A New Class of Indirect Estimators and Bias Correction

Stelios Arvanitis and Antonis Demos
Athens University of Economics and Business

July 22, 2010

Abstract

In this paper we define a set of indirect estimators based on moment approximations of the auxilary estimators. We provide results that describe higher order asymptotic properties of these estimators. The introduction of these is motivated by reasons of analytical and computational facilitation. We extend this set to a class of multistep indirect estimators that have potentially useful higher order bias properties. Furthermore, the widely employed "feasibly biased corrected estimator" is an one optimazation step approxiamtion of the suggested one.

KEYWORDS: Indirect Estimator, Asymptotic Approximation, Moment Approximation, Higher Order Bias Structure, Binding Function, Local Canonical Representation, Convex Variational Distance.

1 Introduction

Indirect estimators, hereafter abbreviated as IE, are multistep extremum statistics derived in the premises of a (semi-) parametric statistical model (say $\mathcal{M}$) used for the estimation of a particular element of the model, termed as the *true parameter value*.[1] They were formally introduced by Gourieroux Monfort and Renault (1993). They are defined as (potentially measurable selections of approximate) 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 reflects

[1] This is usually a point in a topological space that is the image of the probability distribution with which the undelying probability space is endowed, with respect to a parameterization.

(part of) the structure of a possibly misspecified auxiliary model (say $\mathcal{A}$).[2] The inversion criterion, depends on a function on the set $\mathcal{A}^{\mathcal{M}}$ (or on the set $\mathcal{A}^{\mathcal{M}\times\Omega}$, where $(\Omega, \mathcal{F}, P)$ is a relevant probability space). This is termed *the binding function*. Minimization of the inversion criterion, which usually has the form of a stochastic norm, essentially inverts the binding function, thereby obtaining an estimator with values in $\mathcal{M}$.

Given an auxiliary estimator, IE differ due to relevant 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 that satisfies some identification condition. 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 of the cases considered may be different across the particular IE, essentially due to the differences on the involved sequences on binding functions.

Moreover, it is usually the case that the binding functions are not analytically known, hence are *approximated numerically*. In some instances the derivation of particular IE involves *nested* numerical optimization procedures that impose a large numerical cost, a fact that potentially creates, among practitioners, unattractiveness towards them. The same IE under a more involved assumption framework also have attractive high order asymptotic properties,[3] that are not exploited due to the aforementioned numerical burden.

Part of the scope of the present paper, is the introduction of a class of (potentially multistep) IE, in which cases the binding functions depend on approximations of moments of the auxiliary estimator. These approximations when are analytically known essentially reduce the numerical cost of computation of the estimator. This can also remain the case when the particular moment approximations are also approximated numerically. Under a relevant assumption framework, higher order asymptotic properties of these estimators are potentially similar to the ones mentioned in the previous paragraph. Hence this class of estimators can surpass the computational burden without sacrificing useful properties. It can also be verified, that these properties partly reside on a particular algebraic structure that the set of these estimators is "naturally" equipped with.

[2] $\mathcal{A}$ could simply be a reparameterization of $\mathcal{M}$.

[3] See e.g. Gourieroux and Monfort (1996), and Gourieroux, Renault and N. Touzi (2000).

The analysis of higher order asymptotic properties of the aforementioned class of IE, along with already established results, provides us with an interesting unification of distinct procedures of (potentially approximate) bias correction.

First, it is already established by Gourieroux, Renault and Touzi (2000) that the indirect estimator proposed by Gourieroux et al. (1993), derived as the solution of $\theta_n - E_\theta\theta_n = \mathbf{0}$, is approximately unbiased, while it is exactly unbiased if $E_\theta\theta_n$ is linear w.r.t. θ. Analogous properties hold for the estimator defined as $\theta_n - E_{\theta_n}\theta_n$ which can, under relevant conditions, be approximated by a bootstrap procedure. Gourieroux et al. (2000) show that the latter coincides with the estimator derived by the sequential Newton-Raphson approximation of the solution of $\theta_n - E_\theta\theta_n = \mathbf{0}$ when it is restricted to halt upon the completion of the first step. Hence they interpret the bootstrap estimator as a one step numerical approximation of the indirect one with equivalent second order properties.

In a direct analogy, when the previous framework is considered, the IE proposed in this paper are essentially derived as solutions of $\theta_n - \theta - K(\theta, a) = \mathbf{0}$, where $\theta + K(\theta, a)$ is an approximation of $E_\theta\theta_n$ in an appropriate sense. Under relevant conditions, $K(\theta, a)$ would converge uniformly as $a \to \infty$ to $E_\theta\theta_n$, hence these IE would converge in the appropriate sense to the one proposed by Gourieroux et al. (1993). Under the same conditions the former is also approximately unbiased of the same order. Again, a widely used estimator in the econometric literature when $K(\theta, a)$ is available, is $\theta_n - K(\theta_n, a)$, which is also approximately unbiased. It can be easily seen that under the same conditions, and as $a \to \infty$, due to the aforementioned uniform convergence $\theta_n - K(\theta_n, a)$ would converge to the bootstrap estimator, while it can also interpreted as a one step numerical approximation of the zero of $\theta_n - \theta - K(\theta, a)$. Hence if the one step Newton-Raphson approximation is considered as an appropriate self function on the relevant space of estimators we obtain that the diagram shown below, with the obvious choice of notation, commutes, while the relevant higher order properties of **zero** $(\theta_n - E_\theta\theta_n)$ are retained across it.

$$\begin{array}{ccc} \mathbf{zero}(\theta_n - \theta - K(\theta, a)) & \xrightarrow{a\to\infty} & \mathbf{zero}(\theta_n - E_\theta(\theta_n)) \\ {\scriptstyle 1-NR}\big\downarrow & & \big\downarrow{\scriptstyle 1-NR} \\ \theta_n - K(\theta_n, a) & \xrightarrow{a\to\infty} & \theta_n - E_{\theta_n}(\theta_n) \end{array} \tag{1}$$

Second, under appropriate conditions, $E_\theta\theta_n$ can be expressed using an auxiliary reparameterization that depends on n, as the identity function when restricted at an open neighborhood of the *true parameter value.* In this case the IE proposed by Gourieroux et al. (1993) is unbiased. However the (sequence of) auxiliary reparametrization(s) is (are) usually analytically intractable. The same is true for $K(\theta, a)$, while it can be shown that when $K(\theta, a)$ is locally the identity for any a, when $a \to \infty$ converges to the aforementioned local canonical representation of $E_\theta \theta_n$. We approximate the canonical representations of $K(\theta, a)$ using multistep procedures of indirect estimation, where the number of steps depend on a. In this respect, although the arbitrarily close approximation of the unbiased IE remains infeasible, we are able to construct estimators that are approximately unbiased of any given order.

Before the discussion of the framework on which the current results are based upon, in section 2, notice that indirect inference algorithms were initially used by Smith (1993), were formally introduced by Gourieroux et al. (1993), complemented by Gallant and Tauchen (1996) and extended by Calzolari, Fiorentini and E. Sentana (2004). Properties similar to those studied here were more or less algebraically studied in Gourieroux et al. (2000) and more formally in Arvanitis and Demos (2010). In section 3 we define the estimators and derive their asymptotic properties in the following one. In section 5 we extend the procedures to multi step ones, and apply them in two examples presented in section 6. Conclusions are gathered in section 7 and in the appendix A we collect all proofs. In appendix B we present some useful tools concerning the derivation of our results and in appendix C we gather the calculations of the expansions employed in our examples.

2 General Framework

In this paragraph a general assumption framework is described, that facilitates the presentation of the already defined IE. This assumption framework can be generalized in particular ways, some of which are locally remarked. Then, two already defined IE are presented along with some of their properties and relations, that rely upon the particular assumption canvas.

Given two metric spaces, (X, d_X) and (Y, d_Y) we denote the set of the Lipschitz continuous functions from the first to the second, suppressing its dependence on the metrics, by $\mathbf{Lip}(X, Y)$. The symbol $B_\varepsilon(\theta)$ will denote the ε-ball around the point θ in a relevant metric space. We denote with D^r, the r^{th}- order derivative operator on a relevant function space that maps to the space of the algebraic element containing all the r^{th}-order partial derivatives of the first.

For a matrix W, $\|W\|$ will denote a *submultiplicative* matrix norm,[4]

[4]Notice that due to the fact that finite dimensional matrix spaces are identified with

such as the Frobenius norm (i.e. $\|W\| = \sqrt{trW'W}$). The relevant metric space of r-dimensional square real matrices is denoted by $M(\mathbb{R}, r)$. We let $\mathcal{PD}(\mathbb{R}, r) \subset M(\mathbb{R}, r)$ be the cone of positive definite real matrices of dimension r.

When suprema with respect to parameters, of derivatives are discussed these are obviously considered where the differentiated function is differentiable. Finally for $a \in A \doteqdot \left\{\frac{i}{2}, i = 0, 1, \ldots\right\}$, $s \in \{i + 1, i = 0, 1, \ldots\}$, and let $d = \max(2a + 2, 3)$.

Assumption A.1 The following characterize the basic framework:

1. Θ denotes a compact subset of $\mathbb{R}^p$ for some $p \in \mathbb{N}$, equipped with the relevant subspace topology. Let $\theta_0 \in Int(\Theta)$.

2. For each n except potentially for a finite subset of $\mathbb{N}$ and for a probability space $(\Omega, \mathcal{F}, P)$, let $x_n : \Omega \times \Theta \to \Omega_n$, be $\mathcal{F}_n/\mathcal{F}$ measurable functions $\forall \theta \in \Theta$, where $(\Omega_n, \mathcal{F}_n)$ are measurable spaces, $P_{\theta,n}(A) \doteqdot P \circ x_n^{-1}(A, \theta)$ for any $A \in \mathcal{F}_n$.

3. $M_n = \{P_{\theta,n}, \theta \in \Theta\}$ is topologized by the topology of total variation w.r.t. which it is homeomorphic to Θ and consists of the statistical model at hand.

4. The limit binding function (lbf) $b \in Lip(\Theta, \mathbb{R}^q)$ and corresponds to the relevant notion discussed previously. Also, we let $B = b(\Theta)$ and suppose that $b(\theta_0) = b(\theta)$ iff $\theta = \theta_0$, and for some $\varepsilon_1 > 0$, the restriction $b|_{B_{\varepsilon_1}(\theta_0)} : B_{\varepsilon_1}(\theta_0) \to B$ is invertible. Moreover, for some $\varepsilon_1 \geq \varepsilon_2 > 0$, the restriction $b|_{B_{\varepsilon_2}(\theta_0)}$ is a $d+1-$diffeomorphism.

5. There exists a function $c_n : \Omega_n \times B \to \mathbb{R}^l$ for some $l \in \mathbb{N}$ that is $\mathcal{B}_{\mathbb{R}^l}/(\mathcal{F}_n \otimes \mathcal{B}_B)-$measurable, and $Ec_n(x_n(\omega, \theta), \beta) = \mathbf{0}_{l\times 1}$, iff $\beta = b(\theta)$.

6. Let $W_n(\cdot, \beta)$ be $\mathcal{B}_{M(\mathbb{R},l)}/(\mathcal{B}_{\mathcal{F}_n} \otimes \mathcal{B}_\Theta)-$measurable and $P_{\theta_0,n}$-almost surely positive definite, for every $\beta \in B_\varepsilon(b(\theta_0))$ for some $\varepsilon > 0$.

7. Let $W_n^*(\cdot, \theta)$ be $\mathcal{B}_{M(\mathbb{R},q)}/(\mathcal{B}_{\mathcal{F}_n} \otimes \mathcal{B}_\Theta)-$measurable and $P_{\theta_0,n}$-almost surely positive definite, for every $\theta \in B_{\varepsilon_2}(\theta_0)$.

Remark R.1 We denote with $E_{\theta_1} f = \int f(x_n(\omega, \theta_1), \theta_2)\, dP(\omega)$ for any appropriately measurable f and any $\theta_2 \in \Theta$.

---

finite dimensional Euclidean spaces, the norm equivalence theorem applies.

Remark R.2 Usually Ω_n is homeomorphic to $\mathbb{R}^{nm}$ for some m in $\mathbb{N}$ and $\mathcal{F}_n$ is the Borel algebra with respect to the Euclidean topology.

Remark R.3 Since B and Θ are compact subsets of finite dimensional Euclidean spaces they are totally bounded. Also note that due to the fact that the spaces Θ and B are separable, *suprema* of real random elements over these spaces are *measurable*. Obviously the lbf is bounded.

Remark R.4 It is implied that $q \geq p$ and that $rank\left(\frac{\partial b}{\partial \theta'}\right) = p, \forall \theta \in B_{\varepsilon_2}(\theta_0)$. Also due to the local diffeomorphism assumption for the lbf, it is moreover implied that $\sup_\theta \|D^r b(\theta)\| < M_r, \forall r = 2, \ldots, d+1$ for $\theta \in B_{\varepsilon_3}(\theta_0)$, for some $\varepsilon_3 \leq \varepsilon_2$, with $M_r > 0$. Boundeness of the first derivatives follows also from $b(\theta)$ being Lipschitz on Θ and consequently on $B_{\varepsilon_3}(\theta_0)$.

Remark R.5 The estimating equations $c_n(x_n, \beta) = \mathbf{0}_{l\times 1}$ can be implied by some part of the structure of a, potentially, misspecified statistical model. This in turn is a locally differentiable parametric statistical model defined on the same measurable space, usually termed as *auxiliary model*, with B as its parameter space. In this case the lbf is a parametric representation of a relevant function between the two sets of probability measures. It should also be noted that the lbf and some of its posited properties, can be locally retrieved from conditions on $E_\theta c_n(x, b(\theta))$ that facilitate application of relevant implicit function theorems. For example, $b(\theta_0)$ is identified as the unique solution of $E_{\theta_0} c_n(x, \beta) = \mathbf{0}_{l\times 1}$, which along with the further local differentiability assumptions implies that $l \geq q$, and $rank\left(E_\theta \frac{\partial c_n(b(\theta))}{\partial \beta'}\right) = q$, $\forall \theta \in B_{\varepsilon_2}(\theta_0)$.

Remark R.6 In accordance with remark R.2 $c_n(x_n, \beta)$ is often of the form $\frac{1}{n}\sum_{i=1}^{n} c(x_i, \beta)$, for $x_i : \Omega \times \Theta \to \mathbb{R}^m$ for any i, and $c : \mathbb{R}^m \times B \to \mathbb{R}^l$. The same is often true for any of the stochastic matrices considered in assumption A.1.5,6. For example $W_n(x, \beta) = \frac{1}{n}\sum_{i=1}^{n} W(x_i, \beta)$ for $W : \mathbb{R}^m \times B \to M(\mathbb{R}, l)$ etc.

In the following we suppress the dependence of the aforementioned binding functions on Ω_n where unnecessary. We also let θ_n^+ and β_n^+ denote random elements with values in Θ and B respectively. We consider the following real function on $\mathbb{R}^r \times M(\mathbb{R}, r)$

$$(x, W) \to (x'Wx)^{1/2}$$

for a given $W \in M(\mathbb{R}, r)$. This defines a pseudo-norm on $\mathbb{R}^R$ which becomes a norm if $W \in \mathcal{PD}(\mathbb{R}, r)$.

Definitions and Properties of Already Known IE We can now define the already known auxiliary, GMR1, and GMR2 estimators. These were initially formalized by Gourieroux et al. (1993).

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

$$\|c_n(\beta_n)\|_{W_n(\beta_n^+)} = \inf_{\beta \in B} \|c_n(\beta)\|_{W_n(\beta_n^+)}$$

Remark R.7 In view of assumption A.1.3,5,6 and by remark AR.1 (in the Appendix) the above estimator is well defined. Notice that when $l = q$ and if $W_n(\beta_n^+) \in B_\varepsilon(W)$ with probability that tends to unity, $\forall \varepsilon > 0$ for $W \in \mathcal{PD}(\mathbb{R}, r)$, then β_n is asymptotically independent of the weighting matrix, with probability that tends to unity.

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

$$\|\beta_n - b(\text{GMR1})\|_{W_n^*(\theta_n^+)} = \inf_{\theta \in \Theta} \|\beta_n - b(\theta)\|_{W_n^*(\theta_n^+)}$$

Definition D.3 Let $b_n(\theta) = E_\theta \beta_n$, then the *GMR2* estimator is defined as

$$\|\beta_n - b_n(\text{GMR2})\|_{W_n^*(\theta_n^+)} = \inf_{\theta \in \Theta} \|\beta_n - b_n(\theta)\|_{W_n^*(\theta_n^+)}$$

Remark R.8 The GMR1 and GMR2 estimators are defined as $q\left(W_n^*\left(\theta_n^+\right), b(\theta), \beta_n\right)$ and $q\left(W_n^*\left(\theta_n^+\right), b_n(\theta), \beta_n\right)$ respectively where

$$q(A, k(\theta), c) \doteqdot \arg\min_{\theta \in \Theta} \circ J(c, k(\theta), A)$$

and

$$J(c, k(\theta), A) \doteqdot \|c - k(\theta)\|_A$$

Their existence is justified by remark AR.1 (in the Appendix) in view of assumption A.1.3,5,6. The computation of the estimators relies on the analytical knowledge of b and b_n which is in most cases unavailable. Hence the estimators are usually approximated by the use of Monte Carlo simulations, which itself involves *nested numerical optimizations* that is of potentially large computational cost especially in the case of the second estimator.

When $p = q = l$, $c(x, \beta_n) = h(x) - E_\beta h(x) = h(x) - g(\beta)$ and g is linear then $GMR1 = GMR2$ by lemma 2.3 of Arvanitis and Demos (2010). For reasons that will become apparent in the next section, we denote the $GMR1$ estimator by $\theta_n(0)$ and the $GMR2$ estimator by $\theta_n(\infty)$.

Assumptions Specific to a New Class of IE Let $a^* = \frac{s-1}{2}$ for $s \geq 2(a+1)$. Let $m_n(\theta)$ be $\left((\beta_n(x_n(\theta)) - b(\theta))', (\theta_n(0)(x_n(\theta)) - \theta)', (\theta_n^+(x_n(\theta)) - \theta)'\right)'$. Let also $\mathcal{EDG}_{\theta,a^*}$ denote the Edgeworth measure of order s defined on $\mathbb{R}^{q+2p}$ and $\mathcal{B}_C$ the collection of measurable convex sets on the same space.

Assumption A.2

$$\begin{aligned} \sup_{A\in\mathcal{B}_C} \left|P_{\theta,n}\left(\sqrt{n}m_n(\theta) \in A\right) - \mathcal{EDG}_{\theta,a^*}(A)\right| &= \\ \sup_{A\in\mathcal{B}_C} \left|P\left(\sqrt{n}m_n(x_n(\theta), \theta) \in A\right) - \mathcal{EDG}_{\theta,a^*}(A)\right| &= o\left(n^{-a^*}\right) \end{aligned}$$

for any θ on a bounded open set, say Θ', that contains Θ hence there exist $k_{i+1}(z,\theta)$, $i = 1, \ldots, 2a$ such that

$$\left\|E_\theta\beta_n - b(\theta) - \sum_{i=1}^{2a} \frac{1}{n^{(i+1)/2}} Ek_{i+1}(z,\theta)\right\| = o\left(n^{-a-\frac{1}{2}}\right) \tag{2}$$

where $z \sim N(\mathbf{0}_q, \mathbf{Id}_q)$ and $Ek_{i+1}(z,\theta)$ is $d+1$ differentiable on $B_{\varepsilon_2}(\theta_0)$.

