ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF ECONOMICS

WORKING PAPER SERIES **04-2011**

ON THE EXISTENCE OF STRONGLY CONSISTENT INDIRECT ESTIMATORS WHEN THE BINDING FUNCTION IS MULTIVALUED

Stelios Arvanitis

76 Patission Str., Athens 104 34, Greece
Tel. (++30) 210-8203911 - Fax: (++30) 210-8203301
www.econ.aueb.gr

On the Existence of Strongly Consistent Indirect Estimators when the Binding Function is Multivalued

Stelios Arvanitis
Athens University of Economics and Business

March 18, 2011

Abstract

In this paper we establish the definition of IE and their strong consistency, when the binding function is a compact valued correspondence under mild conditions. These results are generalizations of the analogous results in the relevant literature, hence permit a broader scope of statistical models. We provide some examples that concern linear models with weak instruments, and conditionally heteroskedastic ones.

KEYWORDS: Indirect estimator, lower semicontinuous function, random set, normal integrand, measurable selection, upper topology, Fell topology, epi convergence, binding correspondence, indirect identification, weak instuments, conditional heteroskedasticity.

1 Introduction

The set of indirect estimators (henceforth IE) is a subset of the set of M-estimators defined in the context of (semi-) parametric statistical models, associated with the requirement that their derivation involves strictly more than one optimization procedures. They are minimizers of criteria (inversion criterion) that are functions of an auxiliary estimator, itself derived as an extremum estimator. The latter minimizes a criterion function (auxiliary criterion) that partially reflects the structure of a possibly misspecified auxiliary statistical model. The inversion criterion depends on the auxiliary estimator, as well as on a function defined on the parameter space of the statistical model that "approximates" properties of the aforementioned estimator. The latter is usually termed *binding function*. Minimization of the inversion criterion, which usually has the form of a stochastic norm, essentially inverts the binding function thereby obtaining the indirect estimator.[1]

In the present paper, we are concerned with the issue of the existence and strong consistency of an IE, under more general conditions than the usual ones employed for analogous establishments in the econometric literature. More specifically, we are occupied with the particular question *without* necessitating the uniform almost sure convergence on compacta of the auxiliary criterion to a real valued function that possesses a unique minimizer.

Our generalization is thus threefold. *First*, using mild assumptions on the structure of the aforementioned criterion functions, we are occupied with weaker notions of convergence of the relevant sequences of criterion functions, that essentially concern the almost sure asymptotic behavior of their epigraphs and are suitable for the study of the asymptotic behavior of their minimizers.

Secondly, we allow for the analogous limit functions to have values on the extended real line. This also generalizes the set of the statistical models that are in accordance with these conventions, hence the analogous set of estimators under this scope.

Finally, we allow for the set of minimizers of the relevant limit functions to be generally non empty and compact valued and therefore, we are concerned with the issue of the definition and the asymptotic behavior of indirect estimation procedures, when the aforementioned binding function is actually a compact valued *correspondence*. This is essentially the representation of a function defined on the parameter space of the statistical model at hand, with values on the hyperspace of the compact subsets of the parameter space of the auxiliary model.

It can be perceived that all the above are generalizations of the analogous results residing in the relevant literature which in fact is quite limited. For the sake of completeness we notice that indirect inference algorithms were initially employed in [14], formally introduced by [5], complemented by [6] and extended by [4].

The structure of the paper is as follows. We first describe briefly some general notions that are essentially used in the sequel and formulate our general set up. We then define and study the asymptotic behavior of the auxiliary estimator, the binding correspondence and finally of the IE. No-

[1] The set of IE can be enlarged when the binding function itself, in the inversion criterion, is approximated in some relevant sense by a possibly random function defined on the parameter space of the statistical model.

tice that notions concerning the asymptotic behavior of random sets that emerge as the arg min correspondences of random semicontinuous functions are briefly described locally. We conclude with some examples and pose some questions for future research.

2 General Notions, Assumptions and Main results

2.1 Some General Notions

Fell and Upper Topology

Let (E, τ_E) denote a general topological space. We identify the space with E when there is no risk of confusion. We denote with $F(E)$ the set of closed subsets of E, when endowed with the Fell topology which is defined by the use of the following subbase.

Definition D.1 *$F(E)$ is generated by the subbase consisting of*

1. $F_G = \{\mathcal{F} \text{ closed} : \mathcal{F} \cap G \neq \varnothing\}$, *$\forall G \in \tau_E$, and*
2. $F^K = \{\mathcal{F} \text{ closed} : \mathcal{F} \cap K = \varnothing\}$, *$\forall K \subset E$ and compact.*

Due to Theorem B.6 of [9] we have that when E is locally compact, $\mathcal{F}_n \to \mathcal{F}$ with respect to the Fell topology *iff* $\mathcal{F} = \liminf_n \mathcal{F}_n = \limsup_n \mathcal{F}_n$, where $\liminf_n \mathcal{F}_n$ is the set comprised of the limit points of any possible $x_n \in \mathcal{F}_n$, and $\limsup_n \mathcal{F}_n$ is the one comprised of the analogous cluster points. Hence, in this case this type of convergence coincides with the Painleve-Kuratowski convergence (see among others, Appendix B of [9]). Moreover if E is locally compact and Hausdorff (LCHS), the Fell topology is *metrizable*.

When the subbasic sets are restricted we obtain another useful topology on the set of closed sets.

Definition D.2 *The upper topology $U(E)$ is generated by the subbase consisting of*

$$F_G = \{\mathcal{F} \text{ closed} : \mathcal{F} \cap G \neq \varnothing\}, \forall G \in \tau_E$$

The upper topology is extremely useful for the analysis of the asymptotic behavior of sequences of sets of minimizers.

Remark R.1 *When E is compact $U^*(E) = U(E) - \{\varnothing\}$ is hemimetrizable due to Proposition 4.2.2 of [13].*

Epigraphs of Semicontinuous Functions and Epiconvergence

We consider now the case that E is LCHS, we let $\overline{\mathbb{R}}$ denote the two point compactification of $\mathbb{R}$, equipped with the final topology that makes the relevant inclusion continuous, i.e. the *extended* real line, and $c : E \to \overline{\mathbb{R}}$.

Definition D.3 *The epigraph of c is*

$$\mathbf{epi}\,(c) = \{(x,t) \in E \times \mathbb{R} : c\,(t) \leq t\}$$

We note that, despite the fact that the image of c may include non real numbers, $\mathbf{epi}\,(c)$ is by definition a subset of $E \times \mathbb{R}$. If c is lower semicontinuous (lsc) we have that due to Proposition A.2 of [9], $\mathbf{epi}\,(c) \in F\,(E \times \mathbb{R})$ with respect to the obvious product topology. Hence any relevant lsc function can be identified with its epigraph, which in turn lies in a space endowed with Fell topology, which in turn implies a notion of convergence.

