal{Z}(p, t) = \sup_{\theta \in H(p, \Theta_I)} \sup_{\lambda \in \{\lambda : \varphi(\mathcal{M}(\theta, \lambda), W(\theta)) \leq t\}} \langle p, \lambda \rangle. \tag{3.4.2}$$

The representation above specifies how the limiting process $\mathcal{Z}(\cdot, t)$ depends on the weak epilimit $\varphi(\mathcal{M}(\theta, \lambda), W(\theta))$. Note that the asymptotic distribution of $\mathcal{Z}(\cdot, t)$ depends non-trivially on the identified set Θ_I.

3.4.2 A Closed Form for the Limiting Process and the Equivalence of Wald and QLR Statistics

In the previous section, we provided general conditions for moment inequality models that ensure the high level assumptions in section 3. In this section, we develop further results that rely on the properties of CHT's quadratic criterion function.

The goal of this section is to show that (i) a closed form for the limiting process $\mathcal{Z}(\cdot, t)$ can be derived; (ii) for each p, the limiting process $\mathcal{Z}(p, t)$ depends only on the active moment inequalities at $\theta \in H(p, \Theta_I)$; (iii) a certain choice of weighting matrix $W(\theta)$ makes the limiting process take the form $\mathcal{Z}(p, t) =$

$\mu(t) + \mathcal{Z}^*(p)$; and (iv) The Wald statistic (squared directed Hausdorff distance) and CHT's QLR statistic are asymptotically equivalent, under this choice of the weighting matrix and some additional assumptions.

We introduce some further notation to denote active and slack moment inequalities. For each $\theta \in \partial\Theta_I$, let $\mathcal{J}(\theta) \subseteq \{1, \cdots, J\}$ be the set of indices associated with active moment inequalities, i.e., $E(m_{j,\theta}) = 0$ for all $j \in \mathcal{J}(\theta)$. We denote by $J(\theta)$ the number of elements in $\mathcal{J}(\theta)$. Similarly, let $\mathcal{J}^c(\theta) \subseteq \{1, \cdots, J\}$ collect indices associated with slack moment inequalities at $\theta \in \partial\Theta_I$, i.e., $E(m_{j,\theta}) < 0$ for all $\mathcal{J}^c(\theta)$.

Let $\Pi_{\mathcal{J}(\theta)}(\theta)$ denote the $J(\theta) \times d$ matrix that stacks rows of $\Pi(\theta)$ whose indices belong to $\mathcal{J}(\theta)$. Similarly, let $\mathbb{G}_{\mathcal{J}(\theta)}$ denote the $J(\theta) \times 1$ vector of Gaussian processes that stacks components of $\mathbb{G}$ whose indices belong to $\mathcal{J}(\theta)$. Let $W_{\mathcal{J}(\theta)}$ denote the $J(\theta) \times J(\theta)$ matrix that collects (i, j) elements of $W(\theta)$ for $i, j \in \mathcal{J}(\theta)$.

We consider the following problem, which is a part of the optimization problem that defines $\mathcal{Z}(\cdot, t)$ in Eq. (3.4.2), while fixing $p \in \mathbb{S}^{d-1}$, $\theta \in H(p, \Theta_I)$, and $t \in \mathbb{R}_+$.

$$\begin{aligned} \sup_{\lambda} \quad & \langle p, \lambda \rangle \\ s.t. \quad & \|W^{1/2}_{\mathcal{J}(\theta)}(\theta)[\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda]\|_+^2 \leq t. \end{aligned} \tag{3.4.3}$$

Note that the constraint involves only selected rows of $\mathcal{M}(\theta, \lambda)$ whose indices are in $\mathcal{J}(\theta)$. This is because $\varsigma_j(\theta) = -\infty$ if $E(m_{j,\theta}) < 0$, and the index function φ truncates such components. The rows of $\mathcal{M}(\theta, \lambda)$ with indices belonging to $\mathcal{J}^c(\theta)$, therefore, do not marginally affect the constraint. Note also that the $\varsigma_j(\theta)$'s no longer appear in the constraint because $\varsigma_j(\theta) = 0$ for $j \in \mathcal{J}(\theta)$.

To obtain a closed form for $\mathcal{Z}(\cdot, t)$, we assume the following further conditions.

Assumption 3.4.5: *(i) For each $\theta \in \partial\Theta_I$, $rank(\Pi_{\mathcal{J}(\theta)}) = J(\theta)$, i.e. the rows of the Jacobian matrices are linearly independent; (ii) For each $\theta \in \partial\Theta_I$ and $p \in \mathbb{S}^{d-1}$, there exists a vector $\eta \in \mathbb{R}_+^{J(\theta)} \setminus \{0\}$ such that $p = \Pi'_{\mathcal{J}(\theta)}\eta$.*

Assumption 3.4.5 (i) is a linear independence constraint qualification condition. This ensures the solution to the problem in eq (3.4.3) satisfies the Karush-Kuhn-Tucker (KKT) conditions given in the mathematical appendix. Assumption 3.4.5 (ii) is not restrictive, as it usually holds as a necessary condition for the following auxiliary optimization problem, which can be used to characterize the boundary points of the identified set:

$$\begin{aligned} \sup \quad & \langle p, \theta \rangle \\ s.t. \quad & E(m_{j,\theta}) \leq 0, \quad \text{for } j = 1, \cdots, J. \end{aligned}$$

Using these additional assumptions, we can explicitly solve the optimization problem in Eq. (3.4.3) to obtain the following result.

Corollary 3.4.2: *Suppose the conditions of Theorem 3.4.1 and Assumption 3.4.5 hold. Suppose* $\varphi(x,w) = \|w^{1/2}x\|_+^2$. *Then the process* $\mathcal{Z}(\cdot, t)$ *in Corollary 3.4.1 can be represented as*

$$\mathcal{Z}(p,t) = \sup_{\theta \in H(p,\Theta_I)} \left\{ \|\mathcal{R}(p,\theta)\| \, t^{1/2} - \left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)} \right\rangle \right\}, \tag{3.4.4}$$

where

$$\mathcal{R}(p,\theta) := W^{-1/2}_{\mathcal{J}(\theta)} \left(\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)'\right)^{-1} \Pi_{\mathcal{J}(\theta)}(\theta)p.$$

Furthermore, suppose $W(\theta)$ *satisfies* $W_{\mathcal{J}(\theta)}(\theta) = [\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)']^{-1}$ *for any* $\theta \in \partial\Theta_I$. *Then the limiting process takes the form* $\mathcal{Z}(p,t) = \mu(t) + \mathcal{Z}^*(p)$ *with* $\mu(t) = t^{1/2}$ *and* $\mathcal{Z}^*(p) = \sup_{\theta \in H(p,\Theta_I)} -\langle [\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)']^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p, \mathbb{G}_{\mathcal{J}(\theta)}(\theta)\rangle$.

Equation 3.4.4 shows the limiting process $\mathcal{Z}(\cdot, t)$ depends on the multivariate Gaussian process $\mathbb{G}$, but again we note that the only selected components of $\mathbb{G}$ are relevant. Therefore, for each $p \in \mathbb{S}^{d-1}$, the asymptotic distribution of the normalized support function depends only on the active moment inequalities at each boundary point of the identified set. This is a common feature of the statistics studied in the literature (e.g. Rosen, 2008, Andrews and Soares, 2009).

If the weighting matrix satisfies $W_{\mathcal{J}(\theta)}(\theta) := [\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)']^{-1}$ at each boundary point, then straightforward algebra shows $\|\mathcal{R}(p,\theta)\| = 1$, which makes the first term in Eq. (3.4.4) independent of θ. With this choice of weighting matrix, the limiting process takes the form $\mathcal{Z}(p,t) := t^{1/2} + \mathcal{Z}^*(p)$[22].

We now make use of the representation result above to compare the weak limit of the Wald statistic with that of CHT's QLR statistic: $\sup_{\Theta_I} a_n Q_n(\theta)$. The QLR statistic can be written as

$$\sup_{\theta\in\Theta_I} a_n Q_n(\theta) = \max\left\{\sup_{\theta\in\partial\Theta_I} a_n Q_n(\theta), \sup_{\theta\in\Theta_I^o} a_n Q_n(\theta)\right\}.$$

As the second term on the right hand side asymptotically vanishes by Assumption 3.4.3 (ii-d), it suffices to study the first term. Using the local process $\tilde{\zeta}_n$, define

$$\mathcal{L}_n(p,u) := \sup_{\theta\in H(p,\Theta_I)} \sup_{\lambda\in K^-_{u,p}} \tilde{\zeta}_n(\theta,\lambda),$$

where $K^-_{u,p} := \{\lambda \in \mathbb{R}^d : \langle p,\lambda\rangle \le u\}$. Note that $\sup_{p\in\mathbb{S}^{d-1}} \mathcal{L}_n(p,0) = \sup_{\theta\in\partial\Theta_I} a_n Q_n(\theta)$. We therefore study the asymptotic behavior of the process $\mathcal{L}_n(\cdot,u)$ to study that of the QLR statistic. The following theorem establishes the weak convergence of $\mathcal{L}_n(\cdot,u)$. The regularity conditions for this theorem are given in the mathematical appendix.

Theorem 3.4.2: *Suppose the conditions of Corollary 3.4.2 hold. Suppose Assumption 3.B.1 holds. Then $\mathcal{L}_n(\cdot,u) \overset{u.d.}{\to} \mathcal{L}(\cdot,u)$ for each u, and the process $\mathcal{L}$ can be represented as*

$$\mathcal{L}(p,u) = \sup_{\theta\in H(p,\Theta_I)} \|\mathcal{R}(p,\theta)\|^{-1}\left(\left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\right\rangle + u\right)^2_+.$$

Based on this theorem, the following corollary establishes two equivalence results. The first result is the equivalence of the distributional limits of the Wald and the QLR statistics. The second result is the equality of the levels of the

[22]In sample, one may use a sample analog $\hat{W}_{n,\mathcal{J}_n(\theta)}(\theta) := (\hat{E}_{n,\mathcal{J}_n(\theta)}[\nabla m_\theta]\hat{E}_{n,\mathcal{J}_n(\theta)}[\nabla m_\theta]')^{-1}$ to construct Q_n. Here, for each n, $\mathcal{J}_n(\theta)$ is a mapping from Θ to a subset of $\{1,\cdots,J\}$ that selects (approximately) binding sample moment conditions at θ. Such moment selection mechanisms are studied in Andrews and Soares (2009).

criterion function used by the Wald approach and the QLR approach to construct confidence sets. Recall that $t^*_{1-\alpha} := \inf\{t : P(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\} \leq 0) \geq 1-\alpha\}$, and $\tau^*_{1-\alpha}$ is the asymptotic $1-\alpha$ quantile of the QLR statistic.

Corollary 3.4.3 (Asymptotic Equivalence for Moment Inequalities): *Suppose the conditions of Theorem 3.4.2 hold. Suppose $W(\theta)$ satisfies $W_{\mathcal{J}(\theta)}(\theta) = [\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)']^{-1}$ for each $\theta \in \partial\Theta_I$. Suppose Θ_I is strictly convex. For each $p \in \mathbb{S}^{d-1}$, let $\theta_I(p) \in \partial\Theta_I$ be the boundary point of Θ_I such that $H(p,\Theta_I) = \{\theta_I(p)\}$.*

Then, (i)

$$\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}_n(p,t)+t^{1/2}\}^2_+ \xrightarrow{d} \mathbf{Z} \quad \textit{and} \quad \sup_{\Theta_I} nQ_n(\theta) \xrightarrow{d} \mathbf{Z},$$

where

$$\mathbf{Z} := \sup_{p\in\mathbb{S}^{d-1}} \left\langle \left(\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))'\right)^{-1}\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))p, \mathbb{G}_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\right\rangle^2_+ .$$

(ii) $t^*_{1-\alpha} = \tau^*_{1-\alpha}$.

Corollary 3.4.3 shows that our Wald statistic (squared directed Hausdorff distance) and CHT's QLR statistic are asymptotically equivalent in the sense that they converge in distribution to the same limit, the supremum of a truncated and squared Gaussian process. The second result also has important consequeces. It implies the asymptotic equivalence of the Wald and QLR confidence sets for Θ_I. This is due to Theorem 3.3.6. When $t^*_{1-\alpha} = \tau^*_{1-\alpha}$, Theorem 3.3.6 implies

$$d_H\left(\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)}(t), \hat{\Theta}_n(\tau^*_{1-\alpha})\right) = o_p(a_n^{-1/\gamma}),$$

for any $0 \leq t \leq \tau^*_{1-\alpha}$. The first argument of d_H on the left hand side is the Wald confidence set, which is an expansion of the set estimator. The second argument is the QLR confidence set, which is a level set that uses an asymptotic quantile of the QLR statistic as a level. Despite the fundamental difference in ways these confidence sets are constructed, they are asymptotically equivalent in terms of the Hausdorff metric. These are fundamental results that establish the relationship between the Wald and QLR approaches.

3.5 Examples, Monte Carlo Experiments, and Applications

3.5.1 Examples

In this section, we analyze two examples studied in the literature using our inference method. The first model has an identified set that is a closed interval. Using this example, we illustrate our equivalence results in more detail and give a new interpretation to the results established by BM.

EXAMPLE 3.5.1 (Interval Identified Model): *Let X be an unobserved random variable with mean $\theta = E(X)$. Let X_1 and X_2 be observable random variables that satisfy the moment inequalities $E(X_1) \leq \theta \leq E(X_2)$.*

Let $\theta_1 = E(X_{1i})$ and $\theta_2 = E(X_{2i})$. The identified set for θ is a closed interval $\Theta_I = [\theta_1, \theta_2]$. Following the analysis in section 3.4, Θ_I can be characterized as a set of minimizers of the criterion function

$$Q(\theta) = \|W(\theta)^{1/2} E(m_\theta)\|_+^2,$$

where $m_\theta = (X_1 - \theta, \theta - X_2)'$. Define the sample criterion function by

$$Q_n(\theta) = \|\hat{W}_n(\theta)^{1/2} \hat{E}_n(m_\theta)\|_+^2.$$

For simplicity, we set $W(\theta)$ and $\hat{W}_n(\theta)$ to the identity matrix.

It is straightforward to show that these population and sample criterion functions satisfy Assumptions 3.4.1, 3.4.2, 3.4.3, and 3.4.4. The following results follow immediately from Corollaries 3.4.1 and 3.4.2, Corollary 3.5.1: *Let ι be a 2-by-1 vector of ones. Let $t \in \mathbb{R}_+$. Suppose*

$$(\sqrt{n}(\hat{E}_n(X_{1i}) - \theta_1), \sqrt{n}(\hat{E}_n(X_{2i}) - \theta_2))' \xrightarrow{d} N(0, \Omega)$$

and $-\infty < \theta_1 < \theta_2 < \infty$.

Then $\sqrt{n} d_H(\hat{\Theta}_n(t), \Theta_I) \xrightarrow{d} \max\{|\mathcal{Z}(-1,t)|, |\mathcal{Z}(1,t)|\}$ *and* $\sqrt{n}\vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) \xrightarrow{d} \max\{-\mathcal{Z}(-1,t)_+, -\mathcal{Z}(1,t)_+\}$, *where* $\mathcal{Z}(p,t)$ *is a Gaussian process on* $\mathbb{S}^0 = \{-1, 1\}$ *with mean* $t^{1/2}\iota$ *and covariance kernel* $E[\mathcal{Z}(-1,t)\ \mathcal{Z}(-1,t)] = \Omega_{11}$, $E[\mathcal{Z}(1,t)\mathcal{Z}(1,t)] = \Omega_{22}$, *and* $E[\mathcal{Z}(-1,t)\ \mathcal{Z}(1,t)] = -\Omega_{12}$.