Remark R.9 We implicitly assume that $Ek_1(z,\theta) = \mathbf{0}_{q\times 1}$ on Θ. This is essentially proven in lemma 3.5 of Arvanitis and Demos (2010) and it is attributed on the structure of the auxiliary estimating equations and the fact that θ_0 is in the interior of Θ. In the case that θ_0 is a boundary point, this could seize to be true.

Remark R.10 The relationship between Θ and Θ' implies that the moment approximation is also valid on the boundary of Θ.

Remark R.11 Due to assumption A.2 and lemma 2 of Magdalinos (1992) we have that,

$$P\left(\sqrt{n}\left\|m_n(\theta_0)\right\| > C_m\sqrt{\frac{\ln n}{n}}\right) = o\left(n^{-a^*}\right), \text{ for some } C_m > 0$$

Then it trivially follows that $P\left(\left\|m_n(\theta_0)\right\| > \varepsilon\right) = o\left(n^{-a^*}\right)$ $\forall\varepsilon > 0$.

Our next assumption concerns the asymptotic behavior of the sequence of stochastic weighting matrices described in A.1.5,6.

Assumption A.3 $W_n^*(x,\theta)$ is $d+1$-differentiable P_{0_n}-almost surely $\forall\theta \in B_{\varepsilon_2}(\theta_0)$. Moreover, there exists a $M(\mathbb{R}, l)$ valued function denoted by $W^*(\theta)$, defined on Θ which is positive definite and $d+1$-continuously differentiable $\forall \theta \in B_{\varepsilon_2}(\theta_0)$, such that

$$P\left(\sup_{\theta \in B_{\varepsilon_2}(\theta_0)} \|D^r W_n^*(\theta) - D^r W^*(\theta)\| > \varepsilon\right) = o\left(n^{-a^*}\right), \forall \varepsilon > 0, \forall r = 0, 1, \ldots, d+1$$

Remark R.12 If $c_n(x_n, \beta)$ is as in remark R.6, each of the aforementioned convergence is essentially a local ULLN that can be obtained by restrictions on the dependence of the elements of x_n, moment conditions imposed on the involved derivatives in the spirit of Lemma 3 of Andrews (2002) as well as conditions that imply uniform asymptotic equicontinuity with probability $1 - o\left(n^{-a^*}\right)$. More general cases could be reduced to the latter along with further uniformity conditions. In addition lemma AL.1 of Arvanitis and Demos (2010) along with assumption A.3 and remark R.11 imply that

$$P\left(\|W_n^*(\theta_n^*) - W^*(\theta_0)\| > \varepsilon\right) = o\left(n^{-a^*}\right), \forall \varepsilon > 0$$

The following assumptions, enable the *stochastic approximation* of the moment approximations in (2). This can facilitate the definition of the IE that depend on the latter, in the case where $Ek_{i+1}(z, \theta)$ are analytically unknown for some i, due to the structure of M_n that could involve the presence of nuisance parameters, analytically unknown moments in the framework of non linear models etc. We suppose the existence of another probability space that enables the possibility of stochastic approximation via Monte Carlo simulations.

Assumption A.4 The following characterize the basic framework:

1. There exists a probability space $(\Omega', \mathcal{F}', P')$, and $y_n : \Omega' \to \Omega'_n$, $\mathcal{F}'_n/\mathcal{F}'$ measurable functions, where $(\Omega'_n, \mathcal{F}'_n)$ are measurable spaces.

2. For each $i = 1, \ldots, 2a$, there exist $\zeta_{i+1_n} : \Omega_n \times \Omega'_n \times \Theta \to \mathbb{R}^q$, that is $\mathcal{B}_{\mathbb{R}^q}/\left(\mathcal{F}_n \otimes \mathcal{F}'_n \otimes \mathcal{B}_\Theta\right)$ Q-almost surely continuous on Θ and Q-almost surely $d+1$ differentiable on $B_{\varepsilon_2}(\theta_0)$, where $Q = P \times P'$.

3. $Q\left(\sup_{\theta \in \Theta} \left\|\zeta_{i+1_n}(x_n(\theta_0), y_n, \theta)\right\| > M_i\right) = o\left(n^{-a^*}\right)$, for $M_i > 0$, $\forall i = 1, \ldots, 2a$.

4. $Q\left(\sup_{\theta \in B_{\varepsilon_2}(\theta_0)} \left\|D^r \zeta_{i+1_n}(x_n(\theta_0), y_n, \theta)\right\| > M'_i\right) = o\left(n^{-a^*}\right)$, for $M_i > 0$, $\forall i = 1, \ldots, 2a$, for $r = 1, 2$.

y_n can be thought of as a simulated random element, which along with the "observed" sample $x_n(\theta_0)$ constitutes a generalized sample that can be employed to approximate the relevant expectations. The space Ω'_n can also depend on some index that indicates the number of simulated paths which is suppressed. In our framework we are only interested in the case that the number of simulated paths remains bounded, since we only consider the asymptotic theory that emerges as $n \to \infty$.

Remark R.13 Assumption A.4.3,4 essentially require global and/or local asymptotic uniform boundeness with probability $1 - o\left(n^{-a^*}\right)$ for $\left\{\zeta_{i+1_n}\right\}$ and the first two derivatives respectively. These could follow from analogous considerations to the ones in remark R.12 when applied to the extended random element $(x_n(\theta_0), y_n)$. Now, the dependence of ζ_{i+1_n} on $x_n(\theta_0)$ is general enough to allow for cases in which ζ_{i+1_n} is computed on initial estimators of θ_0, and/or on estimators of nuisance parameters. Similarly the dependence on $(x_n(\theta_0), y_n)$ allows for cases in which $Ek_{i+1}(z, \theta)$ depends on analytically intractable moments and/or moments that do not belong in the structure of the statistical model at hand. These are generally functions of θ and are approximated either by analogous sample moments w.r.t. relevant functions of y_n and θ, or their value at θ_0 is approximated by analogous sample moments of $x_n(\theta_0)$. This allows also for approximations of $Ek_{i+1}(z, \theta)$ when the latter is *partially computed* at θ_0, enabling the derivation of estimators that emerge from *partial optimization.*

Remark R.14 In the trivial case where $\zeta_{i+1_n} = Ek_{i+1}$ for any i, assumption A.4.3 follows from a relevant almost sure continuity assumption on the highest order derivative of the criterion function from which the auxiliary estimator emerges, or from an identical assumption on a criterion function $(J_n^*(\beta))$ such that $P\left(\sup_{\beta \in B} |J_n(\beta) - J_n^*(\beta)| > \upsilon_n\right) = o(n^{-a-1})$, for a real sequence $\upsilon_n = o(n^{-a-1})$ such that the metric space carried by the set $\{J_n, J_n^* : n \in \mathbb{N}\}$ is compact. Notice that due to the fact that Θ is compact, $\sup_{\theta \in \Theta} \|Ek_{i+1}(z, \theta)\| < M_i$, $\forall i = 1, \ldots, 2a$ and $Ek_{i+1}(z, \theta)$ is uniformly continuous on Θ, $\forall i = 1, \ldots, 2a$. Furthermore, assumption A.4.4 attributes the Lipschitzian property to the $D^r Ek_{i+1}(z, \theta)$, $\forall i = 0, \ldots, 2a$, for $r = 1, \ldots, d$ on $B_{\varepsilon_2}(\theta_0)$. Hence, by lemma 3.3 of Andrews (2002) there exists a $M_W > 0$ such that $P(\|W_n^*(\theta_n^*)\| > M_W) = o\left(n^{-a^*}\right)$.

We complete our assumption framework by an extension of assumption A.2 that allows for analogous moment approximations of the estimators to be defined in the next section. Let $f_n(\theta)$ be the vector containing the elements of $W_n^*(\theta) - W^*(\theta)$ and their derivatives up to order $d+1$ for any $\theta \in B_{\varepsilon_2}(\theta_0)$. Let $q_n(\theta)$ be the vector containing the elements of $\zeta_{i+1_n}(x_n(\theta), y_n, \theta) -$

$Ek_{i+1}(z,\theta)$ and their derivatives up to order $d+1$ at θ for all $i=1,\ldots,2a$, for any $\theta \in B_{\varepsilon_2}(\theta_0)$. Let also $m_n^*(\theta_0)$ be $(m_n'(\theta_0), f_n'(\theta_0), q_n'(\theta_0))'$.

Assumption A.5 $\sqrt{n}m_n^*(\theta_0)$ admits a valid Edgeworth approximation of order s.

Remark R.15 It is trivial that under assumption A.5

$$Q\left(\left\|D^r\zeta_{i+1_n}(\theta_0) - D^r Ek_{i+1}(z,\theta_0)\right\| > \varepsilon\right) = o\left(n^{-a^*}\right),\ \forall\varepsilon > 0,\ r = 0,1,2$$

In what follows we suppress the dependence of the approximating functions ζ_{i+1_n} on the generalized sample space for notational convenience.

3 Definition of the GMR2* (a) Estimators

We are now ready to define a new class of IE based on these moment approximations. For notational convenience let $\zeta_n(\theta,a) = \left(\zeta_{2_n}(\theta),\ldots,\zeta_{2a+1_n}(\theta)\right)$ and $b_n(\theta,\zeta_n(\theta,a)) = b(\theta) + \sum_{i=1}^{2a}\frac{1}{n^{(i+1)/2}}\zeta_{i+1_n}(\theta)$.

Definition D.4 The GMR2* (a) estimator is defined by

$$\left\|\beta_n - b_n(\theta_n(a),\zeta_n(\theta_n(a),a))\right\|_{W_n^*(\theta_n^+)} = \inf_{\theta\in\Theta}\left\|\beta_n - b_n(\theta,\zeta_n(\theta,a))\right\|_{W_n^*(\theta_n^+)}$$

hence GMR2* $(a) = q\left(W_n^*\left(\theta_n^+\right), b_n(\theta,\zeta_n(\theta,a)),\beta_n\right)$.

Remark R.16 The existence of $GMR2^*(a)$ is facilitated by assumptions A.4 and A.3, and remark AR.1 in the appendix.

Remark R.17 Due to remark R.9 we identify the $GMR1$ estimator with the $GMR2^*(0)$ one, and this justifies the relevant choice of notation.

Remark R.18 Due to the fact that the analytical derivation of $b_n(\theta,\zeta_n(\theta,a))$ for finite a is generally easier than the analogous task for $b_n(\theta)$ the $GMR2^*(a)$ estimators can surpass the nested optimization burden associated with the GMR2 estimator. Of course it increases the analytical burden, but this is a shank cost.

Remark R.19 In the case that $\beta_n = \theta_n(0)$, and $b(\theta) = \theta$, we consider a variant of the GMR2* (a), defined as
$\theta_n^*(a) = \theta_n(0) - \sum_{i=1}^{2a}\frac{1}{n^{(i+1)/2}}\zeta_{i+1_n}(\theta_n(0))$, almost surely, the computation of which is of minimal arithmetic burden. In this case $\theta_n^*(a)$ admits another interesting characterization. Consider without loss of generality the

issue of minimization of $\left\|\theta_n(0) - \theta - \sum_{i=1}^{2a} \frac{1}{n^{(i+1)/2}} \zeta_{i+1_n}(\theta_n(0))\right\|^2$. Due to the structure of the problem, the solution could be characterized as a limit of a Newton recursion scheme, in which the i^{th}-term of the recursion would be defined as $\theta_n^{(i)} = \theta_n(0) - \sum_{i=1}^{2a} \frac{1}{n^{(i+1)/2}} \zeta_{i+1_n}\left(\theta_n^{(i-1)}\right)$, for $i = 0, 1, 2, \ldots$, and $\theta_n^{(i-1)} = \theta_n(0)$. It is obvious that $\theta_n^*(a) = \theta_n^{(1)}$, hence it is an one-computational step approximation of GMR2$^*(a)$. $\theta_n^*(a)$ is widely used in the statistical literature in the case where $a = \frac{1}{2}$, and in this case it is called "feasibly bias correction" of $\theta_n(0)$. We will consider how some of its properties are related to the analogous ones of $GMR2^*(a)$ in subsequent sections. In this instance we note only the following:

1. it is possible that for some n and some measurable subset of $\mathbb{R}^m$ of positive probability, $\theta_n^*(a) \notin \Theta$ or it will be in the boundary of Θ with positive Q probability, as it will be the case in some of the examples considered later.

2. there is a direct analogy between the $GMR2^*(a)$ and $\theta_n^*(a)$ as its one-computational step approximation, and the $GMR2$ and the bootstrap estimator as its one-computational step approximation.

4 Higher Order Asymptotic Theory

In this section the first part of the results are presented. This part concerns the asymptotic properties of the newly defined estimator. Consistency, asymptotic tightness, Edgeworth and moment approximations are established in that order.

4.1 Consistency

It is proven that the GMR2$^*(a)$ is contained in an arbitrary neighborhood of θ_0 with probability $1 - o\left(n^{-a^*}\right)$. It is also shown, that given consistency, the particular estimator has a very convenient characterization as a near minimizer of the GMR1 and GMR2 criteria. Analogous relations are established between GMR2$^*(a)$ and GMR2$^*(a')$, for any a, a' in A.

Lemma 4.1 Under assumptions A.4, A.3 and A.2 $\forall \varepsilon > 0$,

$$Q\left(\sup_{\theta}\left|J\left(\beta_n, b_n(\theta, \zeta_n(\theta, a)), W_n^*\left(\theta_n^+\right)\right) - J\left(b(\theta_0), b(\theta), W^*(\theta_0)\right)\right| > \varepsilon\right) = o\left(n^{-a^*}\right)$$

and therefore

$$P\left(\|\theta_n(a) - \theta_0\| > \varepsilon\right) = o\left(n^{-a^*}\right)$$

Remark R.20 In the light of lemma 4.1 it is evident that for example, θ_n^* could be defined as $\theta_n^*(a)$ for some choice of the weighting matrix sequence (e.g. $W_n = \mathrm{Id}_{q\times q}$).

The $GMR2$ estimator θ_n is defined by

$$J\left(\beta_n, E_{\theta_n}(\beta_n), W_n^*\left(\theta_n^+\right)\right) = \inf_{\theta} J\left(\beta_n, E_{\theta}(\beta_n), W_n^*\left(\theta_n^+\right)\right)$$

From lemma 4.1 we obtain the following results. These concern possible characterizations of the estimator under examination. We employ first of the following proposition.

Proposition 4.2 If $\{b_n\}$ is equi-Lipschitz on Θ, then

$$Q\left(\sup_{\theta\in\Theta}\|b_n(\theta) - b_n(\theta, \zeta_n(\theta, a))\| > \varepsilon\right) = o\left(n^{-a}\right), \forall\varepsilon > 0$$

Remark R.21 The assumption of b_n being equi-Lipschitz on Θ, follows from assumption A.1 (see Arvanitis and Demos 2010, lemma 3.3).

Corollary 4.3 Under the assumptions of lemma 4.1 and proposition 4.2 we have that

$$J\left(\beta_n, E_{\theta_n(a)}(\beta_n), W_n^*\left(\theta_n^+\right)\right) \leq J\left(\beta_n, E_{\theta_n}(\beta_n), W_n^*\left(\theta_n^+\right)\right) + \eta_n$$

with $P(\eta_n > \varepsilon) = o\left(n^{-a^*}\right)$, $a^* = \frac{s-1}{2}$, $\forall\varepsilon > 0$ and η_n is almost surely non negative.

Remark R.22 The examined estimator is essentially an η_n-GMR2 estimator (*approximate minimizer of the GMR2 criterion*). The $\theta_n(\infty)$ estimator (if it exists) is almost surely equal to the GMR2 estimator for every n greater than some $n^* \in \mathbb{N}$. In the same respect, and in the light of paragraph 1.5 of Gourieroux et al. (2000), when β_n is a consistent estimator of θ_0, i.e. the binding function is, at least locally, the identity, we obtain that the $\theta_n^*(\infty)$ (if it exists) is almost surely equal to the bootstrap estimator for every n greater than some $n^* \in \mathbb{N}$. Hence, we obtain an analogy in which *the GMR2 estimator can be perceived as a limiting GMR2* estimator, and the bootstrap estimator, which is an one computational step approximation of the former is a limit of the one step computational approximation of the latter* (see also remark R.19.2).

Remark R.23 We cannot be more informative on the minimum rate of convergence to zero of any real sequence that bounds η_n with probability

$1-o\left(n^{-a^*}\right)$, due to the lack of information with respect to the analogous rate of uniform convergence of $b_n(\theta)$ to $b(\theta)$. However, if

$$\left\|E_\theta\beta_n - b(\theta) - \sum_{i=1}^{\infty}\frac{1}{n^{\frac{i+1}{2}}}Ek_{i+1}(z,\theta)\right\| = o\left(n^{-a}\right)$$

for any $a \in A$, uniformly on Θ, then $\sup_{\theta\in\Theta}\|b_n(\theta) - b_n(\theta,\zeta_n(\theta,a))\| = o(n^{-a})$ for $\zeta_{i+1_n}(\theta) = Ek_{i+1}(z,\theta)$, for any $i = 1,\ldots,2a$, and therefore it is easy to see that $P(\eta_n > \gamma_n) = o\left(n^{-a^*}\right)$ for $\gamma_n = o(n^{-a})$. It follows that if $a \to \infty$, hence $a^* \to \infty$, $P(\eta_n > \gamma_n) = o(n^{-a})$ for $\gamma_n = o(n^{-a})$ for all a and therefore $\{GMR2\}_n$ is asymptotically indistinguishable as a sequence from $\{GMR2^*(\infty)\}_n$ hence we obtain the characterization of the $GMR2$ estimator as a $GMR2^*(\infty)$ one, with the obvious abuse of terminology. An analogous asymptotic relationship can be established between the sequences of the first order approximations of the aforementioned estimators, thereby identifying $\theta_n^*(\infty)$ with the bootstrap estimator. In this respect we justify the commutative diagram presented in the introduction.

The previous reasoning can also establish analogous relations between $GMR2^*(a)$ and $GMR2^*(a')$ estimators, for $a \neq a'$ with a more detailed description of he structure of the error of the analogous approximation. Without loss of generality, let $a > a'$.

Corollary 4.4 Under the assumptions of 4.1, for both a and a', there exists a real sequence $\gamma_n = o\left(n^{-\delta-\frac{1}{2}}\right)$ such that

$$J\left(\beta_n, b_n(\theta_n(a'),a), W_n^*\left(\theta_n^+\right)\right) \leq J\left(\beta_n, b_n(\theta_n(a),a), W_n^*\left(\theta_n^+\right)\right) + \eta_n'$$

with $P(\eta_n' > \gamma_n) = o\left(n^{-a^*}\right)$, where $\delta = \begin{cases} \frac{1}{2}+\varepsilon \text{ if } a = \frac{1}{2} \\ a' \text{ if } a > \frac{1}{2} \end{cases}$ with $0 < \varepsilon < \frac{1}{2}$.

