ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF ECONOMICS

WORKING PAPER SERIES **06-2013**

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

January 28, 2013

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 examine simple examples involving Levy and 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 multistep M-estimators defined in the context of (semi) parametric inference. 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 is usually a (possibly stochastic) distance function evaluated on the auxiliary estimator as well as on some functional approximation of a mapping between the statistical models involved that is termed *binding function*. This is constructed by some limiting argument concerning the auxiliary criterion. The IE is finally defined by minimization of the inversion criterion. This definition is conceptually justified by properties of the binding function that guarantee *indirect identification* and the subsequent use of the *analogy principle*. Given the auxiliary criterion differences between IEs hinge on differences on the distance functions, and/or the approximations of the binding function and/or the optimization errors involved.

Indirect inference algorithms were initially employed in [31], formally introduced by [20], complemented by [16] and extended by [8]. Furthermore econometric applications of these estimators have become increasingly popular. They have been applied to stochastic volatility and equity return models (e.g. [14], [17], [1]), exchange rate models (e.g. [6], and [11]), commodity price and storage models (e.g. [25]), dynamic panel data (e.g. [21]), stochastic differential equation models (e.g. [15] and [19]), and in ARMA models (e.g. [10], [18], [12], and [28]).

In the present paper, we are concerned with the issue of the *existence* of strongly consistent IE allowing for cases where the binding 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 motivation lies in cases where the auxiliary criterion is a quasi likelihood function involving a class of stationary-ergodic volatility processes defined by some GARCH or SV type model that represent the statistical model at hand, and an auxiliary class of stationary-ergodic invertible processes living in the premises of a possibly misspecified analogous model. In such frameworks the limit criterion could assume extended real values due to the possible existence of parameter values that imply non existence of relevant moments. Furthermore since the limit criterion is of the form of a statistical divergence between the two classes of processes the binding function can in principle be multivalued due to the geometry involved. Notice that even when this is true, it is possible to study single valued reductions of it via *measurable selections*. This could however imply stricter conditions for indirect identification. Furthermore since these frameworks are generally suboptimal w.r.t. asymptotic efficiency, actual multi-valuedness *could* lead to efficiency gains.

Our study has the form of a calculus of escalating weak sufficient conditions that enable the *definition* of IE in this framework and the proof of *existence* of strongly consistent ones. First, using mild assumptions on the structure of the auxiliary criterion functions, we are occupied with a weaker than the uniform notion of convergence 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. This form of convergence has been studied in the statistical literature (see among others [13], [22], [24]) and it enables the definition of the binding function and the determination of the limiting relation between this function and the auxiliary estimator.

Then we strengthen our assumptions in order to obtain a form of continuity of the binding function that enables the definition of IE derived from this function and the proof that these exist as appropriate random elements. Finally the imposition of a relatively weak condition of indirect identification on the behavior of the binding function along with the limiting relation already established, enable the proof of the existence of strongly consistent IE via the use of the same limit arguments also used for the pseudo consistency of the auxiliary estimator and the subsequent definition of the binding function. This framework readily enables the description of conditions concerning the behavior of any approximation of the binding function that could also be used for the definition of IE in a similar manner.

Hence we manage to extend the framework for the definition of IE in a threefold manner. We allow for the auxiliary and/or the inversion criteria and\or their appropriate limits to assume extended real values. We study their asymptotic behavior via the use of the weaker known topology associated with convergence of minimizers and we allow for the binding function to be a correspondence that assumes non empty and compact values. This incorporates the definitions used in the existing literature but simultaneously generalizes the set of the statistical models that are in accordance with these conventions.

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 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

In the following $\overline{\mathbb{R}}$ denotes the two point compactification of $\mathbb{R}$, equipped with the final topology that makes the relevant inclusion continuous, i.e. the *extended* real line. 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*

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

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 [23]) by $\delta_u : F_0(E) \times F_0(E) \to \overline{\mathbb{R}}$, defined by $\delta_u(A, B) = \inf\{\varepsilon > 0 : B \subset N_\varepsilon(A)\}$ where $N_\varepsilon(A) = \{x \in E : d(x, A) < \varepsilon\}$.[1] 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.*

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 [26], paragraph 1.1, and [23], 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. $F_G = \{\digamma \text{ closed} : \digamma \cap G \neq \varnothing\}$, *$\forall G \in \tau_E$, non empty, and*
2. $F^K = \{\digamma \text{ closed} : \digamma \cap K = \varnothing\}$, *$\forall K \subset E$ non empty and compact.*

From Theorems 4.5.3-5 of [23] we have that when E is locally compact and Hausdorff then $(F_0(E), \mathcal{T}_F)$ is locally compact and, $\digamma_n \to \digamma$ with respect to the Fell topology *iff* $\digamma = \mathrm{Li}\,\digamma_n = \mathrm{Ls}\,\digamma_n$ where $\mathrm{Li}\,\digamma_n$ is the set comprised of the limit points of any possible sequence (x_n) such that $x_n \in \digamma_n$, and $\mathrm{Ls}\,\digamma_n$ 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 [26], or Definition 3.1.4. of [23]). 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 the Hausdorff extended metric defined via a symmetrization of δ_u i.e. $\delta(A, B) = \max\{\delta_u(A, B), \delta_l(A, B)\}$ where $\delta_l(A, B) \doteqdot \delta_u(B, A)$. In this case we can prove the following lemma.

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

Lemma 2.4 *If E is compact and metrized by d, then δ_u is a lower semicontinuous (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 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 ($\mathrm{Level}_{\leq a}(c) \doteqdot \{x \in E : c(x) \leq a\}$ for $a \in \mathbb{R}$) are compact. $\inf$-compactness follows trivially when c is lsc and E is itself compact.

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 $E \times \mathbb{R}$. Iff c is lower semicontinuous (lsc) we have that due to Proposition A.2 of [26], $\mathbf{epi}(c) \in F(E \times \mathbb{R})$ 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. Furthermore the relevant set of lsc is closed w.r.t. the Fell topology. 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 [26] along with Theorem 7.1.4 of [23], Definition D.1 and Proposition D.2 of [26]).