This result is closely related to that presented by BM (Theorem 3.1), which shows that the normalized support function of their set average estimator weakly converges to a zero-mean Gaussian process that has the same covariance kernel as $\mathcal{Z}(\cdot, t)$. In fact, if we set $t = 0$, the level set estimator is analytically identical to their set-average estimator for this class of problems. An additional interesting result is that, under this choice of the weighting matrix, the squared directed Hausdorff distance is asymptotically equivalent to CHT's QLR-statistic. We summarize these equivalence results as follows:

Theorem 3.5.1: *Let the assumptions of Theorem 3.5.1 hold. Let* $\mathcal{W}_n := \sqrt{n}\vec{d}_H(\Theta_I, \hat{\Theta}_n(0))$. *Let* $\mathcal{QLR}_n := \sup_{\theta \in \Theta_I} nQ_n(\theta)$ *be CHT's QLR statistic. Let* $\tilde{\mathcal{W}}_n := \sqrt{n}\vec{d}_H(\Theta_I, \tilde{\Theta}_n)$ *be BM's Wald statistic, where* $\tilde{\Theta}_n = n^{-1} \bigoplus_{i=1}^n F_i$ *and* $F_i = [X_{1i}, X_{2i}]$ *for* $i = 1, \cdots, n$. *Let* $\mathcal{Z}$ *be the process given in Corollary 3.5.1. Then*

$$\mathcal{W}_n^2 \xrightarrow{d} \max\{(-\mathcal{Z}(-1,0))_+^2, (-\mathcal{Z}(1,0))_+^2\} \tag{3.5.1}$$

$$\mathcal{QLR}_n \xrightarrow{d} \max\{(-\mathcal{Z}(-1,0))_+^2, (-\mathcal{Z}(1,0))_+^2\} \tag{3.5.2}$$

$$\tilde{\mathcal{W}}_n^2 \xrightarrow{d} \max\{(-\mathcal{Z}(-1,0))_+^2, (-\mathcal{Z}(1,0))_+^2\}. \tag{3.5.3}$$

The asymptotic equivalence of CHT's QLR statistic and BM's Wald statistic in equations (3.5.2) and (3.5.3) is due to BM's Theorem 3.1. Here, Theorem 3.5.1 adds eq. (3.5.1).

As we have seen in the previous section, the squared directed Hausdorff distance becomes asymptotically equivalent to CHT's QLR statistic when the weighting matrix satisfies the conditions of Corollary 3.4.3. As the identity matrix satisfies these, the asymptotic equivalence of $\mathcal{W}_n^2$ and $\mathcal{QLR}_n$ follows[23]. Further, for this example, the set-average estimator is a set of minimizers of the truncated squared loss function; this therefore becomes a level-set estimator with $t = 0$. Thus, the "exact" equivalence of $\mathcal{W}_n^2$ and $\tilde{\mathcal{W}}_n^2$ holds. In sum, the asymptotic equivalence result formerly presented by BM can be understood as a combination of (i) the asymptotic equivalence of the Wald statistic and the QLR statistic for the class of moment inequality models and (ii) the equivalence of the level-set estimator and the set-average estimator under the specific choice of criterion function.

In this example, we may interpret BM's set-average estimator as a set-valued quasi maximum likelihood estimator (QMLE) of Θ_I, where the quasi-log likelihood function is the truncated squared loss used by CHT. This is analogous to the point identified case, where the sample average is the QMLE for the location parameter, under the specification that the data are randomly sampled from a normal distribution, which gives a squared error loss function. It is of interest to extend this notion to a more general class of problems.

The second example studies a regression model with interval-valued outcome variables. Our Monte Carlo experiments will be based on this example.

Example 3.5.2 (Regression with Interval-Censored Outcome): *Let* $\theta \in \Theta \subset \mathbb{R}^d$. *Consider the DGP*:

$$Y_i = X_i'\theta + \epsilon_i, \quad i = 1, 2, \cdots, n,$$

where $E[\epsilon_i|X_i] = 0$ *for all* $i = 1, \cdots, n$. *The outcome variable* Y_i *is not observed but the outcome interval* $[Y_{1i}, Y_{2i}]$ *is observed for each* $i = 1, \cdots, n$. *The outcome interval satisfies the following moment inequalities*

$$E[Y_{1i}|X_i] \leq X_i'\theta \leq E[Y_{2i}|X_i], \quad a.s.$$

[23] Here, we use $W(\theta) = I_2$, the identity matrix. Since the two constraints don't bind at the same time, the weighting matrix for the equivalence should satisfy $W_1(\theta_1) = (\Pi_1(\theta_1)\Pi_1(\theta_1)')^{-1} = 1$ and $W_2(\theta_1) = (\Pi_2(\theta_2)\Pi_2(\theta_2)')^{-1} = 1$. Obviously, the identity matrix satisfies this condition.

The identified set can be defined as the set of minimizers of the following criterion function:

$$Q(\theta) := \int (E(Y_{1i}|X_i = x) - x'\theta)_+^2 + (x'\theta - E(Y_{2i}|X_i = x))_+^2 dP(x).$$

Following Chernozhukov, Hong, and Tamer (2004), we use the following minimum-distance type sample criterion function

$$Q_n(\theta) := \frac{1}{n}\sum_{i=1}^{n}(\hat{E}_n(Y_1|X_i) - X_i'\theta)_+^2 + (X_i'\theta - \hat{E}_n(Y_2|X_i))_+^2,$$

where $\hat{E}_n(Y_1|X_i)$ and $\hat{E}_n(Y_2|X_i)$ are estimators of $E(Y_{1i}|X_i)$ and $E(Y_{2i}|X_i)$, respectively.

In our Monte Carlo experiments, we study the cases where X_i is a vector of discrete random variables supported on $\{x_1, \cdots, x_J\}$, $J \in \mathbb{N}$. For these cases, we use $\hat{E}_n(Y_1|X_i = x_j) = n_j^{-1}\sum_{i:X_i=x_j} Y_{1i}$ and $\hat{E}_n(Y_2|X_i = x_j) = n_j^{-1}\sum_{i:X_i=x_j} Y_{2i}$ as estimators, where $n_j = \sum_i 1\{X_i = x_j\}$.

The sample criterion function can be alternatively written as

$$Q_n(\theta) = \|\hat{W}_n^{1/2}(\theta)\hat{E}_n(m_\theta)\|_+^2,$$

where m_θ is a $2J$-dimensional vector whose components are

$$m_{j,\theta} := \begin{cases} \frac{n}{n_j}(Y_{1i} - X_i'\theta)1_{\{X_i=x_j\}} & \text{for } j = 1, \cdots, J \\ \frac{n}{n_j}(X_i'\theta - Y_{2i})1_{\{X_i=x_j\}} & \text{for } j = J+1, \cdots, 2J. \end{cases}$$

The weighting matrix $\hat{W}_n(\theta)$ is a $2J \times 2J$ diagonal matrix whose j-th diagonal element is n_j/n. Therefore, the example above can be studied within the framework presented in section 3.4. It can be shown that this model also satisfies the conditions of Theorem 3.4.1. We use this example to evaluate our inference method.

3.5.2 Monte Carlo Experiments

We conduct a Monte Carlo experiment using Example 3.5.2 to examine the performance of our inference methods. In both designs, the regressor $X_i = (1, X_{2i})$ consists of a constant and a real random variable X_{2i}.

We use the same data as Chernozhukov, Hong, and Tamer (2004). The original data is taken from March 2000 wave of the Current Population Survey (CPS) with 13,290 observations on income and education[24]. Chernozhukov, Hong, and Tamer (2004) bracketed each individual's log-income into 15 different categories. We let Y_{1i} (Y_{2i}) be the lower (upper) bound of the bracketed log-income of each individual. Chernozhukov, Hong, and Tamer (2004, sec 4) provide further details of the construction of the data. We randomly draw samples of size $n = 1,000$ or $2,000$ from this CPS population and check the coverage probabilities of the Wald confidence set under different values of the subsample size b.

For this exercise, we use a grid of points $\{p_l = (\cos(w_l), \sin(w_l))', l = 1, \cdots, L\}$ with $L = 100$, where each w_l is taken from an equally spaced grid of points on the interval $[0, 2\pi]$. We use Algorithm 3.2.1 to compute the support function of the set estimator. This algorithm is quite fast. For samples of size $n = 1,000$ and $2,000$, it takes only 0.051 and 0.054 seconds respectively to compute the support function and approximate boundary $\{\hat{\theta}_n(p_l, t), l = 1, \cdots, L\}$ of the set estimator $\hat{\Theta}_n(t)$[25].

The initial choice of the level t is made in a similar manner to Chernozhukov, Hong, and Tamer (2004). First, we consider an auxiliary point-identified model, where the lower and upper bounds for the individual log-income are $\tilde{Y}_{1i} = \tilde{Y}_{2i} = (Y_{1i} + Y_{2i})/2$ for any i. We use the criterion function Q_n applied to the data $\{(\tilde{Y}_{1i}, \tilde{Y}_{2i}), i = 1, \cdots, n\}$ and compute quantiles of the statistic $n(Q_n(\theta_0^a) - Q_n(\hat{\theta}_n^a))$, where θ_0^a is the minimizer of the population criterion function and $\hat{\theta}_n^a$ its estimator. Let $\hat{t}_{a_0}$ be the $100 \times a_0\%$ quantile of the statistic. Chernozhukov, Hong, and Tamer

[24] We use the dataset that is distributed with a Matlab package by Beresteanu, Molinari, and Wang (2009).

[25] The reported values are the average elapsed time to compute the support function and approximated boundary of the set estimator from simulated samples of size $n = 1,000$ or 2,000 drawn for $S = 2,000$ times. Computation was implemented by a code written in R (and partly in C) on a computer with Intel Core 2 Quad CPU 2.5 Ghz and 6GB memory.

(2004) recommends using t_{a_0} with properly chosen a_0 as an initial level. If $\hat{t}_{a_0}$ is too large, the resulting set estimator may not be expanded because its coverage probability is likely to exceed $1-\alpha$. Therefore, it is desirable to set a_0 to a value less than the nominal level $1-\alpha$.

Table 3.1 reports the coverage probabilities of the Wald confidence set under different values of b and a_0. We also include the coverage probabilities of the QLR confidence set reported by Chernozhukov, Hong, and Tamer (2004) for comparison. The nominal level is $1-\alpha = 0.95$. The QLR benchmark is reported in the third row. We first set $a_0 = 0.5$ (median) for choosing the initial level. Overall, the Wald confidence set's coverage probabilities are close to those of the QLR confidence set, which supports our theoretical results. We also report the coverage probabilities for the case $a_0 = 0.75$. Under this initial choice of level, the Wald confidence sets' coverage probabilities are closer to the nominal level in every case.

We note that the QLR confidence sets' coverage probabilities improve as we move from $n = 1{,}000$ to $n = 2{,}000$. This behavior is not so apparent for the Wald confidence sets.

3.6 Conclusion

In this paper, we introduce an inference framework for partially identified econometric models that unifies two general approaches recently proposed in the literature: the criterion function approach and the support function approach. This yields inference tools that have the wide applicability of the criterion function approach and the computational tractability of the support function approach.

We consider the general case where the identified set Θ_I is the set of minimizers of a criterion function, estimated as an appropriate level set of a sample criterion function, following CHT, and represented as a support function, as in BM. This yields Wald-type inference methods, significantly extending recent work of BM and Bontemps, Magnac, and Maurin (2007), each of which studied special classes of econometric models. Specifically, given a compact convex set Θ_0 or a point θ_0, we present tests for set equality $H_0 : \Theta_I = \Theta_0$, set inclusion $H_0 : \Theta_0 \subseteq \Theta_I$, and point inclusion $H_0 : \theta_0 \in \Theta_I$.

The test for set equality can be inverted to construct a confidence collection that contains the identified set as an element, with a specified confidence level. This type of inference is as yet unavailable within CHT's framework. The test for set inclusion can be inverted to construct another confidence collection, containing each subset of the identified set as an element. Taking the union of the elements of this collection yields a confidence set that covers the identified set, comparable to CHT's confidence set. We provide a new, practical step-up algorithm for selecting the level t used to construct this confidence set. This removes the arbitrariness in the choice of t characterizing previous methods. The test for point inclusion can be inverted to construct a confidence set for each point in the identified set, comparable to methods of Imbens and Manski (2004), CHT, Romano and Shaikh (2008), and Andrews and Guggenberger (2009).

We also contribute to the literature on moment inequality models by establishing the asymptotic equivalence of our Wald statistic and CHT's QLR statistic. We show that this implies the asymptotic equivalence of the Wald confidence set and CHT's confidence set. This equivalence suggests that further investigation into the general relationship between these two approaches, beyond the moment inequality framework, is an interesting topic for future research.

Another interesting direction for further research is the development of Lagrange Multiplier (LM)-type analogs of the Wald-type statistics analyzed here. One may expect that under suitable conditions, LM- and Wald-type statistics may also be asymptotically equivalent in partially identified models, and that under further conditions, these may be asymptotically equivalent to QLR-type statistics. Obtaining these equivalence conditions is an interesting direction for future research.

For testing hypotheses and constructing confidence collections and confidence sets, we propose a general subsampling procedure. This procedure is valid pointwise, as we derive our results under a fixed probability measure. As Romano and Shaikh (2008, 2009) and Andrews and Guggenberger (2009) point out, however, establishing the uniform asymptotic validity of subsampling is important for partially identified models and is one of our future tasks.

3.A Tables

Table 3.1: Coverage Probabilities ($1-\alpha$=.95) of Wald and QLR Confidence Sets

	Subsample Size				
n=1,000					
	$b=50$	$b=80$	$b=120$	$b=200$	$b=300$
Wald ($a_0=0.5$)	0.878	0.887	0.886	0.889	0.909
Wald ($a_0=0.75$)	0.957	0.962	0.953	0.956	0.956
QLR	0.851	0.880	0.872	0.931	0.912
n=2,000					
	$b=200$	$b=300$	$b=400$	$b=500$	$b=600$
Wald ($a_0=0.5$)	0.874	0.873	0.874	0.880	0.888
Wald ($a_0=0.75$)	0.937	0.925	0.925	0.930	0.928
QLR	0.861	0.882	0.905	0.950	0.933

Note: Empirical coverage probabilities of the Wald and QLR confidence sets under different values of subsample size b. The coverage probabilities of the QLR confidence set are taken from Chernozhukov, Hong, and Tamer (2004). Monte Carlo simulations $m=2,000$, subsample replications $B=2,000$, significance level $\alpha=0.05$.

3.B Mathematical Appendix

3.B.1 Consistency and Rate of Convergence of the Level Set Estimator

We summarize below CHT's consistency and the rate of convergence result. Assumption 3.2.3 (i) requires one-sided uniform convergence of Q_n to its population counterpart, which is slightly more general than usual uniform convergence $\sup_{\theta\in\Theta}|Q(\theta)-Q_n(\theta)|=o_p(1)$. Assumption 3.2.3 (ii) is one of the key conditions utilized by CHT, requiring the sample criterion function to approximate the population counterpart at $1/a_n$ rate over Θ_I. This condition ensures that their QLR-statistic $\sup_{\theta\in\Theta_I}a_nQ_n(\theta)$ is nondegenerate. Assumption 3.2.3 (iii) requires the existence of a polynomial function in the distance from Θ_I, which stochastically minorizes (bounds from below) the sample criterion function in a neighborhood of the identified set. It is then immediate from CHT's Theorem 3.1 that the following results hold.