Definition D.4 *A sequence $\{c_n\}_n$ of lsc functions epiconverges to c ($c_n \xrightarrow{e} c$) iff $\mathbf{epi}\,(c_n) \to \mathbf{epi}\,(c)$ with respect to the Fell topology.*

A sequential characterization of epiconvergence that is described in Proposition 3.2 of [9] dictates that the notion is equivalent to that, $\forall x \in E$:

1. $\liminf_{n\to\infty} c_n\,(x_n) \geq c\,(x)$ for any sequence such that $x_n \to x$ and
2. $\limsup_{n\to\infty} c_n\,(x_n) \leq c\,(x)$ for at least one sequence such that $x_n \to x$.

It is also true that the epi-limit function is also lsc. The notion of epiconvergence is particularly suitable for the description of the asymptotic behavior of the set of minimizers of sequences of lsc functions. Theorem 3.4 of [9] dictates that if $c_n \xrightarrow{e} c$ then $\limsup_{n\to\infty} \arg\min_{x\in E} c_n \subset \arg\min_{x\in E} c$ and hence, $\arg\min_{x\in E}$ is $U\,(E)$ continuous as a function on the space of lower semicontinuous functions equipped with the topology of epiconvergence. This result can be easily extended to near minimizers using Theorem 7.31.b of [7]. If $\{c_n\}_n$ and c are cofinitely proper, i.e. they do not assume the value $-\infty$ while they are not constant on $+\infty$ and cofinitely inf-compact, i.e. their level sets are compact, then the corresponding sets of minimizing point are non-empty and compact, i.e. belong to the space $K\,(E)$ comprised of the non empty compact subsets of E with the subspace Fell topology. Inf-compactness follows readily in the case that E is itself compact.

Closed and Compact Valued Correspondences-Random Closed Sets

A closed valued correspondence is by definition a representation of an underlying function c from a set Ω to $F(E)$ (i.e. a closed valued multifunction), when this is considered as a relation in $X \times E$. A correspondence is usually abbreviated as **cor** $:X \rightrightarrows E$, while the benefit of not directly considering the underlying function, is the fact that we can consider the graph of **cor** as the set $\{(\omega, x) : x \in c(\omega)\}$ with values in $\Omega \times E$ instead of the set $\{(\omega, \mathcal{F}) : \mathcal{F} = c(\omega)\}$ with values in $\Omega \times F(E)$. When $c(\omega)$ is compact on Ω, then the correspondence in obviously termed as compact valued. In this sense, $\mathbf{epi}\,(c_n)$ defined in the previous paragraph, can be identified by a closed valued correspondence that is compact valued when inf-compactness holds. In the following we do not make explicit distinction between the correspondence and the underlying multifunction.

Since $F(E)$ is actually a topological space (usually termed as a *hyperspace*), it also defines a Borel algebra which we abbreviate by $\mathcal{B}(F)$ and is usually termed as *Effron algebra.* If $(\Omega, \mathcal{J})$ is a measurable space, then c is a random closed *iff* $\{\omega \in \Omega : c(\omega) \in \overline{\mathcal{F}}\} \in \mathcal{J}$ for any $\overline{\mathcal{F}} \in \mathcal{B}(F)$.

2.2 Assumptions and Main results

General Assumptions and the Structure of the Statistical Problem

We are now ready to state our framework and describe the underlying statistical problem. Let the triad $(\Omega, \mathcal{J}, P)$ denote a *complete* probability space. Let also (Θ, d_Θ) and (B, d_B) denote two *separable compact metric spaces,* and the relevant metric topologies by τ_Θ and τ_B analogously. Let $\mathcal{B}(\Theta)$, $\mathcal{B}(B)$ denote the corresponding Borel algebras respectively, and denote with $\mathcal{B}(\mathbb{R})$ and $\mathcal{B}(\overline{\mathbb{R}})$ the Borel algebras of the real and the extended real numbers with respect to the usual topologies analogously. Consider a sequence of functions $c_n(\omega, \theta, \beta) : \Omega \times \Theta \times B \to \overline{\mathbb{R}}$. For a sequence of functions $y_n : \Omega \times \Theta \to K_n$, where K_n is a topological space, c_n could be defined as $q_n(y_n, \beta)$ where $q_n : K_n \times B \to \overline{\mathbb{R}}$. We abbreviate with P **a.s.** any statement that concerns elements of $\mathcal{J}$ of unit probability.

Assumption A.1 *Let the following hold:*

1. c_n *is* $\mathcal{B}(\mathbb{R}) / \mathcal{J} \otimes \mathcal{B}(\Theta) \otimes \mathcal{B}(\Phi)$ *measurable.*

2. $c_n(\omega, \theta, \cdot) : B \to \overline{\mathbb{R}}$ *is lower semicontinuous and proper* P ***a.s.,*** $\forall \theta \in \Theta$.

Remark R.2 *By assumption A.1.2* $\arg\min_B c_n(\omega, \theta, \beta)$ *is non empty and compact due to theorem 1.9 of [7]* $\forall n$, $\forall \theta \in \Theta$, $c_n(\omega, \theta, \cdot)$ P ***a.s.*** *due to the* *fact that c_n is inf-compact P **a.s.** $\forall\theta \in \Theta$. This follows from the fact that B is compact Hausdorff. $\forall n$, $\forall\theta \in \Theta$, $c_n(\omega,\theta,\cdot)$ is usually termed as the auxiliary criterion.*

Consider the family of θ-parametrized correspondences $\mathbf{epi}_n(\omega,\theta) \doteqdot \mathbf{epi}(c_n(\omega,\theta,\cdot))$. Due to the fact that B is locally compact, $\mathbf{epi}_n(\omega,\theta)$ is a random closed set in the sense of the previous paragraph, i.e. a $\mathcal{B}(F(B))/\mathcal{J}\otimes\mathcal{B}(\Theta)$-measurable correspondence. Hence $\mathbf{epi}_n(\omega,\cdot)$ is an $\mathcal{B}(F(B))/\mathcal{J}$-measurable correspondence due to the measurability of the relevant projection. Using the metrizability of the relevant Fell topology, we denote a corresponding metric be denoted by $\mathcal{D}_{F(B)}$ and the open ball of center $\mathbf{epi}_n(\omega,\theta)$ and radius $\varepsilon > 0$ by $B_{\mathcal{D}_{F(B)}}(\mathbf{epi}_n(\omega,\theta),\varepsilon)$.

Remark R.3 *q_n can be implied by some part of the structure of an **auxiliary model,** which in turn is a statistical model defined on the same measurable space, with B as its parameter space. It could be a reparametrization of the underlying statistical model.*

Definition D.5 *Let θ_0 be an arbitrary element of Θ. The statistical problem in question concerns the existence of strongly consistent indirect estimators of θ_0 from $c_n(\omega,\theta_0,\beta)$.*

Auxiliary Estimator

We are now ready to define and explore properties of the auxiliary estimator.

