

## Research Article

# On the Existence of Strongly Consistent Indirect Estimators When the Binding Function Is Compact Valued

**Stelios Arvanitis**

*Athens University of Economics and Business, Patision 76, 10434 Athens, Greece*

Correspondence should be addressed to Stelios Arvanitis; [stelios@aueb.gr](mailto:stelios@aueb.gr)

Received 2 July 2013; Accepted 5 September 2013

Academic Editor: Mike Tsionas

Copyright © 2013 Stelios Arvanitis. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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

## 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 [1], formally introduced by [2], complemented by [3], and extended by [4]. Furthermore econometric applications of

these estimators have become increasingly popular. They have been applied to stochastic volatility and equity return models (e.g., [5–7]), exchange rate models (e.g., [8, 9]), commodity price and storage models (e.g., [10]), dynamic panel data (e.g., [11]), stochastic dif and only iferential equation models (e.g., [12, 13]), and in ARMA models (e.g., [14–17]).

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 quasilielihood function involving a class of stationary-ergodic volatility processes defined by some GARCH or SV type models 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 nonexistence of relevant moments. Furthermore since the limit criterion is 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 with respect to asymptotic efficiency, actual multivaluedness *could* lead to efficiency gains.

This 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. This is termed epiconvergence and essentially concerns the almost sure asymptotic behavior of their epigraphs. By construction it is suitable for the study of the asymptotic behavior of their minimizers. This form of convergence has been extensively studied in the statistical literature (see among others [18–20]) 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 verification of their existence. 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 enables the proof of the existence of strongly consistent IE via the use of the same limit arguments also used for the pseudoconsistency 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 with values on the collection of nonempty and compact subsets of the relevant parameter space. 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 formulate our setup and 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. In the appendix we briefly describe some general notions that are essentially used in main body.

## 2. Assumptions and Main Results

*2.1. General Setup.* We construct our framework and describe the underlying statistical problem. Let the triad  $(\Omega, \mathcal{F}, P)$  denote a *complete* probability space. Let also  $(\Theta, d_\Theta)$  and  $(B, d_B)$  denote two *compact separable metric spaces*. Let also

$\mathcal{B}(\Theta)$  and  $\mathcal{B}(B)$  denote the corresponding Borel algebras.  $\overline{\mathbb{R}}$  denotes the *extended* real line. Again  $\mathcal{B}(\overline{\mathbb{R}})$  denotes the corresponding Borel algebra.

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

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

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 enables the definition of the IE, and *finally* consistency.

*2.2. 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 with respect to  $\theta$  and  $P$  a.s. with respect to  $\omega$  continuity and some condition concerning the facilitation of minimization. All these conditions are weak enough so that their verification is easy in many cases. Remember that a function with values in the extended real line is called *proper*, if it does not take the value  $-\infty$  and its image contains at least a real number. It is called *inf-compact*, if its level sets (for  $a \in \mathbb{R}$ , the level set of  $c : E \rightarrow \overline{\mathbb{R}}$  with respect to  $a$  ( $\text{Level}_{\leq a}(c)$ ) is defined by  $\{x \in E : c(x) \leq a\}$ ) are compact. *inf-compactness* follows trivially when the function is lsc and its domain is compact.

*Assumption 1.* Let the following holds.

- (1)  $c_n$  is  $\mathcal{B}(\overline{\mathbb{R}})/\mathcal{F} \otimes \mathcal{B}(\Theta) \otimes \mathcal{B}(B)$  measurable.
- (2)  $c_n(\omega, \theta, \cdot) : B \rightarrow \overline{\mathbb{R}}$  is lower semicontinuous (lsc) and proper  $P$  a.s., for all  $\theta \in \Theta$ . (It is obvious that the element of  $\mathcal{F}$  of unit probability with respect to which Assumption 1 holds can depend on  $\theta$ ).

In our examples presented in Section 3, the issue of *joint* measurability is handled easily due to the fact that the  $c_n$ 's considered are in fact Caratheodory functions, that is, jointly continuous (with respect to  $(\theta, \beta)$ ) and pointwise measurable. Separability of  $\Theta \times B$  and Lemma 4.51 of Aliprantis and Border [21] 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 quasilielihood 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 (see Appendix) as in Section 2 due to the compactness of its arguments.

*Remark 2.* The joint measurability and the pointwise semi-continuity imply that  $c_n(\cdot, \theta, \cdot)$  is a normal integrand (see Definition 3.5 and Proposition 3.6 in Chapter 5 of [22]). The compactness of  $B$  then implies that  $c_n$  is inf-compact  $P$  a.s. for all  $\theta \in \Theta$ .

We are now ready to define and prove existence for the auxiliary estimator. We remind the reader that the Appendix provides with the definition of a measurable compact valued correspondence.

*Definition 3.* The auxiliary correspondence  $\beta_n^\#(\omega, \theta, \varepsilon_n)$  satisfies

$$\begin{aligned} \beta_n^\#(\omega, \theta, \varepsilon_n) &= \varepsilon_n\text{-arg min}_B q_n(y_n, \beta) \\ &= \varepsilon_n\text{-arg min}_B c_n(\omega, \theta, \beta) \\ &\doteq \left\{ \beta \in B : c_n(\omega, \theta, \beta) \leq \inf_B c_n(\omega, \theta, \cdot) + \varepsilon_n \right\}, \end{aligned} \quad (1)$$

where  $\varepsilon_n$  is a  $P$  a.s. nonnegative random variable defined on  $\Omega$ .

Notice that the dependence of  $\beta_n^\#$  on  $\theta$  is due to the assumed dependence of  $y_n$  on  $\theta$ . In practice  $\beta_n^\#$  is evaluated by the minimization appearing in the first equality of the previous display.

**Proposition 4.** *Under Assumption 1  $\beta_n^\#(\omega, \theta, \varepsilon_n)$  is  $\mathcal{B}(B_F)/\mathcal{F}$ -measurable and  $P$  a.s. non empty compact valued for all  $\theta \in \Theta$ .*

*Proof.* For any  $\theta \in \Theta$   $\beta_n^\#(\omega, \theta, \varepsilon_n)$  is non empty and compact due to Assumption 1(2) and the compactness of  $B$ . (For the  $\omega$  for which  $\varepsilon_n(\omega) \neq 0$  nonemptiness of  $\beta_n^\#(\omega, \theta, \varepsilon_n(\omega))$  does not necessitate Assumption 1(2) due to the properties of the inf). Then the joint measurability of  $c_n(\cdot, \theta, \cdot)$  due to Assumption 1 along with Proposition 3.10(iii) in Chapter 5 of [22] guarantees the measurability for  $\inf_B c_n(\omega, \theta, \cdot) + \varepsilon_n$ . The result follows from the separability of  $B$ .  $\square$

Obviously  $\beta_n^\#(\omega, \theta, 0) = \arg \min_B c_n(\omega, \theta, \beta)$ . In the following, dependence on  $\Omega$  will henceforth be suppressed (where possible) for notational simplicity. Dependence on  $B$ ,  $\Theta$ , and the ‘‘optimization error’’  $\varepsilon_n$  will be kept. (The fundamental selection theorem (Theorem 2.13 of [22]) implies the existence of a measurable selection, i.e., a  $\mathcal{B}(B)/\mathcal{F} \otimes \mathcal{B}(\Theta)$ -measurable random element  $\beta_n^* : \Omega \times \Theta \rightarrow \mathbb{R}$  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.)

*2.3. Epilimits and Existence of a Fell Consistent Auxiliary Correspondence.* Assumption 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 and thereby permits this construction. It indicates the almost sure epiconvergence of the auxiliary criterion to a proper, semicontinuous asymptotic counterpart. The first part is essentially the sequential characterization of this form of convergence. For the topological definition of epiconvergence along with some properties please see the Appendix. For the equivalence between the topological and the sequential definitions see among others [20]. The second part enforces to the limit criterion a property that are is not generally invariant with respect to epiconvergence.

*Assumption 5.* There exists a function  $c : \Theta \times B \rightarrow \overline{\mathbb{R}}$  such that

- (1)  $\forall \theta \in \Theta, P$  a.s. for any  $\beta$ :
  - (i)  $\liminf_{n \rightarrow \infty} c_n(\omega, \theta, \beta_n) \geq c(\theta, \beta)$  for all  $\beta_n$  such that  $\beta_n \rightarrow \beta$ , and
  - (ii)  $\limsup_{n \rightarrow \infty} c_n(\omega, \theta, \beta_n) \leq c(\theta, \beta)$  for some  $\beta_n$  such that  $\beta_n \rightarrow \beta$ ,
- (2)  $c(\theta, \cdot)$  is proper  $\forall \theta \in \Theta$ .

*Remark 6.* For any  $\theta, c(\theta, \cdot) : B \rightarrow \overline{\mathbb{R}}$  is lower semicontinuous since the local compactness of  $B \times \mathbb{R}$  and the subsequent coincidence of convergence with respect to the Fell topology with the Painleve-Kuratowski one (see Appendix), the epigraph of  $c(\theta, \cdot)$  is closed. In the case that  $c_n = (1/n) \sum_{i=1}^n m_i(\omega, \theta, \beta)$  and  $(m_i(\theta, \beta))_{i \in \mathbb{Z}}$  is ergodic for any  $(\theta, \beta)$ , the assumed epiconvergence would follow if for any  $\theta$  there exists an open cover of  $B$ , so that  $E |\inf_{\beta \in A} m_0(\omega, \theta, \beta)| < +\infty$  for any  $A$  in the cover (condition  $\mathcal{E}_0$  and Theorem 2.3 of [23]). 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, [24] Part 1, (ii) in association with Part 2 of the proof of Theorem 5.3.1, where  $c_n$  is a quasilielihood function and  $\Theta$  coincides with  $B$  for the former or Section 2 for the latter case). inf-compactness follows from the compactness of  $\Theta$ .

