Stochastic Expansions and Moment Approximations for Three Indirect Estimators

Full Paper (pdf)

Stelios Arvanitis and Antonis Demos

Athens University of Economics and Business

Fall 2011

Extended Abstract

1 Introduction

This paper is concerned with the *derivation* of higher order asymptotic properties of indirect estimators with a view towards their characterization in terms of their approximate *bias and mean squared error (*MSE*).*

Indirect estimators (IE) usually emerge from two-step optimization procedures. They were formally introduced by Gourieroux, Monfort and Renault [4]. They are defined as (potentially measurable selections of approximate) minimizers of criteria (inversion criterion) that are functions of an auxiliary estimator (denoted by β_n), itself derived as an extremum estimator. The latter minimizes a criterion function (auxiliary criterion), that reflects (part of) the structure of a possibly misspecified auxiliary model. The inversion criterion, depends on a function connecting the underlying statistical models and termed as *the binding function.* Minimization of the inversion criterion, which usually has the form of a stochastic norm, essentially inverts the binding function.

Given an auxiliary estimator, IE differ due to differences in the inversion criteria that hinge on differences between the binding functions that each one involves. Among the IE involving the same auxiliary estimator, the consistent ones depend on sequences of binding functions that converge appropriately to a common limit binding function (denoted by b) that satisfies some identification condition.[1] In these cases, the auxiliary estimator, also converges in a similar manner to the value of the limit binding function at the true parameter value, hence consistency follows from identification. More refined asymptotic properties may be different across the particular IE, essentially due to differences between the involved sequences of binding functions.

2 The Estimators

For a measurable space $(\Omega, \mathcal{F})$, we suppose that the statistical model (SM) is a compact family of probability distributions on $\mathcal{F}$ when equipped with the topology of weak convergence. Furthermore, we assume that there exists a homeomorphism $\text{par}(\cdot)$ onto $\Theta \subset \mathbb{R}^p$ for some $p \in \mathbb{N}$. $\theta_0 = \text{par}(P_0) \in \text{Int}(\Theta)$, for P_0 in SM.

The auxiliary estimator is defined as a minimizer of a criterion formed as the norm of a measurable function Q_n with values on a finite dimensional

[1] For example let b be injective.

Euclidean space.

Definition D.1 *The auxiliary estimator is defined as*

$$\beta_n = \arg\min_{\beta \in B} Q_n(\beta)$$

Q_n could be a likelihood function, a GMM or more generally, a distance type criterion like the ones appearing in the following definitions. Given β_n the indirect estimators are defined as minimum distance ones. In our setup the relevant distances are represented by norms with respect to positive definite matrices (denoted by W_n, W_n^* and W_n^{**}). As in the context of GMM estimation, we allow these to be stochastic, and/or depend on initial estimators, say β_n^* or θ_n^*. We term this general framework as *stochastic weighting.* We consider the following IE.

Definition D.2 *The* GMR1 *estimator is defined as*

$$\theta_n = \arg\min_{\theta \in \Theta} \|\beta_n - b(\theta)\|_{W_n^*(\theta_n^*)}$$

Given appropriate assumptions $\|E_\theta \beta_n\| < \infty$ on Θ. This enables the definition of GMR2.

Definition D.3 *The* GMR2 *estimator is defined as*

$$\theta_n = \arg\min_{\theta \in \Theta} \|\beta_n - E_\theta \beta_n\|_{W_n^*(\theta_n^*)}$$

The last one denoted by GT and proposed by Gallant and Tauchen [3] is definable when Q_n is differentiable on B for $P_\theta-$almost every $\omega \in \Omega$. We denote with c_n the derivative of Q_n except for the case where $Q_n = \|c_n(\beta)\|_{W_n(\beta_n^*)}$, where $c_n : \Omega \times B \to \mathbb{R}^l$ is appropriately measurable. Under suitable conditions $\|E_\theta(c_n(\beta_n))\|$ is well defined. Consequently:

Definition D.4 *The* GT *estimator is defined as*

$$\theta_n = \arg\min_{\theta \in \Theta} \|E_\theta(c_n(\beta_n))\|_{W_n^{**}(\theta_n^*)}$$

The usual definition of the aforementioned estimator is given only when the auxiliary estimator is the MLE of the auxiliary model. The current one is obviously an extension. The computation of all three estimators relies on the analytical form of the binding function or the engaged expectations, which are usually intractable. Hence in applications these estimators are usually approximated by numerical procedures featuring re-sampling. It is easily seen that the Monte Carlo (or bootstrap) counterpart of the GMR2 estimator is the one associated with the maximal numerical burden among the three. In the following θ_n denotes either GMR1 or GMR2 or GT unless otherwise specified.