Theorem 3.B.1 (Consistency and Convergence rate): *Let t be a positive finite constant. Let $t_n = t\kappa_n$ where κ_n is a positive slowly increasing sequence such that $\kappa_n \to \infty$ and $\kappa_n/a_n = o_p(1)$. Suppose Assumptions 3.2.1, 3.2.2, and 3.2.3 (i), (ii) hold. Then, with probability approaching 1, $\hat{\Theta}_n(t) \subseteq \Theta_I$ and $\Theta_I \subseteq \hat{\Theta}_n(t_n)$. Furthermore, $d_H(\hat{\Theta}_n(t_n), \Theta_I) = o_p(1)$. Suppose, in addition, Assumption 3.2.3 (iii) holds. Then, $r_n d_H(\hat{\Theta}_n(t_n), \Theta_I) = O_p(1)$ with $r_n = (a_n/\max\{1, \kappa_n\})^{1/\gamma}$.*

For the proof, see CHT's Theorem 3.1.

3.B.2 Proof of Lemma 3.3.1, 3.3.2 , and Lemma 3.3.3

Definition 3.B.1 (Level boundedness): *The function $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ is* level-bounded *if the level sets $\{x : f(x) \leq \alpha\}$ are bounded for any $\alpha \in \mathbb{R}$.*

If a function $f : \Omega \times \mathbb{R}^d \to \bar{\mathbb{R}}$ is such that $f(\omega, \cdot)$ is level bounded for all $\omega \in F \in \mathfrak{F}$, $P(F) = 1$, then we say f is level bounded almost surely ($a.s.$).

Proof of Lemma 3.3.1. Note that Assumption 3.2.1 ensures that ζ_n is proper. In addition, the compactness of Θ and the assumption that $Q_n(\omega, \theta) = \infty$ $a.s.$ for $\theta \notin \Theta$ ensure that ζ_n is level-bounded almost surely. For each $\omega \in \{\omega : \zeta_n \text{ is lsc}\}$, we have

$$\begin{aligned}
s(p, \hat{\Theta}_n(t)) < u &\Leftrightarrow \sup_{\theta \in \hat{\Theta}_n(t)} \langle p, \theta \rangle < u \\
&\Leftrightarrow \langle p, \theta \rangle < u, \quad \forall \theta \in \hat{\Theta}_n(t) \\
&\Leftrightarrow \hat{\Theta}_n(t) \subseteq \Theta \setminus K_{u,p} \\
&\Leftrightarrow K_{u,p} \cap \Theta \subseteq \Theta \setminus \hat{\Theta}_n(t) \\
&\Leftrightarrow \zeta_n(\theta) > t, \quad \forall \theta \in K_{u,p} \cap \Theta \\
&\Leftrightarrow \inf_{\theta \in K_{u,p} \cap \Theta} \zeta_n(\theta) > t,
\end{aligned}$$

where the second equivalence follows from the compactness of $\hat{\Theta}_n(t)$, which is implied by the lower semicontinuity and the level-boundedness of ζ_n, and the last equivalence follows from the properness and the lower semicontinuity of ζ_n and the compactness of $K_{u,p} \cap \Theta$. □

Proof of Lemma 3.3.2. Note first that, under our assumptions, $\tilde{\zeta}_n$ inherits the almost sure properness, lower semicontinuity, and level-boundedness from ζ_n. For any $0 < \epsilon < \bar{\epsilon}$, let $D^*_{n,\epsilon} := \{(\theta, \lambda) \in R_{u,p} : \tilde{\zeta}_n(\theta, \lambda) < \inf_{R_{u,p}} \tilde{\zeta}_n(\theta, \lambda) + \epsilon\}$ and $D^*_\epsilon := \liminf_{n\to\infty} D^*_{n,\epsilon}$. By hypothesis, D^*_ϵ is nonempty. For a given $\delta > 0$, let $D^*_{\epsilon,\delta}$ be an open δ-envelope of D^*_ϵ defined by $D^*_{\epsilon,\delta} := \{(\theta, \lambda) : d((\theta, \lambda), D^*_\epsilon) < \delta\}$.

For any $\delta > 0$, $R_{u,p} \cap D^*_{\epsilon,\delta} \neq \varnothing$ implies that there exists $N_\epsilon \in \mathbb{N}$ such that $R_{n,u,p} \cap D^*_{\epsilon,\delta} \neq \varnothing$ for all $n \geq N_\epsilon$ as $R_{n,u,p} \to R_{u,p}$ in the Painlevé-Kuratowski sense (Theorem 4.5, Rockafellar and Wets, 2005). For $n \geq N_\epsilon$, let $E_{n,\epsilon} := \arg\min_{R_{n,u,p} \cap D^*_{\epsilon,\delta}} \tilde{\zeta}_n(\theta, \lambda)$. Let $D_n := \arg\min_{R_{n,u,p}} \tilde{\zeta}_n(\theta, \lambda)$. As $E_n \neq \varnothing$, $E_{n,\epsilon} \subseteq D^*_{\epsilon,\delta}$, and $E_{n,\epsilon} \subseteq D_n$, we have $D_n \cap D^*_{\epsilon,\delta} \neq \varnothing$ for all $n \geq N_\epsilon$ and $\delta > 0$.

Now suppose that the conclusion of the lemma does not hold. Then, there exists a subsequence $\{(\tilde{\zeta}_{n_k}, R_{n_k}), k = 1, 2, \cdots\}$ such that

$$P\left(\left|\inf_{R_{n_k,u,p}} \tilde{\zeta}_{n_k}(\theta, \lambda) - \inf_{R_{u,p}} \tilde{\zeta}_{n_k}(\theta, \lambda)\right| \geq 2\epsilon\right) > 0.$$

for all k. Then, along this subsequence, we have $P(D_{n_k} \cap D^*_{n_k,\epsilon} = \varnothing) > 0$. This implies $D_{n_k} \cap D^*_{\epsilon,\delta} = \varnothing$ for all k with positive probability, which is a contradiction. □

Proof of Lemma 3.3.3. Let $\epsilon > 0$ be arbitrary. For each $(u, p) \in \mathbb{R} \times \mathbb{S}^{d-1}$, take $L_{u,p}$ to be a compact set such that $D^*_\epsilon \subseteq L_{u,p}$. This is possible by the equi-inf-compactness. Now take $\tilde{R}_{u,p} = R_{u,p} \cap L_{u,p}$. Then, by construction,

$$P\left(\left|\inf_{\tilde{R}_{u,p}} \tilde{\zeta}_n(\theta, \lambda) - \inf_{R_{u,p}} \tilde{\zeta}_n(\theta, \lambda)\right| \geq \epsilon\right) \leq \epsilon \tag{3.B.1}$$

for sufficiently large n. Given the discussions preceding this lemma, we have

$$\begin{aligned}\mathcal{Z}_n(p,t) < u &\Leftrightarrow \inf_{R_{n,u,p}} \zeta_n(\theta + \lambda/a_n^{1/\gamma}) > t \\ &\Leftrightarrow \inf_{R_{n,u,p}} \tilde{\zeta}_n(\theta,\lambda) > t.\end{aligned}$$

Since this holds for any finite m-tuple $\{(u_j,p_j)\}_{j=1}^m$, we have

$$\begin{aligned}&P(\mathcal{Z}_n(p_1,t) < u_1, \cdots, \mathcal{Z}_n(p_m,t) < u_m) \\ &\qquad = P\left(\inf_{R_{n,u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \cdots, \inf_{R_{n,u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t\right). \qquad (3.B.2)\end{aligned}$$

Note that

$$\begin{aligned}P&\left(\inf_{\tilde{R}_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t+\epsilon, \cdots, \inf_{\tilde{R}_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t+\epsilon\right) \\ &\le P\left(\max_{1\le j\le m}\left|\inf_{\tilde{R}_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda) - \inf_{R_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda)\right| \ge \epsilon/2\right) \\ &\quad + P\left(\inf_{R_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t+\epsilon/2, \cdots, \inf_{R_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t+\epsilon/2\right) \\ &\le P\left(\max_{1\le j\le m}\left|\inf_{\tilde{R}_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda) - \inf_{R_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda)\right| \ge \epsilon/2\right) \qquad (3.B.3) \\ &\quad + P\left(\max_{1\le j\le m}\left|\inf_{R_{u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda) - \inf_{R_{n,u_j,p_j}} \tilde{\zeta}_n(\theta,\lambda)\right| \ge \epsilon/2\right) \\ &\quad + P\left(\inf_{R_{n,u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \cdots, \inf_{R_{n,u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t\right),\end{aligned}$$

where we used the fact that, for any random vectors $Y_n, X_n : \Omega \to \mathbb{R}^m$, an open set $G \subset \mathbb{R}^m$, and its ϵ-contraction $G_\epsilon := \{x \in G : \rho(x, G^c) \ge \epsilon\}$, we have $P(Y_n \in G_\epsilon) \le P(\rho(X_n,Y_n) \ge \epsilon) + P(X_n \in G)$. Specifically, we used the metric $\rho(X_n,Y_n) = \max_{1\le j\le m} |X_{j,n} - Y_{j,n}|$ and the open set $G = (t,\infty)^m$.

Lemma 3.3.2 and (3.B.1) ensure that the first two terms on the right hand side of (3.B.3) become arbitrarily small as n gets large. Therefore,

$$\liminf_{n\to\infty} P\left(\inf_{\tilde{R}_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t+\epsilon, \cdots, \inf_{\tilde{R}_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t+\epsilon\right)$$
$$\leq \liminf_{n\to\infty} P\left(\inf_{R_{n,u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \cdots, \inf_{R_{n,u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t\right).$$

By letting $\epsilon \downarrow 0$, we obtain

$$\liminf_{n\to\infty} P\left(\inf_{\tilde{R}_{u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \cdots, \inf_{\tilde{R}_{u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t\right)$$
$$\leq \liminf_{n\to\infty} P\left(\inf_{R_{n,u_1,p_1}} \tilde{\zeta}_n(\theta,\lambda) > t, \cdots, \inf_{R_{n,u_m,p_m}} \tilde{\zeta}_n(\theta,\lambda) > t\right)$$
$$= \liminf_{n\to\infty} P(\mathcal{Z}_n(p_1,t) < u_1, \cdots, \mathcal{Z}_n(p_m,t) < u_m),$$

where the last equality follows from Eq. (3.B.2). □

3.B.3 Proof of Theorem 3.3.1 (i), (ii) and Auxiliary Lemmas

Our first goal in this sections is to show that the stochastic process $\mathcal{Z}(\cdot,t)$ given in Eq. (3.3.3) in Theorem 3.3.1 satisfies $\{\omega : \mathcal{Z}(p,t) < u\} = \{\omega : \inf_{\tilde{R}_{u,p}} \tilde{\zeta}(\theta,\lambda) > t\}$ for any $u,p \in \mathbb{R}\times\mathbb{S}^{d-1}$. For this, we need to show the almost sure upper semicontinuity of the map $g : \theta \mapsto s(p,\hat{\Lambda}(t,\theta))$. In the following, we introduce a regularity condition for the criterion function and two lemmas that are useful for establishing the desired result. We then prove Theorem 3.3.1 (i) and (ii).

Definition 3.B.2 (Level-boundedness for parametric optimization): *A function $f : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ with values $f(x,u)$ is level-bounded in x locally uniformly in u if for each $\bar{u} \in \mathbb{R}^m$ and $\alpha \in \mathbb{R}$ there is a neighborhood $V \in \mathcal{N}(\bar{u})$ along with a bounded set $B \subset \mathbb{R}^n$ such that $\{x | f(x,u) \leq \alpha\} \subset B$ for all $u \in V$; or equivalently, there is a neighborhood $V \in \mathcal{N}(\bar{u})$ such that the set $\{(x,u) | u \in V, f(x,u) \leq \alpha\}$ is bounded in $\mathbb{R}^n \times \mathbb{R}^m$.*

Lemma 3.B.1: *Consider*

$$\psi(u) := \inf_x f(x,u)$$

in the case of a proper, lsc function $f : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ *such that* $f(x,u)$ *is level bounded in* x *locally uniformly in* u*. Then, the function* ψ *is proper and lsc on* $\mathbb{R}^m$*.*

Proof. See Theorem 1.17 in Rockafellar and Wets (2005) □

Lemma 3.B.2: *Suppose that* $\tilde{\zeta}_n(\theta,\lambda)$ *satisfies the conditions of Theorem 3.3.1. For each* $t \in \mathbb{R}_+$ *and* $p \in \mathbb{S}$*, let* g *be a stochastic process defined by* $g : \theta \mapsto s\left(p, \hat{\Lambda}(t,\theta)\right)$*. Then, there is a representation of* g*, which is upper semicontinuous (usc) almost surely.*

Proof. First, let $\delta_A : \mathbb{R}^d \to \bar{\mathbb{R}}_+$ be the optimization theory indicator function that takes 0 if $x \in A$ and ∞ otherwise. For each θ, let $h(\theta,\lambda) := -\langle p,\lambda\rangle + \delta_{\hat{\Lambda}(t,\theta)}(\lambda)$ and $\tilde{g}(\theta) := \inf_\lambda h(\theta,\lambda)$. As $g(\theta) = -\tilde{g}(\theta)$, it suffices to show the lower semicontinuity of $\tilde{g}(\theta)$ for the conclusion of the lemma[26]. For establishing the lower semicontinuity of $\tilde{g}$, we make use of Lemma 3.B.1 by taking $\psi = \tilde{g}$, $f = h$, and $(x,u) = (\theta,\lambda)$. Below, we show that h is almost surely proper, lsc, and level bounded in λ locally uniformly in θ.

By our hypothesis, $\hat{\Lambda}(t,\theta)$ is nonempty *a.s.* for any $\theta \in \partial\Theta_I$ and $t \in \mathbb{R}_+$. Therefore, $\delta_{\hat{\Lambda}(t,\theta)}$ is proper, which implies that h is proper. In the following, using Skrokhod representation, we take a version of $\tilde{\zeta}_n$ that is epiconverging almost surely to a version of $\tilde{\zeta}$ that are defined on some common probability space. This is possible since the space of proper lsc functions equipped with a metric that metrizes the topology of epiconvergence is complete and separable (cf. Rockafellar and Wets, 2005). The almost sure epiconvergence of lsc functions $\{\tilde{\zeta}_n, n \geq 1\}$ implies that $\tilde{\zeta}$ is lsc *a.s.* (Attouch, 1984, Theorem 2.1). Therefore, the level set $\hat{\Lambda}(t,\theta)$ of the lsc function $\tilde{\zeta}(\theta,\cdot)$ is closed *a.s.* Note that $-\langle p,\lambda\rangle$ is continuous and $\delta_{\hat{\Lambda}(t,\theta)}$ is lsc by the closedness of $\hat{\Lambda}(t,\theta)$. So, h is lsc *a.s.*

For each $p \in \mathbb{S}^{d-1}$ and $\bar{\theta} \in H(p,\Theta_I)$, let $\mathcal{N}(\bar{\theta})$ be a collection of neighbor-

[26] We follow the convention that $\sup_{x\in C} f(x) = -\infty$ if C is an empty set.

hoods at $\bar{\theta}$. Let $\alpha \in \mathbb{R}$. Take $\theta \in V \in \mathcal{N}(\bar{\theta})$. Define the set

$$C := \{\lambda : \tilde{\zeta}(\theta, \lambda) \leq t, \quad \langle p, \lambda \rangle \geq -\alpha\}.$$

The fact that $\tilde{\zeta}_n$ is equi-inf-compact implies that $\tilde{\zeta}$ is level bounded (Exercise 7.32 (b) in Rockafellar and Wets (2005)), and therefore $\hat{\Lambda}(t, \theta)$ is bounded *a.s.* As $C \subseteq \hat{\Lambda}(t, \theta)$, C is bounded *a.s.* Now, we can rewrite

$$\begin{aligned} C &= \{\lambda : \delta_{\Lambda(t,\theta)}(\lambda) = 0, \quad -\langle p, \lambda \rangle \leq \alpha\} \\ &= \{\lambda : h(\theta, \lambda) \leq \alpha\}. \end{aligned}$$

Therefore, $h(\theta, \lambda)$ is level bounded in λ locally uniformly in θ. By Lemma 3.B.1, $\tilde{g}(\theta)$ is lsc almost surely. □

Given the results above, we fist prove the statement of Theorem 3.3.1 (ii).