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 $F_0(E)$ (i.e. a closed valued multifunction with domain the set Ω), when this is considered as a relation in $\Omega \times E$. The benefit of not directly working with the underlying function, is the fact that we can consider the graph of the correspondence as the set $\{(x, y) : y \in c(x)\}$ which resides in $\Omega \times E$ instead of the set $\{(\omega, \digamma) : \digamma = c(\omega)\}$ inside $\Omega \times F_0(E)$. When $c(x)$ is compact for any x, then the correspondence in obviously termed as compact valued. In the following we do not make explicit distinction between the correspondence and the underlying multifunction.

The Borel σ-algebra on $F_0(E)$ generated by $\mathcal{T}_F$ will be abbreviated by $\mathcal{B}(E_F)$ and is usually termed *Effros algebra (see Paragraph 1.1 of [26])*. If $(\Omega, \mathcal{J})$ is a measurable space, then c is a random closed set *iff* $\{\omega \in \Omega : c(\omega) \in \overline{\digamma}\} \in \mathcal{J}$ for any $\overline{\digamma} \in \mathcal{B}(E_F)$. Analogously we abbreviate by $\mathcal{B}(E_U)$ the Borel σ-algebra on $F_0(E)$ generated by $\mathcal{T}_U$ and by $\mathcal{B}(E_U \times E_F)$ the Borel σ-algebra on $F_0(E) \times F_0(E)$ generated by the product topology described in lemma 2.4. Finally denote with $\mathcal{B}(\overline{\mathbb{R}})$ the Borel σ-algebra of the extended real numbers with respect to the usual topology.

Lemma 2.5 *If E is compact, separable and metrized by d, then δ_u is $\mathcal{B}\left(\overline{\mathbb{R}}\right)/\mathcal{B}\left(E_U\right)\otimes\mathcal{B}\left(E_F\right)$ measurable.*

Proof. The separability of E implies the separability of $\left(F_0\left(E\right),\mathcal{T}_U\right)$ and $\left(F_0\left(E\right),\mathcal{T}_F\right)$ for if $\{x_n, n=0,1,\ldots\}$ is dense in E then the countable subset of $F_0\left(E\right)$, $\{\{x_n\}, n=0,1,\ldots\}$ intersects any basic open set w.r.t. to either topology. This implies the separability of $F_0\left(E\right)\times F_0\left(E\right)$ when equipped with the topology discussed in lemma 2.4. This in turn implies that the Borel σ-algebra w.r.t. to the product topology on $F_0\left(E\right)\times F_0\left(E\right)$ coincides with $\mathcal{B}\left(E_U\right)\otimes\mathcal{B}\left(E_F\right)$ by Lemma 1.4.1. of [33]. The rest follows by lemma 2.4 along with the fact that the sets in the subbase of the upper topology of $\overline{\mathbb{R}}$ generate $\mathcal{B}\left(\overline{\mathbb{R}}\right)$.[2] ∎

3 Assumptions and Main results

General Set Up

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*. Let also $\mathcal{B}\left(\Theta\right)$, $\mathcal{B}\left(B\right)$ denote the corresponding Borel algebras.

The *auxiliary criterion* is a function $c_n\left(\omega,\theta,\beta\right):\Omega\times\Theta\times B\to\overline{\mathbb{R}}$, that is of the form $q_n\left(y_n,\beta\right)$ 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, usually homeomorphic to $\mathbb{R}^{mn}$ for some $m>0$. q_n reflects part of the structure of an *auxiliary model*, a statistical model defined on the measurable space $(K_n,\mathcal{F}_{K_n})$, with B as its parameter space (e.g. it can be a likelihood function or a GMM type criterion-see section 4).[3] $y_n\left(\cdot,\theta\right)$ is measurable for any θ, and thereby *represents* the *underlying statistical model* which is essentially the set $\{P\circ y_n^{-1}\left(\cdot,\theta\right),\theta\in\Theta\}$. These two models need not coincide.

We abbreviate with P **a.s.** any statement that concerns elements of $\mathcal{J}$ of unit probability. When not nessesary we avoid notating the potential dependence of those elements on the parameters. 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.

In the following we provide with an escalating description of a set of sufficient conditions that enable *first*, the existence of the auxiliary estimator, *second* the construction of the binding function and the description of the asymptotic relation between the two, *third* an appropriate form of continuity of the binding function which along with the previous enable the definition of the IE and *finally* consistency.

Definition and Existence of the Auxiliary Estimator

We begin with a sufficient weak assumption on the behavior of c_n that enables the definition and the existence of the auxiliary estimator. It comprises of a joint measurability condition along with a pointwise w.r.t. θ and P **a.s.** w.r.t. ω continuity and some condition concerning the facilitation of minimization. All these conditions are weak enough so that their verification to be easy in many cases.

Assumption A.1 *Let the following hold:*

[2] It is also possible to prove that in the context of separability $\mathcal{B}\left(E_U\right)=\mathcal{B}\left(E_F\right)$.

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

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

2. $c_n\left(\omega,\theta,\cdot\right):B\rightarrow\overline{\mathbb{R}}$ *is* lsc *and proper* P **a.s.,** $\forall\theta\in\Theta$.[4]

In our examples presented in section 4, the issue of *joint* measurability is handled easily due to the fact that the c_n's considered are in fact Caratheodory functions, i.e. jointly continuous (w.r.t. (θ,β)) and pointwise measurable. Separability of $\Theta\times B$ and lemma 4.51 of Aliprantis and Border [2] implies the required measurability. Properness is an ad hoc consideration that is easily established in many cases. For instance, when c_n has the form of a quasi likelihood function it P **a.s.** does not attain extended values. This is also the case in instances where c_n has the form of a hemimetric as in section 3 due to the compactness of its arguments.

Remark R.1 *The joint measurability and the pointwise semicontinuity imply that* $c_n\left(\cdot,\theta,\cdot\right)$ *is a normal integrand (see Definition 3.5 and Proposition 3.6 in Chapter 5 of [26]). The compactness of* B *then implies that* c_n *is* inf*-compact* P **a.s.** $\forall\theta\in\Theta$.

We are now ready to define and acquire existence of the auxiliary estimator.