Remark R.24 Again, any $GMR2^*(a')$ is an approximate $GMR2^*(a)$ for any $a > a'$. This is particularly valid when $a' = 0$, since $GMR1 = GMR2^*(0)$.

4.2 Asymptotic Tightness and Validity of Edgeworth Approximation

In this paragraph, we are concerned with the higher order approximation of the distribution of $GMR2^*(a)$. We essentially rely on the previous results, the local differentiability of the criterion from which it emerges and lemma AL.2 presented at the appendix.

Lemma 4.5 Under the assumptions of corollary 4.4, there exists an $\{\eta_n''\}_n$, with $P\left(\sqrt{n}\,\|\eta_n''\| > \gamma_n'\right) = o\left(n^{-a^*}\right)$, and $\gamma_n' = o(n^{-\varepsilon})$ for some $\varepsilon > 0$, and $\sqrt{n}\left(\theta_n(a) - \theta_n(0)\right) = \eta_n''$ with probability $1 - o\left(n^{-a^*}\right)$.

The validity of the Edgeworth expansion of $\sqrt{n}\left(\theta_n\left(a\right)-\theta_0\right)$ of order $s = 2a^* + 1$ can now be established by assumption A.2, lemma 4.5 and corollary AC.1 presented in the appendix. In this case the sequence of distributions of the aforementioned estimator is also approximated in the $o\left(n^{-a^*}\right)$- convex variational distance by the relevant sequence of distributions of an sequence of random vectors that are polynomial in a standard normal random vector and in $\frac{1}{\sqrt{n}}$.

Lemma 4.6 Under the assumptions of lemma 4.5, the $GMR2^*\left(a\right)$ admits an Edgeworth expansion of order $s = 2a^* + 1$.

Lemma 4.7 Under the assumptions of corollary 4.6, there exists a $C^* > 0$ such that $P\left(\sqrt{n}\left\|\theta_n\left(a\right)-\theta_0\right\| > C^*\ln^{1/2}n\right) = o\left(n^{-a^*}\right)$.

4.3 Valid First Moment Approximation

Lemma 4.6 in the light of lemma AL.4 along with assumption AL.2 (see appendix B) and due to the fact that $a^* > a$, provide with an approximation of the sequence of first moments of the defined estimator, and therefore with an analogous approximation of the bias. In order to facilitate the presentation, we make the following definition.

Definition D.5 Let $\{x_n\}$ and $\{y_n\}$ denote two sequence of random elements with values in an normed space. We denote the relation $x_n \underset{a}{\sim} y_n$ when $\left\|E\left(x_n - y_n\right)\right\| = o\left(n^{-a}\right)$.

Remark R.25 Due to the positive definiteness of the norm and the triangle inequality $\underset{a}{\sim}$ is an equivalence relation on the set of sequences of random elements whose first moments converge to the same limit.

Now, under additional conditions, similar to integrability ones, the sequence of first moments of $\sqrt{n}\left(\theta_n\left(a\right)-\theta_0\right)$ is again approximated, in the relevant sense, by the sequence of moments of the aforementioned polynomial sequence. These are summarized in the following lemma.

Lemma 4.8 Under the assumptions of lemma 4.6, there exists a sequence of polynomial functions of z and $\frac{1}{\sqrt{n}}$, say $g\left(z, \frac{1}{\sqrt{n}}, \theta_0\right)$, such that

$$\left\|E_{\theta_0}\left(\sqrt{n}\left(\theta_n\left(a\right)-\theta_0\right)\right) - E\left(g\left(z, \frac{1}{\sqrt{n}}, \theta_0\right)\right)\right\| = o\left(n^{-a}\right)$$

We immediately obtain the following corollary.

Corollary 4.9 Under the assumptions of corollary 4.6, the expansion of $\sqrt{n}\left(\theta_n\left(a\right)-\theta_0\right)$ coincides with the formal expansion.

Remark R.26 $g\left(z, \frac{1}{\sqrt{n}}, \theta_0\right)$ is computed by the inversion of the Taylor expansion of the first order conditions that define the estimator with respect to $\sqrt{n}\left(\theta_n\left(a\right)-\theta_0\right)$, the replacement of the terms that admit an Edgeworth expansion by relevant polynomials of z emerging by the same reasoning, and by the grouping of terms of the same asymptotic order due to the previous corollary.

For the particular case of $a=\frac{1}{2}$, we may use the following. The relevant approximation concerning the auxiliary estimator, in the case where $a=\frac{1}{2}$, and $a^*=1$ in view of assumption A.2 is of the form $g\left(z, \frac{1}{\sqrt{n}}, \theta\right)=k_1\left(z, \theta\right)+\frac{k_2(z,\theta)}{\sqrt{n}}$, where as commented above $Ek_1\left(z,\theta\right)=\mathbf{0}_{q\times 1}$ on Θ, which also provides the means upon which the definition of GMR2$^*\left(\frac{1}{2}\right)$ is based. The result can be partially retrieved from lemma 4.3 of Arvanitis and Demos (2010), where its validity is also discussed. Similarly, due to the particular assumption we have that $\sqrt{n}q_n\left(\theta_0\right)\underset{\frac{1}{2}}{\sim}k_1^q\left(z,\theta_0\right)+\frac{k_2^q(z,\theta_0)}{\sqrt{n}}$, where $Ek_1^q\left(z,\theta_0\right)=\mathbf{0}_{k\times 1}$. Finally the same is true for θ_n^*, whereas $\sqrt{n}\left(\theta_n^*-\theta_0\right)\underset{\frac{1}{2}}{\sim}q_1^*\left(z,\theta_0\right)+\frac{q_2^*(z,\theta_0)}{\sqrt{n}}$, where $Eq_1^*\left(z,\theta_0\right)=\mathbf{0}_{p\times 1}$.

In view of corollary 4.9, we are able to prove the analogous result for GMR2$^*\left(\frac{1}{2}\right)$, using a procedure analogous to the one in R.26.

Lemma 4.10 Under the assumptions of corollary 4.9 and for $a=\frac{1}{2}$, then $g\left(z, \frac{1}{\sqrt{n}}, \theta_0\right)=q_1\left(z,\theta_0\right)+\frac{1}{\sqrt{n}}q_2\left(z,\theta_0\right)$ where

$$q_1=q_1\left(z,\theta_0\right)=BW^*\left(\theta_0\right)k_1$$

and

$$\begin{aligned} q_2 &= q_2\left(z,\theta_0\right)=BW^*\left(\theta_0\right)\left(\left(k_2-Ek_2\right)-\frac{1}{2}\left[trq_1q_1'\frac{\partial^2 b_j\left(\theta_0\right)}{\partial\theta\partial\theta'}\right]_{j=1,\dots,q}\right) \\ &+\left(\left(\frac{\partial b'\left(\theta_0\right)}{\partial\theta}W^*\left(\theta_0\right)\frac{\partial b\left(\theta_0\right)}{\partial\theta'}\right)^{-1}\left[\frac{\partial b_j\left(\theta_0\right)}{\partial\theta\partial\theta'}q_1\right]W^*\left(\theta_0\right)\right)Ak_1 \\ &+\left(Bw^*\left(z,\theta_0\right)+B\left[\frac{\partial}{\partial\theta^{/}}W_{rj}^*\left(\theta_0\right)q_1^*\right]_{r,j=1,\dots,q}\right)Ak_1 \end{aligned}$$

$A=\left(Id_q-\frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\right)$,
$B=\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}$, and $k_1=k_1(z,\theta_0)$, $k_2=k_2(z,\theta_0)$, and $q_1^*=q_1^*(z,\theta_0)$ are as in the proceeding paragraph.

Remark R.27 Lemma 4.8 is in accordance with the well known result that the second order bias of estimators of this sort hinges on a) non linearity of the estimating equations, b) difference in the relevant dimensions and c) stochastic weighting (see for example [15]).

Remark R.28 Notice that neither q_1 nor q_2 depend on k_1^q or k_2^q. We would not expect this to hold in higher order expansions concerning $\theta_n\left(\frac{1}{2}\right)$.

When $p=q$, then $A=\mathbf{0}_{q\times q}$ and $BW^*(\theta_0)=\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}$. Consequently, we trivially get the following corollary.

Corollary 4.11 Under the assumptions of lemma 4.10 and for $p=q$ we obtain

$$q_1=\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}k_1$$

and

$$q_2=\left(\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left(k_2-Ek_2-\frac{1}{2}\left[trq_1q_1'\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right]_{j=1,\dots,q}\right)$$

Furthermore, if $\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}=0$, e.g. $b(\theta)$ is linear, we trivially get:

Corollary 4.12 If in addition to the provisions of the previous corollary $\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}=\mathbf{0}_{p\times p}$ $\forall j=1,\dots,q$, $E_{\theta_0}q_2=\mathbf{0}_p$, hence the estimator is approximately unbiased of order $s=2$.

Remark R.29 As it will become apparent in the next section, the previous result can easily be extended in the case of the $\theta_n^*(a)$ estimator for any $a\geq\frac{1}{2}$. That is, if $\left\|\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right\|$, vanishes $\forall j$, then $\theta_n^*(a)$ becomes second order unbiased at θ_0.

For the geometric definition of the local canonical form, see Arvanitis and Demos (2010), section 4.2, which is derived by theorem 10.2 of Spivak (1999) (p. 44). This notion essentially depends on the local isomorphism property of the binding function and essentially concerns the choice of the auxiliary coordinates so that the binding function becomes canonical around θ_0, hence locally linear. In this case, the $GMR2^*\left(\frac{1}{2}\right)$ estimator is locally second order unbiased, as the next corollary demonstrates, if the weighting matrix is non-stochastic.

Corollary 4.13 If $b(\theta_0)$ is in local canonical form and $W_n^*(\theta_0) = W^* = \begin{pmatrix} W_{1,p\times p} & W_{3,p\times q-p} \\ W_3' & W_{2,q-p\times q-p} \end{pmatrix}$ then

$$q_1 = \begin{pmatrix} Id_{p\times p} & W_{1,p\times p}^{-1} W_{3,p\times q-p} \end{pmatrix} k_1$$

and

$$q_2 = \begin{pmatrix} Id_{p\times p} & W_{1,p\times p}^{-1} W_{3,p\times q-p} \end{pmatrix} (k_2 - Ek_2)$$

Remark R.30 The $GMR2^*\left(\frac{1}{2}\right)$ estimator is (locally) second order unbiased even in cases where $q > p$, when there is non stochastic weighting given that the binding function is in local canonical representation. However, *given an admissible auxiliary statistical model, there always exists an auxiliary parameterization such that the previous result is valid*, proviso the relevant weighting structure.

Remark R.31 We now consider the case of the one-computational step approximation of $GMR2^*\left(\frac{1}{2}\right)$, named $\theta_n^*\left(\frac{1}{2}\right)$ and described in remark R.19. It can be verified using the results of Andrews (2002) that $\sqrt{n}\left(\theta_n^*\left(\frac{1}{2}\right) - \theta_0\right) \underset{1/2}{\backsim} k_1 + \frac{1}{\sqrt{n}}(k_2 - Ek_2)$, thereby it is second order equivalent to $\sqrt{n}\left(GMR2^*\left(\frac{1}{2}\right) - \theta_0\right)$, due to transitivity, whilst it is of minimal computational burden. However, we make the following observations:[5]

1. Due to remark R.19.1 $\theta_n^*\left(\frac{1}{2}\right)$ could be non-definable for small n, on subsets of the sample space of positive probability.

2. They could be non-equivalent with respect to higher order relations, whereas the analogous expansions could favor $GMR2^*\left(\frac{1}{2}\right)$, with respect to its higher order bias structure.

3. The same could be true even with respect to the second order relation, when θ_0 lies on the boundary of the parameter space, in which case Ek_1 could be different from zero. We suspect that in this case $GMR2^*\left(\frac{1}{2}\right)$ would possess a more favorable second order bias structure than its one step computational approximation. The validation of this statement is out of the scope of the present paper, as it requires a theory of higher order approximations of distributions of M-estimators when the parameter is on the boundary.

[5]Notice that analogous ascertainments could hold with respect to the issue of the k^{th} order comparison between $\theta_n^*(a)$ and $\theta_n(a)$ for arbitrary a, k.

5 Recursive Indirect Estimation

In the current section we are concerned with the issue of extending the notion of indirect estimation in order to allow for procedures that potentially involve an arbitrary number of auxiliary steps. This will enable the construction of multistep IE that are approximately unbiased of some prescribed order without explicit reparameterizations. These will provide a procedure of recursive bias correction of any desired order of an arbitrary estimator of θ_0 that admits a valid moment approximation of the same order. The question addressed in this section can now be stated as follows:

Problem 1 Given the validity of assumptions A.2-A.5 for $s \geq 2(a+1)$ is it possible to define an approximately unbiased indirect estimator of order $2a+1$?

We first distinguish between additional and stronger notions of approximate unbiaseness to the one discussed in the previous section that is obviously concerning only θ_0.

Definition D.6 An estimator admitting a moment expansion such as the ones considered in the previous sections, will be termed:

1. approximately unbiased of order $s = 2a+1$ *at* θ_0 if the relevant expansion is valid, and $\sqrt{n}(\theta_n - \theta_0) \underset{a}{\sim} g\left(z, \frac{1}{\sqrt{n}}, \theta_0\right)$, where $E\left(g\left(z, \frac{1}{\sqrt{n}}, \theta_0\right)\right) = o(n^{-a})$,

2. *locally* approximately unbiased of order $s = 2a+1$ around θ_0, if the relevant expansion is valid, and $\sqrt{n}(\theta_n - \theta) \underset{a}{\sim} g\left(z, \frac{1}{\sqrt{n}}, \theta\right)$, where $E\left(g\left(z, \frac{1}{\sqrt{n}}, \theta\right)\right) = o(n^{-a})$ in an *neighborhood* of θ_0 and

3. approximately *globally* unbiased if it is *locally* unbiased around θ_0 for every neighborhood of θ_0.

We have therefore three notions of approximate unbiaseness that are presented in the order of increasing strength. We make the following assumption.

Assumption A.6 β_n is an estimator of θ_0, i.e. the binding function is, at least locally, the identity, that satisfies assumption A.2 for some $s^* \geq 2(a+1)$.

Notice that this does not pose any loss of generality compared to the previous sections, since β_n could itself be an IE of θ_0, given an auxiliary model that does not coincide with the one at hand. In this respect any concern about the asymptotic behavior of sequences of weighting matrices becomes asymptotically irrelevant.

In order to economize the presentation, we also make the following definition.

Definition D.7 For any $a^+ \leq a$ we denote with $Z_n(a^+, \beta_n)$ the set of functions of the form $\zeta_n(\theta, a^+) = (\zeta_{2_n}(\theta_0), \ldots, \zeta_{2a^++1_n}(\theta_0))$ where for each $i = 1, \ldots, 2a^+$, ζ_{i+1_n} satisfies assumptions A.4 and A.5 *given* assumption A.6 for β_n, under the convention that if β_n is s_*^{th}-order locally unbiased for any $s_* \leq s$, then $\zeta_{i+1_n}(\theta) = 0 \; Q - \mathbf{a}.\mathbf{s}.$ for any $1 < i \leq s_* - 1$ and any θ at the particular neighborhood of θ_0, for any $\zeta_n(\theta, a^+) \in Z_n(a^+, \beta_n)$.

This convention is motivated by the fact that $Z_n(0, \beta_n) = \{\mathbf{0}_\Theta\}$ due to remark R.9 and the definition of the examined estimators. In the same respect we denote with $GMR^*(\zeta_n(\theta, a^+))$ either $GMR2^*(a^+)$ or $\theta_n^*(a^+)$ (which is discussed in remark R.18) w.r.t. $\zeta_n(\theta, a^+) \in Z_n(a^+, \beta_n)$.[6]

Definition D.8 Given assumption A.6 and $a_1, a_2 \leq a$, let

$$GMR^*(\zeta_n(\theta, a_2)) \circledast GMR^*(\zeta_n(\theta, a_1))$$

denote the indirect estimator emerging as follows:

1. $GMR^*(\zeta_n(\theta, a_1))$ is derived using $\zeta_n(\theta, a_1) \in Z_n(a_1, \beta_n)$, and
2. $GMR^*(\zeta_n(\theta, a_2))$ is derived using $\zeta_n(\theta, a_2) \in Z_n(a_2, GMR^*(\zeta_n(\theta, a_1)))$.

In this respect the $GMR^*(\zeta_n(\theta, a_2)) \circledast GMR^*(\zeta_n(\theta, a_1))$ is an indirect estimator emerging in essentially three steps, the first one being β_n. Obviously such estimators can be derived by making the number of steps arbitrary, yet finite. Hence, in general

$$\circledast_{i=1}^{K} GMR^*(\zeta_n(\theta, a_f(i))) \doteqdot GMR^*(\zeta_n(\theta, a_f(K))) \circledast \left(\circledast_{i=1}^{K-1} GMR^*(\zeta_n(\theta, a_f(i)))\right)$$

where in the $(K+1)^{th}$ step the $GMR^*(\zeta_n(a_f(K)))$ is derived using as an auxiliary the $\circledast_{i=1}^{K-1}\mathrm{GMR}^*(\zeta_n(\theta, a_f(i)))$, for $K \in \mathbb{N}$, and $a_f : \{1, 2, \ldots, K\} \rightarrow$

[6] Considering the notions that follow, it would be more appropriate to define $Z(a)$ as the set of equivalence classes of approximating functions, with respect to the relation that renders two such functions equivalent, i.e. iff they define the same $GMR2^*$ estimator. We choose to disregard this detail for notational convenience.

$\{0,\ldots,a\}$. Notice that $[GMR^{*}(\zeta_n(\theta,a_3)) \circledast GMR^{*}(\zeta_n(\theta,a_2))] \circledast GMR^{*}(\zeta_n(\theta,a_1))$ is non definable, a fact that does not permit the set of estimators emerging via $\circledast$ to be closed under $\circledast$, hence prevents this set from obtaining a relevant algebraic structure.

Remark R.32 It is trivial to see that in the present framework

$$GMR^{*}(\zeta_n(\theta,a)) \circledast GMR1 = GMR1 \circledast GMR^{*}(\zeta_n(\theta,a)) = GMR^{*}(\zeta_n(\theta,a))$$

for any a.

Remark R.33 The definition of $GMR^{*}(\zeta_n(a_2)) \circledast GMR^{*}(\zeta_n(a_1))$ essentially depends on the validity of assumption A.2 for $GMR^{*}(\zeta_n(\theta,a_1))$. A.2 follows from the results of the previous sections for $\theta_n(a_1)$ and $\theta_n^{*}(a_1)$ along with the results of Andrews (2002) and remark R.19 for the latter, for $a^{*} > a \geq a_1$ and the fact that $GMR1 \circledast \beta_n = \beta_n$ due to previous remark and assumption A.6. The assumptions on the differentiability of the approximations could follow from analogous assumptions on continuous differentiability of adequate order of the criterion from which β_n emerges. These along with obvious generalizations of A.4 and A.5 would make $Z_n(a_2, GMR^{*}(\zeta_n(\theta,a_1)))$ non empty.