Proof of Theorem 3.3.1 (ii). For each $(u, p) \in \mathbb{R} \times \mathbb{S}^{d-1}$, take $L_{u,p}$ to be a compact set such that $\hat{\Lambda}(t, \theta) \subset L_{u,p}$ *a.s.* For a given $\theta \in H(p, \Theta_I)$, it is straightforward to show

$$s(p, \hat{\Lambda}(t, \theta)) < u \Leftrightarrow \inf_{\lambda \in K_{u,p} \cap L_{u,p}} \tilde{\zeta}(\theta, \lambda) > t, \tag{3.B.4}$$

using an argument similar to the proof of Lemma 3.3.1. By the compactness of $H(p, \Theta_I)$ and Lemma 3.B.2,

$$\begin{aligned} \mathcal{Z}(p, t) < u &\Leftrightarrow \sup_{\theta \in H(p, \Theta_I)} s(p, \hat{\Lambda}(t, \theta)) < u \\ &\Leftrightarrow s(p, \hat{\Lambda}(t, \theta)) < u, \quad \forall \theta \in H(p, \Theta_I). \end{aligned} \tag{3.B.5}$$

Combining Eqs. (3.B.4) and (3.B.5), we obtain

$$\mathcal{Z}(p, t) < u \Leftrightarrow \inf_{(\theta, \lambda) \in \tilde{R}_{u,p}} \tilde{\zeta}(\theta, \lambda) > t,$$

where $\tilde{R}_{u,p} = H(p, \Theta_I) \times (K_{u,p} \cap L_{u,p})$. Therefore, for any finite m-tuple $\{(u_j, p_j)\}_{j=1}^m$,

$$\begin{aligned}&\{\omega : \mathcal{Z}(p_1, t) < u_1, \cdots, \mathcal{Z}(p_m, t) < u_m\} \\ &\qquad = \left\{\omega : \inf_{(\theta,\lambda)\in\tilde{R}_{u_1,p_1}} \tilde{\zeta}(\theta, \lambda) > t, \cdots, \inf_{(\theta,\lambda)\in\tilde{R}_{u_m,p_m}} \tilde{\zeta}(\theta, \lambda) > t\right\}.\end{aligned}$$

Take probability both sides. Then, the conclusion of Theorem 3.3.1 (ii) follows. □

Proof of Theorem 3.3.1 (i). Consider a finite m-tuple $\{(u_j, p_j)\}_{j=1}^m$. Since $\tilde{\zeta}_n \overset{e.d.}{\to} \tilde{\zeta}$, for any $\{(u_j, p_j)\}_{j=1}^m$, we have

$$\begin{aligned}&P\left(\inf_{(\theta,\lambda)\in\tilde{R}_{u_1,p_1}} \tilde{\zeta}(\theta, \lambda) > t, \cdots, \inf_{(\theta,\lambda)\in\tilde{R}_{u_m,p_m}} \tilde{\zeta}(\theta, \lambda) > t\right) \\ &\quad \le \liminf_{n\to\infty} P\left(\inf_{(\theta,\lambda)\in\tilde{R}_{u_1,p_1}} \tilde{\zeta}_n(\theta, \lambda) > t, \cdots, \inf_{(\theta,\lambda)\in\tilde{R}_{u_m,p_m}} \tilde{\zeta}_n(\theta, \lambda) > t\right).\end{aligned}$$

This result, together with Lemma 3.3.3 and Theorem 3.3.1 (ii) proved above, implies that

$$\begin{aligned}&P(\mathcal{Z}(p_1, t) < u_1, \cdots, \mathcal{Z}(p_m, t) < u_m) \\ &\qquad \le \liminf_{n\to\infty} P\left(\mathcal{Z}_n(p_1, t) < u_1, \cdots, \mathcal{Z}_n(p_m, t) < u_m\right).\end{aligned}$$

By the portmanteau theorem, the process $\mathcal{Z}_n(\cdot, t)$ weakly converges to $\mathcal{Z}(\cdot, t)$ in finite dimension. This completes the proof of part (i). □

3.B.4 Proof of Theorem 3.3.1 (iii)

For establishing $\mathcal{Z}_n(\cdot, t) \overset{u.d.}{\to} \mathcal{Z}(\cdot, t)$, we make use of the three lemmas below. Lemmas 3.B.3 and 3.B.4 will be used to show the tightness of the sequence $\{\mathcal{Z}_n(\cdot, t), n = 1, 2, \cdots\}$. Lemma 3.B.3 states that, for the tightness, it suffices to show the stochastic equicontinuity of the process, and Lemma 3.B.4 gives a sufficient condition for the stochastic equicontinuity.

LEMMA 3.B.3 (Tightness Characterization): *Let $\mathbb{E}$ be a metric space. A sequence of stochastic processes $\{\xi_n(x), n \geq 1\}$ is tight in $(\mathcal{C}(\mathbb{E}), d_{\mathcal{C}(\mathbb{E})})$ if and only if $\xi_n(x) = O_p(1)$ for all $x \in \mathbb{E}$ and the stochastic equicontinuity holds. That is, For every $\epsilon, \eta > 0$ there exists random $\Delta_n(\epsilon, \eta)$ and a constant $N_{\epsilon,\eta}$ such that for $n \geq N_{\epsilon,\eta}$, $P(|\Delta_n(\epsilon, \eta)| > \epsilon) < \eta$ and for each $y \in \mathbb{E}$, there is an open set $V(y, \epsilon, \eta)$ containing θ with*

$$\sup_{x \in V(y,\epsilon,\eta)} |\xi_n(x) - \xi_n(y)| \leq \Delta_n(\epsilon, \eta), \quad n \geq N_{\epsilon,\eta}.$$

Proof. See Newey (1991). □

LEMMA 3.B.4: *Let $h : [0, \infty) \to [0, \infty)$ be a function such that $h(0) = 0$ and h is continuous at 0. There is B_n such that $B_n = O_p(1)$. If for all $x, y \in \mathbb{E}$, $|\xi_n(x) - \xi_n(y)| \leq B_n h(\|x - y\|)$, then $\{\xi_n\}$ is stochastically equicontinuous.*

Proof. The result immediately follows from Assumption 3A and Corollary 2.2 in Newey (1991). □

The lemmas above imply that showing that $\{\mathcal{Z}_n(\cdot, t), n \geq 1\}$ satisfies the regularity conditions in Lemma 3.B.4 suffices for the desired result. For this, we make use of the following definition and the lemma.

DEFINITION 3.B.3 (Strict Continuity): *Let $\mathcal{S} \subseteq \mathbb{R}^d$. A function $f : \mathcal{S} \to \mathbb{R}$ is strictly continuous at $\bar{x} \in \mathcal{S}$ if $\bar{x} \in \mathcal{S}^o$ and if the Lipshitz modulus, lip$f(\bar{x}) :=$ $\limsup_{\substack{x,x' \to \bar{x} \\ x \neq x'}} \frac{|f(x') - f(x)|}{\|x' - x\|}$, is finite. A function is strictly continuous on $\mathcal{S}$ if it is strictly continuous at every point in $\mathcal{S}$.*

We say a function $f : \Omega \times \mathcal{S} \to \bar{\mathbb{R}}$ is strictly continuous on $\mathcal{S}$ almost surely $(a.s.)$ if $f(\omega, \cdot)$ is strictly continuous for all $\omega \in F \in \mathfrak{F}, P(F) = 1$. For single-valued mappings, the strict continuity is equivalent to local Lipschitz property, i.e., the function is Lipschitz on a neighborhood of each point (Rockafellar and Wets, 2005).

LEMMA 3.B.5 (Extended Mean Value Theorem): *Suppose f is convex and strictly continuous on an open convex set $O \subset \mathbb{R}^d$, and let x_0 and x_1 be points of*

O. Then there exist $x_\tau = (1-\tau)x_0 + \tau x_1, \tau \in (0,1)$ and $v \in \mathbb{R}^d$ satisfying

$$f(x_1) - f(x_0) = \langle v, x_1 - x_0 \rangle, \quad v \in \partial f(x_\tau),$$

where $\partial f(x)$ is the subdifferential of f at x, defined by

$$\partial f(x) := \{v \in \mathbb{R}^d : f(y) \geq f(x) + \langle v, y - x \rangle, \quad \forall y \in \mathbb{R}^d\}.$$

Proof. See Rockafellar and Wets (2005), Theorem 10.48. □

Proof of Theorem 3.3.1 (iii). We first show the required conditions for Lemma 3.B.4 using an expansion of the support function based on Lemma 3.B.5. In the following, we take $\omega \in F \in \mathfrak{F}$, $P(F) = 1$, for which $\tilde{\zeta}_n(\omega, \theta, \lambda)$ is lsc and convex. Under our assumptions, Θ_I is a compact convex set, and $\hat{\Theta}_n(t)$ is a compact convex set almost surely. For each bounded closed set, its support function is Lipschitz (Theorem F.1. in Molchanov (2005)). This implies that $s(p, \Theta_I)$ is strictly continuous, and $s(p, \hat{\Theta}_n(t))$ is strictly continuous *a.s.* Furthermore, the support function of a compact set is sublinear, and therefore it is convex (Molchanov, 2005, p.421). This further implies that $s(p, \Theta_I)$ is convex, and $s(p, \hat{\Theta}_n(t))$ is convex *a.s.*

Now, take an open convex set O such that $\mathbb{S}^{d-1} \subset O$. Let $p, q \in \mathbb{S}^{d-1}$. Then, by Lemma 3.B.5, for some $\bar{p}_n$ and $\bar{p}$ on the line segment that connects p and q, there exist $\hat{v}_n \in \partial s(\bar{p}_n, \hat{\Theta}_n(t))$ and $w \in \partial s(\bar{p}, \Theta_I)$ such that

$$s(p, \hat{\Theta}_n(t)) - s(q, \hat{\Theta}_n(t)) = \langle \hat{v}_n, p - q \rangle \tag{3.B.6}$$

$$s(p, \Theta_I) - s(q, \Theta_I) = \langle w, p - q \rangle \tag{3.B.7}$$

For any compact convex set F, the subdifferential $\partial s(p, F)$ of the support function at p coincides with its support set $H(p, F)$. Therefore, $\partial s(\bar{p}_n, \hat{\Theta}_n(t)) = H(\bar{p}_n, \hat{\Theta}_n(t))$ and $\partial s(\bar{p}, \Theta_I) = H(\bar{p}, \Theta_I)$. So, we can write

$$\mathcal{Z}_n(p, t) - \mathcal{Z}_n(q, t) = a_n^{1/\gamma} \langle \hat{v}_n - w, p - q \rangle \tag{3.B.8}$$

for some $\hat{v}_n \in H(\bar{p}_n, \hat{\Theta}_n(t))$ and $w \in H(\bar{p}, \Theta_I)$.

Note that, Assumption 3.2.3 (ii) implies $\mathcal{Z}_n(p,t) = O_p(1)$ for any $p \in \mathbb{S}^{d-1}$. Therefore $a_n^{1/\gamma}\langle \hat{v}_n - w, p - q\rangle = \mathcal{Z}_n(p,t) - \mathcal{Z}_n(q,t) = O_p(1)$ for any $p, q \in \mathbb{S}^{d-1}$. Since this holds for any p and q, each component of $a_n^{1/\gamma}(\hat{v}_n - w)$ must be $O_p(1)$. Therefore, $a_n^{1/\gamma}\|\hat{v}_n - w\| = O_p(1)$.

Applying the Cauchy-Schwartz inequality to (3.B.8), we obtain

$$|\mathcal{Z}_n(p,t) - \mathcal{Z}_n(q,t))| \leq a_n^{1/\gamma}\|\hat{v}_n - w\|\|p - q\|.$$

Now, we apply Lemma 3.B.4 with $B_n = a_n^{1/\gamma}\|\hat{v}_n - w\|$ and $h(x) = x$. Then, $\{\mathcal{Z}_n(\cdot,t), n \geq 1\}$ is stochastically equicontinuous. Further, we apply Lemma 3.B.3 to conclude that $\{\mathcal{Z}_n(\cdot,t), n \geq 1\}$ is tight. Note that a tight sequence that is weakly converging in finite dimension weakly converges in the uniform metric (van der Vaart and Wellner, 2000). Thus, we obtain $\mathcal{Z}_n(\cdot,t) \xrightarrow{u.d.} \mathcal{Z}(\cdot,t)$.

Since ζ_n is convex $a.s.$ in a neighborhood of Θ_I, its lower contour sets (in the neighborhood of Θ_I) are convex almost surely. Thus the Hörmander's embedding theorem implies

$$a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) = \|\mathcal{Z}_n(\cdot,t)\|_{\mathcal{C}(\mathbb{S})}.$$

By the continuous mapping theorem, we get $a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \xrightarrow{d} \|\mathcal{Z}(\cdot,t)\|_{\mathcal{C}(\mathbb{S})}$. Similarly, by Theorem 3.2.3, the directed Hausdorff distance satisfies $a_n^{1/\gamma}\vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) = \sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}_n(p,t)\}_+$. The continuous mapping theorem implies $a_n^{1/\gamma}$ $\vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) \xrightarrow{d} \sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+$ □

3.B.5 Proof of Theorems and Corollaries in Section 3.3.2

Proof of Theorem 3.3.2. By Theorem 3.2.2, $a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) = a_n^{1/\gamma}\|s(\cdot, \hat{\Theta}_n(t)) - s(\cdot, \Theta_I)\|_{\mathcal{C}(\mathbb{S}^{d-1})}$. Here, $s(\cdot, \hat{\Theta}_n(t))$ is a random element that takes values in a normed linear space $(\mathcal{C}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathcal{C}(\mathbb{S}^{d-1})})$. Let $F_n(x,t)$ be the cdf of $a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I)$. Note that Theorem 3.3.1 ensures F_n converges weakly to F. Now apply Politis, Romano, and Wolf's (1999) Theorem 2.5.2. with $\tau_n = a_n^{1/\gamma}$, $\hat{\theta}_n = s(\cdot, \hat{\Theta}_n(t))$, and $\theta(P) = s(\cdot, \Theta_I)$. Then, all the results follow. □

Proof of Corollary 3.3.1. (i) First, the consistency of $\hat{c}_{n,b,1-\alpha}(t)$ follows from Theorem 3.3.2 and Lemma 11.2.1 in Lehmann and Romano (2005). Under the null hypothesis, $T_n(t)$ converges in distribution to $F(x,t)$, and by the result above, $\hat{c}_{n,b,1-\alpha}(t) = c_{1-\alpha}(t) + o_p(1)$. Then, by Corollary 11.2.3 in Lehmann and Romano (2005), $\lim_{n\to\infty} P(T_n(t) \leq \hat{c}_{n,b,1-\alpha}(t)) = F(c_{1-\alpha}(t), t) = 1-\alpha$.