The assumption essentially enables the use of the fact that the argmin correspondence is upper continuous as a function defined on the relevant space of lsc functions equipped with the topology of epiconvergence. Analogous assumptions have been used for the establishment of strong consistency of various estimators. See among others [18–20, 25]. Hence it makes possible the definition of the binding function as

$$b(\theta) \doteq \arg \min_B c(\theta, \beta). \quad (2)$$

The following proposition provides its existence and is essentially similar to Proposition 4.

**Proposition 7.** *Under Assumptions 1 and 5 the binding correspondence  $b$  is nonempty compact valued.*

*Proof.* It follows from Remark 6.  $\square$

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. For its construction the hemimetric  $\delta_u$  is needed that is defined by  $\delta_u(A_1, A_2) = \inf\{\varepsilon > 0 : A_2 \subset N_\varepsilon(A_1)\}$  where  $N_\varepsilon(A_1) = \{x \in E : d(x, A_1) < \varepsilon\}$ , where  $d(x, A_1) \doteq \inf_{y \in A_1} d(x, y)$  and  $A_1, A_2$  are nonempty closed subsets of some topological space. For several continuity and measurability properties of  $\delta_u$  see the Appendix.

The first and last implications of the following proposition are already well known. Its second implication is a partial generalization of Theorems 7.30, 7.32 of [26] in our setting. Given the definitions of the upper and the Fell topology in the Appendix, the first essentially establishes the upper pseudoconsistency and the other two the Fell pseudoconsistency of  $\beta_n^\#$  with respect to  $b$ .

**Proposition 8.** *Under Assumptions 1 and 5 the following holds:*

- (1) *for any  $\varepsilon_n$  such that  $\varepsilon_n \rightarrow 0$   $P$  a.s. then  $\delta_u(b(\theta), \beta_n^\#(\theta, \varepsilon_n)) \rightarrow 0$   $P$  a.s.,*
- (2) *there exists a nonnegative random variable,  $\varepsilon_n^*$ , such that  $\varepsilon_n^* \rightarrow 0$   $P$  a.s. and  $\delta(b(\theta), \beta_n^\#(\theta, \varepsilon_n)) \rightarrow 0$   $P$  a.s.,*
- (3) *if  $b(\theta)$  is singleton then for any  $\varepsilon_n$  such that  $\varepsilon_n \rightarrow 0$   $P$  a.s. then  $\delta(b(\theta), \beta_n^\#(\theta, \varepsilon_n)) \rightarrow 0$   $P$  a.s.*

For its proof we will use the following lemmata. Let  $a_\theta \doteq \inf_B c(\theta, \beta)$  which is well defined due to Remark 6 and the compactness of  $B$ . Remember that for  $(F_n)$  a sequence of non empty sets  $\text{Li}F_n$  is the set comprised of the limit points of any possible sequence  $(x_n)$  such that  $x_n \in F_n$ , and  $\text{Ls}F_n$  is the one comprised of the analogous cluster points. Also  $\text{Level}_{\leq a}(c_n(\theta, \cdot)) \doteq \{x \in B : c(\theta, x) \leq a\}$  for  $a \in \mathbb{R}$ .

**Lemma 9.** *Under Assumptions 1 and 5,*

$$\limsup_n \inf_B c_n(\theta, \beta) \leq \inf_B c(\theta, \beta) \text{ } P \text{ a.s.} \quad (3)$$

*Proof.* Consider the family of  $\theta$ -parameterized correspondences  $\text{epi}_n(\omega, \theta) \doteq \text{epi}(c_n(\omega, \theta, \cdot))$ . Due to the fact that  $B$  is locally compact,  $\text{epi}_n(\omega, \theta)$  is a random closed set in the sense of the previous paragraph, that is, a  $\mathcal{B}(B_F)/\mathcal{F} \otimes \mathcal{B}(\Theta)$ -measurable correspondence. Hence  $\text{epi}_n(\omega, \cdot)$  is an  $\mathcal{B}(B_F)/\mathcal{F}$ -measurable correspondence due to the measurability of the relevant projection. Now due to Assumption 5 (see Section 4) we have that for large  $n$  and for all  $\omega$  in an element of  $\mathcal{F}$  of unit probability  $\text{epi}_n(\omega, \theta) \cap B \times (-\infty, a_\theta) \neq \emptyset$  since  $B \times (-\infty, a_\theta)$  is open in the relevant product topology. Hence  $\inf_B c_n(\omega, \theta, \beta) \leq \inf_B c(\theta, \beta)$  for all  $\omega$  described previously.  $\square$

The next result will be used for the proof of Proposition 8(2)-(3).

**Lemma 10.** *Under Assumptions 1 and 5 there exists a sequence of random variables defined on  $\Omega$ , say  $a_n^*$  (It is obvious from the proof that  $a_n^*$  depends also on  $\theta$ ). Such that  $a_n^* \rightarrow \inf_B c(\theta, \beta)$   $P$  a.s. and*

$$\text{Li}(\text{Level}_{\leq a_n^*}(c_n(\theta, \cdot))) \supseteq b(\theta) \text{ } P \text{ a.s.} \quad (4)$$

*Proof.* From the sequential implication of epiconvergence in Remark 6, we have that, for any  $x \in b(\theta)$ , there exists a measurable  $x_n$  such that  $x_n \rightarrow x$   $P$  a.s. Obviously for  $a_{n,x} = c_n(\theta, x_n)$  which is measurable, we have that  $x_n \in \text{Level}_{\leq a_{n,x}}(c_n(\theta, \cdot))$ . Since  $b(\theta)$  is compact, it is totally bounded and therefore, for any  $\varepsilon > 0$ , there exist an  $m(\varepsilon) \in \mathbb{N}$  and  $\{y_i, i = 1, \dots, m(\varepsilon)\} \subset b(\theta)$ , such that the collection of balls (in  $B$ )  $\mathcal{O}(y_i, (\varepsilon/2))$  covers  $b(\theta)$ . For some real sequence  $q_n \rightarrow 0$ , consider  $\{y_i, i = 1, \dots, m(q_n)\}$  and extract analogously random sequences  $(y_{i,n})$  such that  $y_{i,n} \rightarrow y_i$   $P$  a.s. Define  $a_{n,y_i} = c_n(\theta, y_{i,n})$  and  $n^*(n) = \min\{n' : |a_{n',y_i} - a_\theta| \leq (q_n/2), P \text{ a.s. for all } i = 1, \dots, m(q_n)\}$  which is well defined due to Egoroff's theorem and the fact that  $m(q_n)$  is finite. Obviously  $n^*$  is nondecreasing in  $n$ . Then define  $a_n^* = \max\{a_{n',y_i}, i = 1, \dots, m(q_n)\}$  which is measurable and  $a_n^* \rightarrow a_\theta$   $P$  a.s. Then for  $x \in b(\theta)$ , it follows that, for any  $\varepsilon > 0$ , there exist an  $n$  and a measurable  $x_n \in \text{Level}_{\leq a_n^*}(c_n(\theta, \cdot))$  such that  $d(x_n, x) < \varepsilon$   $P$  a.s.  $\square$

*Proof of Proposition 8.* For (1) we have first that for any measurable nonnegative  $\varepsilon_n$  that need not converge to zero if  $x_n \in \beta_n^\#(\theta, \varepsilon_n)$  measurable, such that a subsequence  $x_{n_k} \rightarrow x$   $P$  a.s.

$$\begin{aligned} c(\theta, x) &\leq \liminf_{n_k} c_{n_k}(\theta, x_{n_k}) \text{ } P \text{ a.s.} \\ &= \liminf_{n_k} (c_{n_k}(\theta, x_{n_k}) + \varepsilon_{n_k}) \text{ } P \text{ a.s.} \\ &\leq \limsup_{n_k} (c_{n_k}(\theta, x_{n_k}) + \varepsilon_{n_k}) \text{ } P \text{ a.s.} \quad (5) \\ &= \limsup_{n_k} \left( \inf_B c_{n_k}(\theta, \beta) + \varepsilon_n \right) \text{ } P \text{ a.s.} \\ &\leq \inf_B c(\theta, x) + \varepsilon_n \text{ } P \text{ a.s.,} \end{aligned}$$

where that last inequality follows from Lemma 9. This establishes that for any nonnegative random variable  $\varepsilon_n$

$$\text{Ls}(\beta_n^\#(\theta, \varepsilon_n)) \subseteq \varepsilon_n\text{-arg min}_B c(\theta, \beta) \text{ } P \text{ a.s.} \quad (6)$$

Now (1) follows from the fact that  $\varepsilon\text{-arg min}_B \subseteq \varepsilon'\text{-arg min}_B$  if  $\varepsilon \leq \varepsilon'$ . For (2) notice that, from the definition of the Fell topology in Section 4 for any  $\varepsilon > 0$ , we have that for large  $n$ ,  $\text{epi}_n(\omega, \theta) \cap B \times [a_\theta - 2\varepsilon, a_\theta - \varepsilon] = \emptyset$   $P$  a.s. since  $B \times [a_\theta - 2\varepsilon, a_\theta - \varepsilon]$  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 Lemma 9 that  $\inf_B c_n(\theta, \beta) \rightarrow a_\theta$   $P$  a.s. Then using Lemma 10 set  $\varepsilon_n^* = a_n^* - a_\theta$  which is obviously measurable and converges to zero  $P$  a.s. (3) follows from (1) the single valuedness of  $b(\theta)$  and the compactness of  $B$ .  $\square$

The proof of Lemma 10 implies that the sequence  $(\varepsilon_n^*)_{n \in \mathbb{N}}$  that appears in Proposition 8(2) is nonunique. However the fact that this implication does not hold for any sequence of nonnegative 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.