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

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

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

Proposition 3.1 *Under assumption A.1* $\beta_n^{\#}\left(\omega,\theta,\varepsilon_n\right)$ *is* $\mathcal{B}\left(B_F\right)/\mathcal{J}$*-measurable and* P **a.s.** *non empty-compact valued* $\forall\theta\in\Theta$.

Proof. For any $\theta\in\Theta$ $\beta_n^{\#}\left(\omega,\theta,\varepsilon_n\right)$ is non empty and compact due to A.1.2 and the compactness of B.[5] Then the joint measurability of $c_n\left(\cdot,\theta,\cdot\right)$ due to assumption A.1 along with Proposition 3.10.(iii) in Chapter 5 of [26]) guarantees the measurability for $\inf_B c_n\left(\omega,\beta\right)+\varepsilon_n$. The result follows from the separability of B. ■

Obviously $\beta_n^{\#}\left(\omega,\theta,0\right)=\arg\min_B c_n\left(\omega,\theta,\beta\right)$ 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.[6]

[4] It is obvious that the element of $\mathcal{J}$ of unit probability w.r.t. which A.1 holds can depend on θ.

[5] For the ω for which $\varepsilon_n\left(\omega\right)\neq 0$ non emptyness of $\beta_n^{\#}\left(\omega,\theta,\varepsilon_n\left(\omega\right)\right)$ does not nessesitate A.1.2 due to the properties of the g.l.b.

[6] The fundamental selection theorem (Theorem 2.13 of [26]) implies the existence of a measurable selection, i.e. a $\mathcal{B}\left(B\right)/\mathcal{J}\otimes\mathcal{B}\left(\Theta\right)$-measurable random element $\beta_n^{*}:\Omega\times\Theta\rightarrow\mathbb{R}$ termed as auxiliary selection, defined by

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

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

Assumption A.1 is not sufficient for the construction of the binding function as an appropriate limit of the auxiliary correspondence. The following assumption facilitates the investigation of the issue of (pseudo-) consistency for the auxiliary correspondence. It indicates the almost sure epiconvergence of the auxiliary criterion to a proper, semicontinuous asymptotic counterpart. Hence 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 equiped with the topology of epi-convergence. Analogous assumptions have been used for the establishment of strong consistency of various estimators. See among others [13], [22], [24] and [27].

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 [24] the analogous sequential characterization dictates that for any θ, P* **a.s.** *for any β:*

1. *$\liminf_{n\to\infty} c_n(\omega, \theta, \beta_n) \geq c(\theta, \beta)$ for all β_n such that $\beta_n \to \beta$, and*
2. *$\limsup_{n\to\infty} c_n(\omega, \theta, \beta_n) \leq c(\theta, \beta)$ 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 since by the local compactness of $B \times \mathbb{R}$ and the subsequent coincidence of convergence w.r.t. the Fell topology with the Painleve-Kuratowski one, the epigraph of $c(\theta, \cdot)$ is closed. In the case that $c_n = \frac{1}{n}\sum_{i=1}^{n} m_i(\omega, \theta, \beta)$ for $(m_i(\theta, \beta))_{i\in\mathbb{Z}}$ is ergodic for any (θ, β), then the assumed epiconvergence would follow if for any θ there exists an open cover of B, so that $E\left|\inf_{\beta\in A} m_0(\omega, \theta, \beta)\right| < +\infty$, for any A in the cover (condition $\mathcal{C}_0$ and Theorem 2.3 of [9]). In the considered cases c is proper since it is either an expectation that cannot assume the value $-\infty$, and there exists at least some parameter value for which it is finite, or it is defined by composition with a (hemi-) metric, and assumes the value 0 for at least one parameter value (see, for example, [32] 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 for the former or section 3 for the latter case). $\inf$-compactness follows from the compactness of Θ. Finally, assumptions A.1, A.2 along with theorem 2.3.5 of [26], the separability and sequential completeness of B and the completeness of the underlying probability space imply that for any θ P **a.s.** *for any β:*

1. *$\liminf_{n\to\infty} c_n(\omega, \theta, \beta_n(\omega)) \geq c(\theta, \beta)$ for all measurable β_n such that $\beta_n \to \beta$ P* **a.s.,** *and*
2. *$\limsup_{n\to\infty} c_n(\omega, \theta, \beta_n(\omega)) \leq c(\theta, \beta)$ for some measurable β_n such that $\beta_n \to \beta$ P* **a.s.**[7]

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 $\forall \theta \in \Theta$.*

[7]This formulation implies that the elements of $\mathcal{J}$ of unit probability for which this convergence holds generally depend on θ and the choice of the stochastic B-valued sequence.

Proof. It follows from R.2. ■

Both the auxiliary and the binding correspondence, will be used for the definition of the IE via some intuition that utilizes an analogy principle. The following result explores their asymptotic relation. Its first and last implications are already known. Its second implication is a partial generalization of Theorems 7.30, 7.32 of [30] in our setting. The first establishes the upper pseudo-consistency and the other two the Fell pseudo-consistency of the auxiliary to the binding correspondence.

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

1. *for any ε_n such that $\varepsilon_n \to 0$ P* **a.s.** *then $\delta_u\left(b(\theta), \beta_n^{\#}(\theta, \varepsilon_n)\right) \to 0$ P* **a.s.**,
2. *there exists a non negative random variable, ε_n^* such that $\varepsilon_n^* \to 0$ P* **a.s.** *and $\delta\left(b(\theta), \beta_n^{\#}(\theta, \varepsilon_n)\right) \to 0$ P* **a.s.**,
3. *if $b(\theta)$ is singleton then for any ε_n such that $\varepsilon_n \to 0$ P* **a.s.** *then $\delta\left(b(\theta), \beta_n^{\#}(\theta, \varepsilon_n)\right) \to 0$ P* **a.s.**

