ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF ECONOMICS

WORKING PAPER SERIES **02-2012**

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

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 Compact Valued

Stelios Arvanitis
Athens University of Economics and Business

March 16, 2012

Abstract

We provide sufficient conditions for the definition and the existence of strongly consistent indirect estimators when the binding function is a compact valued correspondence. These are generalizations of the analogous results in the relevant literature, hence permit a broader scope of statistical models. We provide simple examples involving Levy or ergodic conditionally heteroskedastic processes.

KEYWORDS: Indirect estimator, lower semicontinuous function, random set, normal integrand, upper topology, Fell topology, epi convergence, binding correspondence, cluster points, indirect identification, linear model, Levy processes, ergodicity, conditional heteroskedasticity, ARCH model, QARCH model.

1 Introduction

Indirect estimators (henceforth IE) are optimization based 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]

Indirect inference algorithms were initially employed in [28], formally introduced by [18], complemented by [14] and extended by [6]. Furthermore, applications of these estimators have become increasingly popular. They have been applied to stochastic volatility and equity return models (e.g. [12], [15], [1]), exchange rate models (e.g. [4], and [9]), commodity price and storage models (e.g. [23]), dynamic panel data (e.g. [19]), stochastic differential equation models (e.g. [13] and [17]), and in ARMA models (e.g. [8], [16], [10], and [26]).

In the present paper, we are concerned with the issue of the *existence* of strongly consistent IE allowing for cases where the aforementioned function is compact valued (hence possibly multivalued). Therefore we perform our study in a more general framework than the ones employed in the relevant literature.

Our generalization is threefold. *First*, using mild assumptions on the structure of the aforementioned criterion functions, we are occupied with a weaker notion of convergence (in comparison to the uniform) of the relevant sequence of criterion functions, that essentially concerns the almost sure asymptotic behavior of their epigraphs and is 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.

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. Next, we define and study the asymptotic behavior of the

[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.

auxiliary estimator, the binding correspondence and finally of the IE. We then exhibit some of our results by a set of simple examples. We conclude posing some questions for future research.

2 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 $F_0(E)$ the set of closed non empty subsets of E. We next describe two topologies on $F_0(E)$ using τ_E and the inclusion partial order on 2^E.

Definition D.1 *The upper topology $\mathcal{T}_U$ on $F_0(E)$ is generated by the subbase consisting of*

$$[\cdot, G] = \{F \text{ closed} : F \subset G\}, \forall G \in \tau_E, \text{ non empty}$$

The upper topology is extremely useful for the analysis of the asymptotic behavior of sequences of sets of minimizers. If τ_E is generated by a metric (say d) w.r.t. which E is compact then $\mathcal{T}_U$ is hemimetrizable (see Proposition 4.2.2 of [21]) by $\delta_u : F_0(E) \times F_0(E) \to \mathbb{R} \cup \{+\infty\}$, defined by $\delta_u(A, B) = \inf\{\varepsilon > 0 : B \subset N_\varepsilon(A)\}$ where $N_\varepsilon(A) = \{x \in E : d(x, A) < \varepsilon\}$.[2] Obviously, when $B \subseteq A$ then $\delta_u(A, B) = 0$.

Lemma 2.1 $\delta_u(A, B) = 0$ iff $B \subseteq A$.

Proof. Since A is closed if $x \in B$ and $x \notin A$, then $d(x, A) = \delta > 0$. But then $B \nsubseteq N_{\delta/2}(A)$ and therefore $\delta_u(A, B) > \frac{\delta}{2}$. ■

Lemma 2.2 δ_u *is a lower semicontinuous (*lsc*) real function w.r.t. the first argument.*

Proof. If $A_n \to A$ with respect to the upper topology on $F_0(E)$, then $\delta_u(A, B) \leq \delta_u(A, A_n) + \delta_u(A_n, B)$, hence $\liminf_n \delta_u(A_n, B) \geq \delta_u(A, B)$. ■

Lemma 2.3 δ_u *is an upper semicontinuous (*usc*) real function w.r.t. the second argument.*

[2] where $d(x, A) \doteqdot \inf_{y \in A} d(x, y) < \varepsilon$.

Proof. If $B_n \to B$ with respect to the upper topology on $F_0(E)$, then $\delta_u(A, B_n) \leq \delta_u(A, B) + \delta_u(B, B_n)$ establishing that $\limsup_n \delta_u(A, B_n) \leq \delta_u(A, B)$. ■

The second topology on $F_0(E)$, known as the Fell topology, is defined by the use of the following subbase (see [24], paragraph 1.1, and [21], Definition 4.5.1).

Definition D.2 *The Fell topology, say $\mathcal{T}_F$, is the smallest topology on $F_0(E)$ consisting of both*

1. $F_G = \{F \text{ closed} : F \cap G \neq \varnothing\}$, $\forall G \in \tau_E$, *non empty and*

2. $F^K = \{F \text{ closed} : F \cap K = \varnothing\}$, $\forall K \subset E$ *non empty and compact.*

From Theorems 4.5.3-5 of [21] we have that when E is locally compact and Hausdorff then $(F_0(E), \mathcal{T}_F)$ is locally compact and, $F_n \to F$ with respect to the Fell topology *iff* $F = \liminf_n F_n = \limsup_n F_n$.[3] Hence, in this case this type of convergence coincides with the Painleve-Kuratowski convergence (see among others, Appendix B of [24], or Definition 3.1.4. of [21]). If E is also separable then the Fell topology is *metrizable*. If furthermore E is compact and metrized by d, then the Fell topology is actually metrized by $\delta(A, B) = \max\{\delta_u(A, B), \delta_l(A, B)\}$ where $\delta_l(A, B) \doteqdot \delta_u(B, A)$. In this case we can prove the following lemma.

Lemma 2.4 *If E is compact and metrized by d, then δ_u is a lower semi-continuous (*lsc*) real function w.r.t. the product topology on $F_0(E) \times F_0(E)$, when the first factor is endowed with $\mathcal{T}_U$ and the second with $\mathcal{T}_F$.*

Proof. If $(A_n, B_n) \to (A, B)$ with respect to the aforementioned product topology on $F_0(E) \times F_0(E)$, then $\delta_u(A, B) \leq \delta_u(A, A_n) + \delta_u(A_n, B_n) + \delta_l(B, B_n)$ establishing that $\liminf_n \delta_u(A_n, B_n) \geq \delta_u(A, B)$. ■

Epigraphs of Semicontinuous Functions and Epiconvergence

Consider now the case where E is locally compact and Hausdorff, 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 $c : E \to \overline{\mathbb{R}}$. Call c proper, if it does not assume the value $-\infty$ and its image contains at least a real number, and inf-compact, if its level sets $(\text{Level}_{\leq a}(c) \doteqdot \{x \in E : c(x) \leq a\}$ for $a \in \mathbb{R})$ are compact. Inf-compactness follows trivially when c is lsc and E is itself compact.

[3] $\liminf_n F_n$ is the set comprised of the limit points of any possible sequence (x_n) such that $x_n \in F_n$, and $\limsup_n F_n$ is the one comprised of the analogous cluster points.

Definition D.3 *The epigraph of c is*

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

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 [24], $\mathbf{epi}\,(c) \in F\,(E \times \mathbb{R})$ 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) 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.*

It is easy to see that uniform convergence implies epiconvergence. This notion is particularly suitable for the description of the asymptotic behavior of the set of minimizers of sequences of lsc functions (see Theorem 3.4 of [24] along with Theorem 7.1.4 of [21], Definition D.1 and Proposition D.2 of [24]).

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 X to $F_0\,(E)$ (i.e. a closed valued multifunction with domain the set X), when this is considered as a relation in $X \times E$. A correspondence is usually abbreviated as $\mathbf{cor} : X \rightrightarrows E$, while the benefit of not directly considering the underlying function, is the fact that we can consider the graph of $\mathbf{cor}$ as the set $\{(x,y) : y \in c\,(x)\}$ which resides in $X \times E$ instead of the set $\{(x,F) : F = c\,(x)\}$ inside $X \times F_0\,(E)$. When $c\,(x)$ is compact for any x, 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.