Definition D.6 *For a non negative random variable ε_n, the auxiliary correspondence $\beta_n^{\#}(\omega,\theta,\varepsilon_n)$ is defined as*

$$\beta_n^{\#}(\omega,\theta,\varepsilon_n) = \left\{\beta \in B : c_n(\omega,\theta,\beta) \leq \inf_B c_n(\omega,\theta,\cdot) + \varepsilon_n\right\}$$

Lemma 2.1 *$\beta_n^{\#}(\omega,\theta,\varepsilon_n)$ is $\mathcal{B}(F(B))/\mathcal{J}\otimes\mathcal{B}(\Theta)$-measurable, hence $\mathcal{B}(F(B))/\mathcal{J}$-measurable $\forall\theta \in \Theta$, and P **a.s.** non empty-compact valued $\forall\theta \in \Theta$.*

Proof. First $\beta_n^{\#}(\omega,\theta,\varepsilon_n)$ is non empty due to A.1. Second, from separability of B and the joint measurability of c_n due to assumption A.1, the result follows from Proposition 3.10.(i) of [9] which itself applies due to the fact that $\inf_B c_n(\omega,\beta)$ is a random variable due to separability of B and the joint measurability of c_n, and Proposition 3.10.(i)) that guarantees compactness and measurability for $a = \inf_B c_n(\omega,\beta)$ in the first case and $a = \inf_B c_n(\omega,\beta)+\varepsilon_n$ in the second. Pointwise measurability then follows. ■

Remark R.4 $\beta_n^{\#}(\omega,\theta,0) = \arg\min_B c_n(\omega,\theta,\beta)$ P ***a.s.***

Lemma 2.2 *There exists a $\mathcal{B}(B)/\mathcal{J}\otimes\mathcal{B}(\Theta)$-measurable, $\forall\theta\in\Theta$, random element $\beta_n(\varepsilon_n):\Omega\times\Theta\rightarrow\mathbb{R}$ termed as auxiliary selection, defined as*

$$c_n(\omega,\beta_n(\varepsilon_n)(\omega,\theta)) \leq \inf_B c_n(\omega,\theta,\beta) + \varepsilon_n$$

Proof. The result follows from lemma 2.1 and the fundamental selection theorem (Theorem 2.13 of [9]). ■

Epi-Limit Objective and Characterization

The following assumption facilitates the aforementioned asymptotic concern.

Assumption A.2 *There exists a function $c:\Theta\times B\rightarrow\overline{\mathbb{R}}$ with the relevant epigraph correspondence denoted as* $\mathbf{epi}(\theta) = \mathbf{epi}(c(\theta,\cdot))$ *such that*

1. *$\forall\theta\in\Theta$, $c_n \xrightarrow{e} c$ P* ***a.s.,***
2. *$c(\theta,\cdot)$ is proper $\forall\theta\in\Theta$, and*
3. *$\theta\rightarrow c(\theta,\cdot)$ is epicontinuous on Θ which means that is a continuous mapping on Θ into the space of lower semicontinuous real functions equipped with the Fell topology, i.e. it satisfies that $\forall\theta\in\Theta$, $\forall\theta_n\rightarrow\theta$, $\forall\beta\in B$*
   - *(a) $\forall\beta_n\rightarrow\beta$, $\liminf_{n\rightarrow\infty} c(\theta_n,\beta_n) \geq c(\theta,\beta)$ and*
   - *(b) $\exists\beta_n\rightarrow\beta$, such that $\limsup_{n\rightarrow\infty} c(\theta_n,\beta_n) \leq c(\theta,\beta)$.*

Remark R.5 *$\forall\theta\in\Theta$, $c(\theta,\cdot):B\rightarrow\overline{\mathbb{R}}$ is lower semicontinuous due to the fact that epiconvergence preserves this type of continuity (see proposition 7.4.a of [7]). In the case that c_n is ergodic and in the form of an average then the assumed epiconvergence would follow from pointwise convergence and a condition of the form $-\infty < E(\inf_B c_n(\omega,\theta,\beta)) < \infty$ (see [8]). Properness is actually an ad hoc consideration (see, for example, in [15] Part 1, (ii) in association with Part 2 of the proof of Theorem 5.3.1, where c_n is a quasi likelihood function and Θ coincides with B. Inf-compactness follows from the compactness of Θ.*

Remark R.6 *Notice that for the $\theta\rightarrow c(\theta,\cdot)$ continuity, only sequences and not generally nets on Θ and B are considered, due to the fact that both of these spaces are first countable. The required continuity can then be obtained from conditions that guarantee:*

1. *The almost everywhere continuity of* $\theta \to c_n(\omega, \theta, \beta)$ *which in turn would follow from A.2.3.a-b applied to* c_n *almost everywhere for any* n*, which is equivalent to the almost everywhere continuity of* $\mathcal{D}_{F(B)}(\mathbf{epi}_n(\omega, \theta), \mathbf{epi}_n(\omega, \theta'))$ *for any* $\theta, \theta' \in \Theta$*. If* $\sup_B |c_n| < \infty$ P ***a.s.,*** *for any* θ*, and due to the fact that the Fell topology is weaker than the topology of uniform convergence (see [7] Theorem 7.14), the aforementioned almost everywhere continuity would follow from the almost everywhere continuity of* $\sup_B |c_n(\omega, \theta, \beta) - c_n(\omega, \theta', \beta)|$*, for any* $\theta, \theta' \in \Theta$*.*

2. *Given (1), it is then sufficient that* $\sup_{\theta\in\Theta} \mathcal{D}_{F(B)}(\mathbf{epi}_n(\omega, \theta), \mathbf{epi}(\theta))$ *converges almost surely to zero due to corollary 46.6 of [10], which in turn applies due to the metrizability of the Fell topology from the local compactness of* B*, and the local compactness of* Θ*. In the case that* $\sup_{\Theta\times B} |c_n| < \infty$ P ***a.s.***, *and due to the fact that the Fell topology is weaker than the topology of uniform convergence (see [7] Theorem 7.14), the aforementioned continuity would follow if* $\sup_{\theta\in\Theta}\sup_{\beta\in B} |c_n(\omega, \theta, \beta) - c(\theta, \beta)|$ *converges* P ***a.s.*** *to zero.*

Corollary 2.3 *If assumption A.2.1-3 is valid then*

1. *The correspondence* $b(\theta) = \arg\min_B c(\theta, \beta)$ *is non empty-compact valued* $\forall\theta \in \Theta$ *and,*

2. $b(\cdot)$ *is* $U(B)/\tau_\Theta$*-continuous.*

Proof. Non emptiness and compact valuedness follows from R.5. Now, due to A.2.2 $\theta \to c(\theta, \beta)$ is a continuous mapping from Θ to the space of lower semicontinuous functions equipped with the Fell topology due to A.2.2, and the $\arg\min_B$ correspondence is $T_{\mathcal{U}}/T_{\mathcal{F}}$-continuous at c on Θ (see equation 3.1 of Theorem 5.3.4 and proposition Appendix.D.2 of [9]). The result follows from $b(\theta) = \arg\min_B \circ c(\theta, \beta)$. ∎

Remark R.7 *If* $b(\theta_0)$ *is single valued, then from the previous corollary follows that* $\limsup_{\theta\to\theta_0} b(\theta) = \{b(\theta_0)\}$ *since* B *is compact.*

Upper Pseudo-Consistency of the Auxiliary Correspondence

We translate the almost sure asymptotic inclusion of the set of cluster points of the auxiliary correspondence at θ_0 to $b(\theta_0)$, to the asymptotic behavior of the sequence of auxiliary selections. We first make the following assumption that concerns the almost sure convergence of optimization "error" to zero.