3 Higher Order Properties

The higher order asymptotic properties of the aforementioned estimators are established when their distributions are locally (around θ_0) uniformly, *weakly* approximated by Edgeworth measures. The n^{th} element of a sequence of Edgeworth measures of order $s^* \in \mathbb{N}$, depending on θ, is a measure on $\mathbb{R}^p$ that has a density of the form $\left(1+\sum_{i=1}^{s^*}\frac{\pi_i(z,\theta)}{n^{\frac{i}{2}}}\right)\varphi_{V(\theta)}(z)$ where π_i are polynomials in z, and $\varphi_{V(\theta)}$ is the normal density with zero mean and variance the $p\times p$ matrix $V(\theta)$.[2] If $a^* = \frac{s^*-1}{2}$, $\overline{\mathcal{O}}(\theta_0)$ is a closed neighborhood of θ_0 and $\mathcal{B}$ is the Borel algebra on $\mathbb{R}^p$, then the aforementioned locally uniform weak approximation of order s^* is defined by

$$\sup_{\theta\in\overline{\mathcal{O}}(\theta_0)}\left|P_\theta\left(\sqrt{n}\left(\theta_n-\theta\right)\in A\right)-\int_A\left(1+\sum_{i=1}^{s^*}\frac{\pi_i(z,\theta)}{n^{\frac{i}{2}}}\right)\varphi_{V(\theta)}(z)\,dz\right| = o\left(n^{-a^*}\right)$$

for any $A\in\mathcal{B}$.[3] The validity of such approximations can be established by conditions on the dependence of the random elements involved, the existence of integrals of appropriate functions w.r.t. the underlying probability measures and the smoothness of the procedures under which the auxiliary and the IE are defined. Given an assumption framework of this form-denote it as AFEV (s^*)-we prove that:

Lemma 3.1 *Under* AFEV (s^*) $\sqrt{n}\left(\theta_n-\theta\right)$ *has an Edgeworth approximation of order* s^* *uniformly on* $\overline{\mathcal{O}}_\varepsilon(\theta_0)$*.*

The Edgeworth measures involved generally *differ* between the employed IE, thereby establishing differing higher order asymptotic properties. If s^* is large enough and since Θ is bounded we can prove the following:

Lemma 3.2 *Suppose that* K *is a* m*-linear real function on* $\mathbb{R}^p$*, under* AFEV (s^*) *when* $s^* = 2a+m+1$ *then*

$$\sup_{\theta\in\overline{\mathcal{O}}(\theta_0)}\left|\int_{\mathbb{R}^p}K\left(z^m\right)\left(dP_n(\theta)-\left(1+\sum_{i=1}^{s}\frac{\pi_i(z,\theta)}{n^{\frac{i}{2}}}\right)\varphi_{V(\theta)}dz\right)\right| = o\left(n^{-a}\right)$$

where $P_n(\theta)$ *denotes the distribution of* $\sqrt{n}\left(\theta_n-\theta\right)$ *under* P_θ*.*

This enables the approximation of moments of $\sqrt{n}\left(\theta_n-\theta\right)$ from the analogous moments of the Edgeworth measures. We compute these approximations for $a=\frac{1}{2}$. One case is of particular importance:

[2]It is not generally a probability measure.

[3]It is obvious that when it exists, such an approximation is not unique.

Lemma 3.3 *Under* AFEV (s^*) *for* $s^* \geq 3$ *when* W^* *is non stochastic and independent of* θ *and* b *is affine then*

$$\sup_{\theta \in \overline{\mathcal{O}}(\theta_0)} \left\| E_\theta \sqrt{n} \left(\theta_n - \theta\right) - \frac{\xi\left(\theta\right)}{\sqrt{n}} \right\| = o\left(n^{-\frac{1}{2}}\right)$$

where $\xi(\theta) = 0_p$ iff $\theta_n = \text{GMR2}$*.*

Hence under these circumstances the GMR2 is second order locally uniformly unbiased[4] an important property not shared by its counterparts. The following lemma whose validity also emerges from 3.2 implies that this property is *not* at a cost of an augmented approximate MSE of the same order.

Lemma 3.4 *Under* AFEV (s^*) *for* $s^* \geq 4$ *when* W^* *is non stochastic and independent of* θ *and* b *is affine then*

$$\sup_{\theta \in \overline{\mathcal{O}}(\theta_0)} \left\| E_\theta \left(n \left(\theta_n - \theta\right) \left(\theta_n - \theta\right)'\right) - H_1\left(\theta\right) - \frac{H_2\left(\theta\right)}{\sqrt{n}} \right\| = o\left(n^{-\frac{1}{2}}\right)$$

where $H_1(\theta)$ *and* $H_2(\theta)$ *do not depend on the IE.*

The previous establish the superiority of the GMR2 estimator making it a suitable candidate for bias correction when the underlying statistical models coincide and β_n is consistent, whence b is the identity function, in which case the H_1 and H_2 are equal to the analogous of β_n.