Lemma 5.1 Under assumption A.6 and if β_n is approximately locally unbiased of order $(2a_1+1)$, for $a_1 \leq a$, then $\mathrm{GMR}^{*}(\zeta_n(\theta,a_2)) \circledast \beta_n$ is approximately locally unbiased of the same order, $\forall a_2 \leq a_1$.

Lemma 5.2 If β_n is approximately locally unbiased of order $(2a_1+1)$, for $a_1 \leq a$, then $\mathrm{GMR}^{*}(\zeta_n(\theta,a_2)) \circledast \beta_n$ is approximately locally unbiased of order $2(a_1+1)$, $\forall a_2 > a_1$.

Hence a solution to the posed problem emerges from the following algorithm.

Algorithm Suppose that β_n is approximately locally unbiased of order $(2a_1+1)$, for $a_1 < a$:

- set $\theta_n^{(0)} = \beta_n$ and $a^{(0)} = a_1$,
- for $a^{(i)} = a^{(i-1)} + \frac{1}{2}$, $i = 1,\ldots,2(a-a_1)$, set $\theta_n^{(i)} = \mathrm{GMR}^{*}\left(\zeta_n\left(\theta, a^{(i)}\right)\right) \circledast \theta_n^{(i-1)}$ where $\zeta_n\left(\theta,a^{(i)}\right) \in Z_n\left(a^{(i)}, \theta_n^{(i-1)}\right)$. The expansions needed for the derivation of $\theta_n^{(i)}$ can be obtained from the initial one and calculations similar to the proof of lemma 5.2, due to which it is approximately locally unbiased of order $\left(2a^{(i)}+1\right)$.

Then $\theta_n^{(2(a-a_1))}$ is approximately locally unbiased of order $2a^{(2(a-a_1))}+1 = 2a+1$ as required due to lemma 5.2. Obviously the above construction generalizes the case where $a=\frac{1}{2}$ and $a_1=0$, as implied by the results of the previous section.

Remark R.34 In the case the $\theta_n^*\left(a^{(i)}\right)$ is needed for some i, remarks R.19 and R.31 would also hold. Global approximate unbiaseness can be obtained by strengthening the analogous property for β_n.

Remark R.35 Notice that if we are interested in estimators that are approximately locally unbiased of order $(2a+1)$ at θ_0, then remark R.33 for $\beta_n = \theta_n^{(s-1)}$ can be reformulated so that $m_n^*(\theta)$ has the desired asymptotic approximation only at θ_0.

Remark R.36 The aforementioned recursive procedure can be perceived as an approximation of the local canonical representation of $E_\theta(\beta_n)$ as $a \to \infty$, something which is very important for practical purposes.

Let us now turn our attention to two examples.

6 Examples

In this section we apply the suggested estimators to, approximately, correct the bias of various estimators in the $MA(1)$ model as well as correct the bias of the $MLEs$ of the parameters in an $ARCH(1)$ process. As an additional verification of our theoretical results we perform a small simulation exercise, for each example. Let us consider the first one.

6.1 MA(1)

Assume the invertible $MA(1)$ process

$$y_t = u_t + \theta u_{t-1}, \quad t = ..., -1, 0, 1, ..., \quad |\theta| < 1, \quad u_t \overset{iid}{\backsim} (0, \sigma^2).$$

In this case, the $GMR1$ estimator of θ is given by $\frac{1-\sqrt{1-4\beta_n^2}}{2\beta_n}$, where β_n is the $QMLE$ of the $AR(1)$ coefficient of an $AR(1)$ auxiliary model (see Gourieroux et al. 1993, and Demos and Kyriakopoulou 2008). This is also the $\theta_n(0)$ estimator (see section 2 above). Notice that in this case Θ is a compact subset of $(-1,1)$, $b(\theta) = \frac{\theta}{1+\theta^2}$, and B is a compact subset of $\left(-\frac{1}{2}, \frac{1}{2}\right)$.

From the calculations in appendix C we have that

$$E\left[\sqrt{n}\left(GMR1-\theta\right)\right]=\frac{1}{\sqrt{n}}\theta\frac{1+5\theta^2+2\theta^4+\theta^6-\theta^8}{\left(1-\theta^2\right)^3} \tag{3}$$

as in Demos and Kyriakopoulou (2008). As expected the $GMR1$ is not 2^{nd} order *unbiased* as the binding function is not linear.

Now a third step estimator of θ, $GMR2S$, simply solves the equation $\beta_n=\frac{\theta}{1+\theta^2}-\frac{1}{n}\left(\theta^4+2\theta^3-2\theta^2+2\theta+1\right)\frac{\theta^2+\theta+1}{\left(\theta^2+1\right)^3}$. In this case, the binding function is the identity and consequently $GMR2S$ is 2^{nd} order *unbiased* (see appendix C for details). In fact, this is the resulting estimator by applying the $GMR2^*\left(\frac{1}{2}\right)$ on $GMR1$ (see section 5).

Alternatively, as a second step estimator, one can consider the application of $GMR2^*\left(\frac{1}{2}\right)$ on β_n, named $GMR2R$. As the binding function is not linear, this estimator is not 2^{nd} order unbiased (see appendix C for calculations), apart from $\theta=0$, i.e. is locally 2nd order unbiased at 0 by the terminology of section 5. Hence

$$E\left[\sqrt{n}\left(GMR2R-\theta\right)\right]=\frac{1}{\sqrt{n}}\frac{\theta^2+4\theta^4+\theta^6+\theta^8+1}{\left(\theta^2+1\right)}\frac{\theta\left(3-\theta^2\right)}{\left(1-\theta^2\right)^3}. \tag{4}$$

However, applying $GMR2^*\left(\frac{1}{2}\right)$ on $GMR2R$, name it $GMR2RS$, we have that, as the binding function is the identity in this case, $GMR2RS$ is 2^{nd} order *unbiased*. Finally, estimating θ by the $GMR2$ we have that

$$E\sqrt{n}\left(GMR2-\theta\right)=\frac{1}{\sqrt{n}}\frac{\theta\left(3-\theta^2\right)}{1-\theta^2}\frac{\theta^2+4\theta^4+\theta^6+\theta^8+1}{\left(1+\theta^2\right)^5}, \tag{5}$$

i.e. the $GMR2$ is not 2^{nd} order *unbiased*, as expected due to nonlinearities in the binding function. Comparing equation (4) with (5) it is obvious that $GMR2$ is less 2^{nd} order *biased* than $GMR2R$ for all values of θ, apart from $\theta=0$ in which case both are 2^{nd} order locally *unbiased*.

In terms of simulations, we draw a random sample of $n\in\{50,100,150,250,500,750,1000,1500,3000\}$ observations from a non-central Student-t distribution with non-centrality parameter $\eta=1$ and $\nu=20$ degrees of freedom, standardized appropriately so that they have zero mean and unit variance. For each random sample, we generate the $MA(1)$ process y_t for $\theta\in\{-0.4,0.4\}$. We evaluate β_n and if the estimate is in the $[-0.499999,0.499999]$ interval we estimate all estimators, otherwise we throw away the sample and draw another one. For each retained sample we evaluate eight estimators, i.e. the $GMR1$, $GMR2$, $GMR2S$, $GMR2R$, $GMR2RS$, the $QMLE$ of θ, say $QMLE$, the second step $GMR2^*\left(\frac{1}{2}\right)$ on the $QMLE$, say $QMLES$, as well as the feasibly bias corrected estimator of $GMR1$, $BCGMR1$, where the estimated value of θ is employed in equation 3 for bias correction, i.e.

$$BCGMR1 = \theta_n^*(0) - \frac{1}{n}\theta_n^*(0)\frac{1 + 5[\theta_n^*(0)]^2 + 2[\theta_n^*(0)]^4 + [\theta_n^*(0)]^6 - [\theta_n^*(0)]^8}{\left(1 - [\theta_n^*(0)]^2\right)^3}.$$

Out of these estimators only $GMR2S$, $GMR2RS$, $QMLES$ and $BCGMR1$ are 2^{nd} order *unbiased*. We set the number of replications to 100000.

For the $GMR2$ estimator an additional question arises from the presence of $E_\theta\beta_n$ in its objective function (see section 2). In general this expectation is unknown and consequently is approximated by an average of, say H, monte carlo replications (see Gourieroux et al. 1993). Of course under the assumptions in Gourieroux et al. (1993) as $H \to \infty$ we have that the average converges to the expected value. Nevertheless, in practice a finite number of H is employed. Consequently, it could be of interest to compare the theoretical results, i.e. when $H = \infty$, with those in practice, i.e. when H is finite. Clearly, the larger H is the better the approximation is and the more cputime is needed per iteration within the maximization routine. The second effect is of course undesirable. Furthermore, on this point, one expects the $GMR2S$ to be faster than the $GMR2$, however how much faster is an open question.

Consequently, we employ two values of H, i.e. $H = 10$ and $H = 200$, denoting them by $GMR2(10)$ and $GMR2(200)$, respectively. Taking the average over the 100000 replications, in figure 1 we present the absolute value of the biases of the estimators, multiplied by n, i.e. $nE|GMR2(i) - \theta|$, $i = 10, 200$, where the true θ is -0.4. It is obvious that for $H = 10$ the bias of the estimator is far away from the approximate, up to $o\left(\frac{1}{n}\right)$, absolute bias which equals to 0.816 for this value of θ (see equation 5). Consequently, in what follows we consider only the $GMR2(200)$ one.

In figure 2 we present the absolute biases, multiplied by n, of the biased estimators. It seems that, apart from the $GMR2R$ and $GMR2$, 250 observations are enough for the estimators to reach their asymptotic approximate bias. For $\theta = -0.4$ these are 1.252 and 0.4 for the $GMR1$ and $QMLE$, respectively. For the $GMR2R$, 500 observations are needed to reach its asymptotic bias (2.094), whereas 3000 are needed for the $GMR2$.

In figure 3 the absolute biases, multiplied by n, of the unbiased estimators are presented. It is obvious that, apart from the $BCGMR1$ estimator, all estimators are by all means unbiased for sample size bigger or equal to 250. The same is true for the $BCGMR1$ one but for sample size bigger or equal to 500. It is worth noticing that, as expected, in almost all sample size cases the multistep bias corrected estimators ($GMR2S$ and $GMR2RS$) are less biased than the feasibly bias corrected $GMR1$ estimator ($BCGMR1$).

It is worth noticing that, for $n = 250$, the average cpu time per iteration for the $GMR2S$ estimator is 2.47×10^{-4} seconds, whereas the equivalent time for the $GMR2$ estimator is 2.88 seconds. Consequently, the suggested indirect estimator is not only 2^{nd} order *unbiased* but the procedure is very fast, as well, at least for this model.[7]

The results for $\theta = 0.4$ are qualitatively the same and are not presented to conserve space. Let us now turn our attention to the second example.

6.2 ARCH(1)

Consider the second order stationary $ARCH\,(1)$ model

$$\begin{aligned} y_t &= u_t^{1/2} z_t, \quad u_t = \theta_1 + \theta_2 y_{t-1}^2, \quad t = ..., -1, 0, 1, ..., \\ \theta_1 &> 0, \quad \theta_2 \in (0,1) \quad z_t \overset{iid}{\backsim} N(0,1). \end{aligned}$$

For the above model we have, from Iglesias and Linton (2007), and Iglesias and Phillips (2005), that

$$E\left(\widehat{\theta_1} - \theta_1\right) = n^{-1}G + o\left(n^{-1}\right), \quad E\left(\widehat{\theta_2} - \theta_2\right) = n^{-1}G^* + o\left(n^{-1}\right)$$

where $\widehat{\theta_1}$ and $\widehat{\theta_2}$ are the $MLEs$ of θ_1 and θ_2, n is the sample size of the observed process y_t, and G and G^* are given in the appendix.

We draw a random sample of $n \in \{150, 300, 500, 750, 1000, 1500, 2000\}$ observations, plus 250 for initialization, from a standard normal distribution. We perform 10000 replications. For each random sample, we generate the $ARCH(1)$ process y_t with $\theta_1 = 1.0$ and $\theta_2 = 0.5$, and we find the $MLEs$ of the two parameters, as well as the feasibly bias corrected ones as suggested in Iglesias and Linton (2007), named IL, and the indirect estimator suggested here, named AD. As the H_i and H_i^* terms, for $i = 2, ..., 7$, (see appendix C) involve summations up to the sample size, we truncate them in 10 and 40 and call these estimators $IL - 10$, $IL - 40$, and $AD - 10$ and $AD - 40$, for the feasibly corrected estimators and the indirect ones, respectively (see Iglesias and Linton 2007, and Iglesias and Phillips 2005).

In fact, these terms are evaluated from a long simulation with $n = 100000$, where the MLE estimates are employed to generate the $ARCH\,(1)$ process in the case of the two IL estimators. For the two AD ones the summation terms are treated as nuisance parameters, implicitly depending on the estimated

[7] All simulations have been performed to a computer with Intel i7 processor.

parameters. Under the distributional assumptions of our experiment, the validity of the above mentioned procedure, as well as the expansions, are justified (see Corradi and Iglesias 2008). This experiment elucidates remark R.13. It is in this case that the Ek_{i+1} are analytically intractable as functions of θ. Hence they are approximated in the manner described above. Notice that in the spirit of the same remark, a variety of approximations could also be used, that could additionally involve approximations of some (or all) of the unknown moments involved in the expansions using the observed sample (instead of or in addition to the Monte Carlo sampling), as well as the computation of some of the approximating functions on the $QMLE$ etc. We did not employ such cases that can be easily adopted in the framework of assumption A.4 for reasons of presentational convenience.

In few cases the feasibly bias corrected estimator of θ_2 turns out to be either greater than 1 or smaller than 0 (see remark R.19). In these cases we throw away the particular Monte Carlo samples and draw new ones.[8]

In figure 4 the absolute biases, multiplied by n, of the estimators of the constant θ_1 are presented. It is immediately obvious that both estimators, $IL-10$ and $AD-10$, do not correct the bias of the MLE, for $n \leq 750$. For the $40-window$ estimators both partially only correct the bias of the MLE, although for $n = 2000$ the biases of both estimators are close to their MC errors (around 0.987). For the bias-corrected estimators of θ_2 (the ARCH parameter), in figure 5, it is obvious that all four estimators correct the bias of the MLE. With the exemption of $n = 1500$, the $40 - window$ estimators are less biased than the $10-window$ ones and close to their MC error (0.894). Notice also that in almost all sample size cases the $AD-40$ estimator is less biased than the $IL-40$ one.

7 Conclusions

In this paper we define a set of indirect estimators based on moment approximations of the auxiliary estimators and provide results concerning their higher order asymptotic behavior. Our motivation resides on the following properties that these estimators posses:

1. Computational facility as they are derived from procedures avoiding the nested numerical optimization burden that is usually the case with the simulated analog of the GMR2 estimator. This comes at the *fixed*

[8] In fact for $n = 150$ we observed that in 1.33% and 2.10% of the experiments the resulting $IL-10$ and $IL-40$ estimator was greater than 1 or smaller than 0, respectively. Of course for larger n these cases are fewer and for $n \geq 750$ there is none.

cost of the analytical derivation of the approximation. This remark also holds in cases where the analytical form of the approximation is unknown and is in turn numerically approximated.

2. The GMR1 estimator has a convenient interpretation as an approximate minimizer of the criteria from which the considered estimators are derived. This facilitates enormously the analytical derivation of some of the asymptotic properties. Analogous results hold between any pair of the estimators studied.

3. More generally, their asymptotic properties are analytically more tractable than the analogous of the GMR2 estimator. For example, there is no need of imposing rate of convergence conditions on the derivatives of the error of approximation, since the result that would be based on such a condition in the case of the GMR2 estimator, is now based on local boundeness conditions of the parameter functions of the relevant polynomials in $\frac{1}{\sqrt{n}}$.

We extend this class of estimators to multistep indirect estimators that in conjunction with the previously mentioned results identifies subclasses that have potentially useful bias structure of any given order.

We demonstrated that the well known "feasibly biased corrected" estimator is an one-computational step approximation of the suggested estimator. As expected the later performed better, in terms of bias, in two examples. Of course, one could apply the suggested procedures to more complex models than the expository ones employed in this paper.

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] Arvanitis S. and Demos A. (2010), "Stochastic Expansions and Moment Approximations for Three Indirect Estimators", IEES Working Paper No 1004, AUEB.

[3] Demos A. and Kyriakopoulou D. (2008), "Edgeworth and Moment Approximations: The Case of MM and QML Estimators for the MA(1) Models", IEES Working Paper No 1003, AUEB.

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

[5] Corradi, V. and E.M. Iglesias (2008), "Bootstrap refinements for QML estimators of the $GARCH\,(1,1)$ parameters", *Journal of Econometrics*, 144, 500-510.

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

[7] Gourieroux C. and A. Monfort (1996), "*Simualtion-Based Econometric Methods*", Oxford University Press, New York.

[8] 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.

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

[10] Iglesias, E.M. and O.B. Linton, "Higher order Asymptotic Theory when a Parameter is on the Boundary with an Application to GARCH Models", *Econometric Theory*, 2007, 26, pp. 1136–1161.

[11] Iglesias, E.M. and G.D.A. Phillips, "Small Sample Estimation Bias in GARCH Models with any number of Exogenous Variables in the Mean Equation", W.P. Michigan State University.

[12] Klein E., and A.C. Thompson, "Theory of Correspondences", Canadian Mathematical Society, Wiley-Interscience Publication, 1984.

[13] Magdalinos, M.A. (1992), "Stochastic Expansions and Asymptotic Approximations", *Econometric Theory* 8, pp. 343-367.

[14] Molchanov, I. (2005), "Theory of Random Sets, Probability and Its Applications", Springer, 2005.

[15] Newey, W.K., and R.J. Smith (2001), "*Asymptotic Bias and Equivalence of GMM and GEL Estimators*", mimeo, Department of Economics, MIT.

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

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

Appendices

A Proofs of Lemmas, Propositions and Corollaries.