(ii) The proof of part (ii) is very similar to the proof of Theorem 2.2 in BM. □

Proof of Theorem 3.3.3. The first part follows from the equivalence

$$\Theta_I \in \hat{\mathcal{X}}_{n,b,1-\alpha}(t) \quad \Leftrightarrow \quad a_n^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \leq \hat{c}_{n,b,1-\alpha}(t).$$

and Theorem 3.3.2 (iii). Note that for any compact convex set $K \in \mathcal{K}_c$ and $\epsilon > 0$, we have

$$K \oplus B_\epsilon = K^\epsilon,$$

where B_ϵ is a closed ball of radius ϵ centered at the origin, and K^ϵ is a closed ϵ-envelope of K. The rest of the proof is very similar to the proof of Theorem 2.4 in BM.

(iii) Note that, for any $t \in \mathbb{R}_+$, $\Theta_I \in \hat{\mathcal{X}}_{n,b,1-\alpha}(t)$ implies $\Theta_I \subseteq \hat{\Psi}_{n,b,1-\alpha}(t)$, but the converse is not necessarily true. Therefore, by part (i),

$$\lim_{n\to\infty} P\left(\Theta_I \subseteq \hat{\Psi}_{n,b,1-\alpha}(t)\right) \geq \lim_{n\to\infty} P\left(\Theta_I \in \hat{\mathcal{X}}_{n,b,1-\alpha}(t)\right) = 1-\alpha.$$

By part (ii), $\hat{\Psi}_{n,b,1-\alpha}(t) = \hat{\Theta}_n^{\hat{\epsilon}_{n,b,1-\alpha}(t)}(t)$. Now, the conclusion follows. □

Proof of Theorem 3.3.4. Let $U_{n,b}^{\rightarrow}(x,t) := N_{n,b}^{-1} \sum_{k=1}^{N_{n,b}} 1_{\left\{a_b^{1/b} \vec{d}_H(\Theta_I, \hat{\Theta}_{n,b,k}(t)) \leq x\right\}}$ for each $t \in \mathbb{R}_+$. Suppose, for any $\epsilon > 0$,

$$a_b^{1/\gamma} \vec{d}_H(\hat{\Theta}_n(t), \hat{\Theta}_{n,b,k}(t)) = a_b^{1/\gamma} \sup_{p \in \mathbb{S}^{d-1}} \left\{ s(p, \hat{\Theta}_n(t)) - s(p, \hat{\Theta}_{n,b,k}(t)) \right\}_+ \leq x$$

and

$$a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) = a_b^{1/\gamma} \sup_{p \in \mathbb{S}^{d-1}} \left| s(p, \hat{\Theta}_n(t)) - s(p, \Theta_I) \right| \leq \epsilon.$$

Then, $a_b^{1/\gamma}(s(p, \Theta_I) - s(p, \hat{\Theta}_{n,b,k}(t))) \leq x + \epsilon$ for all $p \in \mathbb{S}^{d-1}$. This further implies

$$a_b^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_{n,b,k}(t)) = a_b^{1/\gamma} \sup_{p \in \mathbb{S}^{d-1}} \left\{ s(p, \Theta_I) - s(p, \hat{\Theta}_{n,b,k}(t)) \right\}_+ \leq x + \epsilon.$$

Let $E_{n,b}(t, \epsilon) := \{\omega : a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \leq \epsilon\}$. Then, the arguments above ensure

$$\hat{F}_{n,b}^{\rightarrow}(x, t) 1_{\{E_{n,b}(t,\epsilon)\}} \leq U_{n,b}^{\rightarrow}(x + \epsilon, t). \tag{3.B.9}$$

Now, suppose

$$a_b^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_{n,b,k}(t)) \leq x - \epsilon$$

and

$$a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) \leq \epsilon.$$

Then, we have

$$\begin{aligned} x &\geq a_b^{1/\gamma} d_H(\hat{\Theta}_n(t), \Theta_I) + a_b^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_{n,b,k}(t)) \\ &\geq a_b^{1/\gamma} \vec{d}_H(\hat{\Theta}_n(t), \Theta_I) + a_b^{1/\gamma} \vec{d}_H(\Theta_I, \hat{\Theta}_{n,b,k}(t)) \\ &\geq a_b^{1/\gamma} \vec{d}_H(\hat{\Theta}_n(t), \hat{\Theta}_{n,b,k}(t)). \end{aligned}$$

Therefore, we obtain

$$U_{n,b}^{\rightarrow}(x - \epsilon, t) 1_{\{E_{n,b}(t,\epsilon)\}} \leq \hat{F}_{n,b}^{\rightarrow}(x, t) 1_{\{E_{n,b}(t,\epsilon)\}}. \tag{3.B.10}$$

Since (3.B.9) and (3.B.10) hold for any $\epsilon > 0$ and $E_{n,b}$ has probability tending to one, we have

$$U^{\rightarrow}_{n,b}(x-\epsilon,t) \le \hat{F}^{\rightarrow}_{n,b}(x,t) \le U^{\rightarrow}_{n,b}(x+\epsilon,t), \tag{3.B.11}$$

with probability tending to 1 for any $\epsilon > 0$.

Now it is straightforward to show $U^{\rightarrow}_{n,b}(x-\epsilon,t) = F^{\rightarrow}(x,t) + o_p(1)$ for each continuity point x of $F^{\rightarrow}(\cdot,t)$ by an argument similar to the proof of Theorem 2.2.1 (i) in Politis, Romano, and Wolf (1999). Therefore,

$$F^{\rightarrow}(x-\epsilon,t) - \epsilon \le \hat{F}^{\rightarrow}_{n,b}(x,t) \le F^{\rightarrow}(x+\epsilon,t) + \epsilon,$$

with probability tending to 1 for any $\epsilon > 0$. Now, let $\epsilon \downarrow 0$ so that $x \pm \epsilon$ are continuity points of $F^{\rightarrow}(\cdot,P)$. Then, the conclusion follows.

The proofs of (ii) and (iii) are very similar to those of Theorem 2.2.1 (ii) and (iii) in Politis, Romano, and Wolf (1999). □

Proof of Corollary 3.3.2. (i) As before, the consistency of $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)$ follows from Lemma 11.2.1 in Lehmann and Romano (2005). Under the null, we have $T^{\rightarrow}_n(t) \le a_n^{1/\gamma}\vec{d}_H(\Theta_I,\hat{\Theta}_n(t))$, and $\vec{d}_H(\Theta_I,\hat{\Theta}_n(t))$ converges in distribution to $F^{\rightarrow}(x,t)$. By the results above and by Corollary 11.2.3 in Lehmann and Romano (2005), we have

$$\begin{aligned}\limsup_{n\to\infty} P(T^{\rightarrow}_n(t) > \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)) &\le \lim_{n\to\infty} P(a_n^{1/\gamma}\vec{d}_H(\Theta_I,\hat{\Theta}_n(t)) > \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)) \\ &= 1 - F(c^{\rightarrow}_{1-\alpha}(t),t) = \alpha.\end{aligned}$$

The proof of part (ii) is similar to the proof of Corollary 3.3.1 (ii). □

Proof of Theorem 3.3.5. The first part follows from the equivalence

$$\Theta_I \in \hat{\mathcal{X}}^{\rightarrow}_{n,b,1-\alpha}(t) \quad \Leftrightarrow \quad a_n^{1/\gamma}\vec{d}_H(\Theta_I,\hat{\Theta}_n(t)) \le \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t).$$

and Theorem 3.3.4 (iii). The proof of (ii) is similar to the proof of Theorem 3.3.3 (ii). The proof of (iii) is similar to the proof of Proposition 2.7 in BM. □

Proof of Lemma 3.3.4. First, $c^{\rightarrow}_{1-\alpha}(t^*_{1-\alpha}) = 0$ follows from the definition of $t^*_{1-\alpha}$.

For the conclusion of the lemma, it suffices to show that $P(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+ \leq x)$ is non-decreasing in t for each x. As this is a distributoinal property of the process $\mathcal{Z}(p,t)$, it suffices to show that the statement above holds for the following representation:

$$-\mathcal{Z}(p,t) = -\sup_{\theta\in H(p,\Theta_I)} \sup_{\lambda\in\{\lambda:\tilde{\zeta}(\theta,\lambda)\leq t\}} \langle p,\lambda\rangle.$$

As $\{\lambda : \tilde{\zeta}(\theta,\lambda) \leq t\} \subseteq \{\lambda : \tilde{\zeta}(\theta,\lambda) \leq t'\}$ for any $0 \leq t < t' \leq t^*_{1-\alpha}$ and for each $p \in \mathbb{S}^{d-1}$, $-\mathcal{Z}(p,t)$ is non-increasing in t. This implies that $\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+$ is non-increasing in t for any ω. Thus, $P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+ \leq x\right)$ is non-decreasing in $t \in [0, t^*_{1-\alpha}]$ for each x.

□

We use the following lemma to prove Theorem 3.3.6.

Lemma 3.B.6: *Suppose the conditions of Theorem 3.3.6 hold. Then, for any $\alpha \in (0,1)$ and $0 \leq t < t' \leq t^*_{1-\alpha}$, $c^{\rightarrow}_{1-\alpha}(t) - c^{\rightarrow}_{1-\alpha}(t') = \mu(t') - \mu(t)$.*

Proof of Lemma 3.B.6. First, $c^{\rightarrow}_{1-\alpha}(t)$ can be written as

$$\begin{aligned} c^{\rightarrow}_{1-\alpha}(t) &= \inf\left\{x : P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+ \leq x\right) \geq 1-\alpha\right\} \\ &= \inf\left\{x : P\left(\sup_{p\in\mathbb{S}^{d-1}}\{\mu(t') - \mu(t) - \mu(t') - \mathcal{Z}^*(p)\}_+ \leq x\right) \geq 1-\alpha\right\}. \end{aligned} \tag{3.B.12}$$

Let $\Delta(t,t') := \mu(t') - \mu(t)$. Then, for any $x \geq \Delta(t,t')$, we have

$$\begin{aligned} &P\left(\sup_{p\in\mathbb{S}^{d-1}}\{\mu(t') - \mu(t) - \mu(t') - \mathcal{Z}^*(p)\}_+ \leq x\right) \\ &= P\left(\sup_{p\in\mathbb{S}^{d-1}}\{\Delta(t,t') - \mathcal{Z}(p,t')\}_+ \leq x\right) \\ &= P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t')\}_+ \leq x - \Delta(t,t')\right). \end{aligned} \tag{3.B.13}$$

Substituting Eq. (3.B.13) into Eq. (3.B.12) yields

$$\begin{aligned} c^{\rightarrow}_{1-\alpha}(t) &= \inf\left\{x : P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t')\}_+ \le x - \Delta(t,t')\right) \ge 1-\alpha\right\} \\ &= c^{\rightarrow}_{1-\alpha}(t') + \Delta(t,t'). \end{aligned}$$ □

Proof of Theorem 3.3.6. By Theorem 3.2.2,

$$\begin{aligned} a_n^{1/\gamma} d_H(\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)}(t), &\hat{\Theta}_n(t^*_{1-\alpha})) \\ &= a_n^{1/\gamma} \sup_{p\in\mathbb{S}^{d-1}} |s(p, \hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)}(t)) - s(p, \hat{\Theta}_n(t^*_{1-\alpha}))| \\ &= a_n^{1/\gamma} \sup_{p\in\mathbb{S}^{d-1}} |s(p, \hat{\Theta}_n(t)) + \hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t) - s(p, \hat{\Theta}_n(t^*_{1-\alpha}))| \\ &= \sup_{p\in\mathbb{S}^{d-1}} |a_n^{1/\gamma}[s(p, \hat{\Theta}_n(t)) - s(p, \hat{\Theta}_n(t^*_{1-\alpha}))] + \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)| \\ &= \sup_{p\in\mathbb{S}^{d-1}} |a_n^{1/\gamma}[s(p, \hat{\Theta}_n(t)) - s(p, \Theta_I)] \\ &\qquad - a_n^{1/\gamma}[s(p, \hat{\Theta}_n(t^*_{1-\alpha})) - s(p, \Theta_I)] + \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t)| \\ &= \sup_{p\in\mathbb{S}^{d-1}} |\mathcal{Z}_n(p,t) - \mathcal{Z}_n(p, t^*_{1-\alpha}) + c^{\rightarrow}_{1-\alpha}(t) + o_p(1)| \\ &\overset{(1)}{=} \sup_{p\in\mathbb{S}^{d-1}} |\mu(t) - \mu(t^*_{1-\alpha}) - (c^{\rightarrow}_{1-\alpha}(t^*_{1-\alpha}) - c^{\rightarrow}_{1-\alpha}(t)) + o_p(1)| \\ &= o_p(1), \end{aligned}$$

where we used the fact that $c^{\rightarrow}_{1-\alpha}(t^*_{1-\alpha}) = 0$ in equality (1), and the last equality follows from Lemma 3.B.6. □

Proof of Corollary 3.3.3. The result immediately follows from Theorem 3.3.6 and the triangle inequality:

$$\begin{aligned} d_H\left(\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)}(t), \hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t')}(t')\right) \le \\ d_H\left(\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t)}(t), \hat{\Theta}_n(t^*_{1-\alpha})\right) + d_H\left(\hat{\Theta}_n^{\hat{\epsilon}^{\rightarrow}_{n,b,1-\alpha}(t')}(t'), \hat{\Theta}_n(t^*_{1-\alpha})\right). \end{aligned}$$ □

Proof of Theorem 3.3.7. Let l be the smallest random index for which there is a false rejection. Then, there is $\Theta_0 \in S_l^c \cap \mathcal{K}_{cI}$ such that

$$a_n^{1/\gamma}\vec{d}_H(\Theta_0, \hat{\Theta}_n(t_l)) > \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t_l).$$

As $\Theta_0 \subseteq \Theta_I$, this implies

$$\begin{aligned} a_n^{1/\gamma}\vec{d}_H(\Theta_I, \hat{\Theta}_n(t_l)) > \hat{c}^{\rightarrow}_{n,b,1-\alpha}(t_l) & \\ &\Rightarrow a_n^{1/\gamma}\vec{d}_H(\Theta_I, \hat{\Theta}_n^{\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t_l)/a_n^{1/\gamma}}(t_l)) > 0 \\ &\Rightarrow a_n^{1/\gamma}\vec{d}_H(\Theta_I, \hat{\Theta}_n(t^*_{1-\alpha})) + o_p(1) > 0, \end{aligned}$$

where the last result follows from Theorem 3.3.6. Therefore,

$$\lim_{n\to\infty} FWER \leq P\left(a_n^{1/\gamma}\vec{d}_H(\Theta_I, \hat{\Theta}_n(t^*_{1-\alpha})) + o_p(1) > 0\right) = \alpha. \square$$

3.B.6 Proofs for Theorems and Corollaries in Section 3.3.4

The following result immediately follows from the main theorem.

LEMMA 3.B.7: *Suppose conditions of Theorem 3.1 hold. Suppose that p_0 uniquely maximizes $\langle p, \theta_0\rangle - s(p, \Theta_I)$. Then, for each $t \in \mathbb{R}_+$,*

(i) if $\theta_0 \in \partial\Theta_I$, $T^{\rightarrow}_{n,\theta_0}(t) \xrightarrow{d} \{-\mathcal{Z}(p_0, t)\}_+$;

(ii) if $\theta_0 \in \Theta_I^o$, $T^{\rightarrow}_{n,\theta_0}(t) \xrightarrow{p} 0$;

(iii) if $\theta_0 \notin \Theta_I$, $T^{\rightarrow}_{n,\theta_0}(t) \xrightarrow{p} +\infty$.