Assumption A.3 *$\varepsilon_n(\omega)$ converges to zero P* ***a.s.***

In the following result we utilize the concept of the distance of a point from a closed set in a metric space, that is defined as the infimum between the distances of every point in the set from the particular one. Notice that this infimum is measurable due to the fact that B is separable, and that separability is a hereditary property. In this case we use the notation $d_B(c, A) \doteqdot \inf_{c' \in A} d(c, c')$ where $c \in B$ and $A \in F(B)$. Notice that due to the fact that A is closed, then $d_B(c, A) = 0$ if and only if $c \in A$. A natural convention dictates that $d_B(c, \varnothing) = \infty$.

Lemma 2.4 *If assumptions A.2 and 2.4 are valid then*

$$d_B\left(\beta_n(\varepsilon_n)(\omega, \theta_0), b(\theta_0)\right) \to 0 \ P \ \textbf{\textit{a.s.}}$$

Proof. Due to A.3, A.2 and Theorem 7.31.b of [7], we have that $\limsup_{n\to\infty} \beta_n^{\#}(\omega, \theta_0, \varepsilon_n) \subseteq b(\theta_0)$, and $\limsup_{n\to\infty} \beta_n^{\#}(\omega, \theta_0, \varepsilon_n) \neq \varnothing$, due to compactness.2 P **a.s.** ■

Remark R.8 *In case that $b(\theta_0)$ is a singleton from lemma 2.4 we have that $\limsup_{n\to\infty} \beta_n(\omega, \theta_0, \varepsilon_n) = \{b(\theta_0)\}$ P* ***a.s.*** *due to compactness of B..*

Definition, Existence and Consistency of the Indirect Estimator

We are now ready to define the indirect estimator (IE) and explore the issues of its existence and consistency. Lemma 2.4 allows us concentrate on properties of the real function on $\Omega \times \Theta$, $d_B(\beta_n(\varepsilon_n)(\omega, \theta_0), b(\theta))$ which enables the following definition.

Definition D.7 *Let ε_n^* be a non negative random variable, the indirect estimator $\theta_n(\varepsilon_n^*)(\omega)$ is defined by*

$$d_B\left(\beta_n(\varepsilon_n)(\omega, \theta_0), b\left(\theta_n(\varepsilon_n^*)(\omega)\right)\right) \leq \inf_{\Theta} d_B\left(\beta_n(\varepsilon_n)(\omega, \theta_0), b(\theta)\right) + \varepsilon_n^*$$

We denote a generic auxiliary almost surely convergent subsequence at θ_0 as $\left\{\beta_{n_j}(\varepsilon_{n_j})(\omega, \theta_0), b(\theta_0)\right\}$ and its almost sure limit by $b_j(\theta_0)$. It is obvious that $d_B(b_j(\theta_0), b(\theta_0)) = 0$, and by strengthening this property we will be provided with an asymptotic identification condition for θ_0. We also denote the set of almost sure cluster points of the sequence of auxiliary correspondences at θ_0 by $b^{\#}(\theta_0)$. We are initially concerned with the question of existence of the IE.

Lemma 2.5 *If lemma 2.4 is valid then* $\theta_n(\omega)$ *is* $\mathcal{B}(F(\Theta))/\mathcal{J}$*-measurable almost surely non empty, compact valued correspondence.*

Proof. First, notice that $\inf_{\Theta} d_B(\beta_n(\varepsilon_n)(\omega,\theta_0), b(\theta))$ is P **a.s.** bounded from below by zero, while the function is almost surely proper with domain Θ hence $\inf_{\Theta} d_B(\beta_n(\varepsilon_n)(\omega,\theta_0), b(\theta)) < \infty$ P **a.s.** Secondly, $d_B(\beta_n(\varepsilon_n)(\omega,\theta_0), b(\theta)) = \sup_{\beta\in\beta_n(\varepsilon_n)(\omega,\theta_0)} \inf_{\beta'\in b(\theta)} d_B(\beta,\beta')$, and $\sup_{\beta\in A}\inf_B d_B(\beta,\beta')$ hemi-metrizes the upper topology $U^*(B)$ and is therefore jointly semicontinuous. Hence, due to corollary 2.3.2 $d_B(\beta_n(\varepsilon_n)(\omega,\theta_0), b(\cdot))$ is P **a.s.**almost surely continuous and therefore P **a.s.** almost surely lower semi-continuous. Analogously due to lemma 2.2, $d_B(\beta_n(\varepsilon_n)(\cdot,\theta_0), b(\theta))$ is measurable for any $\theta\in\Theta$ and therefore it is a Caratheodory function. Due to the separability of Θ and Lemma 4.51 of [1], it is jointly measurable and therefore a normal integrand. Due to the compactness of Θ and the P **a.s.** lsc property it is P **a.s.** inf-compact. Hence the correspondence

$$\theta_n^{\#}(\omega,\varepsilon_n^*) \doteqdot \left\{ \begin{array}{c} \theta\in\Theta : d_B(\beta_n(\varepsilon_n)(\omega,\theta_0), b(\theta_n(\varepsilon_n^*)(\omega))) \\ \leq \inf_{\Theta} d_B(\beta_n(\varepsilon_n)(\omega,\theta_0), b(\theta)) + \varepsilon_n^* \end{array} \right\}$$

is $\mathcal{B}(F(\Theta))/\mathcal{J}$-measurable, and almost surely non empty-compact valued $\forall\theta\in\Theta$ due to Proposition 3.10.(i) of [9]. The result follows from the fundamental selection theorem (Theorem 2.13 of [9]). ■

Having established the existence of the IE, we turn to the issue of consistency. We need the following assumption that facilitates the investigation of the issue of the strong consistency of the particular estimator sequence.

Assumption A.4 *If* $\theta\neq\theta_0 \Rightarrow b^{\#}(\theta_0)\cap b(\theta)=\varnothing$.

Remark R.9 *This assertion follows if* $\theta\neq\theta_0\Rightarrow b(\theta)\cap b(\theta_0)=\varnothing$ *due to the fact that* $b^{\#}(\theta_0)\subseteq b(\theta_0)$. *In the case that the binding correspondence is single valued, this reduces to* $\theta\neq\theta_0\Rightarrow b(\theta)\neq b(\theta_0)$.

The main result of the current section follows after the introduction of the following assumption which is directly analogous to A.3.2.