Proof of Lemma 4.1. We have that
$Q\left(\sup_{\theta\in\Theta}\left|J^2\left(\beta_n, b_n\left(\theta,\zeta_n\left(\theta,a\right)\right), W_n^*\left(\theta_n^+\right)\right) - J^2\left(b\left(\theta_0\right), b\left(\theta\right), W^*\left(\theta_0\right)\right)\right| > \varepsilon\right) =$

$$Q\left(\begin{array}{c}\sup_{\theta\in\Theta}\left|J^2\left(\beta_n, b_n\left(\theta,\zeta_n\left(\theta,a\right)\right), W_n^*\left(\theta_n^+\right)\right) - J^2\left(\beta_n, b_n\left(\theta,\zeta_n\left(\theta,a\right)\right), W^*\left(\theta_0\right)\right)\right| \\ +\sup_{\theta\in\Theta}\left|J^2\left(\beta_n, b_n\left(\theta,\zeta_n\left(\theta,a\right)\right), W^*\left(\theta_0^*\right)\right) - J^2\left(b\left(\theta_0\right), b\left(\theta\right), W^*\left(\theta_0\right)\right)\right| > \varepsilon\end{array}\right) \leq$$

$Q\left(\sup_{\theta\in\Theta}\left|J^2\left(\beta_n, b_n\left(\theta,\zeta_n\left(\theta,a\right)\right), W_n^*\left(\theta_n^+\right) - W^*\left(\theta_0\right)\right)\right| > \frac{\varepsilon}{2}\right) +$
$Q\left(\sup_{\theta\in\Theta}\left|J^2\left(\beta_n, b_n\left(\theta,\zeta_n\left(\theta,a\right)\right), W^*\left(\theta_0\right)\right) - J^2\left(b\left(\theta_0\right), b\left(\theta\right), W^*\left(\theta_0\right)\right)\right| > \frac{\varepsilon}{2}\right)$.
Now, due to the triangle inequality, submultiplicativity A.4.3, A.2, R.11 and R.12 we have for the first term of the last sum that it is less than or equal to
$Q\left(\sup_{\theta\in\Theta}\left\|\beta_n - b\left(\theta\right) - \sum_{i=0}^{2a}\frac{1}{n^{(i+1)/2}}\zeta_{i+1_n}\left(\theta\right)\right\|^2\left\|W_n^*\left(\theta_n^+\right) - W^*\left(\theta_0\right)\right\| > \frac{\varepsilon}{2}\right) \leq$

$$Q\left(\begin{array}{c}\sup_{\beta_n\in B\left(b(\theta_0),\varepsilon_1^*\right)}\sup_{\theta\in\Theta}\left\|\beta_n - b\left(\theta\right) - \sum_{i=0}^{2a}\frac{1}{n^{(i+1)/2}}\zeta_{i+1_n}\left(\theta\right)\right\| \\ \left\|W_n^*\left(\theta_n^+\right) - W^*\left(\theta_0\right)\right\|^{1/2} > \frac{\varepsilon}{2}\end{array}\right) +$$

$Q\left(\beta_n \notin B\left(b\left(\theta_0\right), \varepsilon_1^*\right)\right) \leq$

$$Q\left(\begin{array}{c}\left(\left\|\beta_n\right\| + \sup_\theta\left\|b\left(\theta\right)\right\| + \sum_{i=0}^{2a}\frac{1}{n^{(i+1)/2}}\sup_{\theta\in\Theta}\left\|\zeta_{i+1_n}\left(\theta\right)\right\|\right) \\ \left\|W_n^*\left(\theta_n^+\right) - W^*\left(\theta_0\right)\right\| > \frac{\varepsilon}{2}\end{array}\right) +$$

$P\left(\beta_n \notin B\left(b\left(\theta_0\right), \varepsilon_1^*\right)\right) \leq$

$$P\left(\left\|W_n^*\left(\theta_n^+\right) - W^*\left(\theta_0\right)\right\| > \frac{\varepsilon}{2\left(c_1+c_2+\sum_{i=0}^{2a}\frac{M_i}{n^{(i+1)/2}}\right)}\right) +$$

$P\left(\beta_n \notin B\left(b\left(\theta_0\right), \varepsilon_1^*\right)\right) + \sum_{i=0}^{2a} Q\left(\sup_{\theta\in\Theta}\left\|\zeta_{i+1_n}\left(\theta\right)\right\| > M_i\right) = o\left(n^{-a^*}\right)$ due to remark R.11. For the second term we have that due to the continuous mapping theorem $\exists\epsilon > 0$:
$Q\left(\sup_\theta\left|J^2\left(\beta_n, b_n\left(\theta,\delta_n,a\right), W^*\left(\theta_0\right)\right) - J^2\left(b\left(\theta_0\right), b\left(\theta\right), W^*\left(\theta_0\right)\right)\right| > \frac{\varepsilon}{2}\right) \leq$
$Q\left(\sup_\theta\left|J\left(\beta_n, b_n\left(\theta,\delta_n,a\right), W^*\left(\theta_0\right)\right) - J\left(b\left(\theta_0\right), b\left(\theta\right), W^*\left(\theta_0\right)\right)\right| > \epsilon\right)$ and due to the triangle inequality, this is less than or equal to

$$Q\left(\sup_\theta\left\|\left(\beta_n - b\left(\theta\right) - \sum_{i=0}^{2a}\frac{1}{n^{(i+1)/2}}\zeta_{i+1_n}\left(\theta\right)\right) - \left(b\left(\theta_0\right) - b\left(\theta\right)\right)\right\|_{W^*(\theta_0)} > \epsilon\right) \leq$$

$Q\left(\left\|\beta_n - b\left(\theta_0\right)\right\| + \sum_{i=0}^{2a}\frac{1}{n^{(i+1)/2}}\sup_{\theta\in\Theta}\left\|\zeta_{i+1_n}\left(\theta\right)\right\| > \epsilon\right)$. The last term is less than or equal to
$Q\left(\left\|\beta_n - b\left(\theta_0\right)\right\| > \frac{\epsilon}{\sum_{i=0}^{2a}\frac{M_i}{n^{(i+1)/2}}}\right) + \sum_{i=0}^{2a} Q\left(\sup_{\theta\in\Theta}\left\|\zeta_{i+1_n}\left(\theta\right)\right\| > M_i\right)$ which is obviously $o\left(n^{-a}\right)$. The result follows from the continuous mapping theorem and assumption A.1.3 which implies that $J\left(b\left(\theta_0\right), b\left(\theta\right), W^*\left(\theta_0\right)\right)$ is uniquely minimized at θ_0. ■

Proof of Proposition 4.2. We have that $Q\left(\sup_{\theta\in\Theta}\left\|b_n\left(\theta\right) - b_n\left(\theta,\zeta_n\left(\theta,a\right)\right)\right\| > \varepsilon\right) \leq$ $P\left(\sup_{\theta\in\Theta}\left\|b_n\left(\theta\right) - b\left(\theta\right)\right\| > \frac{\varepsilon}{2}\right) + Q\left(\sum_{i=1}^{2a}\frac{1}{n^{i+1/2}}\sup_{\theta\in\Theta}\left\|\zeta_{i+1_n}\left(\theta\right)\right\| > \frac{\varepsilon}{2}\right) \leq$

$P\left(\sup_{\theta\in\Theta}\|b_n(\theta)-b(\theta)\|>\frac{\varepsilon}{2}\right)+Q\left(\sum_{i=1}^{2a}\frac{M_i}{n^{i+1/2}}>\frac{\varepsilon}{2}\right)+\sum_{i=0}^{2a}Q\left(\sup_{\theta\in\Theta}\left\|\zeta_{i+1_n}(\theta)\right\|>M_i\right)$, for some $\varepsilon>0$. Now due to assumption A.1.3 and due to the equi-Lipschitz property of $\{b_n\}$, we have that $\{b_n(\theta)-b(\theta)\}$ is also equi-Lipschitz, hence uniformly equicontinuous, hence $\sup_{\theta\in\Theta}\|b_n(\theta)-b(\theta)\|$ converges to zero uniformly on Θ, due to the fact that it converges pointwise by assumption A.2 and the Arzella-Ascolli theorem. Hence the first two probabilities are exactly zero for large enough n, hence the result follows from remark R.11. ■

Proof of Corollary 4.3. From the definition of the two estimators we obtain that

$J\left(\beta_n,E_{\theta_n(a)}(\beta_n),W_n^*\left(\theta_n^+\right)\right)-J\left(\beta_n,E_{\theta_n}(\beta_n),W_n^*\left(\theta_n^+\right)\right)=$
$\left|J\left(\beta_n,E_{\theta_n(a)}(\beta_n),W_n^*\left(\theta_n^+\right)\right)-J\left(\beta_n,E_{\theta_n}(\beta_n),W_n^*\left(\theta_n^+\right)\right)\right|$
$\leq\left|J\left(\beta_n,E_{\theta_n(a)}(\beta_n),W_n^*\left(\theta_n^+\right)\right)-J\left(\beta_n,b_n(\theta_n(a),\delta_n,a),W_n^*\left(\theta_n^+\right)\right)\right|$
$+\left|J\left(\beta_n,b_n(\theta_n(a),\zeta_n(\theta_n(a),a)),W_n^*\left(\theta_n^+\right)\right)-J\left(\beta_n,E_{\theta_n}(\beta_n),W_n^*\left(\theta_n^+\right)\right)\right|$
$\leq 2\sup_{\theta\in\Theta}\left|J\left(\beta_n,b_n(\theta),W_n^*\left(\theta_n^+\right)\right)-J\left(\beta_n,b_n(\theta,\zeta_n(\theta_n(a),a)),W_n^*\left(\theta_n^+\right)\right)\right|$

and the result follows with

$\eta_n=2\sup_{\theta\in\Theta}\left|J\left(\beta_n,b_n(\theta),W_n^*\left(\theta_n^+\right)\right)-J\left(\beta_n,b_n(\theta,\zeta_n(\theta_n(a),a)),W_n^*\left(\theta_n^+\right)\right)\right|$

due to the fact that $P(\eta_n>\varepsilon)=$

$P\left(\sup_{\theta\in\Theta}\left|J\left(\beta_n,b_n(\theta),W_n^*\left(\theta_n^+\right)\right)-J\left(\beta_n,b_n(\theta,\zeta_n(\theta_n(a),a)),W_n^*\left(\theta_n^+\right)\right)\right|>\varepsilon\right)\leq$

$$P\left(\sup_{\theta\in\Theta}\|(b_n(\theta)-b_n(\theta,\zeta_n(\theta_n(a),a)))\|_{W_n^*\left(\theta_n^+\right)}>\varepsilon\right)\leq$$
$$P\left(\begin{array}{c}\sup_{\theta\in\Theta}\|b_n(\theta)-b_n(\theta,\zeta_n(\theta_n(a),a))\|\left\|W_n^*\left(\theta_n^+\right)-W^*(\theta_0)\right\|^{1/2}\\+\sup_{\theta\in\Theta}\|b_n(\theta)-b_n(\theta,\zeta_n(\theta_n(a),a))\|\left\|W^*(\theta_0)\right\|^{1/2}>\varepsilon\end{array}\right)\leq$$
$$2P\left(\sup_{\theta\in\Theta}\|b_n(\theta)-b_n(\theta,\zeta_n(\theta_n(a),a))\|>\varepsilon_*\right)+P\left(\left\|W_n^*\left(\theta_n^+\right)-W^*(\theta_0)\right\|>K\right)=$$

$o\left(n^{-a^*}\right)$ for $K>0$ and $\varepsilon_*=\frac{\varepsilon}{2}\min\left(\frac{1}{\sqrt{\|W^*(\theta_0)\|}},\frac{1}{\sqrt{K}}\right)$. ■

Proof of Corollary 4.4. As in the previous proof we have that

$J\left(\beta_n,b_n(\theta_n(a'),\zeta_n(\theta_n(a'),a)),W_n^*\left(\theta_n^+\right)\right)-J\left(\beta_n,b_n(\theta_n(a),\zeta_n(\theta_n(a),a)),W_n^*\left(\theta_n^+\right)\right)$
$\leq 2\sup_{\theta\in\Theta}\left|J\left(\beta_n,b_n(\theta,\zeta_n(\theta,a)),W_n^*\left(\theta_n^+\right)\right)-J\left(\beta_n,b_n(\theta,\zeta_n(\theta,a')),W_n^*\left(\theta_n^+\right)\right)\right|$.

Then we have that

$P\left(\sup_{\theta\in\Theta}\left|J\left(\beta_n,b_n(\theta,\zeta_n(\theta,a)),W_n^*\left(\theta_n^+\right)\right)-J\left(\beta_n,b_n(\theta,\zeta_n(\theta,a')),W_n^*\left(\theta_n^+\right)\right)\right|>\frac{\gamma_n}{2}\right)\leq$

$$P\left(\sup_{\theta\in\Theta}\|(b_n(\theta,\zeta_n(\theta,a))-b_n(\theta,\zeta_n(\theta,a')))\|_{W_n^*\left(\theta_n^+\right)}>\frac{\gamma_n}{2}\right)\leq$$
$$P\left(\begin{array}{c}\sup_{\theta\in\Theta}\|b_n(\theta,\zeta_n(\theta,a))-b_n(\theta,\zeta_n(\theta,a'))\|\left\|W_n^*\left(\theta_n^+\right)-W^*(\theta_0)\right\|^{1/2}\\+\sup_{\theta\in\Theta}\|b_n(\theta,\delta_n,a)-b_n(\theta,\zeta_n(\theta,a'))\|\left\|W^*(\theta_0)\right\|^{1/2}>\frac{\gamma_n}{2}\end{array}\right)\leq$$
$$2P\left(\sup_{\theta}\|b_n(\theta,\delta_n,a)-b_n(\theta,\delta_n,a)\|>c_*\frac{\gamma_n}{2}\right)+P\left(\left\|W_n^*\left(\theta_n^+\right)-W^*(\theta_0)\right\|>K\right),$$

for $K>0$ and $c_*=\frac{1}{2}\min\left(\frac{1}{\sqrt{\|W^*(\theta_0)\|}},\frac{1}{\sqrt{K}}\right)$

Now we have that $\left(\sup_\theta \left| \begin{array}{c} \left\|\beta_n - b(\theta) - \sum_{i=1}^{2a} \frac{1}{n^{(i+1)/2}}\zeta_{i+1_n}(\theta)\right\| \\ -\left\|\beta_n - b(\theta) - \sum_{i=1}^{2a'} \frac{1}{n^{(i+1)/2}}\zeta_{i+1_n}(\theta)\right\| \end{array} \right| > c_* \frac{\gamma_n}{2}\right) \leq$
$P\left(\sup_\theta \left\|\sum_{i=2a'+1}^{2a} \frac{1}{n^{(i+1)/2}}\zeta_{i+1_n}(\theta)\right\| > c_* \frac{\gamma_n}{2}\right) \leq$
$P\left(\sum_{i=2a'+1}^{2a} \frac{1}{n^{(i+1)/2}} \sup_{\theta\in\Theta} \left\|\zeta_{i+1_n}(\theta)\right\| > c_* \frac{\gamma_n}{2}\right) = o\left(n^{-a^*}\right)$, hence due to R.10, we can choose $\gamma_n \leq \frac{2}{c_*}\sum_{i=2a'+1}^{2a} \frac{1}{n^{(i+1)/2}} M_i$. Hence in this case let $\eta_n^* = 2\sup_\theta \left|J\left(\beta_n, b_n(\theta, \zeta_n(\theta, a)), W_n^*\left(\theta_n^+\right)\right) - J\left(\beta_n, b_n(\theta, \zeta_n(\theta, a')), W_n^*\left(\theta_n^+\right)\right)\right|$ and the result follows. ■

Proof of Lemma 4.5. Due to lemma 4.1 and assumption A.2 (see remark R.11) we have that $\theta_n(a)$ and $\theta_n(a')$ are in $B_{\varepsilon_2}(\theta_0)$ with probability $1 - o\left(n^{-a^*}\right)$. Hence for some $0 < \varepsilon_3 < \varepsilon_2$, $\theta_n(a') \in B_{\varepsilon_3}(\theta_n(a)) \subset B_{\varepsilon_2}(\theta_0)$ with probability $1 - o\left(n^{-a^*}\right)$. Applying the mean value theorem on the gradient of J with respect to θ, we have that $\sqrt{n}(\theta_n(0) - \theta_n(a)) = \left(D^2 J^2\left(\beta_n, b_n\left(\theta_n^{++}, \zeta_n\left(\theta_n^{++}, a\right)\right), W_n^*\left(\theta_n^+\right)\right)\right)^{-1} \times$
$\sqrt{n} D J^2\left(\beta_n, b_n(\theta_n(0), \zeta_n(\theta_n(0), a)), W_n^*\left(\theta_n^+\right)\right)$.
It suffices to prove that $Q\left(\sqrt{n}\left\|\theta_n(0) - \theta_n(a)\right\| > \gamma_n'\right) = o\left(n^{-a^*}\right)$, for some $\gamma_n' = o\left(n^{-\varepsilon}\right)$ whence the choice of η_n'' is possible. Due to the norm submultiplicativity we have that $Q\left(\sqrt{n}\left\|\theta_n(0) - \theta_n(a)\right\| > \gamma_n'\right) \leq$
$Q\left(\begin{array}{c} \left\|\left(D^2 J\left(\beta_n, b_n\left(\theta_n^{++}, \zeta_n\left(\theta_n^{++}, a\right)\right), W_n^*\left(\theta_n^+\right)\right)\right)^{-1}\right\| \times \\ \left\|\sqrt{n} D J\left(\beta_n, b_n(\theta_n(0), \zeta_n(\theta_n(0), a)), W_n^*\left(\theta_n^+\right)\right)\right\| > \gamma_n' \end{array}\right)$. Now, due to the definition of GMR1, the triangle inequality, norm submultiplicativity, assumptions A.1-A.4 and the subsequent remarks R.4, R.10, R.12 and R.11, and by choosing $0 < \varepsilon < \varepsilon_3$, $\varepsilon^*, \delta_1, \delta_2 > 0$ and positive constants K_b, M_i, M_i' for $i = 1, \ldots, 2a$ we have that
$Q\left(\left\|\sqrt{n} D J^2\left(\beta_n, b_n(\theta_n(0), \delta_n, a), W_n^*\left(\theta_n^+\right)\right)\right\| > 2\rho_n\right) \leq$

$$Q\left(\begin{array}{c} \left\|\frac{\partial b'(\theta_n(0))}{\partial\theta} W_n^*\left(\theta_n^+\right) \sum_{i=1}^{2a} \frac{1}{n^{i/2}}\zeta_{i+1_n}(\theta_n(0))\right\| \\ +\left\|\sum_{i=1}^{2a} \frac{1}{n^{i//2}} \frac{\partial\zeta_{i+1_n}'(\theta_n(0))}{\partial\theta} W_n^*\left(\theta_n^+\right)\left(\beta_n - b(\theta_n(0))\right)\right\| \\ +\left\|\sum_{i=1}^{2a} \frac{1}{n^{(i+1)/2}} \frac{\partial\zeta_{i+1_n}'(\theta_n(0))}{\partial\theta} W_n^*\left(\theta_n^+\right) \sum_{i=1}^{2a} \frac{1}{n^{i/2}}\zeta_{i+1_n}(\theta_n(0))\right\| > \rho_n \end{array}\right) \leq$$

$$Q\left(\begin{array}{c} \left\|\frac{\partial b'(\theta_n(0))}{\partial\theta}\right\| \left\|W_n^*\left(\theta_n^+\right)\right\| \left\|\sum_{i=1}^{2a} \frac{1}{n^{i/2}}\zeta_{i+1_n}(\theta_n(0))\right\| \\ +\left\|\sum_{i=1}^{2a} \frac{1}{n^{i//2}} \frac{\partial\zeta_{i+1_n}'(\theta_n(0))}{\partial\theta}\right\| \times \\ \left\|W_n^*\left(\theta_n^+\right)\right\| \left\|\left(\beta_n - b(\theta_n(0))\right)\right\| \\ +\left\|\sum_{i=1}^{2a} \frac{1}{n^{(i+1)/2}} \frac{\partial\zeta_{i+1_n}'(\theta_n(0))}{\partial\theta}\right\| \times \\ \left\|W_n^*\left(\theta_n^+\right)\right\| \left\|\sum_{i=1}^{2a} \frac{1}{n^{i/2}}\zeta_{i+1_n}(\theta_n(0))\right\| > \rho_n \end{array}\right) \leq$$