The Borel algebra on $F_0\,(E)$ generated by $\mathcal{T}_F$ will be abbreviated by $\mathcal{B}\,(\mathcal{T}_F)$ and is usually termed *Effron algebra (see Paragraph 1.1 of [24])*. If $(\Omega, \mathcal{J})$ is a measurable space, then c is a random closed set *iff* $\{\omega \in \Omega : c\,(\omega) \in \overline{F}\} \in \mathcal{J}$ for any $\overline{F} \in \mathcal{B}\,(\mathcal{T}_F)$. Analogously we abbreviate by $\mathcal{B}\,(\mathcal{T}_U)$ the Borel algebra on $F_0\,(E)$ generated by $\mathcal{T}_U$ and by $\mathcal{B}\,(\mathcal{T}_U \times \mathcal{T}_F)$ the Borel algebra on $F_0\,(E) \times F_0\,(E)$ generated by the product topology described in lemma 2.4. Finally denote with $\mathcal{B}\,(\mathbb{R})$ the Borel algebra of the real numbers with respect to the usual topology.

Lemma 2.5 *If E is compact, separable and metrized by d, then δ_u is $\mathcal{B}(\mathbb{R})/\mathcal{B}(\mathcal{T}_U)\otimes\mathcal{B}(\mathcal{T}_F)$ measurable.*

Proof. The separability of E implies the separability of $(F_0(E), \mathcal{T}_U)$ and $(F_0(E), \mathcal{T}_F)$ for if $\{x_n, n = 0, 1, \ldots\}$ is dense in E then the countable subset of $F_0(E)$, $\{\{x_n\}, n = 0, 1, \ldots\}$ intersects any basic open set w.r.t. to either topology. This implies the separability of $F_0(E) \times F_0(E)$ when equipped with the topology discussed in lemma 2.4. This in turn implies that the Borel algebra w.r.t. to the product topology on $F_0(E) \times F_0(E)$ coincides with $\mathcal{B}(\mathcal{T}_U) \otimes \mathcal{B}(\mathcal{T}_F)$ by Lemma 1.4.1. of [30]. The rest follows by lemma 2.4 along with the fact that the subbasic sets of the upper topology on $\mathbb{R}$ generate $\mathcal{B}(\mathbb{R})$. ■

3 Assumptions and Main results

General Set Up-Existence of the Auxiliary Estimator

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 *compact separable 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 $\mathcal{B}(\overline{\mathbb{R}})$ the Borel algebra of the extended real numbers with respect to the usual topology.

Consider a function $c_n(\omega, \theta, \beta) : \Omega \times \Theta \times B \to \overline{\mathbb{R}}$. In our context c_n is usually of the form $q_n(y_n, \beta)$ with $q_n : K_n \times B \to \overline{\mathbb{R}}$, for $y_n : \Omega \times \Theta \to K_n$, for K_n some appropriate space. q_n usually reflects part of the structure of an **auxiliary model,** a statistical model defined on K_n, with B as its parameter space (e.g. it can be a likelihood function or a GMM type criterion-see section 4).[4] When $y_n(\cdot, \theta)$ is measurable for all θ, then the **underlying** statistical model is the set $\{P \circ y_n^{-1}(\cdot, \theta), \theta \in \Theta\}$. These two models need not coincide. Hence, $c_n(\omega, \theta, \cdot)$ is usually termed as the *auxiliary* criterion.

We abbreviate with P **a.s.** any statement that concerns elements of $\mathcal{J}$ of unit probability. We note that separability and sequential completeness of Θ and B and completeness of the underlying probability space enables the appropriate measurability of inf, sup, arg min etc.

Assumption A.1 *Let the following hold:*

[4] Which in general is a correspondence $B \rightrightarrows \mathcal{P}(K_n)$, with $\mathcal{P}(K_n)$ the set of probability measures on K_n.

1. *c_n is $\mathcal{B}\left(\overline{\mathbb{R}}\right)/\mathcal{J}\otimes\mathcal{B}(\Theta)\otimes\mathcal{B}(\Phi)$ 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.1 *The joint measurability and the pointwise semicontinuity imply that $c_n(\cdot,\theta,\cdot)$ is a normal integrand (see Definition 3.5 and Proposition 3.6 of [24]).* $\arg\min_B c_n(\omega,\theta,\beta)$ *is non empty and compact due to theorem 1.9 of [27], P **a.s.** $\forall\theta\in\Theta$ due to the fact that c_n is inf-compact P **a.s.** $\forall\theta\in\Theta$.*

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

Definition D.5 *The auxiliary correspondence $\beta_n^{\#}(\omega,\theta,\varepsilon_n)$ satisfies*

$$\begin{aligned}\beta_n^{\#}(\omega,\theta,\varepsilon_n) &= \varepsilon_n\text{-}\arg\min_B c_n(\omega,\theta,\beta)\\ &\doteq \left\{\beta\in B: c_n(\omega,\theta,\beta)\leq\inf_B c_n(\omega,\theta,\cdot)+\varepsilon_n\right\}\end{aligned}$$

where ε_n is a non-negative random variable defined on Ω.

Lemma 3.1 *Under assumption A.1 $\beta_n^{\#}(\omega,\theta,\varepsilon_n)$ is $\mathcal{B}(\mathcal{T}_F)/\mathcal{J}\otimes\mathcal{B}(\Theta)$-measurable, hence $\mathcal{B}(\mathcal{T}_F)/\mathcal{J}$-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 [24] 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. ■

Obviously $\beta_n^{\#}(\omega,\theta,0)=\arg\min_B c_n(\omega,\theta,\beta)$ P **a.s.** In the following, dependence on Ω will henceforth be suppressed (where possible) for notational simplicity. Dependence on B, Θ and the "optimization error" ε_n will be kept.[5]

[5] The fundamental selection theorem (Theorem 2.13 of [24]) implies the existence of a measurable selection, i.e. a $\mathcal{B}(B)/\mathcal{J}\otimes\mathcal{B}(\Theta)$-measurable random element $\beta_n^*:\Omega\times\Theta\to\mathbb{R}$ termed as auxiliary selection, defined by

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

We will not use selections to define and explore the subsequent definition of the IE since this would imply stricter conditions for identification.

Epi-Limits and Existence of a Fell Consistent Auxiliary Correspondence

The following assumption facilitates the investigation of the issue of (pseudo-) consistency for the auxiliary correspondence. It indicates the almost sure epi-convergence of the auxiliary criterion to a proper, semicontinuous asymptotic counterpart. It enables the use of the fact that the arg min correspondence is upper continuous as a function defined on the relevant space of lsc functions. Analogous assumptions have been used for the establishment of strong consistency of various estimators. See among others [11], [20], [22] and [25].

Assumption A.2 *There exists a function $c : \Theta \times B \to \overline{\mathbb{R}}$ such that*

1. *$\forall \theta \in \Theta$, $c_n \xrightarrow{e} c$ P* ***a.s.,*** *and*

2. *$c(\theta, \cdot)$ is proper $\forall \theta \in \Theta$.*

Remark R.2 *Following [22] the analogous sequential characterization dictates that for any θ, β:*

1. *$\liminf_{n\to\infty} c_n(\omega, \theta, \beta_n) \geq c(\theta, \beta)$ P* ***a.s.,*** *for all β_n such that $\beta_n \to \beta$**,** and*

2. *$\limsup_{n\to\infty} c_n(\omega, \theta, \beta_n) \leq c(\theta, \beta)$ P* ***a.s.,*** *for some β_n such that $\beta_n \to \beta$**.***

It is easy to see that $\forall \theta \in \Theta$, $c(\theta, \cdot) : B \to \overline{\mathbb{R}}$ is lower semicontinuous (see proposition 7.4.a of [27]). In the case that c_n is an integral w.r.t. an empirical measure of some ergodic random function (say c_i, $i \in \mathbb{Z}$) then the assumed epiconvergence would follow if for any θ there exists a finite open cover of B, such the random variable $\inf_{\beta\in A} c_0(\omega, \theta, \beta)$ is P-integrable, for any A in the cover (condition $\mathcal{C}_0$ and Theorem 2.3 of [7]). Properness is actually an ad hoc consideration (see, for example, in [29] 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. Hence $\inf_B c(\theta, \beta) = \min_B c(\theta, \beta) \in \mathbb{R}$. Finally, assumptions A.1, A.2 along with theorem 2.3.5 of [24], the separability and sequential completeness of B and the completeness of the underlying probability space imply that:

1. *$\liminf_{n\to\infty} c_n(\omega, \theta, \beta_n(\omega)) \geq c(\theta, \beta)$ P* ***a.s.,*** *for all measurable β_n such that $\beta_n \to \beta$ P* ***a.s.,*** *and*

2. *$\limsup_{n\to\infty} c_n(\omega, \theta, \beta_n(\omega)) \leq c(\theta, \beta)$ P* ***a.s.,*** *for some measurable β_n such that $\beta_n \to \beta$ P* ***a.s.***

Proposition 3.2 *Under assumptions A.1, A.2 the binding correspondence* $b(\theta) \doteqdot \arg\min_B c(\theta, \beta)$ *is non empty-compact valued* $\forall\theta \in \Theta$.