Assumption A.5 $\varepsilon_n^*(\omega)$ *converges to zero* P ***a.s.***

Lemma 2.6 *If lemmas 2.4, 2.5, and assumptions A.4- A.5 are valid, then* $\theta_n(\varepsilon_n^*)(\omega)$ *converges* θ_0 P ***a.s.***

Proof. Notice first that $d_B\left(\beta_{n_j}(\varepsilon_{n_j})(\omega,\theta_0), b(\theta)\right)$, and $d_B(b_j(\theta_0), b(\theta))$ are P **a.s.** well defined continuous functions of Θ due to the compactness of

Θ, the continuity of the hemimetric and the upper continuity of the binding correspondence by corollary 2.3. Then we have that

$$\sup_{\theta\in\Theta}\left|d_B\left(\beta_{n_j}\left(\varepsilon_{n_j}\right)(\omega,\theta_0),b(\theta)\right)-d_B\left(b_j(\theta_0),b(\theta)\right)\right|$$
$$\leq\ d_B\left(\beta_{n_j}\left(\varepsilon_{n_j}\right)(\omega,\theta_0),b_j(\theta_0)\right)+d_B\left(b(\theta),b(\theta)\right)$$

due to [13], exercise 4.7.3 and the fact that in the notation of the particular reference $\delta_u(A,B)=\delta_l(B,A)$, where now

$$d_B\left(b(\theta^*),b(\theta)\right)=\delta_u\left(b(\theta^*),b(\theta)\right)\doteqdot\sup_{\beta\in b(\theta^*)}\inf_{\beta'\in b(\theta)}d(\beta,\beta')$$

Since $d_B\left(\beta_{n_j}\left(\varepsilon_{n_j}\right)(\omega,\theta_0),b_j(\theta_0)\right)$ converges by construction P **a.s.** to zero, and $d_B\left(b(\theta),b(\theta)\right)=0$ we have that $d_B\left(\beta_{n_j}\left(\varepsilon_{n_j}\right)(\omega,\theta_0),b\left(\theta^*_{n_j}\right)\right)$ converges uniformly over Θ, P **a.s.** to $d_B\left(b_j(\theta_0),b(\theta)\right)$. From assumption A.4 the latter is uniquely minimized in θ_0, since $d_B\left(b_j(\theta_0),b(\theta)\right)=0$ iff $b_j(\theta_0)\in b(\theta)$ from to the fact that $b(\theta)$ is closed for any θ due to corollary 2.3 and this is true iff $\theta=\theta_0$ due to identification. The result follows. ■

3 Examples

In this section we consider a set of examples that represent the previous results in specific statistical models. We begin with a semi-parametric linear model which contains a set of weak instruments, and continue with a pair of examples that involve conditionally heteroskedastic processes.

Example Semi - Parametric Linear Model with Weak Instruments. Consider the $n\times p$ and $n\times q$ dimensional random matrices $X(\omega)$ and $Z(\omega)$ respectively, where $n\geq q\geq p$. Let $\Theta\in K(\mathbb{R}^p)$ and suppose that $\frac{X'X}{n}\to M_{X'X}$, $\frac{Z'Z}{n}\to M_{Z'Z}$, $\frac{Z'X}{n}\to M_{Z'X}$ almost surely, where $\mathbf{rank}(M_{X'X})=\mathbf{rank}(M_{Z'X})=p$ and $p\leq l\doteqdot\mathbf{rank}(M_{Z'Z})\leq q$. Consider that cosets of $\mathbb{R}^q$, defined for each $\theta\in\Theta$, by the linear systems $M_{Z'Z}\beta=M_{Z'X}\theta$ and denote them as $\theta-Coset(H_{q-l})\doteqdot K\theta+H_{q-l}$, where K represents an injective linear map from $\mathbb{R}^p$ to $\mathbb{R}^q$ due to the rank condition of $M_{Z'X}$ and H_{q-l} is a $q-l$-dimensional subspace of $\mathbb{R}^q$, which is trivial if and only if $l=q$ whereas $K=M^{-1}_{Z'Z}M_{Z'X}$, and maximal in the case that $l=p$. Let $B\in K(\mathbb{R}^q)$ be such that $B\cap\theta-Coset(H_{q-l})\neq\varnothing$, $\forall\theta\in\Theta$, which exists due to the Axiom of Choice. Let $Y(\omega,\theta_0)=X(\omega)\theta_0+\varepsilon(\omega)$, where $u(\omega)$ is a $n\times 1$ random vector. Let the underlying statistical model be the set of "regressions" $\{Y(\omega,\theta) = X(\omega)\theta + \varepsilon(\omega), \theta \in \Theta\}$ which is obviously well specified. Consider for any $\beta \in B$, $c_n(\omega,\theta,\beta) = \frac{1}{n}(Y - Z\beta)'(Y - Z\beta)$, which clearly satisfies assumption A.1 due to continuity with respect to β and the compactness of B. The statistical problem consists of the consistent estimation of θ_0 and c_n can be perceived to emerge from an auxiliary set of regression functions on $Z\beta$, $\beta \in B$. Lemma 2.2 assures the **existence** of an auxiliary selector $\beta_n(\omega,\theta_0)(\varepsilon_n)$ given an appropriate ε_n. Let also $M_{X'\varepsilon}, M_{\varepsilon'\varepsilon} \in \mathbb{R}$ and assume that $\frac{X'\varepsilon}{n} \to M_{X'\varepsilon}$, $\frac{\varepsilon'\varepsilon}{n} \to M_{\varepsilon'\varepsilon}$ and $\frac{Z'\varepsilon}{n} \to 0$ almost surely. In this sense $Z(\omega)$ is interpreted as a matrix of weak instruments. It can be easily seen that $\sup_{\Theta\times B}|c_n(\omega,\theta,\beta) - c(\theta,\beta)|$ almost surely where $c(\theta,\beta) = \beta' M_{X'X}\beta - 2\theta' M_{X'Z}\beta + \beta' M_{Z'Z}\beta + 2\theta' M_{X'\varepsilon} + M_{\varepsilon'\varepsilon}$ and that due to the joint continuity of c on $\Theta \times B$ implied by the uniform convergence $\theta \to c(\theta,\cdot)$, is also epicontinuous on Θ as implied by remark R.6. Also due to the compactness of B it is inf-compact for any θ, hence assumption A.2 applies and therefore corollary 2.3 is verified. Notice that $b(\theta) = B \cap \theta - Coset(H_{p-l})$ due to the fact that $\beta \in \arg\min_B c(\theta,\beta)$ if and only if it satisfies $M_{Z'Z}\beta = M_{Z'X}\theta$, due to the fact that c can be extended to an open set that contains B. Assumption A.3 follows from the compactness of B and the appropriate definition of $\{\varepsilon_n\}$. Hence lemma 2.4 is valid. For an appropriate ε_n^*, the IE is also defined and its existence is assured by lemma **??** due to compactness of Θ. Finally, assumption A.4 is implied by an assumption of the form $K(\theta - \theta_0) \notin H_{q-l}$, if $\theta \neq \theta_0$, which guarantees that $b(\theta) \cap b(\theta_0) = \varnothing$ if $\theta \neq \theta_0$ due to remark R.9, and assumption A.5 follows from an appropriate definition of $\{\varepsilon_n^*\}$. Hence, lemma 2.6 follows.□