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

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

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

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 $\mathcal{B}(B_F)/\mathcal{J} \otimes \mathcal{B}(\Theta)$-measurable correspondence. Hence $\mathbf{epi}_n(\omega, \cdot)$ is an $\mathcal{B}(B_F)/\mathcal{J}$-measurable correspondence due to the measurability of the relevant projection. Now due to A.2 (see section 2) we have that for large n, and for all ω in an element of $\mathcal{J}$ of unit probability $\mathbf{epi}_n(\omega, \theta) \cap B \times (-\infty, a) \neq \varnothing$ since $B \times (-\infty, a)$ is open in the relevant product topology. Hence $\inf_B c_n(\omega, \theta, \beta) \leq \inf_B c(\theta, \beta)$ for all ω described previously. ■

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

Lemma 3.5 *Under assumptions A.1, A.2 there exists a sequence of random variables defined on Ω, say a_n^*,*[8] *such that $a_n^* \to \inf_B(c(\theta, \beta))$ P* **a.s.** *and*

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

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 $b(\theta)$. For some real sequence $q_n \to 0$, consider $\{y_i, i = 1, ..., m(q_n)\}$ and extract analogously random sequences $(y_{i,n})$ such that $y_{i,n} \to y_i$ P **a.s.** Define $a_{n,y_i} = c_n(\theta, y_{i,n})$ and $n^*(n) = \min\left\{n' : |a_{n',y_i} - a_\theta| \leq \frac{q_n}{2}, P \textbf{ a.s.} \text{ for all } i = 1, ..., m(q_n)\right\}$ which well defined due to Egoroff's theorem and the fact that $m(q_n)$ is finite. Obviously n^* is non decreasing in n. Then define $a_n^* = \max\{a_{n^*,y_i}, i = 1, ..., m(q_n)\}$ which is measurable and

[8] It is obvious from the proof that a_n^* depend also on θ.

$a_n^* \to a_\theta$ P **a.s.** Then for $x \in b(\theta)$, it follows that 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 Proposition 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.**

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

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

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

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 $b(\theta) = \{y\}$. ■

The proof of lemma 3.5 implies that the sequence $(\varepsilon_n^*)_{n \in \mathbb{N}}$ that appears in proposition 3.3.2 is non unique. However the fact that this implication does not hold for any sequence of non negative random variables that P **a.s.** converge to zero is the cause of that in what follows we can only prove the *existence* of strongly consistent indirect estimators among the set of the ones to be defined.

Upper Hemi Continuity of the Binding Correspondence

The proposition 3.3 enables the use of $\delta_u\left(b(\theta), \beta_n^{\#}\right)$ as the inversion criterion. The following assumption concerns the upper continuity of the binding correspondence which along with the relevant properties of δ_u would imply the analogous continuity property for the particular inversion criterion and thereby *facilitate* the issue of existence and consistency of the IE to be defined.

Assumption A.3 *b is upper hemicontinuous, i.e. for any θ and $\theta^* \to \theta$, $\delta_u(b(\theta), b(\theta^*)) \to 0$.*

The following proposition 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 be *continuous* on Θ. Notice that its requirements are *also* stricter w.r.t. measurability compared to the ones in assumption A.2 since the former requires that for any θ the relevant set of P unit probability does not depend on the sequence that converges to θ.

Proposition 3.6 *Suppose that for any θ, P **a.s.** for any ω, any $\theta_n \to \theta$ and any β:*

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

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

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 and large enough n, 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 [26] implies that the composite mapping $\theta \to c(\theta,\cdot) \to \arg\min_B c(\theta,\beta)$ is appropriately continuous. ■

Remark R.3 *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 $c(\theta,\beta)$. Since we allow c_n and/or c to assume extended values, the relevant notion of uniform convergence must also be extended as in definition 7.12 of [30]. The following lemma provides with a set of even weaker sufficient conditions than extended jointly uniform P **a.s.** convergence when c_n has the form of an arithmetic mean w.r.t. statonary and ergodic processes.*

Lemma 3.7 *Suppose that $c_n(\omega,\theta,\beta) = \frac{1}{n}\sum_{i=1}^{n} m_i(\omega,\theta,\beta)$, $(m_i(\theta,\beta))_{i\in\mathbb{Z}}$ is ergodic for any (θ,β), $c_n(\omega,\cdot,\cdot) : \Theta\times B \to \overline{\mathbb{R}}$ is jointly continuous P **a.s.**, there exists a finite open cover of $\Theta\times B$, such that $E\left|\inf_{(\theta,\beta)\in A} m_0(\omega,\theta,\beta)\right| < +\infty$, for any A in the cover, $E(m_0(\omega,\theta,\beta^*))$ assumes values in $\overline{\mathbb{R}}$ for any $\theta\in\Theta$ and β^* in a countable dense subset of B. Then proposition 3.6 holds.*

Proof. Condition 3.6.1 follows from the fact that the assumption framework of the lemma implies condition $\mathcal{C}_0$ and thereby Theorem 2.3 of [9], which implies the joint P **a.s.** epiconvergence of c_n to Em_0. For 3.7.2 notice that the separability of B and the P **a.s.** continuity of m_0 implies the existence of a countable dense B^* such that for any β and any $\delta > 0$ there exists a $\beta^* \in B^*$ such that

$$\begin{aligned}
\limsup_{n\to\infty} c_n(\omega,\theta_n,\beta) &\leq \limsup_{n\to\infty}\left(c_n(\omega,\theta_n,\beta) - c_n(\omega,\theta_n,\beta^*)\right) + \limsup_{n\to\infty} c_n(\omega,\theta_n,\beta^*) \ \ P \textbf{ a.s.} \\
&\leq \delta + \limsup_{n\to\infty} c_n(\omega,\theta_n,\beta^*) \ \ P \textbf{ a.s.}
\end{aligned}$$

By assumption the subset of Ω of P unit probability can be chosen independent of δ and δ can be chosen arbitrarily small. Hence, 3.6.2 would be implied for $\beta_n = \beta$, if for any, θ, $\beta^* \in B^*$, P **a.s.** and any $\theta_n \to \theta$, $\limsup_{n\to\infty} c_n(\omega,\theta_n,\beta^*) \leq Em_0(\theta,\beta^*)$ (i) due to the countability of B^*. Notice that $((m_i(\theta,\beta^*)\wedge -\rho)\vee\rho)_{i\in\mathbb{Z}}$ is also stationary-ergodic for any $\rho > 0$ (see for example Proposition 2.1.1. of [32]), hence the uniform version of Birkhoff's LLN implies that

$$\frac{1}{n}\sum_{i=1}^{n}\left((m_i(\omega,\theta_n,\beta^*)\wedge -\rho)\vee\rho\right)$$

converges P **a.s.** to $E(m_0(\omega,\theta,\beta^*)\wedge -\rho)\vee\rho = (Em_0(\omega,\theta,\beta^*)\wedge -\rho)\vee\rho$ for any $\rho > 0$. Due to the separability of $\mathbb{R}$ the subset of Ω of P unit probability can be chosen independent of ρ. Hence from definition 7.12 of [30] we obtain (i). ■