Proof of Lemma 3.B.7. The proof is similar to that of Proposition 16 in Bontemps, Magnac, and Maurin (2007). First, note that $\mathbb{S}^{d-1}$ is compact, p_0 uniquely maximizes $= \langle p, \theta_0\rangle - s(p, \Theta_I)$, $\hat{p}_n$ maximizes $\langle p, \theta_0\rangle - s(p, \hat{\Theta}_n(t))$, and $[\langle p, \theta_0\rangle - s(p, \Theta_I)] - [\langle p, \theta_0\rangle - s(p, \hat{\Theta}_n)] \xrightarrow{p} 0$ uniformly over $\mathbb{S}^{d-1}$. Therefore, $\hat{p}_n \xrightarrow{p} p_0$. Let $\mathcal{A}_n(p,t) := a_n^{1/\gamma}[\langle p, \theta_0\rangle - s(p, \hat{\Theta}_n(t))]$ and $\mathcal{A}(p,t) := a_n^{1/\gamma}[\langle p, \theta_0\rangle - s(p, \Theta_I)]$. First, we show that $\mathcal{A}_n(\hat{p}_n, t) - \mathcal{A}_n(p_0, t) = o_p(1)$. Note that $\mathcal{A}_n(\hat{p}_n, t) - \mathcal{A}_n(p_0, t)$ is bounded from below by 0, as $\hat{p}_n$ maximizes $\mathcal{A}_n(\cdot, t)$. From above, we have

$$\begin{aligned}
0 &\leq \mathcal{A}_n(\hat{p}_n, t) - \mathcal{A}_n(p_0, t) \\
&= \mathcal{A}_n(\hat{p}_n, t) - \mathcal{A}(\hat{p}_n, t) + \mathcal{A}(\hat{p}_n, t) - \mathcal{A}(p_0, t) + \mathcal{A}(p_0, t) - \mathcal{A}_n(p_0, t) \\
&= -\mathcal{Z}_n(\hat{p}_n, t) + \mathcal{A}(\hat{p}_n, t) - \mathcal{A}(p_0, t) + \mathcal{Z}_n(p_0, t) \\
&\leq \mathcal{Z}_n(p_0, t) - \mathcal{Z}_n(\hat{p}_n, t) = o_p(1),
\end{aligned}$$

where the second inequality holds because $\mathcal{A}(\hat{p}_n, t) - \mathcal{A}(p_0, t) \leq 0$ by the construction of p_0 and the last equality follows from the stochastic equicontinuity of $\{\mathcal{Z}_n(\cdot, t)\}$.

(i) Suppose $\theta_0 \in \partial\Theta_I$, then $\langle p_0, \theta_0 \rangle = s(p_0, \Theta_I)$. Therefore, $\mathcal{A}_n(p_0, t) = -\mathcal{Z}_n(p_0, t)$. Now,

$$\begin{aligned}
T_{n,\theta_0}(t) &= \{\mathcal{A}_n(\hat{p}_n, t)\}_+ \\
&= \{\mathcal{A}_n(p_0, t)\}_+ + o_p(1) \\
&= \{-\mathcal{Z}_n(p_0, t)\}_+ + o_p(1).
\end{aligned}$$

By Theorem 3.3.1 (i) and the continuous mapping theorem, $T_{n,\theta_0}(t) \xrightarrow{d} \{-\mathcal{Z}(p_0, t)\}_+$. (ii) Suppose $\theta \in \Theta_I^o$, then $T_{n,\theta_0}(t) = \{-\mathcal{Z}_n(p_0, t) + a_n^{1/\gamma}[\langle p_0, \theta_0 \rangle - s(p_0, \Theta_I)]\}_+ + o_p(1)$ and $a_n^{1/\gamma}[\langle p_0, \theta_0 \rangle - s(p_0, \Theta_I)] \to -\infty$. Therefore, $T_{n,\theta_0}(t) \xrightarrow{p} 0$. (iii) Suppose $\theta \notin \Theta_I$, then $T_{n,\theta_0}(t) = \{-\mathcal{Z}_n(p_0, t) + a_n^{1/\gamma}[\langle p_0, \theta_0 \rangle - s(p_0, \Theta_I)]\}_+ + o_p(1)$ and $a_n^{1/\gamma}[\langle p_0, \theta_0 \rangle - s(p_0, \Theta_I)] \to \infty$. Therefore, $T_{n,\theta_0}(t) \xrightarrow{p} \infty$. □

Proof of Theorem 3.3.8. The proof is very similar to the proof of Theorem 3.3.4, and therefore it is omitted. □

Proof of Corollary 3.3.4. (i) First $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t) = \hat{c}^{\rightarrow}_{n,b,1-\alpha}(p_0, t) + o_p(1)$. Then, the consistency of $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t)$ follows by applying Lemma 11.2.1 in Lehmann and Romano (2005). If $\theta_0 \in \partial\Theta_I$, $T^{\rightarrow}_{n,\theta_0}(t)$ converges in distribution to $F^{\rightarrow}(x, p, t)$ by Lemma 3.B.7 (i), and by the result above, $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t) = c^{\rightarrow}_{1-\alpha}(p_0, t) + o_p(1)$. Then, by Corollary 11.2.3 in Lehmann and Romano (2005), $\lim_{n\to\infty} P(T^{\rightarrow}_{n,\theta_0}(t) \leq \hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t)) = F^{\rightarrow}(c^{\rightarrow}_{1-\alpha}(p_0, t), t) = 1-\alpha$. If $\theta \in \Theta_I^o$, $T^{\rightarrow}_{n,\theta_0}(t) \xrightarrow{p} 0$ by Lemma 3.B.7

(ii), and therefore $\lim_{n\to\infty} P(T^{\rightarrow}_{n,\theta_0}(t) \leq \hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t)) \geq \lim_{n\to\infty} P(\{-\mathcal{Z}_n(\hat{p}_n, t)\}_+ \leq \hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t)) = F^{\rightarrow}(c^{\rightarrow}_{1-\alpha}(p_0, t), t) = 1 - \alpha$.

(ii) The proof of part (ii) is a direct consequence of Lemma 3.B.7 (iii) and $\hat{c}_{n,b,1-\alpha}(\hat{p}_n, t) = O_p(1)$. □

Proof of Theorem 3.3.9. The result simply follows from the equivalence

$$\theta_0 \in \tilde{\Psi}^{\rightarrow}_{n,b,1-\alpha}(t) \quad \Leftrightarrow \quad T^{\rightarrow}_{n,\theta_0} \leq \hat{c}^{\rightarrow}_{n,b,1-\alpha}(\hat{p}_n, t).$$

and Corollary 3.3.4 (i). □

Proof of Theorem 3.3.10. First, we can rewrite the local power as

$$\begin{aligned}
\pi_{n,b,t}(\theta_n) &= P\left(a_n^{1/\gamma}\{\langle \hat{p}_n, \theta_n\rangle - s(\hat{p}_n, \hat{\Theta}_n(t))\}_+ > \hat{c}_{n,b,1-\alpha}(\hat{p}_n, t)\right) \\
&= P\left(a_n^{1/\gamma}\{s(\hat{p}_n, \Theta_I) - s(\hat{p}_n, \hat{\Theta}_n(t)) + \langle \hat{p}_n, \lambda\rangle / a_n^{1/\gamma}\}_+ > \hat{c}_{n,b,1-\alpha}(\hat{p}_n, t)\right) \\
&= P\left(\{-\mathcal{Z}_n(\hat{p}_n, t) + \langle \hat{p}_n, \lambda\rangle\}_+ > \hat{c}_{n,b,1-\alpha}(\hat{p}_n, t)\right).
\end{aligned}$$

By the fact that $\langle \hat{p}_n, \lambda\rangle = \langle p_0, \lambda\rangle + o_p(1) = h + o_p(1)$ and the stochastic equicontinuity of $\mathcal{Z}(\cdot, t)$ and $\hat{c}_{n,b,1-\alpha}(\cdot, t)$, we obtain

$$\pi_{n,b,t}(\theta_n) = P(\{-\mathcal{Z}_n(p_0, t) + h + o_p(1)\}_+ > c^{\rightarrow}_{1-\alpha}(p_0, t) + o_p(1)).$$

Note that $\lim_{n\to\infty} \pi_{n,b,t}(\theta_n) \geq P(\{-\mathcal{Z}(p_0, t)\}_+ > c^{\rightarrow}_{1-\alpha}(p_0, t)) = \alpha$. So, the test has the asymptotic local unbiasedness. Furthermore, using the fact that $\{-x\}_+ > \epsilon \Leftrightarrow x < -\epsilon$ for any $\epsilon \geq 0$, we can write

$$\pi_{n,b,t}(\theta_n) = P(\mathcal{Z}_n(p_0, t) < h - c^{\rightarrow}_{1-\alpha}(p_0, t)) + o(1).$$

By the second duality (Lemma 3.3.3) and the weak epiconvergence

$$\liminf_{n\to\infty} P(\mathcal{Z}_n(p_0,t) < h - c^{\rightarrow}_{1-\alpha}(p_0,t)) \geq \liminf_{n\to\infty} P\left(\inf_{\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0}} \tilde{\zeta}_n(\theta,\lambda) > t\right)$$
$$\geq P\left(\inf_{\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0}} \tilde{\zeta}(\theta,\lambda) > t\right),$$

where $\tilde{R}_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} = \{\theta_0\} \times (K_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0} \cap L_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0})$ with $L_{h-c^{\rightarrow}_{1-\alpha}(p_0,t),p_0}$ properly chosen. □

Proof of Corollary 3.3.5. The proof of part (i) is very similar to the proof of Corollary 3.3.2 (i). For part (ii), we make use of the reverse triangle inequality:

$$\vec{d}_H(\{\theta_0\},\Theta_I) - \vec{d}_H(\Theta_I,\hat{\Theta}_n(t)) \leq \vec{d}_H(\{\theta_0\},\hat{\Theta}_n(t)).$$

Scaling both sides by $a_n^{1/\gamma}$ and taking $\inf_{\Theta_0}$ both sides give

$$a_n^{1/\gamma}\left(\inf_{\theta_0\in\Theta_0} \vec{d}_H(\{\theta_0\},\Theta_I) - \vec{d}_H(\Theta_I,\hat{\Theta}_n(t))\right) \leq \inf_{\theta_0\in\Theta_0} T^{\rightarrow}_{n,\theta_0}(t). \tag{3.B.14}$$

Since $\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t) = c^{\rightarrow}_{1-\alpha}(t) + o_p(1)$ and $\vec{d}_H(\Theta_I,\hat{\Theta}_n(t)) = o_p(1)$, we have

$$\lim_{n\to\infty} P\left(\hat{c}^{\rightarrow}_{n,b,1-\alpha}(t) < a_n^{1/\gamma}\left(\inf_{\theta_0\in\Theta_0} \vec{d}_H(\{\theta_0\},\Theta_I) - \vec{d}_H(\Theta_I,\hat{\Theta}_n(t))\right)\right)$$
$$= \lim_{n\to\infty} P\left(\frac{c^{\rightarrow}_{1-\alpha}(t) + o_p(1)}{a_n^{1/\gamma}} < \inf_{\theta_0\in\Theta_0} \vec{d}_H(\{\theta_0\},\Theta_I) + o_p(1)\right)$$
$$= 1,$$

since $\inf_{\theta_0\in\Theta_0} \vec{d}_H(\{\theta_0\},\Theta_I) > 0$ under the alternative. By the inequality 3.B.14, the conclusion follows. □

3.B.7 Proof of Theorems and Corollaries in Section 3.4

Knight (1999) provides a result that links the weak finite dimensional limit and the weak epilimit, which we summarize below as a lemma. For this result, we introduce the following definition.

Definition 3.B.4 (Stochastic Equi-lowersemicontinuity): *A sequence of random lsc functions* $\{\xi_n, n \geq 1\}$ *on* $\mathbb{R}^d$ *is stochastically equi-lowersemicontinuous (e-lsc) if for each bounded set* B*,* $\epsilon > 0$ *and* $\delta > 0$*, there exist* $x_1, \cdots, x_m \in B$ *and open neighborhoods* $V(x_1), \cdots, V(x_m)$ *of* $x_1, \cdots, x_m$ *such that*

$$B \subset \bigcup_{i=1}^{m} V(x_i)$$

and

$$\limsup_{n\to\infty} P\left(\bigcup_{i=1}^{m}\left\{\inf_{y\in V(x_i)} \xi_n(y) \leq \min\{\epsilon^{-1}, \xi_n(x_i) - \epsilon\}\right\}\right) < \delta.$$

Lemma 3.B.8 (Knight, 1999, Theorem 2): *Let* $\{\xi_n, n \geq 1\}$ *be a stochastically e-lsc sequence of functions and* ξ *be a random lsc function. Then* $\xi_n \overset{f.d.}{\to} \xi$ *if and only if* $\xi_n \overset{e.d.}{\to} \xi$.

Proof of Theorem 3.4.1. Assumption 3.2.1 immediately follows from Assumption 3.4.1. Assumption 3.2.2 immediately follows from Assumption 3.4.2. For the consistency of the level set estimator with the choice of finite nonnegative constant t, we additionally need to show Assumptions 3.2.3 (i), (ii), and 3.2.4 (i).

By Assumption 3.4.3 (i-a,b), we can write

$$
\begin{aligned}
P\left(\sup_{\Theta}\left|\varphi(\hat{E}_n(m_{j,\theta},\hat{W}_n(\theta)) - \varphi(E(m_{j,\theta}),W(\theta))\right| > \epsilon\right) \\
\leq P\Bigg(&\sup_{\Theta}\left|\varphi(\hat{E}_n(m_{j,\theta},\hat{W}_n(\theta)) - \varphi(E(m_{j,\theta}),\hat{W}_n(\theta))\right| \\
&+\sup_{\Theta}\left|\varphi(E(m_{j,\theta}),\hat{W}_n(\theta)) - \varphi(E(m_{j,\theta}),W(\theta))\right| > \epsilon\Bigg) \\
\leq P\Bigg(&\sup_{\Theta}\left|\varphi(\hat{E}_n(m_{j,\theta},\hat{W}_n(\theta)) - \varphi(E(m_{j,\theta}),\hat{W}_n(\theta))\right| > \epsilon/2\Bigg) \\
&+P\Bigg(\sup_{\Theta}\left|\varphi(E(m_{j,\theta}),\hat{W}_n(\theta)) - \varphi(E(m_{j,\theta}),W(\theta))\right| > \epsilon/2\Bigg) \\
\leq P&\left(\sup_{\Theta} L_1 h_1\left(\left\|\hat{E}_n(m_{j,\theta}) - E(m_{j,\theta})\right\|\right) > \epsilon/2\right) \\
&+P\left(\sup_{\Theta} L_2 h_2\left(\max_{i,j}\left|\hat{W}_{n,ij}(\theta) - W_{ij}(\theta)\right|\right) > \epsilon/2\right) \\
\leq P&\left(\sup_{\Theta}\left\|\hat{E}_n(m_{j,\theta}) - E(m_{j,\theta})\right\| > L_1^{-1}h_1^{-1}(\epsilon/2)\right) \\
&+P\left(\sup_{\Theta}\max_{i,j}\left|\hat{W}_{n,ij}(\theta) - W_{ij}(\theta)\right| > L_2^{-1}h_2^{-1}(\epsilon/2)\right) \\
\leq \epsilon&
\end{aligned}
$$

for n sufficiently large. Therefore, Assumptions 3.2.3 (i) holds. In the following, we take $a_n = n^{\gamma/2}$. First, this choice of a_n and the P-donsker property ensure that $\sup_{\Theta_I} a_n Q_n(\theta) = O_p(1)$. Therefore, Assumptions 3.2.3 (ii) holds. Now, let $\eta > 0$ be such that $\sup_{\Theta_I}\max_{i,j}\hat{W}_{n,ij}(\theta) \leq \eta < \infty$, $wp \to 1$. We can write

$$
\begin{aligned}
n^{\gamma/2}Q_n(\theta) &\leq \varphi(\mathbb{G}_n(m_\theta) + \sqrt{n}E(m_\theta), \hat{W}_n(\theta)) \\
&\leq \varphi\left(\mathbb{G}_n(m_\theta) + \sqrt{n}E(m_\theta), \sup_{\Theta_I}\max_{i,j}|\hat{W}_{n,ij}(\theta)|I_J\right) \\
&\leq \varphi\left(O_p(1) - \sqrt{n}C_3\min\{d(\theta,\Theta\setminus\Theta_I),\bar{\epsilon}\}, \eta I_J\right)
\end{aligned}
$$

uniformly over Θ_I $wp \to 1$ by Assumption 3.4.3 (ii-d). We thus have $Q_n(\theta) = 0$ on $\Theta_I^{-\epsilon_n}$ with $\epsilon_n = O_p(1/\sqrt{n})$, and this ensures Assumption 3.2.4 (i).