Proof. It follows from R.2. ■

Both the auxiliary and the binding correspondence, will be used for the definition of the IE. The following result explores their asymptotic relation. We denote by $\mathrm{Ls}(\cdot)$ and $\mathrm{Li}(\cdot)$ the sets of the P **a.s.** cluster and limit points respectively, of any sequence of sets in $F_0(B)$. Its first and last implications are already known. Its second implication is a partial generalization of Theorems 7.30, 7.32 of [27] in our setting.

Lemma 3.3 *Under assumptions A.1, A.2:*

1. *for any* ε_n *such that* $\varepsilon_n \to 0$ P ***a.s.*** *then* $\mathrm{Ls}\left(\beta_n^{\#}(\theta, \varepsilon_n)\right) \subseteq b(\theta)$ P ***a.s.***,

2. *there exists a non negative random variable,* ε_n^* *such that* $\varepsilon_n^* \to 0$ P ***a.s.*** *and* $\mathrm{Li}\left(\beta_n^{\#}(\theta, \varepsilon_n^*)\right) = b(\theta)$ P ***a.s.***,

3. *if* $b(\theta)$ *is singleton then for any* ε_n *such that* $\varepsilon_n \to 0$ P ***a.s.*** *then* $\mathrm{Li}\left(\beta_n^{\#}(\theta, \varepsilon_n)\right) = b(\theta)$ P ***a.s.***

For the proof of the previous lemma, we will need the following propositions.

Proposition 3.4 *Under assumptions A.1, A.2*

$$\limsup_n \inf_B c_n(\theta, \beta) \le \inf_B c(\theta, \beta) \;\; P \;\; \textbf{\textit{a.s.}}$$

Proof. 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. Analogously let the epigraph correspondence of c be denoted as $\mathbf{epi}(\theta) = \mathbf{epi}(c(\theta, \cdot))$. By definition we have that $\mathbf{epi}_n(\omega, \theta) \to \mathbf{epi}(\theta)$ in the Fell topology P **a.s.** Hence for all $\omega \in \Omega_1(\theta)$, with $P(\Omega_1(\theta)) = 1$, $\liminf_n \mathbf{epi}_n(\omega, \theta) \supseteq \mathbf{epi}_n(\theta)$. Let $a_\theta = \inf_B c(\theta, \beta)$. Also from section 2 we have that for large n, $\mathbf{epi}_n(\omega, \theta) \cap B \times (a, +\infty) \neq \varnothing$ since $B \times (a, +\infty)$ is open in the relevant product topology. Hence $\inf_B c_n(\theta, \beta) \le \inf_B c(\theta, \beta)$ for all $\omega \in \Omega_1(\theta)$. ■

The next result will be used for the proof of 3.3.2-3.

Proposition 3.5 *Under assumptions A.1, A.2 there exists a sequence of random variables defined on Ω, say a_n^*, such that $a_n^* \to \inf_B (c(\theta, \beta))$ P **a.s.** and*

$$\mathrm{Li}\left(\mathrm{Level}_{\leq a_n^*}(c_n(\theta, \cdot))\right) \supseteq b(\theta) \;\; P \;\; \textbf{\textit{a.s.}}$$

Proof. Let again $a_\theta = \inf_B c(\theta, \beta)$. From the sequential implication of epiconvergence in remark R.2, we have that for any $x \in b(\theta)$, there exists a measurable x_n such that $x_n \to x$ P **a.s.** Obviously for $a_{n,x} = c_n(\theta, x_n)$ which is measurable, we have that $x_n \in \mathrm{Level}_{\leq a_{n,x}}(c_n(\theta, \cdot))$. Since $b(\theta)$ is compact, it is totally bounded and therefore for any $\varepsilon > 0$, there exists an $m(\varepsilon) \in \mathbb{N}$ and $\{y_i, i = 1, ..., m(\varepsilon)\} \subset b(\theta)$, such that the collection of balls (in B) $\mathcal{O}\left(y_i, \frac{\varepsilon}{2}\right)$ covers $b(\theta)$. Extract analogous (to the aforementioned x_n) sequences $y_{i,n} \to y_i$ P **a.s.** and define $a_{n,y_i} = c_n(\theta, y_{i,n})$ and $n^* = \min\{n : |a_{n,yi} - a_\theta| \leq \varepsilon, P \text{ } \textbf{a.s.} \text{ for all } i = 1, ..., m(q_n)\}$ which well defined due to finiteness and Egoroff's theorem, and $a_n^* = \max\{a_{n^*,y_i}, i = 1, ..., m(q_n)\}$ for all $(n-1)^* < n \leq n^*$ which is measurable and $a_n^* \to a_\theta$ P **a.s.** It follows that for any $x \in b(\theta)$, for any $\varepsilon > 0$, there exists an n and a measurable $x_n \in \mathrm{Level}_{\leq a_n^*}(c_n(\theta, \cdot))$ such that $d(x_n, x) < \varepsilon$ P **a.s.** ■

Proof of Lemma 3.3. For 1. we have first that for any measurable non negative ε_n that need not converge to zero if $x_n \in \beta_n^{\#}(\theta, \varepsilon_n)$ measurable, such that a subsequence $x_{n_k} \to x$ P **a.s.**

$$\begin{aligned}
c(\theta, x) &\leq \liminf_n c_n(\theta, x_n) \;\; P \text{ } \textbf{a.s.} \\
&\leq \liminf_n c_n(\theta, x_n) + \varepsilon_n \;\; P \text{ } \textbf{a.s.} \\
&\leq \limsup_n c_n(\theta, x_n) + \varepsilon_n \;\; P \text{ } \textbf{a.s.} \\
&\leq \limsup_n \inf_B c_n(\theta, x_n) + \varepsilon_n \;\; P \text{ } \textbf{a.s.} \\
&\leq \inf_B c(\theta, x) + \varepsilon_n \;\; P \text{ } \textbf{a.s.}
\end{aligned}$$

where that last inequality follows from proposition 3.4. This establishes that for any non negative random variable ε_n

$$\mathrm{Ls}\left(\beta_n^{\#}(\theta, \varepsilon_n)\right) \subseteq \varepsilon_n\text{-}\arg\min_B c(\theta, \beta) \;\; P \text{ } \textbf{a.s.}$$

Now 1. follows from the fact that ε-$\arg\min_B \subseteq \varepsilon'$-$\arg\min_B$ if $\varepsilon \leq \varepsilon'$. For 2. notice that from the definition of the Fell topology in section 2 for any $\varepsilon > 0$, we have that for large n, $\mathbf{epi}_n(\omega, \theta) \cap B \times [a - \varepsilon, a - 2\varepsilon] = \varnothing$ P **a.s.** since $B \times [a - \varepsilon, a - 2\varepsilon]$ is compact in the relevant product topology. This implies that $\liminf_n \inf_B c_n(\theta, \beta) \geq a_\theta$ P **a.s.** and in conjunction with 3.4

that $\inf_B c_n(\theta,\beta) \to a$ P **a.s**. Then using proposition 3.5 set $\varepsilon_n^* = a_n^* - \inf_B c_n(\theta,\beta)$ which is obviously measurable and converges to zero P **a.s**. 3. follows from the proof of proposition 3.5 since $a_n^* = a_{n,y}$ where for $b(\theta) = \{y\}$. ■

Remark R.3 *Obviously 3.3.1 is equivalent to* $\delta_u\left(b(\theta), \mathrm{Ls}\left(\beta_n^{\#}(\theta,\varepsilon_n)\right)\right) = 0$ P ***a.s.***

Upper Continuity of the Binding Correspondence

The following assumption concerns the behavior of the binding correspondence. It will facilitate both the existence and the strong consistency of the IE to be defined.