**2.4. Upper Hemicontinuity of the Binding Correspondence.** Proposition 8 enables the use of  $\delta_u(b(\theta), \beta_n^\#)$  as the inversion criterion. The following assumption concerns the upper continuity of the binding correspondence which along with the relevant properties of  $\delta_u$  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 11.**  $b$  is upper hemicontinuous that is, for any  $\theta$  and  $\theta^* \rightarrow \theta$ ,  $\delta_u(b(\theta), b(\theta^*)) \rightarrow 0$ .

The following proposition provides with sufficient conditions for this to hold. It essentially strengthens Assumption 5 in that it requires that the relevant  $P$  a.s. epiconvergence be *continuous* on  $\Theta$ . Notice that its requirements are *also* stricter with respect to measurability compared to the ones in Assumption 5 since the former requires that for any  $\theta$  the relevant set of  $P$  unit probability does not depend on the sequence that converges to  $\theta$ .

**Proposition 12.** *If, for any  $\theta$ ,  $P$  a.s. for any  $\omega$ , any  $\theta_n \rightarrow \theta$ , and any  $\beta$ ,*

- (1)  $\liminf_{n \rightarrow \infty} c_n(\omega, \theta_n, \beta_n) \geq c(\theta, \beta)$ , for all  $\beta_n$  such that  $\beta_n \rightarrow \beta$ ,
- (2)  $\limsup_{n \rightarrow \infty} c_n(\omega, \theta_n, \beta_n) \leq c(\theta, \beta)$ , for some  $\beta_n$  such that  $\beta_n \rightarrow \beta$ ,

*then Assumption 11 holds.*

*Proof.* Suppose that  $\mathcal{D}_{F_0}$  metrizes  $\mathcal{T}_F$  on  $F_0(B)$  (see Section 4). Then it is obvious that Proposition 12(1)-(2) is equivalent to the requirement that, for any  $\theta$  and any  $\theta_n \rightarrow \theta$ ,  $\mathcal{D}_{F_0}(\text{epi}_n(\omega, \theta_n), \text{epi}(\theta))$  converges to zero  $P$  a.s. Then, for any  $\theta$  and any  $\theta_n \rightarrow \theta$ ,  $c(\theta_n, \cdot)$  epiconverges to  $c(\theta, \cdot)$  ( $\mathcal{D}_{F_0}(\text{epi}(\theta_n), \text{epi}(\theta)) \rightarrow 0$ ), that is,  $c(\theta, \cdot)$  is epicontinuous on  $\Theta$ . This is due to the following standard argument: for an arbitrary  $\theta$  and  $\varepsilon > 0$ , we have that  $\mathcal{D}_{F_0}(\text{epi}_n(\omega, \theta'), \text{epi}(\theta)) < (\varepsilon/2)$   $P$  a.s. for any  $\theta'$  in some open neighborhood of  $\theta$  and large enough  $n$ , due to the assumed form of convergence and Egoroff's theorem. By the same reasoning  $\mathcal{D}_{F_0}(\text{epi}_n(\omega, \theta'), \text{epi}(\theta')) < (\varepsilon/2)$   $P$  a.s. for any such  $\theta'$ . The result follows from the fact that  $\text{epi}(\theta)$  is independent of  $\omega$ . This along with 3.1 of Theorem 5.3.4 and proposition Appendix D.2 of [22] implies that the composite mapping  $\theta \rightarrow c(\theta, \cdot) \rightarrow \arg \min_B c(\theta, \beta)$  is appropriately continuous.  $\square$

An examination of the proof of Proposition 8 shows that if the conditions (1)-(2) of Proposition 12 hold then Proposition 8 could be restated with  $\delta_u(b(\theta), \beta_n^\#(\theta_n, \varepsilon_n))$

replacing  $\delta_u(b(\theta), \beta_n^\#(\theta, \varepsilon_n))$  when this occurs. This means that the premises of Proposition 12 render  $\beta_n^\#$  strongly continuously (upper in the first case or Fell in the second and third cases of Proposition 8)  $b(\theta)$ -consistent. Furthermore the compactness of  $\Theta$  then implies that these notions are also *uniform* with respect to  $\Theta$ .

**Remark 13.** Proposition 12(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 [26]. 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 with respect to stationary and ergodic processes.

**Lemma 14.** *Suppose that  $c_n(\omega, \theta, \beta) = (1/n) \sum_{i=1}^n m_i(\omega, \theta, \beta)$ ,  $(m_i(\theta, \beta))_{i \in \mathbb{Z}}$  is ergodic for any  $(\theta, \beta)$ ,  $c_n(\omega, \cdot, \cdot) : \Theta \times B \rightarrow \overline{\mathbb{R}}$  is jointly continuous  $P$  a.s., there exists a finite open cover of  $\Theta \times B$ , such that  $E[\inf_{(\theta, \beta) \in A} m_0(\omega, \theta, \beta)] < +\infty$ , for any  $A$  in the cover,  $E(m_0(\omega, \theta, \beta^*))$  assumes values in  $\overline{\mathbb{R}}$  for any  $\theta \in \Theta$ , and  $\beta^*$  in a countable dense subset of  $B$ . Then Proposition 12 holds.*

*Proof.* Proposition 12(1) follows from the fact that the assumption framework of the lemma implies condition  $\mathcal{E}_0$  and thereby Theorem 2.3 of [23], which implies the joint  $P$  a.s. epiconvergence of  $c_n$  to  $Em_0$ . For Proposition 12(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  $\beta$  and any  $\delta > 0$  there exists a  $\beta^* \in B^*$  such that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} c_n(\omega, \theta_n, \beta) \\ & \leq \limsup_{n \rightarrow \infty} (c_n(\omega, \theta_n, \beta) - c_n(\omega, \theta_n, \beta^*)) \\ & \quad + \limsup_{n \rightarrow \infty} c_n(\omega, \theta_n, \beta^*) \quad P \text{ a.s.} \\ & \leq \delta + \limsup_{n \rightarrow \infty} c_n(\omega, \theta_n, \beta^*) \quad P \text{ a.s.} \end{aligned} \tag{7}$$

By assumption the subset of  $\Omega$  of  $P$  unit probability can be chosen independent of  $\delta$  and  $\delta$  can be chosen arbitrarily small. Hence, Proposition 12(2) would be implied for  $\beta_n = \beta$ , if, for any  $\theta$ ,  $\beta^* \in B^*$ ,  $P$  a.s. and any  $\theta_n \rightarrow \theta$ ,

$$\limsup_{n \rightarrow \infty} c_n(\omega, \theta_n, \beta^*) \leq Em_0(\theta, \beta^*) \tag{i}$$

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, e.g., Proposition 2.1.1. of [24]), and hence the uniform version of Birkhoff's LLN implies that

$$\frac{1}{n} \sum_{i=1}^n ((m_i(\omega, \theta_n, \beta^*) \wedge -\rho) \vee \rho) \tag{8}$$

converges  $P$  a.s. to  $\int_{-\rho}^{\rho} m_0(\omega, \theta, \beta^*) dP(\omega)$  for any  $\rho > 0$ . Due to the separability of  $\mathbb{R}$  the subset of  $\Omega$  of  $P$  unit probability

can be chosen independent of  $\rho$ . Hence from Definition 7.12 of [26] we obtain (i).  $\square$

This lemma explores sufficient conditions for the required continuity of  $b$  solely via restrictions on the behavior of  $c_n$  which in applications is generally more analytically tractable than  $c$ . Furthermore it combines joint epi-onvergence with pointwise (on  $B$ ) extended uniform (with respect to  $\theta$ ) 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 13.

**2.5. 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 8 along with the measurability of the auxiliary correspondence and the upper hemicontinuity of  $b$  facilitate the use of  $\delta_u(b(\theta), \beta_n^\#(\theta_0))$  for some distinguished  $\theta_0 \in \Theta$ , for the definition of the IE and the subsequent existence argument. Again an almost surely nonnegative random variable will assume the role of the ‘‘optimization error’’ in this second step of the estimation procedure.

**Definition 15.** The indirect correspondence  $\theta_n^\#(\omega, \theta_0, \varepsilon_n, \varepsilon_n^\#)$  satisfies

$$\begin{aligned} \theta_n^\#(\omega, \theta_0, \varepsilon_n, \varepsilon_n^\#) &= \varepsilon_n^\# \text{-arg min}_{\Theta} \delta_u(b(\theta), \beta_n^\#(\theta_0, \varepsilon_n)) \\ &\doteq \left\{ \theta^* \in \Theta : \delta_u(b(\theta^*), \beta_n^\#(\theta_0, \varepsilon_n)) \right. \\ &\quad \left. \leq \inf_{\Theta} \delta_u(b(\theta), \beta_n^\#(\theta_0, \varepsilon_n)) + \varepsilon_n^\# \right\}, \end{aligned} \quad (9)$$

where  $\varepsilon_n^\#$  is a nonnegative random variable defined on  $\Omega$ .

The dependence of  $\theta_n^\#$  on  $\theta_0$  follows from the analogous dependence of  $\beta_n^\#$  which in turn follows from the dependence of  $y_n$  on  $\theta_0$ . Obviously in practice this dependence is not ‘‘visible.’’ We are initially concerned with the question of existence of the IE. We again suppress the dependence of  $\theta_n$  on  $\omega$  when there is not a risk of confusion.

**Proposition 16.** Under Assumptions 1 and II  $\theta_n^\#$  is  $\mathcal{B}(\Theta_F)/\mathcal{F}$ -measurable,  $P$  a.s. nonempty, compact valued correspondence.