For the rate of convergence of the set estimator, we additionally need to show Assumptions 3.2.3 (iii) and 3.2.4 (ii). For this, we closely follow CHT's proof of Theorem 4.2. Take $\eta' > 0$ such that $\inf_{\Theta_I} \min_{i,j} |\hat{W}_{n,ij}(\theta)| \geq \eta'$, $wp \to 1$. First, we write

$$\begin{aligned} n^{\gamma/2} Q_n(\theta) &= \varphi(\mathbb{G}_n(m_\theta) + \sqrt{n}E(m_\theta), \hat{W}_n(\theta)) \\ &\geq C_2 \|\mathbb{G}_n(m_\theta) + \sqrt{n}E(m_\theta)\|_+^\gamma \\ &\geq C_2 \|\sqrt{n}E(m_\theta)\|_+^\gamma \frac{\|\mathbb{G}_n(m_\theta) + \sqrt{n}E(m_\theta)\|_+^\gamma}{\|\sqrt{n}E(m_\theta)\|_+^\gamma}. \end{aligned}$$

Now, by Assumption 3.4.3 (ii-b), we have $\|\sqrt{n}E(m_\theta)\|_+^\gamma \geq C_1 n^{\gamma/2} \min\{d(\theta, \Theta_I), \delta\}^\gamma$ on Θ for some $C_1 > 0$ and $\delta > 0$. Therefore, for any $\epsilon > 0$, we can choose $(\kappa_\epsilon, n_\epsilon)$ so that for any $n \geq n_\epsilon$ with probability at least $1 - \epsilon$,

$$n^{\gamma/2} Q_n(\theta) \geq C_2 C_1 n^{\gamma/2} \min\{d(\theta, \Theta_I), \delta\}^\gamma,$$

uniformly over $\{\theta \in \Theta : d(\theta, \Theta_I) \geq (\kappa_\epsilon / n^{\gamma/2})^{1/\gamma}\}$, which follows by $\|y+x\|_+/\|x\|_+ \to 1$ as $\|x\|_+ \to \infty$ for any $y \in \mathbb{R}^J$ and by $\sup_{\Theta_I} \|\mathbb{G}_n(m_\theta)\| = O_p(1)$ by the P-Donsker property.

Note that the lower semicontinuity $\tilde{\zeta}_n$ follows the continuity in θ of φ, m, and W. The convexity of $Q_n(\theta)$ in a neighborhood of Θ_I directly follows from Assumption 3.4.4 (ii). Now, we show the weak epiconvergence of $\tilde{\zeta}_n$. First, $\mathbb{G}_n m_\theta \overset{u.d.}{\to} \mathbb{G}(\theta)$ implies $\mathbb{G}_n(m_\theta) \overset{f.d.}{\to} \mathbb{G}(\theta)$. Together with Assumption 3.4.4 (i-a,b), this implies $\mathcal{M}_n(\theta, \lambda) \overset{f.d.}{\to} \mathcal{M}(\theta, \lambda)$. We also have $\hat{W}_n(\theta + \lambda / a_n^{1/\gamma}) \overset{p}{\to} W(\theta)$. Therefore, by the continuous mapping theorem, $\tilde{\zeta}_n = \varphi(\mathcal{M}_n(\theta, \lambda), \hat{W}_n(\theta + \lambda / a_n^{1/\gamma})) \overset{f.d.}{\to} \varphi(\mathcal{M}(\theta, \lambda), W(\theta)) = \tilde{\zeta}$. Now by Assumption 3.4.4 (iii), we can apply Lemma 3.B.8 to conclude that $\tilde{\zeta}_n \overset{e.d.}{\to} \tilde{\zeta}$. □

Proof of Corollary 3.4.1. By Theorem 3.4.1, the conditions required for Theorem 3.3.1 hold. The weak convergence results immediately follow from Theorem 3.3.1, and the representation result follows from the fact that $\tilde{\zeta}(\theta, \lambda) = \varphi(\mathbb{G}(\theta) + \Pi(\theta)\lambda + \varsigma(\theta))$. □

Proof of Corollary 3.4.2. Let $s : \partial\Theta \times \mathbb{R}^d \to \mathbb{R}^{J(\theta)}$ be a vector-valued mapping whose j-th component is $s_j(\theta, \lambda) = 1\{\mathbb{G}_j(\theta) + \langle \Pi_j(\theta), \lambda \rangle > 0\}$. As the linear constraint qualification is satisfied, the solution λ^* to the minimization problem (3.4.3) satisfies the following Karush-Kuhn-Tucker (KKT) conditions with probability 1[27]:

$$\begin{aligned}
&p = 2\mu\Pi_{\mathcal{J}(\theta)}(\theta)'W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)\\
&t \geq \|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)\|^2\\
&0 \leq \mu\\
&0 = \mu(\|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)\|^2 - t),
\end{aligned}$$

where μ is the Lagrange multiplier associated with the constraint in Eq. (3.4.3). By Assumption 3.4.5 (ii), the constraint in (3.4.3) binds, and the conditions above simplify to

$$p = 2\mu\Pi_{\mathcal{J}(\theta)}(\theta)'W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*) \tag{3.B.15}$$

$$t = \|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)\|^2 \tag{3.B.16}$$

$$\mu > 0. \tag{3.B.17}$$

We can solve (3.B.15) to obtain

$$\begin{aligned}
&\left(W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)'W^{1/2}_{\mathcal{J}(\theta)}(\theta)\right)^{-1}W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)p\\
&\qquad\qquad = 2\mu W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*).
\end{aligned} \tag{3.B.18}$$

Let $\mathcal{R}(p, \theta)$ be the left hand side of the equation above. Take squared norms both sides to obtain

$$\begin{aligned}
\|\mathcal{R}(p, \theta)\|^2 &= |2\mu|^2\|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)\|^2\\
&= |2\mu|^2 t,
\end{aligned}$$

[27]The constraint is non-differentiable only at finite number of points, and the probability of $\mathbb{G}(\theta)$ taking these values is 0.

where the second equality follows from (3.B.16). So, we obtain

$$2\mu = \|\mathcal{R}(p,\theta)\| t^{-1/2}. \tag{3.B.19}$$

Plugging this into (3.B.18) gives

$$W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta,\lambda^*) = \frac{\mathcal{R}(p,\theta)}{\|\mathcal{R}(p,\theta)\|} t^{1/2}. \tag{3.B.20}$$

Substituting (3.B.19) and (3.B.20) into (3.B.15) yields

$$p = \Pi'_{\mathcal{J}(\theta)} W^{1/2}_{\mathcal{J}(\theta)} \mathcal{R}(p,\theta). \tag{3.B.21}$$

Now, we can use this result to obtain

$$\begin{aligned}
\mathcal{V}(p,\theta,t) &= \langle p, \lambda^* \rangle \\
&= \left\langle \Pi'_{\mathcal{J}(\theta)} W^{1/2}_{\mathcal{J}(\theta)} \mathcal{R}(p,\theta), \lambda^* \right\rangle \\
&= \left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)} \Pi_{\mathcal{J}(\theta)} \lambda^* \right\rangle \\
&= \left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)} (\Pi_{\mathcal{J}(\theta)} \lambda^* \circ s(\theta,\lambda^*)) \right\rangle \\
&= \left\langle \mathcal{R}(p,\theta), \frac{\mathcal{R}(p,\theta)}{\|\mathcal{R}(p,\theta)\|} t^{1/2} - W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)} \circ s(\theta,\lambda^*)) \right\rangle \\
&= \|\mathcal{R}(p,\theta)\| \, t^{1/2} - \left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta) \mathbb{G}_{\mathcal{J}(\theta)} \right\rangle,
\end{aligned}$$

where the fourth equality follows from the fact that $\mathcal{R}(p,\theta) = \mathcal{R}(p,\theta) \circ s(\theta,\lambda^*)$, and the fifth equality follows from (3.B.20).

If $W(\theta)$ satisfies $W_{\mathcal{J}(\theta)}(\theta) = (\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}$ for any $\theta \in \partial\Theta_I$, then

$$\|\mathcal{R}(p,\theta)\|^2 = p'\Pi_{\mathcal{J}(\theta)}(\theta)'(\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p.$$

Note that Eq. (3.B.21) implies that

$$p'p = p'\Pi_{\mathcal{J}(\theta)}(\theta)'(\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p.$$

As p is in the unit sphere, $p'p = \|p\|^2 = 1$. Combining the results above establishes

$\|\mathcal{R}(p,\theta)\| = 1$. Therefore, the limiting process takes the form $\mathcal{Z}(p,t) := \mu(t) + \mathcal{Z}^*(p)$ with $\mu(t) = t^{1/2}$ and

$$\begin{aligned}\mathcal{Z}^*(p) &= \sup_{\theta\in H(p,\Theta_I)} -\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\rangle \\ &= \sup_{\theta\in H(p,\Theta_I)} -\langle (\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p, \mathbb{G}_{\mathcal{J}(\theta)}(\theta)\rangle. \qquad \square\end{aligned}$$

For Theorem 3.4.2, we require the following regularity conditions.

Assumption 3.B.1 (Local Process Regularity for QLR Statistic): *(i) For any finite sets $U \subset \mathbb{R}$ and $S \subset \mathbb{S}^{d-1}$, $(\sup_{R^-_{u,p}} \tilde{\zeta}_n, (u,p) \in U\times S) \xrightarrow{d} (\sup_{R^-_{u,p}} \tilde{\zeta}, (u,p) \in U \times S)$. (ii) For any $0 < \epsilon$, there exists $\delta > 0$ such that*

$$\lim_{n\to\infty} P\left(\sup_{\|p-q\|<\delta}\left|\sup_{R^-_{u,p}}\tilde{\zeta}_n(\theta,\lambda) - \sup_{R^-_{u,q}}\tilde{\zeta}_n(\theta,\lambda)\right| \geq \epsilon\right) \leq \epsilon,$$

where $R^-_{u,p} := H(p,\Theta_I)\times K^-_{u,p}$.

Assumption 3.B.1 (i) requires that the finite dimensional distribution of the supremum of $\tilde{\zeta}_n$ over a class of compact sets converges to that of $\tilde{\zeta}$. This is analogous to weak epiconvergence. We call this version "weak supconvergence" as it is close in spirit to Condition S.2 of CHT.

Proof of Theorem 3.4.2. First, by the hypothesis that $\tilde{\zeta}_n$ weakly supconverges to $\tilde{\zeta}$, $\mathcal{L}_n(\cdot,u) \xrightarrow{f.d.} \mathcal{L}(\cdot,u)$ where

$$\mathcal{L}(p,u) := \sup_{\theta\in H(p,\Theta_I)}\sup_{\lambda\in K^-_{u,p}} \|W^{1/2}(\theta)(\mathbb{G}(\theta)+\Pi(\theta))\|^2_+.$$

The tightness of $\{\mathcal{L}_n(\cdot,u)\}$ follows from the assumption of the corollary, and these results imply $\mathcal{L}_n(\cdot,u) \xrightarrow{u.d.} \mathcal{L}(\cdot,u)$ for each u.

Now we derive the representation of $\mathcal{L}$ given in the theorem. Below, we fix $p \in \mathbb{S}^{d-1}$ and $\theta \in \partial\Theta_I$. As $\theta\in\partial\Theta_I$, the components of $\mathcal{M}(\theta,\lambda)$ for $j \in \mathcal{J}^c(\theta)$ are irrelevant. To obtain a closed form for $\mathcal{L}$, consider the following optimization problem

$$\mathcal{C}(\theta, p, u) := \sup_{\lambda} \|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda)\|^2_+ \quad \text{s.t.} \langle p, \lambda\rangle \leq u. \tag{3.B.22}$$

Similar to the proof of Corollary 3.4.2, the solution λ^* of the problem above satisfies the following KKT conditions with probability 1.

$$\begin{aligned} \nu p &= 2\Pi_{\mathcal{J}(\theta)}(\theta)' W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*) \\ \langle p, \lambda^*\rangle &\leq u \\ 0 &\leq \nu \\ 0 &= \nu(u - \langle p, \lambda^*\rangle), \end{aligned}$$

where ν is the Lagrange multiplier associated with the constraint in (3.B.22). By Assumption 3.4.5 (ii), the constraint in (3.B.22) binds, and the conditions above simplify to

$$\nu p = 2\Pi_{\mathcal{J}(\theta)}(\theta)' W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*) \tag{3.B.23}$$

$$\langle p, \lambda^*\rangle = u \tag{3.B.24}$$

$$0 < \nu.$$

We can solve (3.B.23) to obtain

$$\nu\mathcal{R}(p, \theta) = 2W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*). \tag{3.B.25}$$

Taking squared norms both sides, we obtain

$$\begin{aligned} \nu^2\|\mathcal{R}(p, \theta)\|^2 &= 4\|W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*)\|^2 \\ &= 4\mathcal{C}(\theta, p, u). \end{aligned} \tag{3.B.26}$$

Plugging in $\nu = 2\mathcal{C}(\theta, p, u)^{1/2}/\|\mathcal{R}(p, \theta)\|$ back to (3.B.23), we obtain

$$p = \|\mathcal{R}(p,\theta)\|\mathcal{C}(\theta,p,u)^{-1/2}\Pi_{\mathcal{J}(\theta)}(\theta)'W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*).$$

Now, substitute this into (3.B.24),

$$\begin{aligned}
u &= \|\mathcal{R}(p,\theta)\|\mathcal{C}(\theta,p,u)^{-1/2} \\
&\qquad \times \left\langle \Pi_{\mathcal{J}(\theta)}(\theta)'W_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*), \lambda^* \right\rangle \\
&= \|\mathcal{R}(p,\theta)\|\mathcal{C}(\theta,p,u)^{-1/2} \\
&\qquad \times \left\langle W^{1/2}_{\mathcal{J}(\theta)}(\theta)(\mathbb{G}_{\mathcal{J}(\theta)}(\theta) + \Pi_{\mathcal{J}(\theta)}(\theta)\lambda^*) \circ s(\theta, \lambda^*), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^* \right\rangle \\
&= \|\mathcal{R}(p,\theta)\|\mathcal{C}(\theta,p,u)^{-1/2} \left\langle \frac{\nu}{2}\mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^* \right\rangle \\
&= \left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^* \right\rangle,
\end{aligned}$$

where the second equality follows from (3.B.25). Using (3.B.25) and the result above, the right hand side of (3.B.26) can be alternatively written as

$$\begin{aligned}
&2\nu \left(\left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta) \right\rangle + \left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)\lambda^* \right\rangle \right) \\
&= 2\nu \left(\left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta) \right\rangle + u \right).
\end{aligned}$$