Assumption A.3 *b is $\mathcal{T}_U/\tau_\Theta$-continuous.*

The following lemma provides with sufficient conditions for this to hold. It essentially strengthens assumption A.2 in that it requires that the relevant P **a.s.** epiconvergence is *continuous* with respect to Θ.

Lemma 3.6 *Suppose that for any θ, β, and any $\theta_n \to \theta$:*

1. $\liminf_{n\to\infty} c_n(\omega,\theta_n,\beta_n) \geq c(\theta,\beta)$ P ***a.s.****, for all β_n such that $\beta_n \to \beta$***,** *and*

2. $\limsup_{n\to\infty} c_n(\omega,\theta_n,\beta_n) \leq c(\theta,\beta)$ P ***a.s.****, for some β_n such that $\beta_n \to \beta$ (independent of θ_n)***,** *then*

assumption A.3 holds.

Proof. Suppose that $\mathcal{D}_{F_0}$ metrizes $\mathcal{T}_F$ on $F_0(B)$ (see section 2). Then it is obvious that 3.6 1-2 are equivalent to the requirement that for any θ, and any $\theta_n \to \theta$, $\mathcal{D}_{F_0}(\mathbf{epi}_n(\omega,\theta_n), \mathbf{epi}(\theta))$ converges to zero P **a.s.** Then, for any θ, and any $\theta_n \to \theta$, $c(\theta_n,\cdot)$ epiconverges to $c(\theta,\cdot)$ ($\mathcal{D}_{F_0}(\mathbf{epi}(\theta_n), \mathbf{epi}(\theta)) \to 0$), i.e. $c(\theta,\cdot)$ is epicontinuous on Θ. This is due to the following standard argument: for an arbitrary θ, and $\varepsilon > 0$, we have that $\mathcal{D}_{F_0}(\mathbf{epi}_n(\omega,\theta'), \mathbf{epi}(\theta)) < \frac{\varepsilon}{2}$ P **a.s.** for any θ' in some open neighborhood of θ due to the assumed form of convergence and Egoroff's Theorem. By the same reasoning $\mathcal{D}_{F_0}(\mathbf{epi}_n(\omega,\theta'), \mathbf{epi}(\theta')) < \frac{\varepsilon}{2}$ P **a.s.** for any such θ'. The result follows from the fact that $\mathbf{epi}(\theta)$ is independent of Ω. This along with equation 3.1 of Theorem 5.3.4 and proposition Appendix.D.2 of [24] implies that the composite mapping $\theta \to c(\theta,\cdot) \to \arg\min_B c(\theta,\beta)$ is $\mathcal{T}_U/\tau_\Theta$-continuous. ■

Remark R.4 *3.6.1-2 would obviously be implied if $c_n(\omega, \theta, \beta)$ is P **a.s.** jointly continuous and converges jointly uniformly P **a.s.** to $c(\theta, \beta)$.*[6] *The following lemma provides with a set of weaker sufficient conditions in the case that c_n is an integral w.r.t. an empirical measure of some ergodic random function (see also remark R.2). The required joint measurability of c_n is not explicitly described.*

Lemma 3.7 *Suppose that $c_n(\omega, \theta, \beta) = \frac{1}{n}\sum_{i=1}^{n} c_i(\omega, \theta, \beta)$, (c_i) is ergodic, and:*

1. *$c_n(\omega, \cdot, \cdot) : \Theta \times B \to \overline{\mathbb{R}}$ is lower semicontinuous P **a.s.**, and there exists a finite open cover of $\Theta \times B$, such the random variable $\inf_{(\theta,\beta)\in A} c_0(\omega, \theta, \beta)$ is P-integrable, for any A in the cover, and*

2. *for any β, $c_n(\omega, \cdot, \beta) : \Theta \to \overline{\mathbb{R}}$ is P **a.s.** continuous, and there exists a countable open cover of Θ, such the random variable $\sup_{\theta\in A} |c_0(\omega, \theta, \beta)|$ is P-integrable, for any A in the cover, then*

assumption A.3 holds.

Proof. It suffices that conditions 3.6.1-2 hold. The first follows from the fact that 3.7.1 implies condition $\mathcal{C}_0$ and thereby Theorem 2.3 of [7], which implies the joint P **a.s.** epiconvergence of c_n to Ec_0. 3.7.1 implies condition $\mathcal{C}_0'$ and thereby Theorem 2.5 of [7], which implies the P **a.s.** uniform convergence of $c_n(\omega, \cdot, \beta)$ to $Ec_0(\cdot, \beta)$ for any β. This implies 3.6.2 for $\beta_n = \beta$. ■

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. Remark R.3 allows us concentrate on properties of the real function on $\Omega \times \Theta$, $\delta_u\left(b(\theta), \beta_n^{\#}(\theta_0)\right)$ for $\theta_0 \in \Theta$, which enables the following definition. Again an almost surely non-negative random variable will assume the role of the "optimization error" in the second step of the estimation procedure.

Definition D.6 *The indirect correspondence $\theta_n^{\#}(\omega, \theta, \varepsilon_n)$ satisfies*

$$\begin{aligned}
\theta_n^{\#}\left(\omega, \theta_0, \varepsilon_n, \varepsilon_n^{\#}\right) &= \varepsilon_n^{\#}\text{-}\arg\min_{\Theta} \delta_u\left(b(\theta), \beta_n^{\#}(\theta_0, \varepsilon_n)\right) \\
&\doteqdot \left\{\theta^* \in \Theta : \delta_u\left(b(\theta^*), \beta_n^{\#}(\theta_0, \varepsilon_n)\right) \leq \inf_{\Theta} \delta_u\left(b(\theta), \beta_n^{\#}(\theta_0, \varepsilon_n)\right) + \varepsilon_n^{\#}\right\}
\end{aligned}$$

where $\varepsilon_n^{\#}$ is a non-negative random variable defined on Ω.

[6] Since we allow c_n and/or c to assume extended real values, the relevant notion of uniform convergence must also be extended as in definition 7.12 of [27].

We are initially concerned with the question of existence of the IE. We again suppress the dependence of θ_n on Ω when there is not a risk of confusion.

Lemma 3.8 *Under assumptions A.1 and A.3* $\theta_n^{\#}$ *is* $\mathcal{B}(\mathcal{T}_F)/\mathcal{J}$*-measurable,* P ***a.s.*** *non empty, compact valued correspondence.*

Proof. First, notice that due to 2.5, 3.1 (implied by A.1), A.3 and the facts that $\beta_n^{\#}(\theta_0, \varepsilon_n)$ is independent of θ and $b(\theta)$ is independent of ω, we obtain that $\delta_u\left(\beta_n^{\#}(\theta_0, \varepsilon_n), b(\theta)\right)$ is $\mathcal{B}(\mathbb{R})/\mathcal{B}(\mathcal{T}_F)\otimes\mathcal{J}$-measurable. Due to 2.3 and 3.1 $\delta_u\left(\beta_n^{\#}(\theta_0, \varepsilon_n), b(\theta)\right)$ is P **a.s.** lsc and therefore a normal intergrand. It is also obviously proper hence the result follows from lemma 3.1 where $c_n = \delta_u\left(\beta_n^{\#}(\theta_0, \varepsilon_n), b(\theta)\right)$, when we consider $B = \Theta$ and $\Theta = \{\theta_0\}$ (the left hand sides correspond to the notation of the latter lemma). ■

The fundamental selection theorem (Theorem 2.13 of [24]) would also enable the definition of the IE as a measurable function with values in Θ. Having established the existence of the IE, we turn to the issue of consistency. We need the following assumption that facilitates this investigation.

Assumption A.4 *If* $\theta \neq \theta_0 \Rightarrow b(\theta_0) - b(\theta) \neq \varnothing$.

Remark R.5 *Notice that this condition is weaker that a condition of the form "If* $\theta \neq \theta_0 \Rightarrow b(\theta_0) \cap b(\theta) = \varnothing$*" and stronger that a condition of the form "If* $\theta \neq \theta_0 \Rightarrow b(\theta_0) \neq b(\theta)$*". The latter cannot be used due to the properties of* δ_u *upon which the definition of the* IE *is based. In the case that the binding correspondence is single valued, these become equivalent.*

The main result of the current section follows. It concerns the existence of strongly consistent IE. Denote with $\theta(\theta_0)$ the $\arg\min_{\Theta}\delta_u(b(\theta), b(\theta_0))$ which is non empty and compact due to the compactness of Θ, the properness of δ_u and lemma 2.2. Obviously, $\theta_0 \in \theta(\theta_0)$, while $\{\theta_0\} = \theta(\theta_0)$ iff assumption A.4 holds.