$$
Q\left(\begin{array}{c}
\left\|\frac{\partial b'(\theta_n(0))}{\partial\theta}\right\|\left\|\sum_{i=1}^{2a}\frac{1}{n^{(i+1)/2}}\zeta_{i+1_n}\left(\theta_n\left(0\right)\right)\right\| \\
+\left\|\sum_{i=1}^{2a}\frac{1}{n^{i//2}}\frac{\partial\zeta'_{i+1_n}(\theta_n(0))}{\partial\theta}\right\|\left\|\left(\beta_n-b\left(\theta_n\left(0\right)\right)\right)\right\| \\
+\left\|\sum_{i=1}^{2a}\frac{1}{n^{(i+1)/2}}\frac{\partial E\zeta'_{i+1_n}(\theta_n(0))}{\partial\theta}\right\|\left\|\sum_{i=1}^{2a}\frac{1}{n^{i/2}}\zeta_{i+1_n}\left(\theta_n\left(0\right)\right)\right\|>\frac{\rho_n}{M_W}
\end{array}\right)
$$

$$
+P\left(\left\|W_n^*\left(\theta_n^*\right)\right\|>M_W\right)\leq
$$

$$
Q\left(\begin{array}{c}
\sup_{\theta\in B_{\varepsilon_2}(\theta_0)}\left\|\frac{\partial b'(\theta)}{\partial\theta}\right\|\times \\
\sup_{\theta\in B_{\varepsilon_2}(\theta_0)}\sum_{i=1}^{2a}\frac{1}{n^{i/2}}\left\|\zeta_{i+1_n}\left(\theta\right)\right\| \\
+\sup_{\theta\in B_{\varepsilon_2}(\theta_0)}\sum_{i=1}^{2a}\frac{1}{n^{i//2}}\left\|\frac{\partial\zeta'_{i+1_n}(\theta)}{\partial\theta}\right\|\times \\
\left(\left\|\beta_n-b\left(\theta_0\right)\right\|+\left\|b\left(\theta_n\left(0\right)\right)-b\left(\theta_0\right)\right\|\right) \\
+\sup_{\theta\in B_{\varepsilon_2}(\theta_0)}\sum_{i=1}^{2a}\frac{1}{n^{(i+1)/2}}\left\|\frac{\partial\zeta'_{i+1_n}(\theta)}{\partial\theta}\right\|\times \\
\sup_{\theta\in B_{\varepsilon_2}(\theta_0)}\sum_{i=1}^{2a}\frac{1}{n^{i/2}}\left\|\zeta_{i+1_n}\left(\theta\right)\right\|>\frac{\rho_n}{M_W}
\end{array}\right)
$$

$$
+P\left(\theta_n\left(0\right)\in B_{\varepsilon_2}\left(\theta_0\right)\right)+P\left(\left\|W_n^*\left(\theta_n^+\right)\right\|>M_W\right)\leq
$$

$$
P\left(\begin{array}{c}
\sup_{\theta\in\overline{B}_{\varepsilon}(\theta_0)}\left\|\frac{\partial b'(\theta)}{\partial\theta}\right\|\times \\
\sup_{\theta\in\overline{B}_{\varepsilon}(\theta_0)}\sum_{i=1}^{2a}\frac{1}{n^{i/2}}\left\|\zeta_{i+1_n}\left(\theta\right)\right\| \\
+\sup_{\theta\in\overline{B}_{\varepsilon}(\theta_0)}\sum_{i=1}^{2a}\frac{1}{n^{i//2}}\left\|\frac{\partial\zeta'_{i+1_n}(\theta)}{\partial\theta}\right\|\times \\
\left(\left\|\beta_n-b\left(\theta_0\right)\right\|+\left\|b\left(\theta_n\left(0\right)\right)-b\left(\theta_0\right)\right\|\right) \\
+\sup_{\theta\in\overline{B}_{\varepsilon}(\theta_0)}\sum_{i=1}^{2a}\frac{1}{n^{(i+1)/2}}\left\|\frac{\partial\zeta'_{i+1_n}(\theta)}{\partial\theta}\right\|\times \\
\sup_{\theta\in\overline{B}_{\varepsilon}(\theta_0)}\sum_{i=1}^{2a}\frac{1}{n^{i/2}}\left\|\zeta_{i+1_n}\left(\theta\right)\right\|>\frac{\rho_n}{M_W}
\end{array}\right)
$$

$$
+P\left(\theta_n\left(0\right)\in B_{\varepsilon_2}\left(\theta_0\right)\right)+P\left(\left\|W_n^*\left(\theta_n^+\right)\right\|>M_W\right)\leq
$$

$$
P\left(\begin{array}{c}
K_b\sum_{i=1}^{2a}\frac{M_i}{n^{(i+1)/2}}+ \\
\sum_{i=1}^{2a}\frac{M'_i}{n^{i//2}}\left\|\beta_n-b\left(\theta_0\right)\right\|+ \\
\sum_{i=1}^{2a}\frac{M'_i}{n^{i//2}}L_b\left\|\theta_n\left(0\right)-\theta_0\right\|+ \\
\sum_{i=1}^{2a}\frac{M'_i}{n^{(i+1)/2}}\sum_{i=1}^{2a}\frac{M_i}{n^{(i+1)/2}}>\frac{\rho_n}{M_W}
\end{array}\right)
$$

$$
+P\left(\theta_n\left(0\right)\in\overline{B}_{\varepsilon}\left(\theta_0\right)\right)+P\left(\left\|W_n^*\left(\theta_n^+\right)\right\|>M_W\right)
$$

$$
+\sum_{i=1}^{2a}Q\left(\left\|\zeta_{i+1_n}\left(\theta\right)\right\|>M_i\right)+\sum_{i=1}^{2a}Q\left(\left\|\frac{\partial\zeta'_{i+1_n}(\theta)}{\partial\theta}\right\|>M'_i\right)\leq
$$

$$
Q\left(\begin{array}{c}
K_b\sum_{i=1}^{2a}\frac{M_i}{n^{i/2}}+ \\
\sum_{i=1}^{2a}\frac{M'_i}{n^{i//2}}\delta_1+ \\
\sum_{i=1}^{2a}\frac{M'_i}{n^{i//2}}L_b\delta_2 \\
+\sum_{i=1}^{2a}\frac{M'_i}{n^{(i+1)/2}}\sum_{i=1}^{2a}\frac{M_i}{n^{i/2}}>\frac{\rho_n}{M_W}
\end{array}\right)
$$

$$
+P\left(\theta_n\left(0\right)\in\overline{B}_{\varepsilon}\left(\theta_0\right)\right)+P\left(\left\|W_n^*\left(\theta_n^+\right)\right\|>M_W\right)
$$

$$
+\sum_{i=1}^{2a}Q\left(\left\|\zeta_{i+1_n}\left(\theta\right)\right\|>M_i\right)+\sum_{i=1}^{2a}Q\left(\left\|\frac{\partial\zeta'_{i+1_n}(\theta)}{\partial\theta}\right\|>M'_i\right)
$$

$$
+P\left(\theta_n\left(0\right)\in B_{\delta_2}\left(\theta_0\right)\right)+P\left(\beta_n\in B_{\delta_1}\left(b\left(\theta_0\right)\right)\right)\leq
$$

$$
Q\left(K_b\sum_{i=1}^{2a}\frac{M_i}{n^{i/2}}+\left(\delta_1+L_b\delta_2\right)\sum_{i=1}^{2a}\frac{M'_i}{n^{i//2}}+\sum_{i=1}^{2a}\frac{M'_i}{n^{(i+1)/2}}\sum_{i=1}^{2a}\frac{M_i}{n^{i/2}}>\frac{\rho_n}{M_W}\right)
$$

$+ P\left(\theta_n(0) \in \overline{B}_{\varepsilon}(\theta_0)\right) + P\left(\left\|W_n^*\left(\theta_n^+\right)\right\| > M_W\right)$
$+ \sum_{i=1}^{2a} Q\left(\left\|\zeta_{i+1_n}(\theta)\right\| > M_i\right) + \sum_{i=1}^{2a} Q\left(\left\|\frac{\partial \zeta'_{i+1_n}(\theta)}{\partial \theta}\right\| > M'_i\right)$
$+ P\left(\theta_n(0) \in B_{\delta_2}(\theta_0)\right) + P\left(\beta_n \in B_{\delta_1}(b(\theta_0))\right) \leq o\left(n^{-a^*}\right)$ for
$\rho_n \leq M_W\left(K_b \sum_{i=1}^{2a} \frac{M_i}{n^{i/2}} + (\delta_1 + L_b\delta_2) \sum_{i=1}^{2a} \frac{M'_i}{n^{i//2}} + \sum_{i=1}^{2a} \frac{M'_i}{n^{(i+1)/2}} \sum_{i=1}^{2a} \frac{M_i}{n^{i/2}}\right) =$
$O\left(n^{-1/2}\right)$. In an analogous manner we can prove that there exists a positive constant C^*, such that $Q\left(\left\|\left(D^2 J\left(\beta_n, b_n\left(\theta_n^{++}, \zeta_n\left(\theta_n^{++}, a\right)\right), W_n^*\left(\theta_n^+\right)\right)\right)^{-1}\right\| > C^*\right) =$
$o\left(n^{-a}\right)$ and therefore we obtain the needed result if we choose $\gamma'_n \leq 2C^*\rho_n$.
■

Proof of Lemma 4.6. The result follows directly from AC.1 in appendix B due to lemma 4.5. ■

Proof of Lemma 4.7. It follows from lemma 4.6 and lemma 2 of Magdalinos (1992). ■

Proof of Lemma 4.8. Lemma 4.6 assures that the $GMR2^*(a)$ estimator admits an Edgeworth expansion of order $s = 2a + 2$, if assumption A.2 is valid for $a^* = a + \frac{1}{2}$. The rest follow from lemma AL.4 presented in the appendix B. ■

Proof of Corollary 4.9. Lemma AL.1 (in appendix B) is valid since, $\theta_n =$GMR2$^*(a)$, $\varphi_n - \varphi_0 = \begin{pmatrix} \beta_n - b(\theta_0) \\ f_n(\theta_0) \\ q_n(\theta_0) \\ \theta_n^* - \theta_0 \end{pmatrix}$ and the application is justified by the fact that provision 1 holds due to A.2, and 4.7, 2 follows from A.2, and A.3 and 3 follows from lemma 5 of [1] and A.2. The result follows from corollary AC.2 of appendix B. ■

Proof of Lemma 4.10. Using the procedure described in R.26 and noting that the derivatives of the estimating equations need not be approximated as in the case of the GMR2 due to their form and assumptions A.2, A.5.4 and A.4. Holding terms of the relevant order, we thus obtain
$\frac{\partial}{\partial\theta} b'_n\left(\theta_n\left(\frac{1}{2}\right), \zeta_n\left(\theta_n\left(\frac{1}{2}\right), a\right)\right) W_n^*\left(\theta_n^+\right) \sqrt{n}\left(\beta_n - b_n\left(\theta_n\left(\frac{1}{2}\right), \zeta_n\left(\theta_n\left(\frac{1}{2}\right), a\right)\right)\right) =$
$\mathbf{0}_p \Rightarrow$

$$\left(\frac{\partial b'(\theta_0)}{\partial \theta} + \frac{1}{\sqrt{n}}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n - \theta_0\right)\right]_{j=1,\ldots,q}\right) \times$$
$$\begin{pmatrix} W_n^*(\theta_0) \\ +\frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\theta'} W_{rj_n}^*(\theta_0)\sqrt{n}\left(\theta_n^+ - \theta_0\right)\right]_{r,j=1,\ldots,q} \end{pmatrix} \times$$

$$\begin{pmatrix} \left(\sqrt{n}\left(\beta_n - b\left(\theta_0\right)\right) - \frac{1}{\sqrt{n}}\zeta_{2_n}\left(\theta_0\right)\right) \\ -\left(\frac{\partial b(\theta_0)}{\partial \theta'} + \frac{1}{n}\frac{\partial \zeta_{2_n}(\theta_0)}{\partial \theta'}\right)\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right) \\ -\frac{1}{2\sqrt{n}}\left[\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)^{/}\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)\right]_{j=1,\dots,q} \end{pmatrix} = \mathbf{0}_p \Rightarrow$$

$$\left(\frac{\partial b'(\theta_0)}{\partial \theta} + \frac{1}{\sqrt{n}}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)\right]_{j=1,\dots,q}\right)$$

$$\times \begin{pmatrix} W^*\left(\theta_0\right) \\ +\frac{1}{\sqrt{n}}w^*\left(z,\theta_0\right) + \frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\theta^{/}}W^*_{rj}\left(\theta_0\right)q_1^*\right]_{r,j=1,\dots,q} \end{pmatrix}$$

$$\begin{pmatrix} \left(\sqrt{n}\left(\beta_n - b\left(\theta_0\right)\right) - \frac{1}{\sqrt{n}}Ek_2\left(\theta_0,z\right)\right) \\ -\frac{\partial b(\theta_0)}{\partial \theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right) \\ -\frac{1}{2\sqrt{n}}\left[\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)^{/}\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)\right]_{j=1,\dots,q} \end{pmatrix} = \mathbf{0}_p \Rightarrow$$

$$\begin{pmatrix} \frac{\partial b'(\theta_0)}{\partial \theta}W^*\left(\theta_0\right) + \frac{1}{\sqrt{n}}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)\right]W^*\left(\theta_0\right) \\ +\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial \theta}w^*\left(z,\theta_0\right) + \frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial \theta}\left[\frac{\partial}{\partial\theta^{/}}W^*_{rj}\left(\theta_0\right)q_1^*\right]_{r,j=1,\dots,q} \end{pmatrix}$$

$$\begin{pmatrix} \left(k_1\left(z,\theta_0,\delta_0\right) + \frac{1}{\sqrt{n}}\left(k_2\left(z,\theta_0,\delta_0\right) - Ek_2\left(z,\theta_0,\delta_0\right)\right)\right) \\ -\left(\frac{\partial b(\theta_0)}{\partial \theta'} + \frac{1}{n}\frac{\partial Ek_2(\theta_0,z)}{\partial \theta'}\right)\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right) \\ -\frac{1}{n}\frac{\partial Ek_2(z,\theta_0,\delta_0)}{\partial \delta'}\sqrt{n}\left(\delta_n - \delta_0\right) \\ -\frac{1}{2\sqrt{n}}\left[\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)^{/}_{j}\frac{\partial^2 b(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)\right]_{j=1,\dots,q} \end{pmatrix} = \mathbf{0}_p \Rightarrow$$

$$\begin{pmatrix} \frac{\partial b'(\theta_0)}{\partial \theta}W^*\left(\theta_0\right) + \frac{1}{\sqrt{n}}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)\right]W^*\left(\theta_0\right) \\ +\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial \theta}w^*\left(z,\theta_0\right) + \frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial \theta}\left[\frac{\partial}{\partial\theta^{/}}W^*_{rj}\left(\theta_0\right)q_1^*\right]_{r,j=1,\dots,q} \end{pmatrix}$$

$$\begin{pmatrix} \left(k_1 + \frac{1}{\sqrt{n}}\left(k_2\left(z,\theta_0,\delta_0\right) - Ek_2\left(z,\theta_0,\delta_0\right)\right)\right) - \frac{\partial b(\theta_0)}{\partial \theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right) \\ -\frac{1}{2\sqrt{n}}\left[\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)^{/}\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)\right]_{j=1,\dots,q} \end{pmatrix} = \mathbf{0}_p \Rightarrow$$

$$\begin{pmatrix} \frac{\partial b'(\theta_0)}{\partial \theta}W^*\left(\theta_0\right) + \frac{1}{\sqrt{n}}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)\right]W^*\left(\theta_0\right) \\ +\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial \theta}w^*\left(z,\theta_0\right) + \frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial \theta}\left[\frac{\partial}{\partial\theta^{/}}W^*_{rj}\left(\theta_0\right)q_1^*\right]_{r,j=1,\dots,q} \end{pmatrix} \times$$

$$\frac{\partial b(\theta_0)}{\partial \theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right) =$$

$$\begin{pmatrix} \frac{\partial b'(\theta_0)}{\partial \theta}W^*\left(\theta_0\right) + \frac{1}{\sqrt{n}}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)\right]W^*\left(\theta_0\right) \\ +\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial \theta}w^*\left(z,\theta_0\right) + \frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial \theta}\left[\frac{\partial}{\partial\theta^{/}}W^*_{rj}\left(\theta_0\right)q_1^*\right]_{r,j=1,\dots,q} \end{pmatrix} \times$$

$$\begin{pmatrix} \left(k_1\left(z,\theta_0,\delta_0\right) + \frac{1}{\sqrt{n}}\left(k_2\left(z,\theta_0,\delta_0\right) - Ek_2\left(z,\theta_0,\delta_0\right)\right)\right) \\ -\frac{1}{2\sqrt{n}}\left[\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)^{/}\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right)\right]_{j=1,\dots,q} \end{pmatrix} = \mathbf{0}_p \Rightarrow$$

$$\left(\begin{array}{c}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}+\frac{1}{\sqrt{n}}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\\ +\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}w^*(z,\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}+\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta^{/}}W^*_{rj}(\theta_0)q_1^*\right]_{r,j=1,\ldots,q}\frac{\partial b(\theta_0)}{\partial\theta'}\end{array}\right)$$

$$\times\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)=$$

$$\left(\begin{array}{c}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)k_1(z,\theta_0,\delta_0)\\ +\frac{1}{\sqrt{n}}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]W^*(\theta_0)k_1(z,\theta_0,\delta_0)\\ +\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}w^*(z,\theta_0)k_1(z,\theta_0,\delta_0)+\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta^{/}}W^*_{rj}(\theta_0)q_1^*(z,\theta_0,\delta_0)\right]_{r,j=1,\ldots,q}k_1(z,\theta_0,\delta_0)\\ +\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\left(k_2(z,\theta_0,\delta_0)-Ek_2(z,\theta_0,\delta_0)\right)\\ -\frac{1}{2\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\left[\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)_j^{/}\frac{\partial^2 b(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]_{j=1,\ldots,q}\end{array}\right)\Rightarrow$$

$$\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}+\frac{1}{\sqrt{n}}\left(\begin{array}{c}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\\ +\frac{\partial b'(\theta_0)}{\partial\theta}w^*(z,\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\\ +\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta^{/}}W^*_{rj}(\theta_0)q_1^*\right]_{r,j=1,\ldots,q}\frac{\partial b(\theta_0)}{\partial\theta'}\end{array}\right)\right)$$

$$\times\sqrt{n}(\theta_n-\theta_0)=$$

$$\left(\begin{array}{c}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)k_1\\ +\frac{1}{\sqrt{n}}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]W^*(\theta_0)k_1(z,\theta_0,\delta_0)\\ +\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}w^*(z,\theta_0)k_1(z,\theta_0,\delta_0)+\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta^{/}}W^*_{rj}(\theta_0)q_1^*\right]_{r,j=1,\ldots,q}k_1\\ +\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\left(k_2(z,\theta_0,\delta_0)-Ek_2(z,\theta_0,\delta_0)\right)\\ -\frac{1}{2\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\left[\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)_j^{/}\frac{\partial^2 b(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]_{j=1,\ldots,q}\end{array}\right)\Rightarrow$$