Therefore, from (3.B.26), we obtain

$$\begin{aligned}
\nu &= 2\|\mathcal{R}(p,\theta)\|^{-1} \left(\left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta) \right\rangle + u \right) \\
&= 2\|\mathcal{R}(p,\theta)\|^{-1} \left(\left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta) \right\rangle + u \right)_+,
\end{aligned}$$

where the second equality follows from the fact $\nu > 0$. As $\mathcal{C}(\theta, p, u) = \|\mathcal{R}(p,\theta)\|\nu^2/4$, we have

$$\mathcal{C}(\theta,p,u) = \|\mathcal{R}(p,\theta)\|^{-1} \left(\left\langle \mathcal{R}(p,\theta), W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta) \right\rangle + u \right)_+^2.$$

Take the supremum over $H(p, \Theta_I)$. The result follows. □

Proof of Corollary 3.4.3. We first analyze the Wald statistic $\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}_n(p,t)+$

$t^{1/2}\}_+^2$. By Corollary 3.4.2, the distributional limit $\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)+t^{1/2}\}_+^2$ of this statistic can be represented as

$$
\begin{aligned}
&\sup_{p\in\mathbb{S}^{d-1}}\left\{-\sup_{\theta\in H(p,\Theta_I)}-\left\langle(\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p,\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\right\rangle\right\}_+^2\\
&=\sup_{p\in\mathbb{S}^{d-1}}\left\{\inf_{\theta\in H(p,\Theta_I)}\left\langle(\Pi_{\mathcal{J}(\theta)}(\theta)\Pi_{\mathcal{J}(\theta)}(\theta)')^{-1}\Pi_{\mathcal{J}(\theta)}(\theta)p,\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\right\rangle\right\}_+^2\\
&=\sup_{p\in\mathbb{S}^{d-1}}\left\langle\left(\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))'\right)^{-1}\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))p,\mathbb{G}_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\right\rangle_+^2\\
&=\mathbf{Z},
\end{aligned}
$$

where we used $H(p,\Theta_I)=\{\theta_I(p)\}$ to obtain the third equality. For the QLR statistic,

$$
\sup_{\theta\in\Theta_I} nQ_n(\theta)\xrightarrow{d}\sup_{p\in\mathbb{S}^{d-1}}\mathcal{L}(p,0)
$$

by Theorem 3.4.2 and the continuous mapping theorem. By Theorem 3.4.2, this limit can be represented as

$$
\begin{aligned}
&\sup_{p\in\mathbb{S}^{d-1}}\sup_{\theta\in H(p,\Theta_I)}\left(\left\langle\mathcal{R}(p,\theta),W^{1/2}_{\mathcal{J}(\theta)}(\theta)\mathbb{G}_{\mathcal{J}(\theta)}(\theta)\right\rangle\right)_+^2\\
&=\sup_{p\in\mathbb{S}^{d-1}}\left\langle\left(\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))'\right)^{-1}\Pi_{\mathcal{J}(\theta_I(p))}(\theta_I(p))p,\mathbb{G}_{\mathcal{J}(\theta_I(p))}(\theta_I(p))\right\rangle_+^2\\
&=\mathbf{Z}.
\end{aligned}
$$

For the second part, note that $\tau^*_{1-\alpha}$ is the $1-\alpha$ quantile of $\mathbf{Z}$. Therefore, it suffices to show that $t^*_{1-\alpha}$ is also the $1-\alpha$ quantile of $\mathbf{Z}$ under our hypotheses. For that, we can write

$$\begin{aligned}
t^*_{1-\alpha} &= \inf\left\{t : P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}(p,t)\}_+ \leq 0\right) \geq 1-\alpha\right\} \\
&= \inf\left\{t : P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-t^{1/2} - \mathcal{Z}^*(p)\}_+ \leq 0\right) \geq 1-\alpha\right\} \\
&= \inf\left\{t : P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}^*(p)\}_+ \leq t^{1/2}\right) \geq 1-\alpha\right\} \\
&= \inf\left\{t : P\left(\sup_{p\in\mathbb{S}^{d-1}}\{-\mathcal{Z}^*(p)\}_+^2 \leq t\right) \geq 1-\alpha\right\} \\
&= \inf\left\{t : P\left(\mathbf{Z} \leq t\right) \geq 1-\alpha\right\},
\end{aligned}$$

where the third equality follows from the fact that for any $x \geq 0$ and a continuous function f, $\sup_{p\in\mathbb{S}^{d-1}}\{-x + f(p)\}_+ \leq 0 \Leftrightarrow \sup_p\{f(p)\}_+ \leq x$. □

3.B.8 Proof of Theorems in Section 3.5

The following lemma is often useful to identify the weak epilimit of a sequence of stochastic processes.

Lemma 3.B.9: *Let $\Gamma(\mathbb{R}^d)$ be the space of convex lsc functions on $\mathbb{R}^d$ that are proper and have effective domains with nonempty interiors (or equivalently are finite on an open set). Suppose that $\{\xi_n, n \geq 1\}$ is a sequence in $\Gamma(\mathbb{R}^d)$ and let $\mathbb{Q}$ be a countable dense subset of $\mathbb{R}^d$. If $\xi_n \overset{f.d.}{\to} \xi$ on $\mathbb{Q}$ where $P(\xi \in \Gamma(\mathbb{R}^d)) = 1$, then $\xi_n \overset{e.d.}{\to} \xi$.*

Proof. See Lemma 3.1. in Geyer (2003). □

Proof of Theorem 3.5.1. Let

$$\begin{aligned}
\tilde{\zeta}_n(\theta,\lambda) &= nQ_n(\theta + \lambda/\sqrt{n}) \\
&= (\sqrt{n}(\hat{E}_n(X_{1i}) - \theta_1) - \lambda + \sqrt{n}(\theta_1 - \theta))_+^2 \\
&\quad + (\sqrt{n}(\theta - \theta_2) + \lambda - \sqrt{n}(\hat{E}_n(X_{2i}) - \theta_2))_+^2 + \infty \times 1_{\theta\notin\Theta}.
\end{aligned}$$

This function is convex in (θ, λ), lsc, and has an effective domain with nonempty interior. Under our hypothesis, the finite dimensional limit of $\tilde{\zeta}_n(\theta, \lambda)$ is

$$\tilde{\zeta}(\theta, \lambda) = (Z_1 - \lambda + \varsigma_1(\theta))_+^2 + (\varsigma_2(\theta) + \lambda - Z_2)_+^2,$$

where $(Z_1, Z_2)' \sim N(0, \Omega)$ and

$$\varsigma_1(\theta) = \begin{cases} \infty & \theta < \theta_1 \\ 0 & \theta = \theta_1 \\ -\infty & \theta > \theta_1 \end{cases}, \qquad \varsigma_2(\theta) = \begin{cases} \infty & \theta > \theta_2 \\ 0 & \theta = \theta_2 \\ -\infty & \theta < \theta_2 \end{cases}.$$

This function is convex and lsc, and finite on an open interval (θ_1, θ_2), and $\tilde{\zeta}_n(\theta, \lambda) \xrightarrow{f.d.} \tilde{\zeta}(\theta, \lambda)$. Therefore, Lemma 3.B.9 is applicable. Thus, the weak epi-limit coincides with the finite dimensional limit.

Using the representation result in Corollary 3.4.2, we can derive a closed form for $\mathcal{Z}$. For example, when $p = -1$ and $\theta \in H(-1, \theta) = \{\theta_1\}$, we have $\mathcal{J}(\theta_1) = 1$, $\mathcal{R}(-1, \theta) = 1$. Therefore,

$$\mathcal{Z}(-1, t) = t^{1/2} - Z_1.$$

Similarly,

$$\mathcal{Z}(1, t) = t^{1/2} + Z_2.$$

Therefore, the limiting process $\mathcal{Z}(p, t)$ has mean $t^{1/2}\iota$ and covariance $E[(\mathcal{Z}(-1,t) - t^{1/2})(\mathcal{Z}(-1,t) - t^{1/2})] = \Omega_{11}$, $E[(\mathcal{Z}(1,t) - t^{1/2})(\mathcal{Z}(1,t) - t^{1/2})] = \Omega_{22}$, and $E[(\mathcal{Z}(-1,t) - t^{1/2})(\mathcal{Z}(1,t) - t^{1/2})] = -\Omega_{12}$. By Corollary 3.4.1,

$$\sqrt{n} d_H(\hat{\Theta}_n(t), \Theta_I) \xrightarrow{d} \max\{|\mathcal{Z}(-1,t)|, |\mathcal{Z}(1,t)|\}$$

and

$$\sqrt{n}\vec{d}_H(\Theta_I, \hat{\Theta}_n(t)) \xrightarrow{d} \max\{-\mathcal{Z}(-1,t)_+, -\mathcal{Z}(1,t)_+\}. \square$$

Proof of Theorem 3.5.1. The result for $\mathcal{W}_n$ follows directly from Theorem 3.5.1. The results for $\mathcal{QLR}_n$ and $\tilde{\mathcal{W}}_n$ are due to Chernozhukov, Hong, and Tamer (2004) and BM respectively. $\square$

Bibliography

Andersen, P. K., and R. D. Gill (1982): "Cox's Regression Model for Counting Processes: A Large Sample Study," *The Annals of Statistics*, 10(4), 1100–1120.

Andrews, D. W. K. (1999): "Estimation When a Parameter is on a Boundary," *Econometrica*, 67(6), 1341–1383.

Andrews, D. W. K., S. Berry, and P. Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," Working Paper, Yale University.

Andrews, D. W. K., and P. Guggenberger (2009): "Validity of Subsampling and 'Plug-In Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25(3), 669–709.

Andrews, D. W. K., and G. Soares (2009): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, forthcoming.

Attouch, H. (1986): *Variational Convergence for Functions and Operators (Applicable Mathematics Series)*. Pitman (Advanced Publishing Program), Boston, MA.

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

Beer, G. A. (1993): *Topologies on Closed and Closed Convex Sets (Mathematics and Its Applications)*. Kluwer Academic Publishers, Dordrecht.

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76(4), 763–814.

Beresteanu, A., F. Molinari, and E. Y. Wang (2009): "SVRV for Matlab: Beta Version: Set Valued Random Variables Estimation and Inference," Duke University.

Bontemps, C., T. Magnac, and E. Maurin (2007): "Set Identified Linear Models," Working Paper, Toulouse School of Economics.

Bugni, F. A. (2009a): "Bootstrap Inference in Partially Identified Models," Working Paper, Duke University.

——— (2009b): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, forthcoming.

Canay, I. A. (2009): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," Working Paper, Northwestern University.

Chernoff, H. (1954): "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 25(3), 573–578.

Chernozhukov, V. (2005): "Extremal Quantile Regression," *Annals of Statistics*, 33(2), 806–839.

Chernozhukov, V., and H. Hong (2001): "Likelihood Inference with Density Jump," Working Paper, Massachusetts Institute of Technology.

——— (2004): "Likelihood Estimation and Inference in a Class of Nonregular Econometric Models," *Econometrica*, 72(5), 1445–1480.

Chernozhukov, V., H. Hong, and E. Tamer (2004): "Inference on Parameter Sets in Econometric Models," Working Paper, Massachusetts Institute of Technology.

——— (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.

Fan, Y., and S. S. Park (2007): "Confidence Sets for Some Partially Identified Parameters," Working Paper, Vanderbilt University.

Galichon, A., and M. Henry (2007): "Dilation Bootstrap: A Natural Approach to Inference in Incomplete Models," Columbia University, Working Paper.

——— (2009): "A test of Non-identifying Restrictions and Confidence Regions for Partially Identified Parameters," *Journal of Econometrics*, 152(2), 186 – 196, Nonparametric and Robust Methods in Econometrics.

GEYER, C. J. (1994): "On the Asymptotics of Constrained M-Estimation," *The Annals of Statistics*, 22(4), 1993–2010.

——— (2003): "On the Asymptotics of Convex Stochastic Optimization," Unpublished Manuscript, University of Minnesota.

GUGGENBERGER, P., J. HAHN, AND K. KIM (2008): "Specification testing under moment inequalities," *Economics Letters*, 99(2), 375–378.

HAHN, J., AND G. RIDDER (2009): "Partial Identification and Confidence Intervals," Working Paper, University of California, Los Angeles.

HAN, C., AND P. C. B. PHILLIPS (2006): "GMM with Many Moment Conditions," *Econometrica*, 74(1), 147–192.

HÖRMANDER, L. (1955): "Sur la Fonction D'Appui des Ensembles Convexes Dans un Espace Localement Convexes," *Arkiv fur Matematik*, 3, 181–186.

HOROWITZ, J. L., AND C. F. MANSKI (1998): "Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations," *Journal of Econometrics*, 84(1), 37–58.

——— (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95(449), 77–84.

IMBENS, G. W., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

KAIDO, H., AND H. WHITE (2008): "A Two-Stage Procedure for Partially Identified Models," Discussion Paper, University of California San Diego.

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

KNIGHT, K. (1999): "Epi-convergence in Distribution and Stochastic Equisemicontinuity," Working Paper, University of Tronto.

LEHMANN, E., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses (Springer Texts in Statistics)*. Berlin: Springer, 3rd edn.

LI, S., Y. OGURA, AND V. KREINOVICH (2002): *Limit Theorems and Applications of Set-Valued and Fuzzy Set-Valued Random Variables (Theory and Decision Library B)*. Berlin: Springer.

MANSKI, C. F. (2003): *Partial Identification of Probability Distributions (Springer Series in Statistics)*. Berlin: Springer.

MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546.

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

MOON, H. R., AND F. SCHORFHEIDE (2009): "Bayesian and Frequentist Inference in Partially Identified Models," Working Paper, University of Southern California.

NEWEY, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59(4), 1161–1167.

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, chap. 36. New York: Elsevier.

PAKES, A., J. PORTER, K. HO, AND J. ISHII (2005): "Moment Inequalities and Their Application," Working Paper, Harvard University.

PFLUG, G. C. (1992): "Asymptotic Dominance and Confidence for Solutions of Stochastic Programs," *Czechoslovak Journal for Operations Ressearch*, 1, 21–30.

——— (1995): "Asymptotic Stochastic Programs," *Mathematics of Operations Research*, 20(4), 769–789.

POLITIS, D., J. ROMANO, AND M. WOLF (1999): *Subsampling*. Berlin: Springer.

ROCKAFELLAR, R. T., AND R. J.-B. WETS (2005): *Variational Analysis*. Berlin: Springer, 2nd edn.

ROMANO, J. P., AND A. M. SHAIKH (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, Special Issue in Honor of T. W. Anderson.

——— (2009): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, forthcoming.

ROSEN, A. M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146(1), 107 – 117.

SALINETTI, G., AND R. J.-B. WETS (1986): "On the Convergence in Distribution of Measurable Multifunctions (Random Sets), Normal Integrands, Stochastic Processes and Stochastic Infima," *Mathematics of Operations Research*, 11(3), 385–419.

SANTOS, A. (2007): "Inference in Nonparametric Instrumental Variables with Partial Identification," Working Paper, University of California San Diego.

STOYE, J. (2009): "More on Confidence Regions for Partially Identified Parameters," *Econometrica*, 77(4), 1299–1315.

VAN DER VAART, A., AND J. WELLNER (2000): *Weak Convergence and Empirical Processes: With Applications to Statistics*. Berlin: Springer.

WHITE, H. (2001): *Asymptotic Theory for Econometricians, Revised Edition*. San Diego: Academic Press.

WOUTERSEN, T. (2006): "A Simple Way to Calculate Confidence Intervals for Partially Identified," Working Paper, Johns Hopkins University.

YILDIZ, N. (2009): "Consistency of Plug-In Estimators of Upper Contour Consistency of Plug-In Estimators of Upper Contour and Level Sets," Working Paper, University of Rochester.