Lemma 3.9 *Under assumptions A.1, A.2, A.3, and if* $\varepsilon_n^{\#} \to 0$ P ***a.s.*** *then:*

1. $\delta_u\left(\theta(\theta_0), \theta_n^{\#}\left(\omega, \theta_0, \varepsilon_n^*, \varepsilon_n^{\#}\right)\right) \to 0$ P ***a.s.*** *where* ε_n^* *is defined in lemma 3.3.2,*
2. *if* $b(\theta_0)$ *is singleton then* $\delta_u\left(\theta(\theta_0), \theta_n^{\#}\left(\omega, \theta_0, \varepsilon_n, \varepsilon_n^{\#}\right)\right) \to 0$ P ***a.s.*** *for any* $\varepsilon_n \to 0$ P ***a.s.,***

if furthermore A.4 holds then

*1** $\delta\left(\{\theta_0\}, \theta_n^{\#}\left(\omega, \theta_0, \varepsilon_n^*, \varepsilon_n^{\#}\right)\right) \to 0$ P ***a.s.*** *where* ε_n^* *is defined in lemma 3.3.2, and*

*2** *if* $b(\theta_0)$ *is singleton then* $\delta\left(\{\theta_0\}, \theta_n^{\#}\left(\omega, \theta_0, \varepsilon_n, \varepsilon_n^{\#}\right)\right) \to 0$ P ***a.s.*** *for any* $\varepsilon_n \to 0$ P ***a.s.***

Proof. First notice that due to lemma 3.3.2, A.3 and lemma 2.4 we have that for any θ and $\theta_n \to \theta$

$$\liminf_n \delta_u\left(b(\theta_n), \beta_n^{\#}(\theta_0, \varepsilon_n^*)\right) \geq \delta_u(b(\theta), b(\theta_0)) \ P \ \textbf{a.s.}$$

and that for any θ and $\theta_n = \theta$, due to lemma 2.3

$$\limsup_n \delta_u\left(b(\theta_n), \beta_n^{\#}(\theta_0, \varepsilon_n^*)\right) \leq \delta_u(b(\theta), b(\theta_0)) \ P \ \textbf{a.s.}$$

hence 1. follows from 3.3.1 for $c_n = \delta_u\left(b(\theta_n), \beta_n^{\#}(\theta_0, \varepsilon_n^*)\right)$ and if we denote with B (in the notation of this lemma) the Θ space and with Θ (in the notation of this lemma) $\{\theta_0\}$. 2. follows in the same manner if we replace any invocation of 3.3.2 with 3.3.3. Finally, notice that if A.4 holds, then $\theta(\theta_0) = \{\theta_0\}$ establishing 1^* and 2^* via another use of 3.3.3. ■

4 Examples

In this section we consider four simple examples that represent some of the previous results. The first concerns the case of a linear semi-parametric model, the second a model comprised of Levy processes and the final two emerge in the context of conditionally heteroskedastic ones. In any of these, Θ is a compact subset of $\mathbb{R}^p$ and B a compact subset of $\mathbb{R}^q$. In the second and the fourth one the binding function is actually single valued (hence a fortiori compact valued) and 1-1 enabling the direct application of $3.9.2^*$. The first and second examples include cases in which the IE can be interpreted as performing "inconsistency" correction to the auxiliary one.

Example Semi - Parametric Linear Model with Linear Auxiliary. Consider the $n \times p$ and $n \times q$ dimensional random matrices $X(\omega)$ and $Z(\omega)$ respectively, where $n \geq q \geq p$. Suppose that $\frac{X'X}{n} \to M_{X'X}$, $\frac{Z'Z}{n} \to M_{Z'Z}$, $\frac{Z'X}{n} \to M_{Z'X}$ P **a.s.**, where $\textbf{rank}(M_{X'X}) = \textbf{rank}(M_{Z'X}) = p$ and $p \leq l \doteqdot \textbf{rank}(M_{Z'Z}) \leq q$. For $u(\omega)$ is a $n \times 1$ random vector, let the underlying statistical model be the set of "regressions" $Y(\omega, \theta) = X(\omega)\theta + u(\omega)$, $\theta \in \Theta$. For B a large enough compact and convex subset of $\mathbb{R}^q$ and any $\beta \in B$, let $c_n(\omega, \theta, \beta) = \frac{1}{n}(Y - Z\beta)'(Y - Z\beta)$, which clearly satisfies assumption A.1

due to continuity with respect to β and the compactness of B. Obviously, c_n is constructed by the auxiliary set of regression w.r.t. Z. Lemma 3.1 ensures the existence of β_n which in the light of the previous can be interpreted as an OLSE in the context of the auxiliary model. Let $\mathcal{P} : \mathbb{R}^q \to M_{Z'Z}B$ be the (generally non-linear) projection defined by the optimization problem

$$\arg \min_{x \in M_{Z'Z}B} \|x - y\|$$

for y in $\mathbb{R}^q$. $\mathcal{P}$ is well defined due to the compactness and the convexity of B and the linearity and continuity of $M_{Z'Z}$ and continuous. Furthermore, for any y in the column space of $M_{Z'Z}$, consider the linear system $M_{Z'Z}x = y$, which is always satisfied by any member of the coset $Ky + H_{q-l}$, where K is a matrix of rank l and H_{q-l} is a $q - l$-dimensional subspace of $\mathbb{R}^q$, which is trivial if and only if $l = q$ whereas $K = M_{Z'Z}^{-1}$, and maximal in the case that $l = p$. Suppose that B contains the closed ball $\mathcal{O}(0_q, \delta)$ for $\delta > \sup_{\theta \in \Theta} \|K\mathcal{P}(M_{Z'X}\theta - M_{Z'u})\|$ which is well defined due to the compactness of Θ and the continuity of $\mathcal{P}$. Let also $M_{X'u}, M_{Z'u}, M_{u'u} \in \mathbb{R}$ and assume that $\frac{X'u}{n} \to M_{X'u}$, $\frac{u'u}{n} \to M_{u'u}$ and $\frac{Z'u}{n} \to M_{Z'u}$ P **a.s.** The previous imply the joint uniform P **a.s.** convergence of c_n to

$$c(\theta, \beta) = \theta' M_{X'X}\theta - 2\beta' M_{Z'X}\theta + \beta' M_{Z'Z}\beta + 2\theta' M_{X'u} - 2\beta' M_{Z'u} + M_{u'u}$$

which implies both assumptions A.2, A.3 (via lemma 3.6 and R.4).Notice that

$$b(\theta) = B \cap (K\mathcal{P}(M_{Z'X}\theta - M_{Z'u}) + H_{q-l})$$

due to the convexity of $c(\theta, \beta)$ w.r.t. β for any θ and the definition of B. If $\mathcal{P}(M_{Z'X}\theta_0 - M_{Z'u}) \neq \mathcal{P}(M_{Z'X}\theta - M_{Z'u})$ and $K\mathcal{P}(M_{Z'X}\theta_0 - M_{Z'u}) - K\mathcal{P}(M_{Z'X}\theta - M_{Z'u}) \notin H_{q-l}$ for any $\theta \neq \theta_0$ then assumption A.4 applies. Hence, lemma 3.9.1* implies the *existence* of a consistent IE for $\theta_0 \in \Theta$. In the special case where $X = Z$ then $b(\theta) = \left\{\theta - M_{Z'Z}^{-1} M_{Z'u}\right\}$ and lemma 3.9.2* implies that any IE defined by D.6 can be perceived as an "inconsistency corrector" of the underlying OLSE.$\square$

We know consider the case of the estimation of the drift of a continuous time cadlag process.