*Proof.* First, notice that due to Lemma A.9, Proposition 4 (implied by Assumption I), Assumption II, and the facts that  $\beta_n^\#(\theta_0, \varepsilon_n)$  is independent of  $\theta$  and  $b(\theta)$  is independent of  $\omega$ , we obtain that  $\delta_u(b(\theta), \beta_n^\#(\theta_0, \varepsilon_n))$  is  $\mathcal{B}(\mathbb{R})/\mathcal{B}(\Theta_F) \otimes \mathcal{F}$ -measurable. Due to Lemma A.4 and Proposition 4  $\delta_u(b(\theta), \beta_n^\#(\theta_0, \varepsilon_n))$  is  $P$  a.s. lsc and therefore a normal integrand. 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 4 where  $c_n = \delta_u(b(\theta), \beta_n^\#(\theta_0, \varepsilon_n))$ , when we consider  $B = \Theta$  and  $\Theta = \{\theta_0\}$  (the left-hand sides correspond to the notation of the latter lemma).  $\square$

The fundamental selection theorem (Theorem 2.13 of [22]) would also enable the definition of the IE as a measurable function with values in  $\Theta$ . Having established existence we turn to the issue of consistency. We need an assumption of indirect identification that is essentially 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 Definition 15 by an analogy principle. Mathematically both the definition and the existence argument do not require the following assumption in order to be valid.

**Assumption 17.** If  $\theta \neq \theta_0 \Rightarrow b(\theta_0) - b(\theta) \neq \emptyset$ .

**Remark 18.** This condition is weaker than a condition of the form ‘‘If  $\theta \neq \theta_0 \Rightarrow b(\theta_0) \cap b(\theta) = \emptyset$ ’’ and stronger than 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  $\delta_u$  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 cannot 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 nonempty and compact due to the compactness of  $\Theta$ , the properness of  $\delta_u$ , the hemicontinuity of  $b$ , and Lemma A.3. Obviously,  $\theta_0 \in \theta(\theta_0)$ , while  $\{\theta_0\} = \theta(\theta_0)$  if and only if Assumption 17 holds. Since we once again are dealing with compact valued correspondences, convergence is metrized by  $\delta_u$  and/or  $\delta$ .

**Proposition 19.** Under Assumptions 1, 5, and II and if  $\varepsilon_n^\# \rightarrow 0$   $P$  a.s. then

(1)  $\delta_u(\theta(\theta_0), \theta_n^\#(\omega, \theta_0, \varepsilon_n^*, \varepsilon_n^\#)) \rightarrow 0$   $P$  a.s. where  $\varepsilon_n^*$  is defined in Proposition 8(2),

(2) if  $b(\theta_0)$  is singleton then  $\delta_u(\theta(\theta_0), \theta_n^\#(\omega, \theta_0, \varepsilon_n, \varepsilon_n^\#)) \rightarrow 0$   $P$  a.s. for any  $\varepsilon_n \rightarrow 0$   $P$  a.s.,

if furthermore Assumption 17 holds then

(1\*)  $\delta(\{\theta_0\}, \theta_n^\#(\omega, \theta_0, \varepsilon_n^*, \varepsilon_n^\#)) \rightarrow 0$   $P$  a.s. where  $\varepsilon_n^*$  is defined in Proposition 8(2),

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

*Proof.* First notice that due to Proposition 8(2), Assumption II, and Lemma A.6 we have that for any  $\theta$  and  $\theta_n \rightarrow \theta$

$$\liminf_n \delta_u(b(\theta_n), \beta_n^\#(\theta_0, \varepsilon_n^*)) \geq \delta_u(b(\theta), b(\theta_0)) \quad P \text{ a.s.} \quad (10)$$

and that, for any  $\theta$  and  $\theta_n = \theta$ , due to Lemma A.4

$$\limsup_n \delta_u(b(\theta_n), \beta_n^\#(\theta_0, \varepsilon_n^*)) \leq \delta_u(b(\theta), b(\theta_0)) \quad P \text{ a.s.} \quad (11)$$

Hence (1) follows from Proposition 8(1) for  $c_n = \delta_u(b(\theta_n), \beta_n^\#(\theta_0, \varepsilon_n^*))$  and if we denote by  $B$  (in the notation of this lemma) the  $\Theta$  space and by  $\Theta$  (in the notation of this lemma)  $\{\theta_0\}$ . (2) follows in the same manner if we replace any invocation of Proposition 8(2) with Proposition 8(3). Finally, notice that if Assumption 17 holds, then  $\theta(\theta_0) = \{\theta_0\}$  establishing (1\*) and (2\*) via another use of Proposition 8(3).  $\square$

If Assumption 17 does not hold then the implications of Proposition 19(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 Proposition 19(1) and (1\*) merely explore the issue of the existence of strongly consistent estimators among those that comply with Definition 15. The properties of the  $\delta_u$  function along with Proposition 8 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.

**2.6. Extension.** In most cases  $b$  is analytically 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 Definition 15 are infeasible. However, it may be the case that (possibly) stochastic and algorithmically feasible approximations of the unknown  $b$  can be used for the construction of several other classes of feasible IE. In such a context the results derived previously could be used to describe properties of such approximations that would imply that these estimators are well defined and among them strongly consistent ones exist. 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  $\kappa_n$  in Definition 15.

**Proposition 20.** *Consider the IE defined by*

$$\vartheta_n^\#(\omega, \theta_0, \varepsilon_n, \varepsilon_n^\#) = \varepsilon_n^\# \text{-arg min}_{\Theta} \delta_u(\kappa_n(\omega, \theta), \beta_n^\#(\theta_0, \varepsilon_n)), \quad (12)$$

where  $\varepsilon_n, \varepsilon_n^\#$  as before.

- (i) *Under Assumption 1 and if  $\kappa_n(\omega, \theta)$  is  $\mathcal{B}(B_F)/\mathcal{F} \otimes \mathcal{B}(\Theta_F)$ -measurable, P.a.s. compact valued and upper hemicontinuous, then  $\vartheta_n^\#$  is  $\mathcal{B}(\Theta_F)/\mathcal{F}$ -measurable, P a.s. nonempty, compact valued correspondence.*
- (ii) *If moreover Assumption 5 holds and P a.s. for any  $\theta$  and any  $\theta_n \rightarrow \theta$ ,  $\delta(b(\theta), \kappa_n(\omega, \theta))$  and  $\delta_u(b(\theta), \kappa_n(\omega, \theta_n))$  converge to zero then the implications of Proposition 19(1)-(2) hold also for  $\vartheta_n^\#$ . If furthermore Assumption 17 holds of then the implications Proposition 19(1\*)-(2\*) hold also for  $\vartheta_n^\#$ .*

*Proof.* (i) As in the proof of Proposition 16 from Lemma A.9, Proposition 4 the  $P$  a.s. upper hemicontinuity of  $\kappa_n$  the facts that  $\beta_n^\#(\theta_0, \varepsilon_n)$  is independent of  $\theta$ , and  $\kappa_n$  is jointly measurable, we obtain that  $\delta_u(\beta_n^\#(\theta_0, \varepsilon_n), b(\theta))$  is  $\mathcal{B}(\mathbb{R})/\mathcal{B}(\Theta_F) \otimes \mathcal{F}$ -measurable. Due to Lemma A.4 and Proposition 4  $\delta_u(\kappa_n(\omega, \theta_n), \beta_n^\#(\theta_0, \varepsilon_n))$  is  $P$  a.s. lsc and therefore a normal integrand. It is also  $P$  a.s. proper due to the fact that  $\kappa_n$  and  $\beta_n^\#$  are  $P$  a.s. compact valued. Hence the result follows from Proposition 4. (ii) It suffices to prove that  $\delta_u(\kappa_n(\omega, \theta), \beta_n^\#(\theta_0, \varepsilon_n)) \xrightarrow{e} \delta_u(b(\theta), b(\theta_0))$   $P$  a.s. when  $\beta_n^\#(\theta_0, \varepsilon_n)$  satisfies the implications (2) or (3) of Proposition 8. The rest would then follow as in the proof of Proposition 20. Notice that for any  $\theta$  and any  $\theta_n \rightarrow \theta$

$$\liminf_n \delta_u(\kappa_n(\omega, \theta_n), \beta_n^\#(\theta_0, \varepsilon_n^*)) \geq \delta_u(b(\theta), b(\theta_0)), P \text{ a.s.} \quad (13)$$

due to Lemma A.6, the definition of  $\beta_n^\#$  and the  $P$  a.s. continuous w.r.t  $\theta$  upper convergence of  $\kappa_n$  to  $b$ . Moreover for  $\theta_n = \theta$

$$\begin{aligned} & \limsup_n \delta_u(\kappa_n(\omega, \theta), \beta_n^\#(\theta_0, \varepsilon_n^*)) \\ & \leq \limsup_n \delta_u(\kappa_n(\omega, \theta), b(\theta)) \\ & \quad + \limsup_n \delta_u(b(\theta), \beta_n^\#(\theta_0, \varepsilon_n^*)), P \text{ a.s.} \quad (14) \\ & \leq \limsup_n \delta(\kappa_n(\omega, \theta), b(\theta)) \\ & \quad + \delta_u(b(\theta), b(\theta_0)), P \text{ a.s.,} \end{aligned}$$

where the first inequality follows from the triangle inequality and the second from the definition of  $\delta$  and the  $P$  a.s. Fell convergence of  $\beta_n^\#$  to  $b(\theta_0)$  and Lemma A.4. Due to the  $P$  a.s. pointwise with respect to  $\theta$  Fell convergence of  $\kappa_n$  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.  $\square$

Notice that this proposition generalizes the results of Propositions 16 and 20. 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  $\theta$  and thereby so is  $\beta_n^\#(\theta, \varepsilon_n)$  for any  $\theta$  and some optimization error  $\varepsilon_n$  independent of  $\theta$ . Then a feasible IE can be defined by the approximate minimization of  $\delta_u(\beta_n^\#(\theta, \varepsilon_n), \beta_n^\#(\theta_0, \varepsilon_n))$  with respect to  $\theta$ . In this case the joint measurability of  $\beta_n^\#$  would follow from the joint measurability of  $c_n$  and  $\varepsilon_n$ , 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 Proposition 8(1) if Assumption 1 is strengthened so that the mapping  $\theta \rightarrow \text{epi}_n(\omega, \theta)$  is  $P$  a.s. Fell continuous. 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 Proposition 8(2) or (3) and the  $P$  a.s. continuous w.r.t  $\theta$  upper convergence to  $b$  would follow if Proposition 12 holds

with the set of unit probability independent of  $\theta$ . It would suffice that  $c_n$  is  $P$  a.s. jointly continuous and converges to  $c$  jointly uniformly.