This lemma explores sufficient conditions for the required continuity of b solely via restrictions on the behavior of c_n which in applications is generaly more analyticaly tractable than c. Furthermore it combines joint epi-convergence with pointwise (on B) extended uniform (w.r.t. θ) almost sure convergence. Finally notice that analogous result would also hold if ergodicity is replaced by any kind of mixing condition that would justify the LLNs used in the previous proof or implied in remark R.3.

Definition, Existence and Consistency of the Indirect Estimator

We are now ready to define the IE and explore the issues of its existence and consistency. Proposition 3.3 along with the measurability of the auxiliary correspondence and the upper hemi continuity of b facilitate the use of $\delta_u\left(b\left(\theta\right), \beta_n^{\#}\left(\theta_0\right)\right)$ for some distinguished $\theta_0 \in \Theta$, for the definition of the IE and the subsequent existence argument. Again an almost surely non-negative random variable will assume the role of the "optimization error" in this second step of the estimation procedure.

Definition D.6 *The indirect correspondence $\theta_n^{\#}\left(\omega, \theta_0, \varepsilon_n, \varepsilon_n^{\#}\right)$ satisfies*

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

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

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.

Proposition 3.8 *Under assumptions A.1 and A.3 $\theta_n^{\#}$ is $\mathcal{B}\left(\Theta_F\right)/\mathcal{J}$-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 $\beta_n^{\#}\left(\theta_0, \varepsilon_n\right)$ is independent of θ and $b\left(\theta\right)$ is independent of ω, we obtain that $\delta_u\left(b\left(\theta\right), \beta_n^{\#}\left(\theta_0, \varepsilon_n\right)\right)$ is $\mathcal{B}\left(\overline{\mathbb{R}}\right)/\mathcal{B}\left(\Theta_F\right) \otimes \mathcal{J}$-measurable. Due to 2.3 and 3.1 $\delta_u\left(b\left(\theta\right), \beta_n^{\#}\left(\theta_0, \varepsilon_n\right)\right)$ is P **a.s.** lsc and therefore a normal intergrand. It is also P **a.s.** proper due to the fact that b and $\beta_n^{\#}$ are P **a.s.** compact valued. Hence the result follows from proposition 3.1 where $c_n = \delta_u\left(b\left(\theta\right), \beta_n^{\#}\left(\theta_0, \varepsilon_n\right)\right)$, 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 [26]) would also enable the definition of the IE as a measurable function with values in Θ. Having established existence we turn to the issue of consistency. We need an assumption of indirect identification that is essentialy derived from the form of the roots of the hemimetric used. Notice that this assumption along with the proof of the following proposition *justifies the* definition of the D.6 by an analogy principle. Mathematicaly both the definition and the existence argument do not require the following assumption in order to be valid.

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

Remark R.4 *This condition is weaker that a condition of the form "If $\theta \neq \theta_0 \Rightarrow b\left(\theta_0\right) \cap b\left(\theta\right) = \varnothing$" and stronger that a condition of the form "If $\theta \neq \theta_0 \Rightarrow b\left(\theta_0\right) \neq b\left(\theta\right)$". 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. This also makes evident the claim that if the auxiliary estimator is defined by a measurable selection of $\beta_n^{\#}$ the corresponding identification condition would not be weaker than the one above.*

The main result of the current section follows. It merely concerns the *existence* of strongly consistent IE inside the established framework. Denote by $\theta(\theta_0)$ 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 the hemi continuity of b and lemma 2.2. Obviously, $\theta_0 \in \theta(\theta_0)$, while $\{\theta_0\} = \theta(\theta_0)$ iff assumption A.4 holds. Since we once again are dealing with compact valued correspondences, convergence is metrized by δ_u and/or δ.

Proposition 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. ■

If assumption A.4 does not hold then the implications of 3.9.1-2 correspond to the fact that the statistical model is only indirectly *set identified* given this framework. They are trivial when $\theta(\theta_0) = \Theta$ and the closer to zero $\delta_u(\{\theta_0\}, \theta(\theta_0))$ is the more informative they become. We once again point out that the implications 3.9.1 and 1^* merely explore the issue of the existence of strongly consistent estimators among those that comply with definition D.6. The properties of the δ_u function along with proposition 3.3 do not permit for a stronger result without strengthening the assumption framework. Finally notice that this framework enables both the definition and the result on consistency of the IE to be derived via the use of exact same notions that were used for the analogous results concerning the auxiliary one.

Extension

In most cases b is analyticaly unknown even if several of its properties, such as some of the ones discussed above can be established. In these cases the estimators defined in D.6 are obviously infeasible and possibly stochastic and algorithmicaly feasible approximations are used for the construction of several other classes of feasible IE. In such a context the results derived previously enable us to describe properties of such approximations that would imply that these estimators are well defined and among them there exist strongly consistent ones. Let $\kappa_n(\omega,\theta)$ denote such an approximation. We readily obtain the following result for the indirect estimator defined by the substitution of b with κ_n in D.6.