Example The drift of a Levy Process with Bounded Jumps.
Let W denote a standard Wiener process and v a finite measure on the Borel algebra of $\mathbb{R} - \{0\}$, such that $v(A) = 0$ when $A \subseteq (-\infty, -C_2) \cup (-C_1, C_1) \cup (C_2, +\infty)$ for $0 < C_1 < C_2$. Obviously v is a Levy measure (see paragraph

1.2.4 of [2]). For $p = 1$ consider the stochastic process on $\mathbb{R}^+$ defined by the following Levy-Ito decomposition (see Theorem 2.4.16 of [2])

$$X_t(\omega) = \theta t + W_t(\omega) + \int_{|x| \in [C_1, C_2]} xN(t, dx)(\omega)$$

where N denotes the independent to W Poisson random measure on $\mathbb{R}^+ \times ([-C_2, C_1] \cup [C_1, C_2])$ the existence of which is established by Theorem 2.3.6 of [2]. Let the underlying statistical model be the set of the previous stochastic processes and for B a large enough compact subset of $\mathbb{R}$ and any $\beta \in B$, let $c_n(\omega, \theta, \beta) = \frac{1}{n}\sum_{t=1}^{n}(y_t - \beta)^2$, where $y_t = \exp(X_t - X_{t-1}) - 1$. This can be perceived to emerge as an approximate likelihood function of the auxiliary model that contains the relevant discretizations of the processes that satisfy the SDE

$$dy_t = \beta y_t dt + y_t dW_t$$

for each $\beta \in B$. Obviously assumption A.1 is satisfied, due to continuity with respect to β and the compactness of B. Lemma 3.1 ensures the existence of β_n which in the light of the previous can be interpreted as an (approximate) MLE in the context of the auxiliary model. Furthermore since

$$\begin{aligned}
\left|\int_{|x| \in [C_1, C_2]} x(N(t, dx) - N(t-1, dx))\right| &\leq \int_{|x| \in [C_1, C_2]} |x|(N(t, dx) - N(t-1, dx)) \\
&\leq C_2(N(t, [C_1, C_2]) - N(t-1, [C_1, C_2]))
\end{aligned}$$

and $N(t, [C_1, C_2]) - N(t-1, [C_1, C_2]) \overset{i.i.d.}{\sim} \text{Poiss}(v([C_1, C_2]))$ independent of W, we have that

$$E\exp(X_t - X_{t-1}) = \exp\left(\theta + \frac{1}{2}\right)C$$

for

$$0 < C \leq \exp(-v([C_1, C_2])(1 - \exp(C_2)))$$

and

$$E(\exp(X_t - X_{t-1}))^2 = \exp(2(\theta + 1))C^*$$

for

$$0 < C^* \leq \exp(-v([C_1, C_2])(1 - \exp(2C_2)))$$

Due to the definition of X the process y is i.i.d. and this along with the compactness of Θ, and B and the existence of the previous moments imply the joint uniform P **a.s.** convergence of c_n to

$$c(\theta, \beta) = \exp(2(\theta + 1))C^* - 2\exp\left(\theta + \frac{1}{2}\right)C(1 + \beta) + (1 + \beta)^2$$

which implies both assumptions A.2, A.3 (via lemma 3.6 and R.4). If $B \supseteq \exp\left(\Theta + \frac{1}{2}\right) C - 1$ then

$$b(\theta) = \left\{\left(\exp\left(\theta + \frac{1}{2}\right) C - 1\right)\right\}$$

In this case assumption A.4 applies and therefore lemma 3.9.2* implies that any IE defined by D.6 is consistent for any $\theta_0 \in \Theta$. When $v([C_1, C_2]) = 0$ (and therefore $v = 0$) whereas $C = 1$ the IE can be perceived as an "inconsistency corrector" of the underlying MLE for the estimation of the drift of a geometric Brownian motion (see for example paragraph 6.1.1 of [17]).$\square$

For the last pair of examples, let $z : \Omega \to \mathbb{R}^{\mathbb{Z}}$ 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 $\mathcal{B}(\Theta) \otimes \mathcal{J}$ with $\sigma_t^2(\theta)$ independent of $(z_i)_{i \geq t}$, $\forall t \in \mathbb{Z}$, $\forall \theta \in \Theta$. 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. Our examples will solely concern ergodic heteroskedastic models.[7]

Example IV **Estimation in Regressions on Squared** ARCH(1) **processes.** Let $\sigma_4 = Ez_0^4 < +\infty$ and $0 < \delta < \frac{1}{\sqrt{\sigma_4}}$. Suppose that $a \in \Theta = [0, \delta]$ and consider the stochastic difference equation

$$\sigma_t^2(a) = 1 + az_{t-1}^2\sigma_t^2(a)$$

Due to the fact that $\sigma_4 > 1$ Theorem 5.2.1. of [29] implies that for any $\alpha \in \Theta$ the equation admits a unique ergodic solution defining the analogous ARCH(1) process. Consider the random vector $Y(a) = (y_t^2(a))_{t\in\{1,\dots,n\}}$, and the $n \times 2$ dimensional random matrices

$$Z(a) = \begin{pmatrix} 1 & y_{-1}^2(a) \\ \vdots & \vdots \\ 1 & y_{n-2}^2(a) \end{pmatrix}, \; X(a) = \begin{pmatrix} 1 & y_0^2(a) \\ \vdots & \vdots \\ 1 & y_{n-1}^2(a) \end{pmatrix} \quad (1)$$

[7] The establishment of the ergodicity is initiated by the analogous establishment for $\left(\sigma_t^2(\theta)\right)_{t\in\mathbb{Z}}$ $\forall \theta \in \Theta$. Sufficient conditions for that are described and employed in a variety of heteroskedastic models in chapter 5 of [29] via Theorem 5.2.1. Then the ergodicity of $(y_t(\theta))_{t\in\mathbb{Z}}$ and $\left(y_t^2(\theta)\right)_{t\in\mathbb{Z}}$ $\forall \theta \in \Theta$ follow from the definition of z, y, the previous assumption and Proposition 2.2.1 of [29].

jointly measurable with respect to $\mathcal{J} \otimes \mathcal{B}(\Theta)$, where $n > 2$ and ergodic for any $a \in \Theta$. For $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \in B = [1, 1+\delta] \times [-\delta, \delta]$, let $c_n(\omega, a, \beta) = \left\| \frac{1}{n} Z'(a)(Y(a) - X(a)\beta) \right\|$ which clearly satisfies assumption A.1 due to joint continuity with respect to (a, β) the compactness of B, the joint measurability and the fact that c_n is defined via composition with a norm.[8] This consideration is motivated from the AR(1) representation of the ARCH(1) process with respect to the martingale difference noise $v_t = (z_t^2 - 1)\sigma_t^2(a)$ (see, for example, [5]) 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 $X\beta$, along with the instrumental variables appearing in the columns of Z where obviously the i^{th} element in any column is clearly orthogonal to v_t for $i \leq t$. Lemma 3.1 assures the existence of β_n which in the light of the previous sentence can be interpreted as an IV estimator in the context of the auxiliary model. Due to the compactness of B, the definition of the ARCH(1) model and the definitions of Θ and σ_4 we have that

$$\begin{aligned} & E \sup_{a,\beta} \left\| \frac{1}{n} Z'(a)(Y(a) - X(a)\beta) \right\| \\ \leq \; & C_1 \left\| \begin{pmatrix} 1 & \frac{1}{1-\delta} \\ \frac{1}{1-\delta} & \frac{1}{(1-\delta)(1-\delta^2\sigma_4)} \end{pmatrix} \right\| + C_2 \left\| \begin{pmatrix} \frac{1}{1-\delta} \\ \frac{1+\delta}{(1-\delta)(1-\delta^2\sigma_4)} \end{pmatrix} \right\| < +\infty \end{aligned}$$

for $C_1, C_2 > 0$ which along with (the uniform version of) Birkhoff's Ergodic Theorem (see for example [29], Theorem 2.2.1) implies both assumptions A.2, A.3 (via lemma 3.6 and R.4) for

$$c(a, \beta) = \left\| \begin{pmatrix} 1 & \frac{1}{1-a} \\ \frac{1}{1-a} & \frac{1}{(1-a)(1-a^2\sigma_4)} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} - \begin{pmatrix} \frac{1}{1-a} \\ \frac{1+a}{(1-a)(1-a^2\sigma_4)} \end{pmatrix} \right\|$$

In fact a simple calculation shows that

$$b(a) = \begin{cases} \left\{ \begin{pmatrix} 1 \\ a \end{pmatrix} \right\}, & a \in (0, \delta] \\ [1, 1+\delta] \times [-\delta, 0], & a = 0 \end{cases}$$

[8] For expositional reasons and in the context of the first paragraph of section 3, we have that $K_n = (\mathbb{R}^+)^{\mathbb{Z}} \times (\mathbb{R}^+)^{\mathbb{Z}} \times (\mathbb{R}^+)^{\mathbb{Z}}$ and $y_n = \begin{pmatrix} \vdots & \vdots & \vdots \\ 0 & 0 & 0 \\ Y(a) & Y_{-1}(a) & Y_{-2}(a) \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{pmatrix}$ (in block form) where Y_{-1} and Y_{-2} denote the second columns of X and Z respectively. In all the other examples the analogous representations are not explicitely derived or presented for reasons of economy of space.

which clearly implies assumption A.4. Hence Lemma 3.9.2* implies that any IE defined by D.6 is consistent if $a_0 \in (0, \delta]$, and Lemma 3.9.1* implies the *existence* of an analogously consistent IE when $a_0 = 0$.$\square$

The final example is about an asymmetric heteroskedastic process.