### 3. Examples

In this section we consider four simple examples that represent some of the previous results. The first concerns the case of a linear semiparametric 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,  $\Theta$  is a compact subset of  $\mathbb{R}^p$  and  $B$  a compact subset of  $\mathbb{R}^q$ . In the second and the fourth ones the binding function is actually single valued (hence a fortiori compact valued) and 1-1 enabling the direct application of Proposition 19(2\*). The first and second examples include cases in which the IE can be interpreted as performing “inconsistency” correction to the auxiliary one.

*Example 21* (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  $(X'X/n) \rightarrow M_{X'X}$ ,  $(Z'Z/n) \rightarrow M_{Z'Z}$ ,  $(Z'X/n) \rightarrow M_{Z'X}$   $P$  a.s., where  $\text{rank}(M_{X'X}) = \text{rank}(M_{Z'X}) = p$  and  $p \leq l \doteq \text{rank}(M_{Z'Z}) \leq q$ . For  $u$  an  $n \times 1$  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  $c_n(\omega, \theta, \beta) = (1/n)(Y - Z\beta)'(Y - Z\beta)$ , which clearly satisfies Assumption 1 due to continuity with respect to  $\beta$  and the compactness of  $B$ . Obviously,  $c_n$  is constructed by the auxiliary set of regression with respect to  $Z$ . Proposition 4 ensures the existence of  $\beta_n$  which in the light of the previous can be interpreted as an OLSE in the context of the auxiliary model. Let  $\mathcal{P} : \mathbb{R}^q \rightarrow M_{Z'Z}B$  be the (generally nonlinear) projection defined by the optimization problem

$$\arg \min_{x \in M_{Z'Z}B} \|x - y\| \quad (15)$$

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  $y \in \text{col}(M_{Z'Z})$ , 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  $(X'u/n) \rightarrow M_{X'u}$ ,  $(u'u/n) \rightarrow M_{u'u}$  and  $(Z'u/n) \rightarrow M_{Z'u}$   $P$  a.s. The previous imply the joint uniform  $P$  a.s. convergence of  $c_n$  to

$$\begin{aligned} c(\theta, \beta) &= \theta' (M_{X'X}\theta + 2M_{X'u}) \\ &+ \beta' (M_{Z'Z}\beta - 2(M_{Z'X}\theta + M_{Z'u})) + M_{u'u} \end{aligned} \quad (16)$$

which implies both Assumptions 5 and 11 (via Proposition 12 and Remark 13). Notice that

$$b(\theta) = B \cap (K\mathcal{P}(M_{Z'X}\theta + M_{Z'u}) + H_{q-l}) \quad (17)$$

due to the convexity of  $c(\theta, \beta)$  with respect to  $\beta$  for any  $\theta$  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 17 applies. (More precisely we have that  $b(\theta) \cap b(\theta_0) = \emptyset$ .) Hence, Proposition 19(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  $b(\theta) = \{\theta - M_{X'X}^{-1}M_{X'u}\}$  and Proposition 19(2) implies that any IE defined by Definition 15 can be perceived as an “inconsistency corrector” of the underlying OLSE for  $\theta$ .

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

*Example 22* (the drift of a levy process with bounded jumps). Let  $W$  denote a standard Brownian motion and  $\nu$  a finite measure on the Borel algebra of  $\mathbb{R} - \{0\}$ , such that  $\nu(A) = 0$  when  $A \subseteq (-\infty, -C_2) \cup (-C_1, C_1) \cup (C_2, +\infty)$  for  $0 < C_1 < C_2$ . Obviously  $\nu$  is a Levy measure (see paragraph 1.2.4 of [27]). For  $p = 1$  consider the stochastic process on  $\mathbb{R}^+$  defined by the following Levy-Ito decomposition (see Theorem 2.4.16 of [27])

$$X_t(\omega) = \theta t + W_t(\omega) + \int_{|x| \in [C_1, C_2]} x N(t, dx)(\omega), \quad (18)$$

where  $N$  denotes the independence to  $W$  Poisson random measure of  $\mathbb{R}^+ \times ([-C_2, C_1] \cup [C_1, C_2])$  the existence of which is established by Theorem 2.3.6 of [27]. 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  $c_n(\omega, \theta, \beta) = (1/n) \sum_{t=1}^n (y_t - \beta)^2$ , 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

$$dy_t = \beta y_t dt + y_t dW_t \quad (19)$$

for each  $\beta \in B$ . Obviously Assumption 1 is satisfied, due to continuity with respect to  $\beta$  and the compactness of  $B$ . Proposition 4 ensures the existence of  $\beta_n$  which in the light of the previous can be interpreted as an (approximate) MLE in the context of the auxiliary model. Furthermore since

$$\begin{aligned} &\left| \int_{|x| \in [C_1, C_2]} x (N(t, dx) - N(t-1, dx)) \right| \\ &\leq \int_{|x| \in [C_1, C_2]} |x| (N(t, dx) - N(t-1, dx)) \\ &\leq C_2 (N(t, [C_1, C_2]) - N(t-1, [C_1, C_2])) \end{aligned} \quad (20)$$

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

$$E \exp(X_t - X_{t-1}) = \exp\left(\theta + \frac{1}{2}\right) C \quad (21)$$

for

$$\begin{aligned} 0 < C &\leq \exp(-\nu([C_1, C_2])) (1 - \exp(C_2)) \\ E(\exp(X_t - X_{t-1}))^2 &= \exp(2(\theta + 1)) C^* \end{aligned} \quad (22)$$

for

$$0 < C^* \leq \exp(-\nu([C_1, C_2]))(1 - \exp(2C_2)). \quad (23)$$

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

$$c(\theta, \beta) = \exp(2(\theta + 1)C^*) - 2 \exp\left(\theta + \frac{1}{2}\right)C(1 + \beta) + (1 + \beta)^2 \quad (24)$$

which implies both Assumptions 5 and II (via Proposition 12 and Remark 13). If  $B \supseteq \exp(\Theta + (1/2)C) - 1$  then

$$b(\theta) = \left\{ \exp\left(\theta + \frac{1}{2}\right)C - 1 \right\}. \quad (25)$$

In this case Assumption 17 applies and therefore Proposition 19(2\*) implies that any IE defined by Definition 15 is consistent for any  $\theta_0 \in \Theta$ . When  $\nu([C_1, C_2]) = 0$  (and therefore  $\nu = 0$ ) 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, e.g., paragraph 6.1.1 of [13]).

For the last pair of examples, let  $z : \Omega \rightarrow \mathbb{R}^{\mathbb{Z}}$  be an i.i.d. (double infinite) sequence of random variables, with  $Ez_0 = 0$  and  $Ez_0^2 = 1$ . Consider a random element  $\sigma^2 : \Theta \times \Omega \rightarrow (\mathbb{R}^+)^{\mathbb{Z}}$ , with the product space  $\Theta \times \Omega$  equipped with  $\mathcal{B}(\Theta) \otimes \mathcal{F}$  with  $\sigma_t^2(\theta)$  independent of  $(z_i)_{i \geq t}$ , for all  $t \in \mathbb{Z}$ , for all  $\theta \in \Theta$ . Analogously, define the random element  $y : \Theta \times \Omega \rightarrow (\mathbb{R})^{\mathbb{Z}}$  as

$$(y_t(\omega)(\theta))_{t \in \mathbb{Z}, \theta \in \Theta} = \left( z_t(\omega) \sqrt{\sigma_t^2(\omega)(\theta)} \right)_{t \in \mathbb{Z}, \theta \in \Theta}. \quad (26)$$

Then for all  $\theta \in \Theta$ ,  $(y_t(\theta))_{t \in \mathbb{Z}}$  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. (The establishment of the ergodicity is initiated by the analogous establishment for  $(\sigma_t^2(\theta))_{t \in \mathbb{Z}}$  for all  $\theta \in \Theta$ . Sufficient conditions for that are described and employed in a variety of heteroskedastic models in Chapter 5 of [24] via Theorem 5.2.1. Then the ergodicity of  $(y_t(\theta))_{t \in \mathbb{Z}}$  and  $(y_t^2(\theta))_{t \in \mathbb{Z}}$  for all  $\theta \in \Theta$  follows from the definition of  $z, y$ , the previous assumption and Proposition 2.2.1 of [24]).