4 Recursive GMR2

The previous section highlights the fact that the second order bias of the GMR2 estimator depends on the local to θ_0 behavior of the binding function. Due to theorem 10.2 of Spivak [6] (p. 44) B *can always be chosen* so that the binding function b is of the form $\begin{pmatrix} \theta \\ \mathbf{0}_{q-p} \end{pmatrix}$ at least in a small enough neighborhood of θ_0. This along with non stochastic weighting imply that there always exists an auxiliary parametrization such that the GMR2 estimator is second order unbiased. Usually, the re-parametrization of the auxiliary model is *analytically intractable*.

[4]We term an IE locally (around θ_0) uniformly unbiased of order s, if

$$\sup_{\theta \in \overline{\mathcal{O}}(\theta_0)} \left\| E_\theta \sqrt{n} \left(\theta_n - \theta\right) \right\| = o\left(n^{-\frac{s-1}{2}}\right)$$

However there exists at least one indirect estimation procedure that can be employed in order to approximate this *canonical* parameterization. Given the GMR1, let $\beta_n' = (\text{GMR1}', 0_{q-p})$and apply the GMR2 estimator to the latter. Then the resulting indirect estimator is derived from a three-step procedure, in the last step of which the binding function is obviously $\left(\theta_1, \theta_2, \ldots, \theta_p, \underbrace{0, \ldots, 0}_{q-p}\right)'$.

An extension of the three step procedure of the previous remark to an arbitrary number of steps, where the i^{th}-step auxiliary estimator is the the GMR2 of the previous step embedded to $\mathbb{R}^q$, can provide an unbiased indirect estimator of arbitrary order when i is large enough. Obviously, the embedding of the auxiliary estimator in any step after the first to $\mathbb{R}^q$ is irrelevant and therefore will be dropped.

We define recursive indirect estimation procedures as follows. Let $\theta_n^{(0}$ denote any estimator of θ.

Definition D.5 *Let $\zeta \in \mathbb{N}$, the recursive $\zeta -$ GMR2 estimator (denoted by $\theta_n^{(\zeta}$) is defined in the following steps:*

1. $\theta_n^{(1} = \arg\min_\theta \left\| \theta_n^{(0} - E_\theta \theta_n^{(0} \right\|$,

2. *for* $\zeta > 1$ $\theta_n^{(\zeta} = \arg\min_\theta \left\| \theta_n^{(\zeta-1} - E_\theta \theta_n^{(\zeta-1} \right\|$.

We prove the following lemma.

Lemma 4.1 *Under* AFEV (s^*) *for* $s^* \geq 2\zeta + 3$ *the* $\zeta -$GMR2 *estimator is of order* $s = 2\zeta + 1$ *unbiased and has the same* MSE *with the* $(\zeta - 1) -$ GMR2*, up to* 2ζ *order, uniformly on* $\overline{\mathcal{O}}_\varepsilon(\theta_0)$.

When $\zeta = 1$ we partially strengthen the previous results, since $1-$GMR2 is actually 3^{rd} order unbiased at θ_0. Furthermore, the $1-$GMR2 has the same second order MSE as the $1-$GMR2 one. Using Andrews [1] and Gourieroux et al. [5] we obtain a characterization of iterative bootstrap procedures as approximations of GMR2 type estimators.

5 Further Research

The previous results motivate some possible further extensions. First, the derivation of the analogous approximations when the true parameter value and/or its image w.r.t. the binding function lie on the boundary of the parameter spaces (see Calzolari et al. [2]). This could also imply the first order asymptotic non-equivalence between the three IE. Second, an application of the Edgeworth approximations could lay in the derivation of higher order properties of indirect testing procedures. Third, the introduction of indirect estimators via the actual *use* of the Edgeworth approximations for the auxiliary one. For example, an indirect estimator could be defined by substituting $E_\theta \beta_n$ with $\int_{\mathbb{R}^p} z \frac{\pi_2(z,\theta)}{n} \varphi_{V(\theta)} dz$ in the definition of the of GMR2 estimator. We leave all these questions for future work.

References

[1] Andrews D.W.K. (2002), "Equivalence of the Higher-order Asymptotic Efficiency of k-step and Extremum Statistics", *Econometric Theory*, 18, pp. 1040–1085.

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

[3] Gallant R.A. and G.E. Tauchen (1996), "Which Moments to Match," *Econometric Theory*, Vol. 12, pp. 657-681.

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

[5] Gourieroux C., E. Renault and N. Touzi, (2000), "Calibration by simulation for small sample bias correction", in "*Simulation-Based Inference in Econometrics: Methods and Applications*", Cambridge University Press, Cambridge.

[6] Spivak Michael (1999), "*A Comprehensive Introduction to DIFFERENTIAL GEOMETRY*", Vol. 1, Publish or Perish, Inc, Texas.