The following examples involve conditionally heteroskedastic processes, some characteristics of which are reviewed, before any example description. We again consider $\Theta \in K(\mathbb{R}^p)$ and $B \in K(\mathbb{R}^q)$. Let also $z : \Omega \to \mathbb{R}^{\mathbb{Z}}$ be an i.i.d. sequence of random variables, with $Ez_0 = 0$, and $Ez_0^2 = 1$. Consider a random element $\sigma^2 : \Theta \times \Omega \to (\mathbb{R}^+)^{\mathbb{Z}}$, with the product space $\Theta \times \Omega$ equipped with $\mathcal{B}(\Theta) \otimes \mathcal{F}$ with $\sigma_t^2(\theta)$ independent of $(z_i)_{i \geq t}$, $\forall t \in \mathbb{Z}$, $\forall \theta \in \Theta$. For arbitrary $\theta \in \Theta$, $\sigma^2(\theta)$ is a $\mathcal{B}^{\mathbb{N}}(\mathbb{R}^+)/\mathcal{J}$-measurable function. Analogously, define the random element $y : \Theta \times \Omega \to (\mathbb{R})^{\mathbb{Z}}$ as

$$(y_t(\omega)(\theta))_{t\in\mathbb{Z},\theta\in\Theta} = \left(z_t(\omega)\sqrt{\sigma_t^2(\omega)(\theta)}\right)_{t\in\mathbb{Z},\theta\in\Theta}$$

Then $\forall \theta \in \Theta$, $(y_t(\theta))_{t\in\mathbb{Z}}$ is called a conditionally heteroskedastic process, while the random element $(y_t(\omega)(\theta))_{t\in\mathbb{Z},\theta\in\Theta}$ a conditionally heteroskedastic model. Notice that $y(\theta) : \Omega \to \mathbb{R}^{\mathbb{Z}}$ is $\mathcal{B}^{\mathbb{Z}}(\mathbb{R}^+)/\mathcal{J}$-measurable $\forall \theta \in \Theta$. We consider $(y_t(\omega)(\theta_0))_{t\in\{1,\ldots,n\}}$ for some $\theta_0 \in \Theta$, and define the statistical problem in question to be the consistent estimation of θ_0.

Assumption A.6 *Theorem 2.6.1. of [15] holds, hence* $(\sigma_t^2(\theta))_{t\in\mathbb{Z}}$ *is stationary ergodic* $\forall\theta\in\Theta$*.*

Remark R.10 *Conditions that ensure assumption A.6 are described and employed in a variety of heteroskedastic models in chapter 4 of [15].*

Corollary 1 $(y_t(\theta))_{t\in\mathbb{Z}}$ *and* $(y_t^2(\theta))_{t\in\mathbb{Z}}$ *are stationary ergodic* $\forall\theta\in\Theta$*.*

Proof. It follows from the definition of z, y, the previous assumption and Proposition 2.2.1 of [15]. ■

Example Regressions on Squared Heteroskedastic Processes. Consider the random vector $Y(\theta) = (y_t^2(\omega)(\theta))_{t\in\{1,\ldots,n\}}$ for any $\theta\in\Theta$ and set $Y = Y(\theta_0)$, and the $n\times q$ dimensional random matrix $Z(\omega,\theta)$, jointly measurable with respect to $\mathcal{B}(\Theta)\otimes\mathcal{F}$, where $n\geq q\geq p$ and ergodic for any $\theta\in\Theta$. Its columns could partly emerge from time shifts (lags) of Y. Similarly set $Z(\theta)=Z(\omega,\theta)$ and $Z=Z(\theta_0)$. Consider for any $\beta\in B\in K(\mathbb{R}^q)$, $c_n(\omega,\theta,\beta)=\frac{1}{n}(Y(\theta)-Z(\theta)\beta)'(Y(\theta)-Z(\theta)\beta)$, where B is to be further specified below, which clearly satisfies assumption A.1 due to continuity with respect to β the compactness of B and the joint measurability of Z. This consideration is motivated from the ARMA(1,1) representations of the GARCH(1,1) model with respect to martingale difference "errors" (see, for example, [2]) and c_n can be perceived to emerge from an auxiliary model that is consisted of the set of "auxiliary" regression functions of Y on $Z\beta$, $\beta\in B$. As in the previous example, Lemma 2.2 assures the **existence** of an auxiliary selector $\beta_n(\omega,\theta_0)(\varepsilon_n)$ given an appropriate ε_n. Let, also $E(h_0(\theta))<\infty$ for any $\theta\in\Theta$, which along corollary 1 and Birkhoff's ergodic LLN implies that $\frac{Y'(\theta)Y(\theta)}{n}\to E(h_0(\theta))$ almost surely. In the same fashion assume that $E(\|Z'(\theta)Z(\theta)\|)<\infty$, for any $\theta\in\Theta$, which along another application of Proposition 2.2.1 of [15] and Birkhoff's ergodic LLN implies that $\frac{Z'(\theta)Z(\theta)}{n}\to M_{Z'Z}(\theta)$, for any $\theta\in\Theta$. Assume that $l(\theta)\doteqdot \mathbf{rank}(M_{Z'Z}(\theta))\leq p$ for any $\theta\in\Theta$, and that $E(h_0(\theta))$, and $M_{Z'Z}(\theta)$ are continuous functions on Θ. The previous moment existence assumptions along with another application of Proposition 2.2.1 of [15] and Birkhoff's ergodic LLN imply that $\frac{Z'(\theta)Y(\theta)}{n}\to M_{Z'Y}(\theta)\in\mathbb{R}^q$ for any $\theta\in\Theta$. Assume that $M_{Z'Y}(\theta)$ is an injective continuous function on Θ. Due to Proposition 2.2.1 of [15], we have that $c_n(\omega,\theta,\beta)$ is also stationary ergodic for any $\theta\in\Theta$ and any $\beta\in B$, the fact that $Ec_n(\omega,\theta,\beta)=E(h_0(\theta))-2E\left(\frac{Y'(\theta)Z(\theta)}{n}\right)\beta+E\left(\beta'\frac{Z'(\theta)Z(\theta)}{n}\beta\right)<\infty$ for any $\theta\in\Theta$ and any $\beta\in B$ and Birkhoff's ergodic LLN $c_n(\omega,\theta,\beta)\to c(\theta,\beta)$ pointwise on $\Theta\times B$, where $c(\theta,\beta)=E(h_0(\theta))-2M_{Z'Y}(\theta)\beta+\beta'M_{Z'Z}(\theta)\beta$. Also, consider, $E\inf_B c_n(\omega,\theta,\beta)=$