*Example 23* (instrumental variables estimation in regressions on squared ARCH(1) processes). Let  $\sigma_4 = Ez_0^4 < +\infty$  and  $0 < \delta < (1/\sqrt{\sigma_4})$ . Suppose that  $a \in \Theta = [0, \delta]$  and consider the stochastic dif and only iference equation

$$\sigma_t^2(a) = 1 + az_{t-1}^2 \sigma_t^2(a). \quad (27)$$

Due to the fact that  $\sigma_4 > 1$  Theorem 5.2.1. of [24] implies that for any  $\alpha \in \Theta$  the equation admits a unique stationary and 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

$$Z(a) = \begin{pmatrix} 1 & y_{-1}^2(a) \\ \vdots & \vdots \\ 1 & y_{n-2}^2(a) \end{pmatrix}, \quad X(a) = \begin{pmatrix} 1 & y_0^2(a) \\ \vdots & \vdots \\ 1 & y_{n-1}^2(a) \end{pmatrix} \quad (28)$$

jointly measurable with respect to  $\mathcal{F} \otimes \mathcal{B}(\Theta)$ , 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  $c_n(\omega, a, \beta) = \|(1/n)Z'(a)(Y(a) - X(a)\beta)\|$  which clearly satisfies Assumption 1 due to joint continuity with respect to  $(a, \beta)$  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 AR(1) representation of the ARCH(1) process with respect to the martingale dif and only iference noise  $v_t = (z_t^2 - 1)\sigma_t^2(a)$  (see, e.g., [28]) 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 4 ensures the existence of  $\beta_n$  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  $\Theta$  and  $\sigma_4$  we have that

$$E \sup_{a, \beta} \left\| \frac{1}{n} Z'(a) (Y(a) - X(a)\beta) \right\| \leq C_1 \left\| \begin{pmatrix} 1 & \frac{1}{1-\delta} \\ \frac{1}{1-\delta} & \frac{1}{(1-\delta)(1-\delta^2\sigma_4)} \end{pmatrix} \right\| + C_2 \left\| \begin{pmatrix} \frac{1}{1-\delta} \\ \frac{1+\delta}{(1-\delta)(1-\delta^2\sigma_4)} \end{pmatrix} \right\| < +\infty \quad (29)$$

for  $C_1, C_2 > 0$  which along with (the uniform version of) Birkhoff’s Ergodic theorem (see, e.g., [24], Theorem 2.2.1) implies both Assumptions 5 and II (via Proposition 12 and Remark 13) for

$$c(a, \beta) = \left\| \begin{pmatrix} 1 & \frac{1}{1-a} \\ \frac{1}{1-a} & \frac{1}{(1-a)(1-a^2\sigma_4)} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} - \begin{pmatrix} \frac{1}{1-a} \\ \frac{1+a}{(1-a)(1-a^2\sigma_4)} \end{pmatrix} \right\|. \quad (30)$$

In fact a simple calculation shows that

$$b(a) = \begin{cases} \left\{ \begin{pmatrix} 1 \\ a \end{pmatrix} \right\}, & a \in (0, \delta], \\ \left\{ (\beta_1, 1 - \beta_1), \beta_1 \in [1, 1 + \delta] \right\}, & a = 0 \end{cases} \quad (31)$$

which clearly implies Assumption 17. Hence Proposition 19(2\*) implies that any IE defined by Definition 15 is consistent if  $a_0 \in (0, \delta]$ , and Proposition 19(1\*) implies the *existence* of an analogously consistent IE when  $a_0 = 0$ . It is easy to see that, given  $\beta_n$ ,

$$\theta_n^\#(\omega, \theta_0, \varepsilon_n^*, 0) = \begin{cases} \{a\} & \text{if } \beta_n \subseteq b(a) \\ \text{mid}(\text{proj}_{\beta_2}(\beta_n)) & \text{if } \beta_n \not\subseteq b(a), \\ & \text{mid}(\text{proj}_{\beta_2}(\beta_n)) \geq 0, \\ \{0\} & \text{if } \beta_n \not\subseteq b(a), \\ & \text{mid}(\text{proj}_{\beta_2}(\beta_n)) < 0, \end{cases} \quad (32)$$

where  $\text{proj}_{\beta_2}$  denotes projection to the  $\beta_2$ -axis and  $\text{mid}(A)$  denotes the midpoint of the smallest interval that contains  $A$ . Notice that in our case  $\text{mid}(\text{proj}_{\beta_2}(\beta_n))$  is well defined due to the fact that  $\beta_n$  is  $P$  a.s. compact valued hence its  $\text{proj}_{\beta_2}(\beta_n)$  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 20 implies also the analogous properties for IE defined by  $\kappa_n$  when this equals the auxiliary estimator derived from bootstrap resampling for any  $a$ .

The final example is about an asymmetric heteroskedastic process.

*Example 24* ( $-c_n$  is the quasilielihood function of an approximate to QARCH(1) model). Let  $E|z_0| < 1$  and  $\delta > 0$ , and consider for  $\gamma \in \Theta = [-\delta, 0]$   $\bar{\omega}_0, a_0 > 0$ ,  $\bar{\omega} = \bar{\omega}_0 + (\delta^2/4a_0)$ , and  $a_0 < \exp(-2E \ln |z_0|)$  the stochastic dif and only inference equation

$$\sigma_t^2(\gamma) = \bar{\omega} + a_0 z_{t-1}^2 \sigma_t^2(\gamma) + \gamma z_{t-1} \sigma_t(\gamma). \quad (33)$$

For any  $\gamma \in \Theta$ , the previous defines a unique stationary and ergodic 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 [29]). 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

$$h_t(\gamma, \gamma^*) = \bar{\omega} + a_0 y_{t-1}^2 + \gamma^* |y_{t-1}| 1_{y_{t-1} < 0} \stackrel{P \text{ a.s.}}{=} \bar{\omega} + a_0 y_{t-1}^2 + \gamma^* |y_{t-1}| 1_{z_{t-1} < 0}, \quad (34)$$

where  $h_t(\gamma, \gamma^*)$  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 [24]. Now consider

$$c_n(\omega, \gamma, \gamma^*) \doteq \frac{1}{n} \sum_{i=1}^n \left( -\ln \frac{y_i^2(\gamma)}{h_i(\gamma, \gamma^*)} + \frac{y_i^2(\gamma)}{h_i(\gamma, \gamma^*)} \right), \quad (35)$$

where  $-c_n$  can be considered as an approximation of (a monotonic transformation of) the conditional quasilielihood function of the auxiliary conditionally heteroskedastic

model defined by 3 and  $B$ . Also the ergodicity of  $(c_n)$  for any  $(\gamma, \gamma^*)$  follows from the previous and Proposition 2.1.1. of [24]. (In practice  $c_n(\omega, \gamma, \gamma^*)$  is unknown but approximated by an analogous  $\hat{c}_n(\omega, \gamma, \gamma^*)$  dependent on nonergodic solutions of the stochastic dif and only inference equation that defines  $h$  based on arbitrary initial conditions. In this case, due to ergodicity, Proposition 5.2.12 of [24] can be employed in order to ensure that  $\sup_B |c_n(\omega, \theta, \beta) - \hat{c}_n(\omega, \theta, \beta)|$  converges almost surely to zero for any  $\theta \in \Theta$  (see the first part of the proof of Theorem 5.3.1 of [24]), thereby facilitating the asymptotic analysis of minimizers of  $\hat{c}_n(\omega, \theta, \beta)$  by the analogous analysis of minimizers of  $c_n(\omega, \theta, \beta)$ .) Assumption 1 follows readily from the form of  $c_n$  and the  $P$  a.s. continuity with respect to  $(\gamma, \gamma^*)$ . Hence  $\beta_n$  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

$$\begin{aligned} E \inf_{A \cap B} \left( -\ln \frac{\bar{\omega} + a_0 y_0^2 + \gamma y_0}{\bar{\omega} + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0 < 0}} \right. \\ \left. + z_0^2 \frac{\bar{\omega} + a_0 y_0^2 + \gamma y_0}{\bar{\omega} + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0 < 0}} \right) \\ \geq -E \sup_{A \cap B} \ln \frac{\bar{\omega} + a_0 y_0^2 + \gamma y_0}{\bar{\omega} + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0 < 0}} \\ + E \left( \frac{\bar{\omega} + a_0 y_0^2 + \gamma y_0}{\bar{\omega} + a_0 y_0^2 + \delta |y_0| 1_{z_0 < 0}} \right) \\ \geq -E \ln \left( \left( 1 + \frac{\gamma y_0}{\bar{\omega} + a_0 y_0^2} \right) 1_{z_0 < 0} \right) \\ + E \left( \left( 1 + \frac{\gamma y_0}{\bar{\omega} + a_0 y_0^2} \right) 1_{z_0 \geq 0} \right) > -\infty \end{aligned} \quad (36)$$

and that

$$\begin{aligned} E \inf_{A \cap B} \left( -\ln \frac{\bar{\omega} + a_0 y_0^2 + \gamma y_0}{\bar{\omega} + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0 < 0}} \right. \\ \left. + z_0^2 \frac{\bar{\omega} + a_0 y_0^2 + \gamma y_0}{\bar{\omega} + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0 < 0}} \right) \\ \leq 1 - E \left( \ln \left( 1 + \frac{\gamma y_0}{\bar{\omega} + a_0 y_0^2} \right) \right) 1_{z_0 \geq 0} \\ + E \left( 1 + \frac{\gamma y_0}{\bar{\omega} + a_0 y_0^2} \right) < +\infty \end{aligned} \quad (37)$$

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

$$E \ln \frac{\bar{\omega} + a_0 y_0^2 + \gamma y_0}{\bar{\omega} + a_0 y_0^2 + \gamma^* |y_0| 1_{z_0 < 0}} \geq E \ln \left( 1 + \frac{\gamma y_0}{\bar{\omega} + a_0 y_0^2} \right) > -\infty. \quad (38)$$

Hence, Remark 6 implies that Assumption 5 holds with

$$\begin{aligned} c(\gamma, \gamma^*) &= -2E \ln |z| - E \ln \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0}{\bar{\omega} + a_0 \gamma_0^2 + \gamma^* |\gamma_0| 1_{z_0 < 0}} \\ &\quad + E \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0}{\bar{\omega} + a_0 \gamma_0^2 + \gamma^* |\gamma_0| 1_{z_0 < 0}}. \end{aligned} \quad (39)$$