$$\begin{aligned}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\quad=\quad&\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)k_1(z,\theta_0,\delta_0)\\ &+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}(\theta_n-\theta_0)\right]W^*(\theta_0)k_1(z,\theta_0,\delta_0)\\ &+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}w^*(z,\theta_0)k_1(z,\theta_0,\delta_0)\\ &+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta^{/}}W^*_{rj}(\theta_0)q_1^*\right]_{r,j=1,\ldots,q}k_1(z,\theta_0,\delta_0)\\ &+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\left(k_2(z,\theta_0,\delta_0)-Ek_2(z,\theta_0,\delta_0)\right)\\ &-\frac{1}{2\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\left[\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)_j^{/}\frac{\partial^2 b(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]_{j=1,\ldots,q}\\ &-\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\\ &\times\left(\begin{array}{c}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\\ +\frac{\partial b'(\theta_0)}{\partial\theta}w^*(z,\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\\ +\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta^{/}}W^*_{rj}(\theta_0)q_1^*(z,\theta_0,\delta_0)\right]_{r,j=1,\ldots,q}\frac{\partial b(\theta_0)}{\partial\theta'}\end{array}\right)\end{aligned}$$

$$
\begin{aligned}
\times & \left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) k_1 \quad \Rightarrow \\
\sqrt{n}\left(\theta_n\left(\tfrac{1}{2}\right)-\theta_0\right) &= \left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) k_1 \\
+ & \frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}(\theta_n-\theta_0)\right] W^*(\theta_0) k_1(z,\theta_0,\delta_0) \\
& +\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} w^*(z,\theta_0) k_1(z,\theta_0,\delta_0) \\
& +\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} \left[\frac{\partial}{\partial\theta'} W^*_{rj}(\theta_0) q_1^*\right]_{r,j=1,\ldots,q} k_1(z,\theta_0,\delta_0) \\
& +\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \left(k_2(z,\theta_0,\delta_0) - Ek_2(z,\theta_0,\delta_0)\right) \\
& -\frac{1}{2\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \left[\sqrt{n}(\theta_n-\theta_0)_j' \frac{\partial^2 b(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\ldots,q} \\
& -\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\tfrac{1}{2}\right)-\theta_0\right)\right] \\
\times & \quad W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) k_1 \\
& -\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} w^*(z,\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'} \\
\times & \left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) k_1(z,\theta_0,\delta_0) \\
& -\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} \left[\frac{\partial}{\partial\theta'} W^*_{rj}(\theta_0) q_1^*(z,\theta_0,\delta_0)\right]_{r,j=1,\ldots,q} \\
\times & \frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) k_1(z,\theta_0,\delta_0) \quad \Rightarrow \\
\sqrt{n}\left(\theta_n\left(\tfrac{1}{2}\right)-\theta_0\right) &= \left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) k_1(z,\theta_0,\delta_0) \\
& +\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \left(k_2(z,\theta_0,\delta_0) - Ek_2(z,\theta_0,\delta_0)\right) \\
& -\frac{1}{2\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \\
\times & \left[\sqrt{n}\left(\theta_n\left(\tfrac{1}{2}\right)-\theta_0\right)_j' \frac{\partial^2 b(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}(\theta_n-\theta_0)\right]_{j=1,\ldots,q} \\
& +\frac{1}{\sqrt{n}}\left(\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\tfrac{1}{2}\right)-\theta_0\right)\right] W^*(\theta_0)\right) \\
\times & \left(Id_q - \frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0)\right) k_1(z,\theta_0,\delta_0) \\
& +\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} w^*(z,\theta_0) \\
\times & \left(Id_q - \frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0)\right) k_1(z,\theta_0,\delta_0) \\
& +\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} \left[\frac{\partial}{\partial\theta'} W^*_{rj}(\theta_0) q_1^*(z,\theta_0,\delta_0)\right]_{r,j=1,\ldots,q} \\
\times & \left(Id_q - \frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0) \frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1} \frac{\partial b'(\theta_0)}{\partial\theta} W^*(\theta_0)\right) k_1(z,\theta_0,\delta_0)
\end{aligned}
$$

which is the required result. ■

Proof of Corollary 4.13. Follows from direct substitutions on the results of lemma 4.10 by noting first that $\frac{\partial b(\theta_0)}{\partial \theta'} = \begin{pmatrix} Id_{p\times p} \\ \mathbf{0}_{q-p\times p} \end{pmatrix}$, $\frac{\partial b_j^2(\theta_0)}{\partial\theta\partial\theta'} = \mathbf{0}_{p\times p}\ \forall j = 1,\ldots,q$, and $w^* = \mathbf{0}_p$. ■

Proof of Lemma 5.1. For $a_2 = 0$ the result follows from remark R.32. For $a_2 \geq \frac{1}{2}$ we have that for large enough n by expanding analogously and keeping terms up to $O\left(n^{-2a_1}\right)$ we obtain

$$\mathbf{0}_{p\times 1} = \sqrt{n}\left(\beta_n - \theta_0\right) - \sum_{i=1}^{2a_2} \frac{1}{n^{\frac{i}{2}}}\zeta_{i+1_n}\left(\theta_0\right) - \mathbf{Id} \times \sqrt{n}\left(\theta_n\left(a_2\right) - \theta_0\right)$$

due to the fact that any partial derivative of any order up to $2a_1$ of ζ_{i+1_n} at θ_0 for any $i = 1,\ldots,2a_2$ is 0 due to the convention of definition D.7, and therefore we obtain that $\sqrt{n}\left(\theta_n\left(a_2\right)-\theta_0\right) \underset{a_1}{\backsim} \sqrt{n}\left(\beta_n - \theta_0\right)$ due to the same convention. The result follows since an analogous expansion would be valid for any θ at a relevant open neighborhood of θ_0 due to local approximate unbiaseness of the assumed order. ■

Proof of Lemma 5.2. We have that for large enough n in the case where the final computation concerns the $\theta_n\left(a^*\right)$, by expanding analogously

$$\begin{aligned}\mathbf{0}_{p\times 1} &= \sqrt{n}\left(\beta_n - \theta_0\right) - \sum_{i=2a_1+1}^{2a_2} \frac{1}{n^{\frac{i}{2}}}\zeta_{i+1_n}\left(\theta_0\right) \\ &- \left(\mathbf{Id} + \sum_{i=2a_1+1}^{2a_2} \frac{1}{n^{\frac{i+1}{2}}}\frac{\partial\zeta_{i+1_n}\left(\theta_0\right)}{\partial\theta'}\right)\sqrt{n}\left(\theta_n\left(a_2\right)-\theta_0\right) \\ &- \ldots\end{aligned}$$

due to the fact that any partial derivative of any order up to $2a_2$ of ζ_{i+1_n} at θ_0 for any $i = 1,\ldots,2a_1$ is 0 due to the convention of definition D.7, and therefore by keeping terms up to $O\left(n^{-a_1-\frac{1}{2}}\right)$ we obtain

$$\begin{aligned}&\sqrt{n}\left(\theta_n\left(a_2\right)-\theta_0\right) \underset{a_1+\frac{1}{2}}{\sim} \sqrt{n}\left(\beta_n - \theta_0\right) - \frac{1}{n^{a_1+\frac{1}{2}}}\zeta_{2a_1+2_n}\left(\theta_0\right) \\ &\underset{a_1+\frac{1}{2}}{\sim} \sum_{i=0}^{2a_1} \frac{1}{n^{\frac{i}{2}}}\left(k_{i+1}\left(z,\theta_0\right) - Ek_{i+1}\left(z,\theta_0\right)\right) + \frac{1}{n^{a_1+\frac{1}{2}}}\left(k_{2a_1+2_n}\left(z,\theta_0\right) - Ek_{2a_1+2_n}\left(z,\theta_0\right)\right)\end{aligned}$$

and the result follows since an analogous expansion would be valid for any θ at a relevant open neighborhood of θ_0 due to local approximate unbiaseness of the assumed order. ■

B Proofs of General Lemmas and Corollaries.

In this appendix we include several results, either directly drawn from the relevant references or simple extensions and/or corollaries of the latter. These are employed throughout the main body of the paper. In the following we denote by θ_n and φ_n (the n^{th} terms of) generic (sequences of) random elements with values in Euclidean spaces, with J_n (the n^{th} term of a sequence of) stochastic functions defined on the product of the aforementioned spaces and by J its pointwise stochastic (in the appropriate sense) limit. Recall also that $d = \max(2a+2, 3)$. Next lemma concerns the derivation of the validity of the Edgeworth expansion in any of the examined cases. It essentially determines that the local approximation of $\sqrt{n}(\theta_n - \theta_0)$ obtained by the inversion of a polynomial approximation of the first order conditions, has an error that is not greater that any $o(n^{-a})$-real sequence with probability $1 - o(n^{-a})$. This result, along with the provisions of corollary AC.1 that follows, establish that these two sequences have the same Edgeworth expansions if any one of them has a valid Edgeworth expansion.

Lemma AL.1 If

1. $P\left(\left\|n^{\frac{1}{2}}(\theta_n - \theta_0)\right\| > C\ln^{1/2} n\right) = o(n^{-a})$, $P\left(\left\|n^{\frac{1}{2}}(\varphi_n - \varphi_0)\right\| > C^* \ln^{1/2} n\right) = o(n^{-a})$ for $C, C^* > 0$,

2. $\frac{\partial J_n(\theta,\varphi)}{\partial\theta}$ is differentiable of order d in a neighborhood of (θ_0, φ_0) and the d order derivative is Lipschitz in this neighborhood (or in a subset of it) the Lipschitz coefficient is bounded with probability $1 - o(n^{-a})$, and $\frac{\partial J^2(\theta_0,\varphi_0)}{\partial\theta\partial\theta'}$ is positive definite,

3. $P\left(\left\|n^{\frac{1}{2}}(\varphi_n - \varphi_0) - n^{\frac{1}{2}}\pi(R_n)\right\| > \omega_n^*\right) = o(n^{-a})$ with π,R_n, and ω_n^* analogous to the relevant quantities of the present lemma (see below) that are derived in an analogous manner with a potentially different J_n,

then there exists a smooth function $\pi^* : \mathbb{R}^m \to \mathbb{R}^p$, that is independent of n such that

$$P\left(\left\|n^{\frac{1}{2}}(\theta_n - \theta_0) - n^{\frac{1}{2}}\pi^*(R_n^*)\right\| > \omega_n\right) = o\left(n^{-a}\right)$$

where R_n^* is the sequence of random elements with values on $\mathbb{R}^m$, with components the distinct components of $\frac{\partial J_n(\theta_0,\varphi_0)}{\partial\theta}$, and $\left\{D^{j_1,j_2}\left(\frac{\partial J_n(\theta,\varphi)}{\partial\theta}\right)|_{(\theta=\theta_0,\varphi=\varphi_0)}\right\}_{\substack{j_1+j_2=i\\,i=1,..,d-1}}$, where $D^{j_1,j_2}\left(\frac{\partial J_n(\theta,\varphi)}{\partial\theta}\right) = D_\varphi^{j_2} \circ D_\theta^{j_1}\left(\frac{\partial J_n(\theta,\varphi)}{\partial\theta}\right)$, $m = \dim(R_n^*)$ and $\omega_n = o(n^{-a})$ deterministic.

Proof. See Arvanitis and Demos (2010), lemma AL.2 or Andrews (2003) lemma 9. ■

Lemma AL.2 Suppose that $\sqrt{n}(\theta_n - \theta_0)$ admits a valid Edgeworth expansion of order $s = 2a + 1$. Let $\{x_n\}$ denote a sequence of random vectors and there exists an $\varepsilon > 0$ and a real sequence $\{a_n\}$, such that $a_n = o(n^{-\varepsilon})$ and $P(\sqrt{n}\|x_n\| > a_n) = o(n^{-a})$. Then any η_n, such that $P(\sqrt{n}(\theta_n - \theta_0 + x_n) = \eta_n) = 1 - o(n^{-a})$, admits a valid Edgeworth expansion of the same order.

Proof. We have that $\sup_{A\in\mathcal{B}_C}|P(\eta_n \in A) - P(\sqrt{n}(\theta_n - \theta_0 + x_n) \in A)| \leq \sup_{A\in\mathcal{B}_C}|P(\eta_n \in A, \sqrt{n}(\theta_n - \theta_0 + x_n) = \eta_n) - P(\sqrt{n}(\theta_n - \theta_0 + x_n) \in A)| + P(\sqrt{n}(\theta_n - \theta_0 + x_n) \neq \eta_n) = o(n^{-a})$, the rest follows as in the proof of lemma AL.3, in Arvanitis and Demos (2010). ■

Corollary AC.1 If $a \leq \varepsilon$ then $\pi_i(z) = \pi_i^*(z)$, $\forall i$, and therefore the resulting Edgeworth distribution coincides with the initial.

We term the Edgeworth expansion of the random sequence $n^{\frac{1}{2}}\pi^*(R_n^*)$ as the *formal* Edgeworth expansion of $\sqrt{n}(\theta_n - \theta_0)$. Notice that the formal expansion can be generally defined as the formal expansion of any other random element whose distance from $\sqrt{n}(\theta_n - \theta_0)$ is bounded by an $o(n^{-a})$-real sequence with probability $1 - o(n^{-a})$, in case that J_n is for example non differentiable.

Corollary AC.2 If $\sqrt{n}(\theta_n - \theta_0)$ has a valid Edgeworth expansion and the provisions of lemma AL.1 hold then this coincides with the formal expansion.

Proof. Trivial consequence of corollary AC.1. ■

Now, denote by P_n the measure $P \circ (\sqrt{n}(\theta_n - \theta_0))^{-1}$. The following lemma provides another asymptotic approximation of $\{P_n\}$ obtained from the validity of an Edgeworth approximation.

Lemma AL.3 Suppose that $\sqrt{n}(\theta_n - \theta_0)$ has an Edgeworth distribution of order $s = 2a+1$. Then for $z \sim N(\mathbf{0}_q, Id_q)$, there exists a polynomial function g_n with respect to z and $\frac{1}{\sqrt{n}}$, with values in $\mathbb{R}^q$ of order $s = 2a + 1$, such that if $Q_n = P \circ g_n(z, \theta_0)^{-1}$ then

$$\sup_{A\in\mathcal{B}_C}|P_n(A) - Q_n(A)| = o\left(n^{-a}\right)$$

Proof. Given a sequence of Edgeworth measures of order $s = 2a + 1$ (say $\mathcal{EDG}_n(\theta_0, a)$), due to the smoothness of the Normal distribution there exists a function g_n^* with values in $\mathbb{R}^q$ that is polynomial in x and $\frac{1}{\sqrt{n}}$ (for x a generic variable in the range of $\sqrt{n}\left(\theta_n-\theta_0\right))$ and is of order $s=2a+1$, such that $\sup_{A\in\mathcal{B}_C}\left|\Phi\left(g_n^*\left(A\right)\right)-\mathcal{EDG}_n\left(a\right)\left(A\right)\right|=o\left(n^{-a}\right)$. If P_n is approximated by the latter sequence in the same order, then due to the triangle inequality

$$\sup_{A\in\mathcal{B}_C}\left|P_n\left(A\right)-\Phi\left(g_n^*\left(A\right)\right)\right|=o\left(n^{-a}\right)$$

Now, $\Phi\left(g_n^*\left(A\right)\right)=P\left(z\in g_n^*\left(A\right)\right)=P\left(\left(g\left(z,\frac{1}{\sqrt{n}},\theta_0\right)+o\left(n^{-a}\right)\right)\in A\right)$ since g_n^* is polynomial and is invertible mod $\frac{1}{n^a}$. $Q_n=P\circ g\left(z,\frac{1}{\sqrt{n}},\theta_0\right)^{-1}$ is a smooth distribution since the Normal distribution is smooth and the fact that g_n is smooth w.r.t. z, therefore by construction $P\left(\left(g\left(z,\frac{1}{\sqrt{n}},\theta_0\right)+o\left(n^{-a}\right)\right)\in A\right)$ $=Q_n\left(A\right)+o\left(n^{-a}\right)$ and the approximation is uniform with respect to the relevant collection of measurable sets due to the fact that that g_n^* does not depend on A. The result follows from the triangle inequality. ■

Given the previous approximation and by strengthening the order of the Edgeworth expansion we obtain the following lemma that is quite useful for the validation of the analogous moment approximations.

Lemma AL.4 If P_n admits an Edgeworth approximation of order $s=2\left(a+1\right)$ then

$$n^a\left\|\int_{\mathbb{R}^q}x\left(dP_n-dQ_n\right)\right\|=o\left(1\right)$$

Proof. See Arvanitis and Demos (2010), lemma 4.1. ■

arg min **Properties**

In the following, let Θ be a compact metric space, and $\left(\Omega\,,\mathcal{F},P\right)$ a complete probability space. Let $\left(\mathcal{K}\left(\Theta\right),\mathcal{H}\right)$ denote that space of compact subsets of Θ, equipped with the Hausdorff metric. Let $\mathcal{B}_{\mathcal{H}}$ denote the corresponding Borel algebra.

Remark AR.1 Let J be a real function on $\Omega\times\Theta$, continuous on Θ for almost every $\omega\in\Omega$ and jointly measurable on the product algebra of $\Omega\times\Theta$. Then due to the compactness of Θ and by theorem 3.10 (iii) of Molchanov (2005) $\arg\min_\theta\circ J$ is non empty, measurable and almost surely compact valued. By theorem 2.13 of Molchanov (2005), $\arg\min_\theta\circ J$ has a measurable selection.

C Examples' Expansions

MA(1) calculations

Given the expansion results in [3], and employing the notation of lemma 4.2 in Arvanitis and Demos (2010) we have that

$$k_1 = \omega z, \quad and \quad k_2 = -2\theta \frac{\theta^4+1}{\left(\theta^2+1\right)^3} - \frac{1}{6}\frac{a_1+3a_3}{\omega^2} + \frac{1}{6}\frac{a_1+3a_3}{\omega^2}z^2$$

where

$$\begin{aligned} \omega^2 &= \frac{\theta^2+4\theta^4+\theta^6+\theta^8+1}{\left(1+\theta^2\right)^4}, \quad a_1 = \frac{6\theta\left(1+\theta^4\right)^2}{\left(1+\theta^2\right)^5} + \frac{\left(1+\theta^4\right)^3+\theta^3\left(1+\theta^2\right)^3}{\left(1+\theta^2\right)^6}\kappa_3^2, \\ a_3 &= -4\frac{\theta\left(\theta^2+4\theta^4+\theta^6+\theta^8+1\right)\left(1+\theta^4\right)}{\left(\theta^2+1\right)^7}, \end{aligned}$$

κ_3 is the third order cumulant of u_t, and z is a standard normal random variable.