Proposition 3.10 *Consider the* IE *defined by*

$$\vartheta_n^{\#}\left(\omega,\theta_0,\varepsilon_n,\varepsilon_n^{\#}\right)=\varepsilon_n^{\#}\text{-}\arg\min_{\Theta}\delta_u\left(\kappa_n(\omega,\theta),\beta_n^{\#}(\theta_0,\varepsilon_n)\right)$$

where $\varepsilon_n,\varepsilon_n^{\#}$ *as before.*

i. under assumption A.1 and if $\kappa_n(\omega,\theta)$ *is* $\mathcal{B}(B_F)/\mathcal{J}\otimes\mathcal{B}(\Theta_F)$*-measurable,* P **a.s.** *compact valued and upper hemicontinuous then* $\vartheta_n^{\#}$ *is* $\mathcal{B}(\Theta_F)/\mathcal{J}$*-measurable,* P **a.s.** *non empty, compact valued correspondence, and*

ii. if moreover A.2 holds and P **a.s.** *for any* θ *and any* $\theta_n\to\theta$*,* $\delta(b(\theta),\kappa_n(\omega,\theta))$ *and* $\delta_u(b(\theta),\kappa_n(\omega,\theta_n))$ *converge to zero then the implications 3.10.1-2 hold also for* $\vartheta_n^{\#}$*. If furthermore assumption A.4 then then the implications 3.10.1*-2* hold also for* $\vartheta_n^{\#}$**.**

Proof. i. As in the proof of 3.8 from 2.5, 3.1 the P **a.s.** upper hemicontinuity of κ_n the facts that $\beta_n^{\#}(\theta_0,\varepsilon_n)$ is independent of θ and κ_n is jointly measurable, we obtain that $\delta_u\left(\beta_n^{\#}(\theta_0,\varepsilon_n),b(\theta)\right)$ is $\mathcal{B}\left(\overline{\mathbb{R}}\right)/\mathcal{B}(\Theta_F)\otimes\mathcal{J}$-measurable. Due to 2.3 and 3.1 $\delta_u\left(\kappa_n(\omega,\theta_n),\beta_n^{\#}(\theta_0,\varepsilon_n)\right)$ is P **a.s.** lsc and therefore a normal intergrand. It is also P **a.s.** proper due to the fact that κ_n and $\beta_n^{\#}$ are P **a.s.** compact valued. Hence the result follows from proposition 3.1. ii. It suffices to prove that $\delta_u\left(\kappa_n(\omega,\theta),\beta_n^{\#}(\theta_0,\varepsilon_n)\right)\xrightarrow{e}\delta_u(b(\theta),b(\theta_0))$ P **a.s.** when $\beta_n^{\#}(\theta_0,\varepsilon_n)$ *satisfies the implications 2 or 3 of proposition* 3.3. The rest would then follow as in the proof of proposition 3.10. Notice that for any θ and any $\theta_n\to\theta$

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

due to lemma 2.4, the definition of $\beta_n^{\#}$ and the P **a.s.** continuous w.r.t θ upper convergence of κ_n to b. Moreover for $\theta_n=\theta$

$$\begin{aligned}&\limsup_n\delta_u\left(\kappa_n(\omega,\theta),\beta_n^{\#}(\theta_0,\varepsilon_n^{*})\right)\\\leq\ &\limsup_n\delta_u(\kappa_n(\omega,\theta),b(\theta))+\limsup_n\delta_u\left(b(\theta),\beta_n^{\#}(\theta_0,\varepsilon_n^{*})\right),\ P\ \textbf{a.s.}\\\leq\ &\limsup_n\delta(\kappa_n(\omega,\theta),b(\theta))+\delta_u(b(\theta),b(\theta_0)),\ P\ \textbf{a.s.}\end{aligned}$$

where the first inequality follows from the triangle inequality and the second from the the definition of δ and the P **a.s.** Fell convergence of $\beta_n^{\#}$ to $b(\theta_0)$ and lemma 2.3. Due to the P **a.s.** pointwise w.r.t. θ Fell convergence of κ_n to b we have that $\limsup_n\delta(\kappa_n(\omega,\theta),b(\theta))=0$ P **a.s.** and therefore we obtain the needed result. ■

Notice that this proposition generalizes the results of propositions 3.8, 3.10. For a simple example consider the case where $\kappa_n=\beta_n^{\#}$. This is possible when by some sort of resampling technique (e.g. bootstrap or Monte Carlo) realizations of the y_n random elements are available to the practitioner for any θ, and thereby so is $\beta_n^{\#}(\theta, \epsilon_n)$ for any θ and some optimization error ϵ_n *independent* of θ. Then a feasible IE can be defined by the approximate minimization of $\delta_u\left(\beta_n^{\#}(\theta, \epsilon_n), \beta_n^{\#}(\theta_0, \varepsilon_n)\right)$ w.r.t. θ. In this case the joint measurability of $\beta_n^{\#}$ would follow from the joint measurability of c_n and ϵ_n, the separability of B and the subsequent joint measurability of the relevant projection. The P **a.s.** upper hemicontinuity of $\beta_n^{\#}$ would follow from an easy extension of the implication 3.3.1 if assumption A.1 is strengthened so that the mapping $\theta \to \mathbf{epi}_n(\omega, \theta)$ is P **a.s.** Fell continuous. Obviously the P **a.s.** joint continuity of c_n would suffice. Then the P **a.s.** pointwise Fell convergence to b would follow as in 3.3.2 or 3 and the P **a.s.** continuous w.r.t θ upper convergence to b would follow if proposition 3.6 holds with the set of unit probability independent of θ. Obviously it would suffice that c_n is P **a.s.** jointly continuous and converges to c jointly uniformly.

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 $\mathbf{rank}(M_{X'X}) = \mathbf{rank}(M_{Z'X}) = p$ and $p \leq l \doteqdot \mathbf{rank}(M_{Z'Z}) \leq q$. For u a $n \times 1$ random vector, let the underlying statistical model be the set of "regressions" $Y(\omega, \theta) = X\theta + u$, $\theta \in \Theta$. 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. Proposition 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 $y \in \mathrm{col}(M_{Z'Z})$, 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 $l = p$. For $M_{X'u} \in \mathbb{R}^p$, $M_{Z'u} \in \mathbb{R}^q$, $M_{u'u} \in \mathbb{R}$ 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