Notice that  $c(0, \gamma^*)$  is uniquely minimized at  $\gamma^* = 0$  (see, e.g., Part 1. of the proof of Theorem 5.3.1. of [24] to obtain the analogous arguments along with the fact that  $((\bar{\omega} + a_0 \gamma_0^2)/(\bar{\omega} + a_0 \gamma_0^2 + \gamma^* |\gamma_0| 1_{z_0 < 0})) \stackrel{P \text{ a.s.}}{=} 1$  if and only if  $\gamma^* = 0$ ). When  $\gamma \neq 0$  then  $c(\gamma, -\gamma) < c(\gamma, 0)$  due to the fact that

$$\begin{aligned} &-E \ln \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0}{\bar{\omega} + a_0 \gamma_0^2 - \gamma |\gamma_0| 1_{z_0 < 0}} + E \ln \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0}{\bar{\omega} + a_0 \gamma_0^2} \\ &= E \left( \ln \left( 1 + \frac{\gamma \gamma_0}{\bar{\omega} + a_0 \gamma_0^2} \right) 1_{z_0 < 0} \right), \\ &E \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0}{\bar{\omega} + a_0 \gamma_0^2 - \gamma |\gamma_0| 1_{z_0 < 0}} - E \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0}{\bar{\omega} + a_0 \gamma_0^2} \\ &= -E \left( \frac{\gamma \gamma_0}{\bar{\omega} + a_0 \gamma_0^2} 1_{z_0 < 0} \right) < 0 \end{aligned} \quad (40)$$

and that when  $x > 0$ , then  $\ln(1 + x) < x$ . Furthermore, using the fact that by 3h is  $P$  a.s. two times dif and only if erentiable with respect to  $\gamma^*$  for  $\gamma^* \neq 0$  and since

$$\begin{aligned} &E \sup_{\gamma^*} \left| \left( \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma^* |\gamma_0| 1_{z_0 < 0}}{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0} - 1 \right) \right. \\ &\quad \left. \times \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0}{(\bar{\omega} + a_0 \gamma_0^2 + \gamma^* |\gamma_0| 1_{z_0 < 0})^2} |\gamma_0| 1_{z_0 < 0} \right| \\ &\leq \left( \frac{\delta}{|\gamma|} - 1 \right) \\ &\quad \times E \left| \frac{\bar{\omega} |\gamma_0| + a_0 |\gamma_0|^3 + \gamma \gamma_0^2}{(\bar{\omega} + a_0 \gamma_0^2)^2} 1_{z_0 < 0} \right| < +\infty, \\ &E \sup_{\gamma^*} \left| \left( \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0}{(\bar{\omega} + a_0 \gamma_0^2 + \gamma^* |\gamma_0| 1_{z_0 < 0})^2} \right) z_0^2 1_{z_0 < 0} \right| \\ &\leq \frac{1}{\bar{\omega}} E \left| \left( 1 + \frac{\gamma \gamma_0}{\bar{\omega} + a_0 \gamma_0^2} \right) 1_{z_0 < 0} \right| < +\infty, \end{aligned}$$

$$\begin{aligned} &E \sup_{\gamma^*} \left| \left( \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma^* |\gamma_0| 1_{z_0 < 0}}{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0} - 1 \right) \right. \\ &\quad \left. \times \frac{(\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0)^2}{(\bar{\omega} + a_0 \gamma_0^2 + \gamma^* |\gamma_0| 1_{z_0 < 0})^3} z_0^2 1_{z_0 < 0} \right| \\ &\leq \frac{1}{\bar{\omega}} \left( \frac{\delta}{|\gamma|} - 1 \right) E \left| \left( 1 + \frac{\gamma \gamma_0}{\bar{\omega} + a_0 \gamma_0^2} \right)^2 1_{z_0 < 0} \right| < +\infty \end{aligned} \quad (41)$$

as well as dominated convergence, we have that

$$\begin{aligned} \frac{\partial c(\gamma, \gamma^*)}{\partial \gamma^*} &= E \left( \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma^* |\gamma_0| 1_{z_0 < 0}}{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0} - 1 \right) \\ &\quad \times \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0}{(\bar{\omega} + a_0 \gamma_0^2 + \gamma^* |\gamma_0| 1_{z_0 < 0})^2} |\gamma_0| 1_{z_0 < 0} \end{aligned} \quad (42)$$

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

$$\begin{aligned} &\frac{\partial^2 c(\gamma, \gamma^*)}{\partial (\gamma^*)^2} \Big|_{\gamma^* = -\gamma} \\ &= E \left( \left( \frac{\bar{\omega} + a_0 \gamma_0^2 + \gamma \gamma_0}{(\bar{\omega} + a_0 \gamma_0^2 - \gamma |\gamma_0| 1_{z_0 < 0})^2} \right) z_0^2 1_{z_0 < 0} \right) > 0 \end{aligned} \quad (43)$$

establishing along with the previous that  $b(\gamma) = \{-\gamma\}$ . This validates simultaneously both Assumptions 11 and 17. Notice that Assumption 11 could also be verified by the use of Lemma 14 due to the continuity of  $\sigma_t^2$  and  $h_t$  with respect to the parameters, the existence of log moments, and the fact that  $h_t$  is uniformly bounded from below away from zero. Then Proposition 19(2\*) implies that any IE defined by Definition 15 is consistent for any  $\gamma_0 \in [-\delta, 0]$ .

## 4. 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 criteria involved in the relevant procedures, a relevant notion of continuity for the binding correspondence and an indirect identification condition that restricts the behavior of the aforementioned correspondence. These results are generalizations of the analogous ones in the relevant literature and 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 [30] 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  $\delta_u(b(\cdot), \beta_n^\#(\theta_0, \varepsilon_n^*))$  have totally bounded level sets and nonempty argmins.

Second, the present generalization is certainly *nonunique*. Again under stricter conditions on the behavior of the auxiliary criteria, possibly relevant to the ones in Proposition 3.42 of [31], the implication Proposition 8(2) could be strengthened to hold for any asymptotically null sequence of optimization errors. In this case in the definition of the IE  $\delta_u$  could be replaced by  $\delta$  and this would initially allow the identification condition in Assumption 17 to be replaced by the weaker “if  $\theta \neq \theta_0 \Rightarrow b(\theta_0) \neq b(\theta)$ .” If Assumption 11 was 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  $\kappa_n$  appearing in Proposition 20 is some sort of integral of  $\beta_n^\#$  (see, e.g., [32]) 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 dif and only iferent properties of potential polynomial approximations for dif and only iferent selections of  $b$  around  $\theta$  and so forth. The establishment of primitive conditions that guarantee the indirect identification Assumption 17 is also of separate interest.

## Appendices

### A. Some General Notions

In this section we briefly describe some general notions that are essentially used in the main body.

*A.1. Fell and Upper Topology.* Let  $(E, \tau_E)$  denote a topological space. We identify it with  $E$  when there is no risk of confusion. We denote by  $F_0(E)$  the set of closed nonempty subsets of  $E$ . We will use two topologies on  $F_0(E)$  constructed via  $\tau_E$  and the inclusion partial order on  $2^E$ . The first is entitled upper topology and we define it as the restriction of the topology in Definition 1.3.1 of [30] on  $F_0(E)$ .

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

$$\{ \cdot, G \} = \{ F \text{ closed} : F \subset G \}, \quad \forall G \in \tau_E, \text{ nonempty.} \quad (\text{A.1})$$

The upper topology is extremely useful for the analysis of the asymptotic behavior of sequences of sets of minimizers. If  $\tau_E$  is generated by a metric (say  $d$ ) with respect to which  $E$  is compact then  $\mathcal{T}_U$  is hemimetrizable (see Definition 4.1.3 for the term hemimetric and Proposition 4.2.2 of [30]) by  $\delta_u : F_0(E) \times F_0(E) \rightarrow \mathbb{R}$ , defined by  $\delta_u(A, B) = \inf\{\varepsilon > 0 : B \subset N_\varepsilon(A)\}$  where  $N_\varepsilon(A) = \{x \in E : d(x, A) < \varepsilon\}$ . (where  $d(x, A) \doteq \inf_{y \in A} d(x, y)$ ). Obviously, when  $B \subseteq A$

then  $\delta_u(A, B) = 0$ . The following properties of  $\delta_u$  will be useful in what follows.

**Lemma A.2.** *consider*

$$\delta_u(A, B) = 0 \quad \text{if and only if } B \subseteq A. \quad (\text{A.2})$$

*Proof.* Since  $A$  is closed if  $x \in B$  and  $x \notin A$ , then  $d(x, A) = \delta > 0$ . But then  $B \not\subseteq N_{\delta/2}(A)$  and therefore  $\delta_u(A, B) > (\delta/2)$ .  $\square$

**Lemma A.3.**  $\delta_u$  is a lower semicontinuous (lsc) real function with respect to the first argument.

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

**Lemma A.4.**  $\delta_u$  is an upper semicontinuous (usc) real function with respect to the second argument.

*Proof.* Suppose that  $B_n \rightarrow B$  with respect to the upper topology on  $F_0(E)$ ; then by the triangle inequality we have that  $\delta_u(A, B_n) \leq \delta_u(A, B) + \delta_u(B, B_n)$  establishing that  $\limsup_n \delta_u(A, B_n) \leq \delta_u(A, B)$ .  $\square$

The second topology on  $F_0(E)$ , known as the Fell topology, is defined by the use of the following subbase (see [22], paragraph 1.1, and [30], Definition 4.5.1).

*Definition A.5.* The Fell topology, say  $\mathcal{T}_F$ , is the smallest topology on  $F_0(E)$  consisting of both

- (1)  $F_G = \{F \text{ closed} : F \cap G \neq \emptyset\}$ ,  $\forall G \in \tau_E$ , nonempty,
- (2)  $F^K = \{F \text{ closed} : F \cap K = \emptyset\}$ ,  $\forall K \subset E$  nonempty and compact.

From Theorems 4.5.3-5 of [30] we have that when  $E$  is locally compact and Hausdorff then  $(F_0(E), \mathcal{T}_F)$  is locally compact and  $F_n \rightarrow F$  with respect to the Fell topology if and only if  $F = \text{Li}F_n = \text{Ls}F_n$  where  $\text{Li}F_n$  is the set comprised of the limit points of any possible sequence  $(x_n)$  such that  $x_n \in F_n$ , and  $\text{Ls}F_n$  is the one comprised of the analogous cluster points. Hence, in this case this type of convergence coincides with the Painleve-Kuratowski convergence (see among others, Appendix B of [22], or Definition 3.1.4. of [30]). 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  $\delta_u$ ; that is,  $\delta(A, B) = \max\{\delta_u(A, B), \delta_l(A, B)\}$  where  $\delta_l(A, B) \doteq \delta_u(B, A)$ . In this case we can prove the following lemma.