$E\inf_B\left|\left(\frac{Y'(\theta)Y(\theta)}{n}-2\frac{Y'(\theta)Z(\theta)}{n}\beta+\beta'\frac{Z'(\theta)Z(\theta)}{n}\beta\right)\right|\leq E\left(h_0\left(\theta\right)\right)+c_1E\left(\left\|\frac{Y'(\theta)Z(\theta)}{n}\right\|\right)+$ $c_2E\left(\frac{Z'(\theta)Z(\theta)}{n}\right)<\infty$, for any $\theta\in\Theta$, due to the previous, for some $c_1,c_2>0$ which exist due to the compactness of B. Hence, by remark R.5 $c_n\overset{e}{\to}c$ for any $\theta\in\Theta$. Due to the joint continuity of c on $\Theta\times B$ implied by the previous continuity assumptions c is also epicontinuous on Θ as implied by remark R.6. Also, due to the compactness of B it is inf-compact for any θ, hence, assumption A.2 applies, and therefore corollary 2.3 is verified. Notice that $b\left(\theta\right)=B\cap Coset\left(H\left(\theta\right)\right)$ due to the fact that $\beta\in\arg\min_B c\left(\theta,\beta\right)$ if and only if it satisfies $M_{Z'Z}\left(\theta\right)\beta=M_{Z'Y}\left(\theta\right)$, since c can be extended to an open set that contains B. Similarly we can assume that $B\cap Coset\left(H\left(\theta\right)\right)\neq\varnothing$, $\forall\theta\in\Theta$, which is possible due to the Axiom of Choice, where again $Coset\left(H\left(\theta\right)\right)\doteqdot K\left(\theta\right)+H\left(\theta\right)$, where $K\left(\theta\right)$ represents an injective linear map from $\mathbb{R}^p$ to $\mathbb{R}^q$ due to the injectivity of $M_{Z'Y}\left(\theta\right)$ and that $l\left(\theta\right)\geq p$ for any $\theta\in\Theta$, and $H\left(\theta\right)$ is a $q-l\left(\theta\right)$-dimensional subspace of $\mathbb{R}^q$, which is trivial if and only of $l\left(\theta\right)=q$ whereas $K\left(\theta\right)=M^{-1}_{Z'Z}\left(\theta\right)M_{Z'Y}\left(\theta\right)$, and maximal in the case that $l\left(\theta\right)=p$. Assumption A.3 follows from the compactness of B and the appropriate definition of $\{\varepsilon_n\}$. Hence lemma 2.4 is valid. For an appropriate ε^*_n, the IE is also defined and its existence is assured by lemma **??** due to compactness of Θ. Finally assumption A.4 is implied by an assumption of the form $K\left(\theta\right)-K\left(\theta_0\right)\notin span\left(H\left(\theta_0\right),H\left(\theta\right)\right)$, if $\theta\neq\theta_0$, which guarantees that $b\left(\theta\right)\cap b\left(\theta_0\right)=\varnothing$ if $\theta\neq\theta_0$ due to remark R.9, and assumption A.5 follows from an appropriate definition of $\{\varepsilon^*_n\}$. Hence, lemma 2.6 follows.□

In the previous examples we have encountered cases in which the image of the limit objective function c is in $\mathbb{R}$. In the final example, we consider a case in which c attains values in $\overline{\mathbb{R}}$ outside of $\mathbb{R}$. We consider the case of the second order non stationary GARCH(1,1) model described as follows. Let $\Theta=B\in K\left(\mathbb{R}^3\right)$ and $\sigma^2_t\left(\theta\right)$ be the stationary solution of the stochastic difference equation $\sigma^2_t\left(\theta\right)=\omega+\left(az^2_{t-1}+b\right)\sigma^2_{t-1}\left(\theta\right)$, where $\theta=\begin{pmatrix}\omega\\ a\\ b\end{pmatrix}$ such that $\omega>0,a,b\geq0$ and $\alpha+b\in[1,1+c]$, for $c>0$ ensuring A.6. This also means that $E\left(\sigma^2_0\left(\theta\right)\right)=\infty$ on Θ, justifying covariance non-stationarity (see [15], section 5.4.2). We also consider the random element $h:\Omega\to\Theta\times B\times\left(\mathbb{R}^+\right)^{\mathbb{Z}}$, with the product space $B\times\left(\mathbb{R}^+\right)^{\mathbb{Z}}$ equipped with $\mathcal{B}\left(\Theta\right)\otimes\mathcal{B}\left(B\right)\otimes\mathcal{B}^{\mathbb{Z}}\left(\mathbb{R}^+\right)$ with $h_t\left(\theta,\beta\right)$ independent of $\left(y_i\left(\theta\right)\right)_{i\geq t}$, $\forall t\in\mathbb{Z}$, $\forall\theta\in\Theta$, $\forall\beta\in B$, defined by the the stochastic difference equation $h_t\left(\theta,\beta\right)=\omega_*+a_*y^2_{t-1}\left(\theta\right)+b_*h_{t-1}\left(\theta,\beta\right)$, where $\beta = \begin{pmatrix} \omega_* \\ a_* \\ b_* \end{pmatrix}$. Due to corollary 1 and Theorem 2.6.1 of [15], $h_t(\theta, \beta)$ is stationary ergodic if $b_* < 1$ and we therefore finalize the description of $\Theta = B$, to be such that $b_* < 1$. Notice that due to compactness of Θ, $\exists k > 0 : \inf_{(\theta,\beta)\in\Theta^2} h_t(\theta, \beta) > k$, $\forall \omega$ where k is independent of ω, and therefore $\inf_{\beta\in\Theta} \inf_{\omega\in\Omega} h_t(\theta, \beta) > k$. k is actually the lower bound of the compact interval that contains the possible values of h. Since $\sigma_t^2(\theta) = h_t(\theta, \theta)$, the same is true for $\inf_{\theta\in\Theta} \sigma_t^2(\theta)$. Then for $n \in \mathbb{N}$, given the aforementioned volatility model we consider the following sequence of real random functions:

$$\begin{aligned} c_n(\omega, \theta, \beta) &\doteqdot \frac{1}{n}\sum_{i=1}^{n} l_i(\omega, \theta, \beta) \\ l_i(\omega, \theta, \beta) &\doteqdot \ln h_i(\theta, \beta) + \frac{y_i^2(\theta)}{h_i(\theta, \beta)} \end{aligned}$$

Remark R.11 *In practice $c_n(\omega, \theta_0, \beta)$ is unknown but approximated by an analogous $\widehat{c}_n(\omega, \theta_0, \beta)$ dependent on non ergodic solutions of the stochastic difference equation that defines h based on arbitrary initial conditions. In this case, due to assumption A.6, Proposition 5.2.12 of [15] can be employed in order to ensure that $\sup_B |c_n(\omega, \theta, \beta) - \widehat{c}_n(\omega, \theta, \beta)|$ converges almost surely to zero for any $\theta \in \Theta$ (see the first part of the proof of Theorem 5.3.1 of [15]), thereby facilitating the asymptotic analysis of minimizers of $\widehat{c}_n(\omega, \theta, \beta)$ by the analogous analysis of minimizers of $c_n(\omega, \theta, \beta)$.*