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

which implies both assumptions A.2, A.3 (via lemma 3.6 and R.3).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 $\mathcal{P}(M_{Z'X}\theta_0 + M_{Z'u}) - \mathcal{P}(M_{Z'X}\theta + M_{Z'u}) \notin K^{-1}H_{q-l}$ for any $\theta \neq \theta_0$ then assumption A.4 applies.[9] Hence, proposition 3.9.1* implies the *existence* of a consistent IE for $\theta_0 \in \Theta$. In the special case where $X = Z$ and $M_{X'u} \in M_{X'X}(B - \theta)$ then $b(\theta) = \left\{\theta - M_{X'X}^{-1}M_{X'u}\right\}$ and lemma 3.9.2* implies that any IE defined by D.6 can be perceived as an "inconsistency corrector" of the underlying OLSE for θ.□

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 Bownian motion 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 [3]). For $p = 1$ consider the stochastic process on $\mathbb{R}^+$ defined by the following Levy-Ito decomposition (see Theorem 2.4.16 of [3])

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

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 [3]. 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. Proposition 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 $N(t, [C_1, C_2]) - N(t-1, [C_1, C_2]) \overset{i.i.d.}{\sim} \mathrm{Poiss}(v([C_1, C_2]))$ 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

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

[9] More precicely we have that $b(\theta) \cap b(\theta_0) = \varnothing$.

for

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

Due to the definition of X the process y is iid 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

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

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

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

In this case assumption A.4 applies and therefore proposition 3.9.2* implies that any IE defined by D.6 is consistent for any $\theta_0 \in \Theta$. When $v\left([C_1, C_2]\right) = 0$ (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 [19]).□

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 $\sigma^2 : \Theta \times \Omega \to \left(\mathbb{R}^+\right)^{\mathbb{Z}}$, with the product space $\Theta \times \Omega$ equipped with $\mathcal{B}\left(\Theta\right) \otimes \mathcal{J}$ with $\sigma_t^2\left(\theta\right)$ independent of $\left(z_i\right)_{i \geq t}$, $\forall t \in \mathbb{Z}$, $\forall \theta \in \Theta$. Analogously, define the random element $y : \Theta \times \Omega \to \left(\mathbb{R}\right)^{\mathbb{Z}}$ as

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

Then $\forall \theta \in \Theta$, $\left(y_t\left(\theta\right)\right)_{t \in \mathbb{Z}}$ is called a conditionally heteroskedastic process, while the random element $\left(y_t\left(\omega\right)\left(\theta\right)\right)_{t \in \mathbb{Z}, \theta \in \Theta}$ a conditionally heteroskedastic model. Our examples will solely concern ergodic heteroskedastic models.[10]

Example IV **Estimation in Regressions on Squared** $\mathrm{ARCH}(1)$ **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

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

Due to the fact that $\sigma_4 > 1$ Theorem 5.2.1. of [32] implies that for any $\alpha \in \Theta$ the equation admits a unique stationary and ergodic solution defining the analogous $\mathrm{ARCH}\left(1\right)$ process. Consider the random vector $Y\left(a\right) = \left(y_t^2\left(a\right)\right)_{t \in \{1, \ldots, n\}}$, and the $n \times 2$ dimensional random matrices

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

jointly measurable with respect to $\mathcal{J} \otimes \mathcal{B}\left(\Theta\right)$, 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 $c_n\left(\omega, a, \beta\right) = \left\|\frac{1}{n}Z'\left(a\right)\left(Y\left(a\right) - X\left(a\right)\beta\right)\right\|$ which

[10] The establishment of the ergodicity is initiated by the analogous establishment for $\left(\sigma_t^2\left(\theta\right)\right)_{t \in \mathbb{Z}}$ $\forall \theta \in \Theta$. Sufficient conditions for that are described and employed in a variety of heteroskedastic models in chapter 5 of [32] via Theorem 5.2.1. Then the ergodicity of $\left(y_t\left(\theta\right)\right)_{t \in \mathbb{Z}}$ and $\left(y_t^2\left(\theta\right)\right)_{t \in \mathbb{Z}}$ $\forall \theta \in \Theta$ follow from the definition of z, y, the previous assumption and Proposition 2.2.1 of [32].

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. This consideration is motivated from the $\mathrm{AR}(1)$ representation of the $\mathrm{ARCH}(1)$ process with respect to the martingale difference noise $v_t = (z_t^2 - 1)\sigma_t^2(a)$ (see, for example, [7]) 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 $i \leq t$. Proposition 3.1 ensures 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 $\mathrm{ARCH}(1)$ model and the definitions of Θ and σ_4 we have that

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

for $C_1, C_2 > 0$ which along with (the uniform version of) Birkhoff's Ergodic Theorem (see for example [32], Theorem 2.2.1) implies both assumptions A.2, A.3 (via proposition 3.6 and R.3) 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

$$b(a) = \begin{cases} \left\{\begin{pmatrix} 1 \\ a \end{pmatrix}\right\}, & a \in (0, \delta] \\ \left\{(\beta_1, 1-\beta_1), \beta_1 \in [1, 1+\delta]\right\}, & a = 0 \end{cases}$$

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

$$\theta_n^{\#}(\omega, \theta_0, \varepsilon_n^*, 0) = \begin{cases} \{a\} \text{ if } \beta_n \subseteq b(a) \\ \mathrm{mid}\left(\mathrm{proj}_{\beta_2}(\beta_n)\right) \text{ if } \beta_n \nsubseteq b(a) \text{ and } \mathrm{mid}\left(\mathrm{proj}_{\beta_2}(\beta_n)\right) \geq 0 \\ \{0\} \text{ if } \beta_n \nsubseteq b(a) \text{ and } \mathrm{mid}\left(\mathrm{proj}_{\beta_2}(\beta_n)\right) < 0 \end{cases},$$

where proj_{β_2} denotes projection to the β_2-axis and $\mathrm{mid}(A)$ denotes the midpoint of the smallest interval that contains A. Notice that in our case $\mathrm{mid}\left(\mathrm{proj}_{\beta_2}(\beta_n)\right)$ is well defined due to the fact that β_n is P **a.s.** compact valued hence its $\mathrm{proj}_{\beta_2}(\beta_n)$ is a P **a.s.** compact subset of the real line. Finally and due to the fact that bootstrap resampling techniques are readily available in the context of this model, proposition 3.10 implies also the analogous properties for IE defined by κ_n when this equals the auxiliary estimator derived from bootstrap resampling for any a.□

The final example is about an asymmetric heteroskedastic process.