**Lemma A.6.** *If  $E$  is compact and metrized by  $d$ , then  $\delta_u$  is a lower semicontinuous (lsc) real function with respect to 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  $(A_n, B_n) \rightarrow (A, B)$  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  $\liminf_n \delta_u(A_n, B_n) \geq \delta_u(A, B)$ .  $\square$

**A.2. Epigraphs of Semicontinuous Functions and Epiconvergence.** Consider now the case where  $E$  is locally compact and Hausdorff and  $c : E \rightarrow \overline{\mathbb{R}}$ .

**Definition A.7.** The epigraph of  $c$  is

$$\text{epi}(c) = \{(x, t) \in E \times \mathbb{R} : c(x) \leq t\}. \quad (\text{A.3})$$

Note that despite the fact that the image of  $c$  may include infinities,  $\text{epi}(c)$  is by definition a subset of  $E \times \mathbb{R}$ . If and only if  $c$  is lower semicontinuous (lsc) we have that, due to Proposition A.2 of [22],  $\text{epi}(c) \in F(E \times \mathbb{R})$  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 A.8.** A sequence  $(c_n)$  of lsc functions epiconverges to  $c$  ( $c_n \xrightarrow{e} c$ ) if and only if  $\text{epi}(c_n) \rightarrow \text{epi}(c)$  with respect to the Fell topology.

It is easy to see that uniform convergence implies epiconvergence (see, e.g., Remark 6 above). Furthermore the relevant set of lsc is closed with respect to 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 [22] along with Theorem 7.1.4 of [30], Definition D.1 and Proposition D.2 of [22]).

**A.3. 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  $\Omega$  to  $F_0(E)$  (i.e., a closed valued multifunction with domain the set  $\Omega$ ), 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  $\{(\omega, F) : F = c(\omega)\}$  inside  $\Omega \times F_0(E)$ . When  $c(x)$  is compact for any  $x$ , then the correspondence is obviously termed as compact valued. In the following we do not make explicit distinction between the correspondence and the underlying multifunction.

The Borel  $\sigma$ -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 [22]). If  $(\Omega, \mathcal{F})$  is a measurable space, then  $c$  is a random closed set if and only if  $\{\omega \in \Omega : c(\omega) \in \overline{F}\} \in \mathcal{F}$  for any  $\overline{F} \in \mathcal{B}(E_F)$ . Analogously we abbreviate by  $\mathcal{B}(E_U)$  the Borel  $\sigma$ -algebra on  $F_0(E)$  generated by  $\mathcal{T}_U$  and by  $\mathcal{B}(E_U \times E_F)$  the Borel  $\sigma$ -algebra on  $F_0(E) \times F_0(E)$  generated by the product topology described in Lemma A.6. Finally denote by  $\mathcal{B}(\mathbb{R})$  the Borel  $\sigma$ -algebra of the extended real numbers with respect to the usual topology.

**Lemma A.9.** *If  $E$  is compact, separable, and metrized by  $d$ , then  $\delta_u$  is  $\mathcal{B}(\overline{\mathbb{R}})/\mathcal{B}(E_U) \otimes \mathcal{B}(E_F)$  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  $\{x_n, n = 0, 1, \dots\}$  is dense in  $E$  then the countable subset of  $F_0(E)$ ,  $\{\{x_n\}, n = 0, 1, \dots\}$  intersects any basic open set with respect to either topology. This implies the separability of  $F_0(E) \times F_0(E)$  when equipped with the topology discussed in Lemma A.6. This in turn implies that the Borel  $\sigma$ -algebra with respect to the product topology on  $F_0(E) \times F_0(E)$  coincides with  $\mathcal{B}(E_U) \otimes \mathcal{B}(E_F)$  by Lemma 1.4.1. of [33]. The rest follows by Lemma A.6 along with the fact that the sets in the subbase of the upper topology of  $\overline{\mathbb{R}}$  generate  $\mathcal{B}(\mathbb{R})$ . (It is also possible to prove that in the context of separability  $\mathcal{B}(E_U) = \mathcal{B}(E_F)$ ).  $\square$

## References

- [1] A. A. Smith, "Estimating nonlinear time-series models using simulated vector autoregressions," *Journal of Applied Econometrics*, vol. 8, no. 1, pp. 63–84, 1993.
- [2] C. Gourieroux, A. Monfort, and E. Renault, "Indirect inference," *Journal of Applied Econometrics*, vol. 8, no. 1, pp. 85–118, 1993.
- [3] R. A. Gallant and G. E. Tauchen, "Which Moments to Match?" Working Papers 95-20, Duke University, Department of Economic, 1995.
- [4] G. Calzolari, G. Fiorentini, and E. Sentana, "Constrained indirect estimation," *Review of Economic Studies*, vol. 71, no. 4, pp. 945–973, 2004.
- [5] A. R. Gallant, D. Hsieh, and G. Tauchen, "Estimation of stochastic volatility models with diagnostics," *Journal of Econometrics*, vol. 81, no. 1, pp. 159–192, 1997.
- [6] R. Garcia, E. Renault, and D. Veredas, "Estimation of stable distributions by indirect inference," *Journal of Econometrics*, vol. 161, no. 2, pp. 325–337, 2011.
- [7] T. G. Andersen, L. Benzoni, and J. Lund, "An empirical investigation of continuous-time equity return models," *Journal of Finance*, vol. 57, no. 3, pp. 1239–1284, 2002.
- [8] R. Bansal, A. R. Gallant, R. Hussey, and G. Tauchen, "Non-parametric estimation of structural models for high-frequency currency market data," *Journal of Econometrics*, vol. 66, no. 1-2, pp. 251–287, 1995.
- [9] C.-S. Chung and G. Tauchen, "Testing target-zone models using efficient method of moments," *Journal of Business & Economic Statistics*, vol. 19, no. 3, pp. 255–277, 2001.
- [10] A. Michaelides and S. Ng, "Estimating the rational expectations model of speculative storage: a Monte Carlo comparison of three simulation estimators," *Journal of Econometrics*, vol. 96, no. 2, pp. 231–266, 2000.
- [11] C. Gouriéroux, P. C. B. Phillips, and J. Yu, "Indirect inference for dynamic panel models," *Journal of Econometrics*, vol. 157, no. 1, pp. 68–77, 2010.
- [12] A. R. Gallant and J. R. Long, "Estimating stochastic differential equations efficiently by minimum chi-squared," *Biometrika*, vol. 84, no. 1, pp. 125–141, 1997.
- [13] C. Gourieroux and A. Monfort, *Simulation-Based Econometric Methods*, CORE Lectures, Ox. Un. Press, 1996.
- [14] R. A. Chumacero, "Estimating ARMA models efficiently," *Studies in Nonlinear Dynamics and Econometrics*, vol. 5, pp. 103–114, 2001.
- [15] I. Ghysels, L. Khalaf, and C. Vodounou, "Simulation based inference in moving average models," *Annales d'Economie et de Statistique*, vol. 69, pp. 85–99, 2003.

- [16] A. Demos and D. Kyriakopoulou, *Edgeworth Expansions for the MLE And MM Estimators of An MA(1) Model*, Communications in Statistics-Theory and Methods, Forthcoming, 2008.
- [17] P. C. B. Phillips, "Folklore theorems, implicit maps, and indirect inference," *Econometrica*, vol. 80, no. 1, pp. 435–454, 2012.
- [18] J. Dupačová and R. Wets, "Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems," *The Annals of Statistics*, vol. 16, no. 4, pp. 1517–1549, 1988.
- [19] C. Hess, "Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator," *The Annals of Statistics*, vol. 24, no. 3, pp. 1298–1315, 1996.
- [20] P. Lachout, E. Liebscher, and S. Vogel, "Strong convergence of estimators as  $\epsilon_n$ -minimisers of optimisation problems," *Annals of the Institute of Statistical Mathematics*, vol. 57, no. 2, pp. 291–313, 2005.
- [21] C. D. Aliprantis and K. C. Border, *Infinite-Dimensional Analysis*, Springer, Berlin, Germany, 2nd edition, 1999.
- [22] I. Molchanov, *Theory of Random Sets*, Springer, London, UK, 2005.
- [23] C. Choirat, C. Hess, and R. Seri, "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, 2003.
- [24] D. Straumann, *Estimation in Conditionally Heteroscedastic Time Series Models*, vol. 181 of *Lecture Notes in Statistics*, Springer, Berlin, Germany, 2005.
- [25] J. Pfanzagl, "On the measurability and consistency of minimum contrast estimates," *Metrika*, vol. 14, pp. 249–272, 1969.
- [26] T. R. Rockafellar and J.-B. Wets, *Variational Analysis*, Springer, 1997.
- [27] D. Applebaum, *Lévy Processes and Stochastic Calculus*, vol. 93 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge, UK, 2004.
- [28] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.
- [29] S. Arvanitis and Al. Louka, "Limit Theory for the QMLE of the GQARCH(1,1) model," Discussion Paper Series, Department of Economics, AUEB, 2012, <http://www.econ.aueb.gr/Discussion-Papers2012.html>.
- [30] E. Klein and A. C. Thompson, *Theory of Correspondences*, Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley & Sons, New York, NY, USA, 1984.
- [31] R. T. Rockafellar and R. J.-B. Wets, "Variational systems, an introduction," in *Multifunctions and Integrands*, vol. 1091, pp. 1–54, Springer, Berlin, Germany, 1984.
- [32] Z. Artstein and J. A. Burns, "Integration of compact set-valued functions," *Pacific Journal of Mathematics*, vol. 58, no. 2, pp. 297–307, 1975.
- [33] R. van der Vaart, W. Aad, and J. A. Wellner, *Weak Convergence and Empirical Processes*, Springer, 2000.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