Example $-c_n$ is the Quasi-Likelihood Function of an Approximate to QARCH (1) **Model**. Let $\varsigma = E|z_0| < 1$ suppose that $P(z_0 = 0) = 0$, and consider for $\gamma \in \Theta = [-\delta, 0]$ where $\delta < 2\sqrt{\varpi_0}(\exp(-E\ln|z|) - a_0\varsigma)$, $\varpi_0, a_0 > 0$, and $\varpi = \varpi_0 + \frac{\delta^2}{4a_0}$, the stochastic difference equation

$$\sigma_t^2(\gamma) = \varpi + a_0 z_{t-1}^2 \sigma_t^2(\gamma) + \gamma z_{t-1}\sigma_t(\gamma)$$

For any $\gamma \in \Theta$, the previous represent a well defined ergodic QARCH (1) process, due to the fact that $\inf_z(a_0 z^2 x + \gamma z\sqrt{x}) = -\frac{\gamma^2}{4a_0}$, which implies the P **a.s.** positivity of the volatility process and the subsequent well definition of $(y_t(\gamma))_{t\in\mathbb{Z}}$, as well as that $\inf_{t,\gamma}\sigma_t^2(\gamma) \geq \varpi_0$ which in turn along with the bound on δ, and Jensen's inequality, imply that $E\ln\left|a_0 z_0^2 + \frac{\gamma z_0}{2\sqrt{\varpi_0}}\right| < 0$ which establishes the ergodicity via Theorem 5.2.1. of [29]. Notice that since $\frac{\exp(-E\ln|z|)}{\varsigma} > 1$, a_0 can assume the value 1 which in turn implies that in this model ergodicity is possible even when $E\sigma_t^2(\gamma) = +\infty$. For $\gamma^* \in B = [0, \delta]$ consider the process defined by

$$\begin{aligned} h_t(\gamma, \gamma^*) &= \varpi + a_0 y_{t-1}^2 + \gamma^* |y_{t-1}| 1_{y_{t-1}<0} \quad (2) \\ &\overset{P\ \mathbf{a.s.}}{=} \varpi + a_0 y_{t-1}^2 + \gamma^* |y_{t-1}| 1_{z_{t-1}<0} \end{aligned}$$

$h_t(\gamma, \gamma^*)$ is well defined due to the definition of B and it is ergodic due to the ergodicity of the QARCH (1) and Proposition 2.1.1. of [29]. Now consider

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

$-c_n$ can be considered as an approximation of (a monotonic transformation of) the conditional quasi likelihood function of the auxiliary conditionally heteroskedastic model defined by 2 and B. Also the ergodicity of (c_n) for any (γ, γ^*) follows from the previous and Proposition 2.1.1. of [29].[9] Assumption A.1 follows readily from the form of c_n and the P **a.s.** continuity with

[9] In practice $c_n(\omega, \gamma, \gamma^*)$ is unknown but approximated by an analogous $\widehat{c}_n(\omega, \gamma, \gamma^*)$

respect to (γ, γ^*). Hence β_n is well defined and can be interpreted as an approximate QMLE in the context of the auxiliary model. Now, consider an arbitrary finite open cover of B and notice that

$$\begin{aligned}
& E \inf_{A\cap B}\left(-\ln \frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}} + z_0^2 \frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}}\right) \\
\geq\ & E \inf_{A\cap B}\left(-\ln \frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}}\right) + E \inf_{A\cap B}\left(\frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}}\right) \\
\geq\ & -E \sup_{A\cap B} \ln \frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}} + E\left(\frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \delta |y_0| 1_{z_0<0}}\right) \\
\geq\ & -E \ln\left(\left(1 + \frac{\gamma y_0}{\varpi + a_0 y_0^2}\right) 1_{z_0<0}\right) + E\left(\left(1 + \frac{\gamma y_0}{\varpi + a_0 y_0^2}\right) 1_{z_0\geq 0}\right) > -\infty
\end{aligned}$$

and that

$$\begin{aligned}
& E \inf_{A\cap B}\left(-\ln \frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}} + z_0^2 \frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}}\right) \\
\leq\ & -E \ln \frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \delta |y_0| 1_{z_0<0}} + E\left(\frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2}\right) \\
\leq\ & 1 - E\left(\ln\left(1 + \frac{\gamma y_0}{\varpi + a_0 y_0^2}\right)\right) 1_{z_0\geq 0} + E\left(1 + \frac{\gamma y_0}{\varpi + a_0 y_0^2}\right) < +\infty
\end{aligned}$$

for A an arbitrary member of the partition. Notice also that $-2E\ln|z| - E\frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}} + E\frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}} > -\infty$ for all (γ, γ^*) due to the fact that

$$E \ln \frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}} \geq E \ln\left(1 + \frac{\gamma y_0}{\varpi + a_0 y_0^2}\right) > -\infty$$

Hence, remark R.2 implies that assumption A.2 holds with $c(\gamma, \gamma^*) = -2E\ln|z| - E\ln\frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}} + E\frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}}$. Notice that $c(0, \gamma^*)$ is uniquely minimized at $\gamma^* = 0$ (see for example the Part 1. of the proof of Theorem 5.3.1. of [29] to obtain the analogous arguments along with the fact that $\frac{\varpi + a_0 y_0^2}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}} \overset{P \text{ a.s.}}{=} 1$ iff $\gamma^* = 0$). When $\gamma \neq 0$ then $c(\gamma, -\gamma) < c(\gamma, 0)$

dependent on non ergodic solutions of the stochastic difference equation that defines h based on arbitrary initial conditions. In this case, due to ergodicity, Proposition 5.2.12 of [29] 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 [29]), thereby facilitating the asymptotic analysis of minimizers of $\widehat{c}_n(\omega, \theta, \beta)$ by the analogous analysis of minimizers of $c_n(\omega, \theta, \beta)$.

due to the fact

$$
\begin{aligned}
& -E\ln\frac{\varpi+a_0y_0^2+\gamma y_0}{\varpi+a_0y_0^2-\gamma\left|y_0\right|1_{z_0<0}}+E\ln\frac{\varpi+a_0y_0^2+\gamma y_0}{\varpi+a_0y_0^2} \\
= \; & E\left(\ln\left(1+\frac{\gamma y_0}{\varpi+a_0y_0^2}\right)1_{z_0<0}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
& E\frac{\varpi+a_0y_0^2+\gamma y_0}{\varpi+a_0y_0^2-\gamma\left|y_0\right|1_{z_0<0}}-E\frac{\varpi+a_0y_0^2+\gamma y_0}{\varpi+a_0y_0^2} \\
= \; & E\frac{\varpi+a_0y_0^2+\gamma y_0}{\varpi+a_0y_0^2-\gamma\left|y_0\right|1_{z_0<0}}1_{z_0<0}+E\frac{\varpi+a_0y_0^2+\gamma y_0}{\varpi+a_0y_0^2-\gamma\left|y_0\right|1_{z_0<0}}1_{z_0\geq 0} \\
& -E\left(\left(1+\frac{\gamma y_0}{\varpi+a_0y_0^2}\right)1_{z_0<0}\right)-E\left(\left(1+\frac{\gamma y_0}{\varpi+a_0y_0^2}\right)1_{z_0\geq 0}\right) \\
= \; & -E\left(\frac{\gamma y_0}{\varpi+a_0y_0^2}1_{z_0<0}\right)<0
\end{aligned}
$$