Example $-c_n$ is the Quasi-Likelihood Function of an Approximate to $\mathrm{QARCH}(1)$ Model. Let $E|z_0| < 1$ and $\delta > 0$, and consider for $\gamma \in \Theta = [-\delta, 0]$ $\varpi_0, a_0 > 0$, $\varpi = \varpi_0 + \frac{\delta^2}{4a_0}$, and $a_0 < \exp(-2E\ln|z_0|)$ 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 define a unique stationary and ergodic $\mathrm{QARCH}(1)$ volatility process with existing $\log$ moments that is uniformly bounded from below away from zero (see Lemmas 2.1 and 3.4 and Remark R.2 of Arvanitis and Louka [5]). Notice that Jensen's inequality allows $a_0 \geq 1$ which in turn implies that $E\sigma_t^2(\gamma) = +\infty$. For $\gamma^* \in B = [0, \delta]$ consider the process defined by

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

$h_t(\gamma, \gamma^*)$ is well defined due to the definition of B and it is stationary and ergodic with existing $\log$ moments due to the previous and Proposition 2.1.1. of [32]. Now consider

$$c_n(\omega, \gamma, \gamma^*) \doteqdot \frac{1}{n}\sum_{i=1}^{n}\left(-\ln\frac{y_i^2(\gamma)}{h_i(\gamma, \gamma^*)} + \frac{y_i^2(\gamma)}{h_i(\gamma, \gamma^*)}\right)$$

$-c_n$ 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 [32].[11] Assumption A.1 follows readily from the form of c_n and the P **a.s.** continuity with 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

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

and that

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

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

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

[11] In practice $c_n(\omega, \gamma, \gamma^*)$ is unknown but approximated by an analogous $\widehat{c}_n(\omega, \gamma, \gamma^*)$ 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 [32] 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 [32]), thereby facilitating the asymptotic analysis of minimizers of $\widehat{c}_n(\omega, \theta, \beta)$ by the analogous analysis of minimizers of $c_n(\omega, \theta, \beta)$.

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 [32] to obtain the analogous arguments along with the fact that $\frac{\varpi + a_0 y_0^2}{\varpi + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0<0}} \overset{P\ \textbf{a.s.}}{=} 1$ iff $\gamma^* = 0$). When $\gamma \neq 0$ then $c(\gamma,-\gamma) < c(\gamma,0)$ due to the fact that

$$\begin{aligned}
& -E\ln\frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 - \gamma |y_0| 1_{z_0<0}} + E\ln\frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2} \\
= \; & E\left(\ln\left(1 + \frac{\gamma y_0}{\varpi + a_0 y_0^2}\right) 1_{z_0<0}\right)
\end{aligned}$$

and

$$\begin{aligned}
& E\frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2 - \gamma |y_0| 1_{z_0<0}} - E\frac{\varpi + a_0 y_0^2 + \gamma y_0}{\varpi + a_0 y_0^2} \\
= \; & -E\left(\frac{\gamma y_0}{\varpi + a_0 y_0^2} 1_{z_0<0}\right) < 0
\end{aligned}$$

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

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

and

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

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

as well as dominated convergence, we have that

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

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

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

establishing along with the previous that $b(\gamma) = \{-\gamma\}$. This validates simultaneously both assumptions A.3 and A.4. Hence proposition 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 generalize the definition of IE and are occupied with the questions of existence and strong consistency. We allow 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 binding 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.

First, notice that our framework could still be extended in the following manner. The established results would remain almost intact if the underlying parameter spaces were only locally compact under more restrictive assumptions on the behavior of the criteria involved. In such a case Proposition 4.2.1.(i) of [23] would permit the validity of the results, except for the compactness of the auxiliary and the binding correspondences, under the additional condition that P **a.s.** $c_n(\omega, \theta, \cdot)$, $c(\theta, \cdot)$ and $\delta_u\left(b(\cdot), \beta_n^{\#}(\theta_0, \varepsilon_n^*)\right)$ have totally bounded level sets and non empty $\arg\min$s.

Second, the present generalization in certainly *non unique*. Again under stricter conditions on the behavior of the auxiliary criteria, possibly relevant to the ones in Proposition 3.42 of [29], the implication 3.3.2 could be strengthened to hold *for any* asymptotically null sequence of optimization errors. In this case in the definition of the IE δ_u could be replaced by δ and this would initially allow the identification condition in assumption A.4 to be replaced by the *weaker* "if $\theta \neq \theta_0 \Rightarrow b(\theta_0) \neq b(\theta)$". If assumption A.3 were also strengthened to require Fell continuity then the strong consistency result would be valid *for any* IE defined in this framework. We leave this for future research.

We also leave for future research the questions of the definition and consistency of IE when the κ_n appearing in proposition 3.10 is some sort of integral of $\beta_n^{\#}$ (see for example [4]) or some (possibly) 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. Notice that this limit theory could in principle be quite complex due to complexities in the analogous theory for $\beta_n^{\#}$ and/or to different properties of potential polynomial approximations for different selections of b around θ e.t.c.

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] Aliprantis C.D. and K.C. Border (1999), "*Infinite Dimensional Analysis. A Hitchhiker's Guide.*", Springer.

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

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

[5] Arvanitis S., and Al. Louka (2012), "Limit Theory for the QMLE of the GQARCH(1,1) model.", Discussion Paper Series: 12-12, Dept. of Econ., AUEB (http://www.econ.aueb.gr/DiscussionPapers2012.html).

[6] 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.

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

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

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

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

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

[12] 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.

[13] 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.

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

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

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

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

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

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

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

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

[22] 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.

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

[24] 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.

[25] 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.

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

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

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

[29] Rockafellar T.R. and Wetts J-B (1984), "Variational Systems, an Introduction.", "*Multifunctions and Integrands*", Ed. G. Salinetti, Springer-Verlag, pp. 1-55.

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

[31] Smith A.A. (1993), "Estimating Nonlinear Time-series Models using Simulated Vector Autoregressions.", *Journal of Applied Econometrics,* Volume 8 Issue 1, pp. 63-84.

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

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