Example $-c_n$ is the Qausi-Likelihood Function of the Heteroskedastic Model.
c_n satisfies assumption A.1, from the continuity with respect to (θ, β) which follows from the continuity of the parameterization and the existence of k, due to the compactness of Θ and the evident joint measurability. Lemma 2.2 assures the **existence** of an auxiliary selector $\beta_n(\omega, \theta_0)(\varepsilon_n)$ given an appropriate ε_n. Due to corollary 1, the ergodicity of h, and Theorem 2.6.1 of [15] $l_i(\omega, \theta, \beta)$ is stationary ergodic $\forall \theta, \beta \in \Theta$, and therefore $c_n(\omega, \theta, \beta)$ epiconverges to a proper lower semicontinuous c **jointly** on $\Theta \times B$, hence for any θ in Θ, due to the reasoning in Part 2 of the proof of Theorem 5.3.1 of [15] which is based on Lemma 3.11 of [11] and the fact that $E \inf_{(\theta,\beta)\in\Theta^2} \left(\ln h_0(\theta, \beta) + \frac{z_0^2 \sigma_0^2(\theta)}{h_0(\theta,\beta)} \right) \leq \inf_{(\theta,\beta)\in\Theta^2} E \left(\ln h_0(\theta, \beta) + \frac{z_0^2 \sigma_0^2(\theta)}{h_0(\theta,\beta)} \right) \leq \inf_{\beta\in\Theta} \left(E \ln h_0(\theta, \beta) + E \frac{\sigma_0^2(\theta)}{h_0(\theta,\beta)} \right) = E \ln h_0(\theta, \theta) + 1$ where the equality follows from Part 1.(iii) of the proof of Theorem 5.3.1 of [15] and $E \ln h_0(\theta, \theta)$

exists due to that $h_0(\theta,\theta) = \sigma_0^2(\theta)$, $E\log^+\sigma_0^2(\theta) < \infty$ on Θ, the applicability of Jensen's inequality on $E\ln\sigma_0^2(\theta)\,1_{\sigma_0^2(\theta)\leq 1}$ and the fact that $\ln h_0(\theta,\beta) + \frac{z_0^2\sigma_0^2(\theta)}{h_0(\theta,\beta)} \geq \ln k$ hence $E\inf_{(\theta,\beta)\in\Theta^2}\left(\ln h_0(\theta,\beta) + \frac{z_0^2\sigma_0^2(\theta)}{h_0(\theta,\beta)}\right) > -\infty$. This result allows the application of remark R.5. Properness follows from the existence of k and the fact that the monotonic transformation of the limit $E\left(-\ln\frac{\sigma_i^2(\theta)}{h_i(\theta,\beta)} + \frac{\sigma_i^2(\theta)}{h_i(\theta,\beta)}\right)$ equals 1 when $\beta = \theta$, which is the infimum of the latter. Since $c(\theta,\beta)$ is jointly lower semicontinuous A.2.3.a follows readily while A.2.3.b follows for $\beta_n = \theta_n$, since $E\ln\sigma_i^2(\theta) < \infty$ and it is continuous $\forall\theta\in\Theta$, hence $\limsup_{n\to\infty} c(\theta_n,\theta_n) = \limsup_{n\to\infty}(E\ln\sigma_i^2(\theta_n)+1) = \lim_{n\to\infty}(E\ln\sigma_i^2(\theta_n)+1)$
$= E\ln\sigma_i^2(\theta) + 1 = c(\theta,\theta)$. Also, due to the compactness of Θ it is inf-compact for any θ, hence, assumption A.2 applies, and therefore corollary 2.3 is verified. Assumption A.3 follows from the compactness of Θ and the appropriate definition of $\{\varepsilon_n\}$. Therefore, lemma 2.4 is valid. For an appropriate ε_n^*, the IE is also defined and its existence is assured by lemma 2.5 due to compactness of Θ. Finally assumption A.4 is implied by the previous remark on the behavior of $E\left(-\ln\frac{\sigma_i^2(\theta)}{h_i(\theta,\beta)} + \frac{\sigma_i^2(\theta)}{h_i(\theta,\beta)}\right)$ and assumption C.4 of [15] (page 100), which is verified when the support of the distribution of z_0 has more than two elements as lemmas 5.4.4-5 of [15] imply. In this case $b(\theta_0)$ is single valued and equals θ_0. Assumption A.5 follows from an appropriate definition of $\{\varepsilon_n^*\}$. Hence lemma 2.6 follows.□

4 Conclusions

In this paper we have established the definition of Indirect Estimators and their strong consistency, when the binding function is a compact valued correspondence under mild conditions. These concern the asymptotic behavior of the epigraphs of the criterion functions involved in the relevant procedures, as well as asymptotic indirect identification that restricts the behavior of the aforementioned correspondence. These results are wide generalizations of the analogous results in the relevant literature, hence permit a broader scope of statistical models.

We leave for future research the issue of further generalization of these results on IE that are defined by possibly random approximations of the binding correspondence, as well as the issues of considering the other stages of first step asymptotic theory, namely the one concerning the establishment of rates of convergence, as well as of the one concerning the establishment of the asymptotic distributions of IE in our general set up.

References

[1] Aliprantis Charalambos D. and Kim C. Border,"Infinite Dimensional Analysis", Springer, 2006.

[2] Bollerslev Tim, "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, Vol. 31, Issue 3, pages 307-327, 1986.

[3] Davidson J., "Stochastic Limit Theory, Advanced Texts In Econometrics", Oxford University Press, 1995.

[4] Calzolari, C., Fiorentini G. and E. Sentana, Constrained Indirect Inference Estimation, 2004, *Review of Economic Studies* **71**, *pp.* 945-973.

[5] Gourieroux C., A. Monfort, and E. Renault , "Indirect Inference", *Journal of Applied Econometrics*, 1993, Volume 8 Issue 1, pp. 85-118.

[6] Gallant A. Ronald, and Tauchen G.E., "Which Moments to Match," Working Papers 95-20, 1995, Duke University, Department of Economics.

[7] Rockafellar T.R. and Wetts J-B, "Variational Analysis", Springer-Verlag, 1997.

[8] Korf, A.L., and Wets J-B, "An Ergodic Theorem for Stochastic Programming Problems", unpublished document.

[9] Molchanov, Ilya, "Theory of Random Sets" ,Probability and Its Applications, Springer, 2004.

[10] James, R. Munkres,"Topology", 2nd Edition, Prentice Hall, 2000.

[11] J. Pfanzagl, "On the Measurability and Consistency of Minimum Contrast Estimates", *Metrika*, 14:249-272, 1969.

[12] Rockafellar T.R. and Wets J-B, "Variational Systems, and Introduction" in "Multifunctions and Intergrands", edited by G. Salinetti, Springer-Verlag, 1984.

[13] Klein Er., and Thompson A.C., "Theory of Correspondences", CMS Series of Monographs, Wiley Interscience, 1984.

[14] Smith A.A., "Estimating Nonlinear Time-series Models using Simulated Vector Autoregressions", *Journal of Applied Econometrics,* Volume 8 Issue 1, pp. 63-84.

[15] Straumann, D., "Estimation in Conditionally Heteroscedastic Time Series Models", Springer, 2004.