and that when $x>0$, then $\ln(1+x)<x$. Furthermore, using the fact that by 2 h is P **a.s.** two times differentiable w.r.t. γ^* for $\gamma^*\neq 0$ and since

$$
\begin{aligned}
& E\sup_{\gamma^*}\left|\left(\frac{\varpi+a_0y_0^2+\gamma^*\left|y_0\right|1_{z_0<0}}{\varpi+a_0y_0^2+\gamma y_0}-1\right)\frac{\varpi+a_0y_0^2+\gamma y_0}{\left(\varpi+a_0y_0^2+\gamma^*\left|y_0\right|1_{z_0<0}\right)^2}\left|y_0\right|1_{z_0<0}\right| \\
\leq \; & \left(\frac{\delta}{|\gamma|}-1\right)E\left|\frac{\varpi\left|y_0\right|+a_0\left|y_0\right|^3+\gamma y_0^2}{\left(\varpi+a_0y_0^2\right)^2}1_{z_0<0}\right|<+\infty
\end{aligned}
$$

and

$$
\begin{aligned}
& E\sup_{\gamma^*}\left|\left(\frac{\varpi+a_0y_0^2+\gamma y_0}{\left(\varpi+a_0y_0^2+\gamma^*\left|y_0\right|1_{z_0<0}\right)^2}\right)z_0^2 1_{z_0<0}\right| \\
\leq \; & \frac{1}{\varpi}E\left|\left(1+\frac{\gamma y_0}{\varpi+a_0y_0^2}\right)1_{z_0<0}\right|<+\infty
\end{aligned}
$$

$$
\begin{aligned}
& E\sup_{\gamma^*}\left|\left(\frac{\varpi+a_0y_0^2+\gamma^*\left|y_0\right|1_{z_0<0}}{\varpi+a_0y_0^2+\gamma y_0}-1\right)\frac{\left(\varpi+a_0y_0^2+\gamma y_0\right)^2}{\left(\varpi+a_0y_0^2+\gamma^*\left|y_0\right|1_{z_0<0}\right)^3}z_0^2 1_{z_0<0}\right| \\
\leq \; & \frac{1}{\varpi}\left(\frac{\delta}{|\gamma|}-1\right)E\left|\left(1+\frac{\gamma y_0}{\varpi+a_0y_0^2}\right)^2 1_{z_0<0}\right|<+\infty
\end{aligned}
$$

as well as dominated convergence, we have that

$$
\frac{\partial c\left(\gamma,\gamma^*\right)}{\partial\gamma^*}=E\left(\frac{\varpi+a_0y_0^2+\gamma^*\left|y_0\right|1_{z_0<0}}{\varpi+a_0y_0^2+\gamma y_0}-1\right)\frac{\varpi+a_0y_0^2+\gamma y_0}{\left(\varpi+a_0y_0^2+\gamma^*\left|y_0\right|1_{z_0<0}\right)^2}\left|y_0\right|1_{z_0<0}
$$

which is zero iff $\gamma^* = -\gamma$, and

$$\frac{\partial^2 c(\gamma,\gamma^*)}{\partial(\gamma^*)^2}|_{\gamma^*=-\gamma} = E\left(\left(\frac{\varpi + a_0 y_0^2 + \gamma y_0}{\left(\varpi + a_0 y_0^2 - \gamma\,|y_0|\,1_{z_0<0}\right)^2}\right) z_0^2 1_{z_0<0}\right) > 0$$

establishing along with the previous that $b(\gamma) = \{-\gamma\}$. This validates simultaneously both assumptions A.3 and A.4. Hence Lemma 3.9.2* implies that any IE defined by D.6 is consistent for any $\gamma_0 \in [-\delta, 0]$.□

5 Conclusions

In this paper we have generalized the definition of IE and were occupied with the questions of their existence and strong consistency, allowing for cases where the binding function is a compact valued correspondence. We have used conditions that concern the asymptotic behavior of the epigraphs of the criterion functions involved in the relevant procedures, a relevant notion of continuity for the correspondence as well as an indirect identification condition that restricts the behavior of the aforementioned correspondence. These results are generalizations of the analogous ones in the relevant literature, hence permit a broader scope of statistical models.

We leave for future research the questions of the definition and consistency of IE when the binding correspondence can only be appropriately approximated. Consider for example the case where $b(\theta)$ is replaced by $E\left(\beta_n^{\#}(\theta,\varepsilon_n)\right)$ in definition D.6 where the latter is interpreted as some sort of integral of $\beta_n^{\#}$ (see for example [3]) or by some stochastic approximation of it. The same holds for the issues of the establishment of the rates of convergence, and the asymptotic distribution of the IE in this general framework.

References

[1] Andersen, T.G., L. Benzoni and J. Lund (2002), "An empirical investigation of continuous-time equity return models", *The Journal of Finance,* LVII, pp. 1239-1284.

[2] Applebaum D., "*Levy Processes and Stochastic Calculus*", Cambridge University Press, 2004.

[3] Artstein Zvi and John A. Burns (1975), "Integration of Compact Valued Set Functions", *Pacific Journal of Mathematics*, Vol. 58, 297-307.

[4] Bansal R.P., A. R. Gallant, R. Hussey and G. Tauchen (1995), "Nonparametric estimation of structural models for high-frequency currency market data" *Journal of Econometrics*, 66, 251-287.

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

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

[7] Choirat C., Hess C., and R. Seri (2003), "A Functional Version of the Birkhoff Ergodic Theorem for a Norma Integrand: A Variational Approach", *The Annals of Probability*, Vol. 31, No 1, pp. 63-92.

[8] Chumacero R. A. (2001), "Estimating ARMA Models Efficiently", *Studies in Nonlinear Dynamics and Econometrics*, 5, 103–114.

[9] Chung, C. and G. Tauchen (2001), "Testing target-zone models using efficient method of moments" *Journal of Business and Economic Statistics*, 19, 255–269.

[10] Demos A. and D. Kyriakopoulou (2008), "Edgeworth expansions for the MLE and MM estimators of an MA(1) model", *Communications in Statistics-Theory and Methods*, forthcoming.

[11] Dupacova J. and Wets, R. J.-B. (1988). "Asymptotic Behavior of Statistical Estimators and of Optimal Solutions of Stochastic Optimization Problems", *The Annals of Statistics*, 16, 1517-1549.

[12] Gallant R.A., D. Hsieh and G. Tauchen (1997), "Estimation of stochastic volatility models with diagnostics", *Journal of Econometrics*, 81, 159-192.

[13] Gallant R.A. and J.R. Long (1997) "Estimating stochastic differential equations efficiently by Minimum Chi- Squared", *Biometrika*, 84, 125-141.

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

[15] Garcia R., E. Renault and D. Veredas (2011) "Estimation of stable distributions by indirect inference", *Journal of Econometrics*, 161, 325-337.

[16] Ghysels I., L. Khalaf and C. Vodounou (2003), "Simulation based inference in moving average models", *Annales d' Econommie et de Statistique*, 69, 85-99.

[17] Gourieroux C.,and A. Monfort (1996), "*Simulation-based Econometric Methods*", CORE Lectures, Ox. Un. Press.

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

[19] Gourieroux C., P.C.B. Phillips and J. Yu (2010), "Indirect inference for dynamic panel models", *Journal of Econometrics*, 157, 68 77.

[20] Hess, C. (1996). "Epi-convergence of Sequences of Normal Integrands and Strong Consistency of the Maximum Likelihood Estimator", *The Annals of Statistics*, 24, 1298-1315.

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

[22] Lachout P., E. Liebsher, and S. Vogel, 2005, "Strong Convergence of Estimators as ε_n-Minimisers of Optimization Problems", *Ann. Inst. Stat. Math.*, Vol. 57-2, pp. 291-313.

[23] Michaelides A. and S. Ng (2000), "Estimating the rational expectations model of speculative storage: A Monte Carlo comparison of three simulation estimators" *Journal of Econometrics*, 96, 231-266.

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

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

[26] Phillips P.C.B. (2011), "Folklore theorems, implicit maps and Indirect Inference" *Econometrica*, Vol. 80, No 1, 425-454.

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

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

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

[30] van der Vaart, Aad W. and J.A. Wellner, "*Weak Convergence and Empirical Processes*", Springer, 2000.