Now from Arvanitis and Demos (2010) we have, for the second step estimator $\theta_n(0)$, that:

$$q_1 = \left(\frac{\partial b_0}{\partial \theta^{/}}\right)^{-1} k_1 = \frac{\sqrt{\theta^2+4\theta^4+\theta^6+\theta^8+1}}{1-\theta^2}z$$

$$\begin{aligned} q_2 &= \left(\frac{\partial b_0}{\partial \theta^{/}}\right)^{-1}\left(k_2 - \frac{1}{2}\left(\frac{\partial b_0}{\partial \theta^{/}}\right)^{-1}\frac{\partial^2 b_0}{\partial\theta\partial\theta'}q_1^2\right) \\ &= -2\frac{\theta}{1-\theta^2}\frac{\theta^4+1}{\theta^2+1} + \frac{\left(1+\theta^2\right)^2}{1-\theta^2}\frac{a_1^{(1)}+3a_3^{(1)}}{6\omega^2}\left(z^2-1\right) - \frac{\theta\left(\theta^2-3\right)}{\theta^2+1}\frac{\theta^2+4\theta^4+\theta^6+\theta^8+1}{\left(1-\theta^2\right)^3}z^2. \end{aligned}$$

Now for $\theta_n\left(\frac{1}{2}\right)$,applying corollary 4.11 again we get that

$$q_1^* = q_1 = \frac{\sqrt{\theta^2+4\theta^4+\theta^6+\theta^8+1}}{1-\theta^2}z,$$

$$q_2^* = \frac{1}{6}\frac{\left(\theta^2+1\right)^2}{1-\theta^2}\frac{a_1+3a_3}{\omega^2}\left(z^2-1\right) + \frac{\theta^2+4\theta^4+\theta^6+\theta^8+1}{\left(1-\theta^2\right)^3}\frac{\theta\left(3-\theta^2\right)}{\left(\theta^2+1\right)}\left(z^2-1\right)$$

and consequently $E\left(\theta_n^*\left(\frac{1}{2}\right)-\theta\right) = o\left(n^{-1}\right)$.

For $GMR2R$ applying corollary 4.11 once more we get:

$$q_1^{**} = \frac{\sqrt{\theta^2+4\theta^4+\theta^6+\theta^8+1}}{1-\theta^2} z$$

and

$$q_2^{**} = \frac{1}{6}\frac{\left(\theta^2+1\right)^2}{1-\theta^2}\frac{a_1^{(1)}+3a_3^{(1)}}{\omega^2}\left(z^2-1\right)+\frac{\theta^2+4\theta^4+\theta^6+\theta^8+1}{\left(1-\theta^2\right)^3}\frac{\theta\left(3-\theta^2\right)}{\left(\theta^2+1\right)}z^2.$$

Taking expectations we get the result in section 6.1.

On the other hand for $GMR2RS$, we have that

$$q_1^{***} = \frac{\sqrt{\theta^2+4\theta^4+\theta^6+\theta^8+1}}{1-\theta^2} z,$$

$$q_2^{***} = \frac{1}{6}\frac{\left(\theta^2+1\right)^2}{1-\theta^2}\frac{a_1^{(1)}+3a_3^{(1)}}{\omega^2}\left(z^2-1\right)+\frac{\theta^2+4\theta^4+\theta^6+\theta^8+1}{\left(1-\theta^2\right)^3}\frac{\theta\left(3-\theta^2\right)}{\left(\theta^2+1\right)}\left(z^2-1\right)$$

and

$$E\sqrt{n}\left(\theta_n^{***}\left(\frac{1}{2}\right)-\theta\right)=o\left(n^{-\frac{1}{2}}\right).$$

Finally, for the $GMR2$ estimator we have that

$$q_1 = \frac{\left(\theta^2+1\right)^2}{1-\theta^2}\omega z,$$

and

$$q_2 = \frac{1}{6}\frac{\left(\theta^2+1\right)^2}{1-\theta^2}\frac{a_1^{(1)}+3a_3^{(1)}}{\omega^2}\left(z^2-1\right)+\frac{\theta\left(3-\theta^2\right)}{1-\theta^2}\frac{\theta^2+4\theta^4+\theta^6+\theta^8+1}{\left(1+\theta^2\right)^5}z^2.$$

Taking expectations we get the result in section 6.1.

ARCH(1) calculations

For the ARCH(1) model we have that

$$\begin{aligned} G &= H_1^{-1}\left[E\left(\frac{y_{t-1}^4}{u_t^2}\right)^2 H_2 - E\left(\frac{y_{t-1}^4}{u_t^2}\right)E\left(\frac{y_{t-1}^2}{u_t^2}\right)(H_3+2H_4)\right] \\ &\quad + H_1^{-1}\left[\begin{array}{c}\left(E\left(\frac{y_{t-1}^2}{u_t^2}\right)^2+E\left(\frac{y_{t-1}^4}{u_t^2}\right)E\left(\frac{1}{u_t^2}\right)\right)H_5 \\ +E\left(\frac{y_{t-1}^2}{u_t^2}\right)^2 H_6 - E\left(\frac{1}{u_t^2}\right)E\left(\frac{y_{t-1}^2}{u_t^2}\right)H_7\end{array}\right],\end{aligned}$$

and

$$G^* = H_1^{-1}\left[E\left(\frac{y_{t-1}^2}{u_t^2}\right)^2 H_3 + E\left(\frac{y_{t-1}^4}{u_t^2}\right)E\left(\frac{1}{u_t^2}\right)H_4 - E\left(\frac{y_{t-1}^4}{u_t^2}\right)E\left(\frac{y_{t-1}^2}{u_t^2}\right)H_2\right]$$
$$+H_1^{-1}\left[-E\left(\frac{y_{t-1}^2}{u_t^2}\right)E\left(\frac{1}{u_t^2}\right)(2H_5+H_6)+\theta_3 E\left(\frac{1}{u_t^4}\right)H_7^*\right],$$

where

$$H_1 = \left[E\left(\frac{1}{u_t^2}\right)E\left(\frac{y_{t-1}^4}{u_t^2}\right) - E\left(\frac{y_{t-1}^2}{u_t^2}\right)^2\right]^2, \quad H_2 = \sum_{i=1}^n E\left(\frac{1}{u_t^2 u_{t-i}} - \frac{y_{t-i}^2}{u_t^2 u_{t-i}^2}\right),$$
$$H_3 = \sum_{i=1}^n E\left(\frac{y_{t-i-1}^2}{u_t^2 u_{t-i}} - \frac{y_{t-i}^2 y_{t-i-1}^2}{u_t^2 u_{t-i}^2}\right), \quad H_4 = \sum_{i=1}^n E\left(\frac{y_{t-1}^2}{u_t^2 u_{t-i}} - \frac{y_{t-1}^2 y_{t-i}^2}{u_t^2 u_{t-i}^2}\right),$$
$$H_5 = \sum_{i=1}^n E\left(\frac{y_{t-1}^2 y_{t-i-1}^2}{u_t^2 u_{t-i}} - \frac{y_{t-1}^2 y_{t-i}^2 y_{t-i-1}^2}{u_t^2 u_{t-i}^2}\right), \quad H_6 = \sum_{i=1}^n E\left(\frac{y_{t-1}^4}{u_t^2 u_{t-i}} - \frac{y_{t-1}^4 y_{t-i}^2}{u_t^2 u_{t-i}^2}\right) \quad \text{and}$$
$$H_7 = \sum_{i=1}^n E\left(\frac{y_{t-1}^4 y_{t-i-1}^2}{u_t^2 u_{t-i}} - \frac{y_{t-1}^4 y_{t-i-1}^2 y_{t-i}^2}{u_t^2 u_{t-i}^2}\right).$$

Now taking into account that

$$\frac{u_{t-i}-\theta_1}{\theta_3} = y_{t-i-1}^2 \quad \text{and} \quad y_t = u_t^{1/2} z_t$$

the above formulae can be simplified to

$$G = (H_1^*)^{-1}\begin{bmatrix} E\left(1+6\theta_1^2\frac{1}{u_t^2}+\theta_1^4\frac{1}{u_t^4}-4\theta_1\left(\frac{1}{u_t}+\theta_1^2\frac{1}{u_t^3}\right)\right)H_2 \\ -E\left(1-2\theta_1\frac{1}{u_t}+\theta_1^2\frac{1}{u_t^2}\right)E\left(\frac{1}{u_t}-\theta_1\frac{1}{u_t^2}\right)(H_3^*+2H_4^*) \end{bmatrix}$$
$$+(H_1^*)^{-1}\begin{bmatrix} \left[2E\left(\frac{1}{u_t^2}\right)-2\theta_1 E\left(\frac{1}{u_t^3}\right)+\theta_1^2 E\left(\frac{1}{u_t^4}\right)-\theta_1 E\left(\frac{1}{u_t^2}\right)E\left(\frac{1}{u_t}\right)\right]H_5^* \\ +E\left(\frac{1}{u_t^2}-2\theta_1\frac{1}{u_t^3}+\theta_1^2\frac{1}{u_t^4}\right)H_6^* \\ +\left[\theta_1 E\left(\frac{1}{u_t^2}\right)-E\left(\frac{1}{u_t}\right)+\right]E\left(\frac{1}{u_t^2}\right)(\theta_1 H_5^*+H_7^*) \end{bmatrix}$$

$$G^* = (H_1^*)^{-1}\begin{bmatrix} \left(E\left(\frac{1}{u_t^2}\right)-2\theta_1 E\left(\frac{1}{u_t^3}\right)+\theta_1^2 E\left(\frac{1}{u_t^4}\right)\right)(\theta_2 H_3^*+H_5^*) \\ +\theta_2\left(1-2\theta_1 E\left(\frac{1}{u_t}\right)+\theta_1^2 E\left(\frac{1}{u_t^2}\right)\right)E\left(\frac{1}{u_t^2}\right)(H_4^*+\theta_1 H_2) \\ -\theta_2\left(1-2\theta_1 E\left(\frac{1}{u_t}\right)+\theta_1^2 E\left(\frac{1}{u_t^2}\right)\right)E\left(\frac{1}{u_t}\right)H_2 \end{bmatrix}$$
$$+\theta_2 (H_1^*)^{-1}\left[-\left(E\left(\frac{1}{u_t}\right)-\theta_1 E\left(\frac{1}{u_t^2}\right)\right)E\left(\frac{1}{u_t^2}\right)(2H_5^*+H_6^*)+E\left(\frac{1}{u_t^4}\right)H_7^*\right]$$

where

$$H_1^* = \theta_1^2\left[2\left(E\left(\frac{1}{u_t^3}\right)-E\left(\frac{1}{u_t^2}\right)E\left(\frac{1}{u_t}\right)\right)+\theta_1\left(E^2\left(\frac{1}{u_t^2}\right)-E\left(\frac{1}{u_t^4}\right)\right)\right]^2,$$

$$H_3^* = nE\left(\frac{1}{u_t^2}\right)-\sum_{i=1}^{n}E\left(\frac{z_{t-i}^2}{u_t^2}\right)-\theta_1 H_2,\quad H_4^*=\sum_{i=1}^{n}\left(E\left(\frac{1}{u_t u_{t-i}}\right)-E\left(\frac{z_{t-i}^2}{u_t u_{t-i}}\right)\right)-\theta_1 H_2,$$

$$H_5^* = nE\left(\frac{1}{u_t}\right)-\sum_{i=1}^{n}E\left(\frac{z_{t-i}^2}{u_t}\right)+\theta_1\sum_{i=1}^{n}\left[E\left(\frac{z_{t-i}^2}{u_t u_{t-i}}\right)-E\left(\frac{1}{u_t u_{t-i}}\right)\right]-\theta_1 H_3^*,$$

$$H_6^* = 2\theta_1\sum_{i=1}^{n}\left[E\left(\frac{z_{t-i}^2}{u_t u_{t-i}}\right)-E\left(\frac{1}{u_t u_{t-i}}\right)\right]+\theta_1^2 H_2,\quad H_7^*=-\theta_1\left(2H_5^*+\theta_1 H_3^*\right)$$

and H_2 as before.

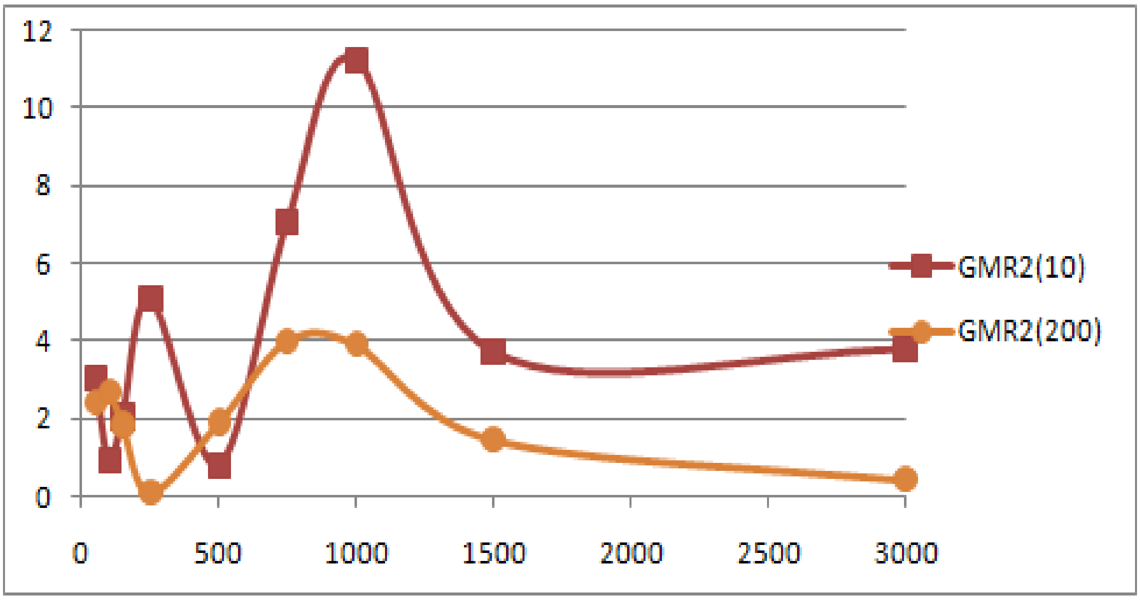

Figure 1: $n\,|E\,(GMR2(i)-\theta)|$, $i=10,200$, $MA(1)$ model, $\theta=-0.4$

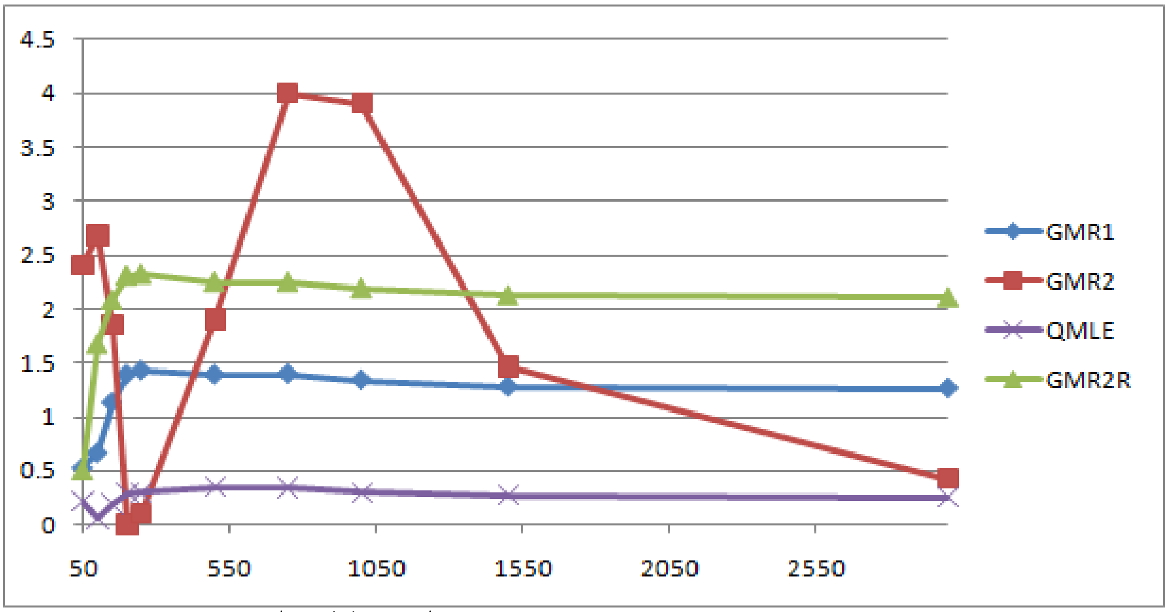

Figure 2: $n\left|E\left(\widehat{\theta}\right)-\theta\right|$ Biased Estimators, $MA(1)$ model, $\theta=-0.4$.

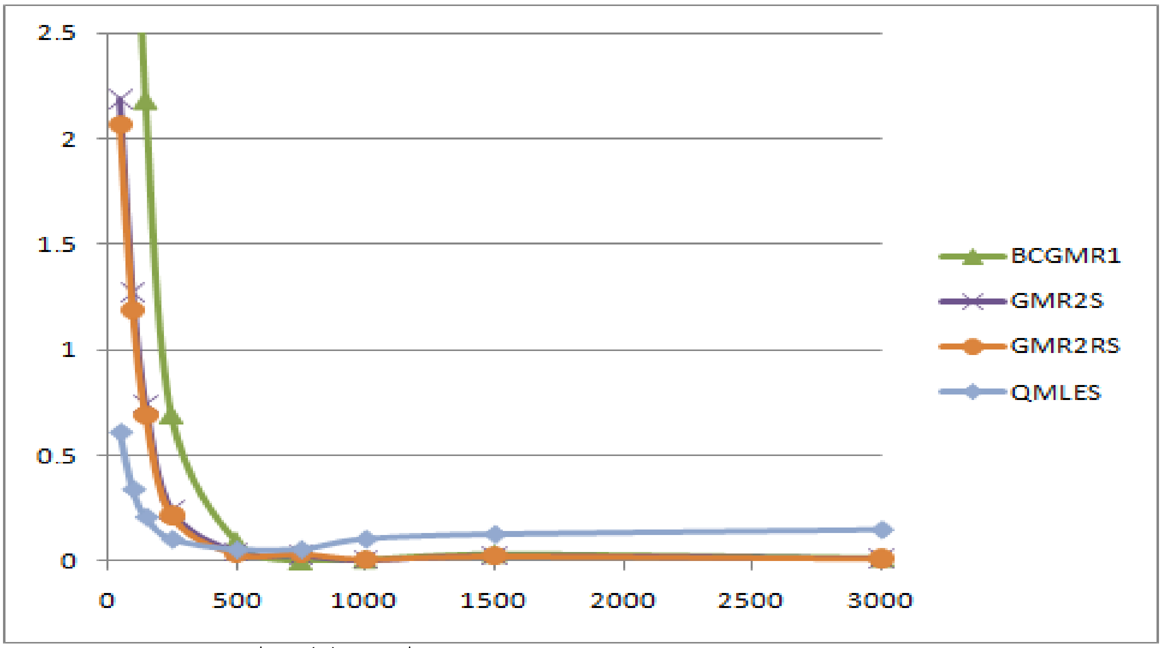

Figure 3: $n\left|E\left(\widehat{\theta}\right)-\theta\right|$ Unbiased Estimators, $MA(1)$ model, $\theta=-0.4$.

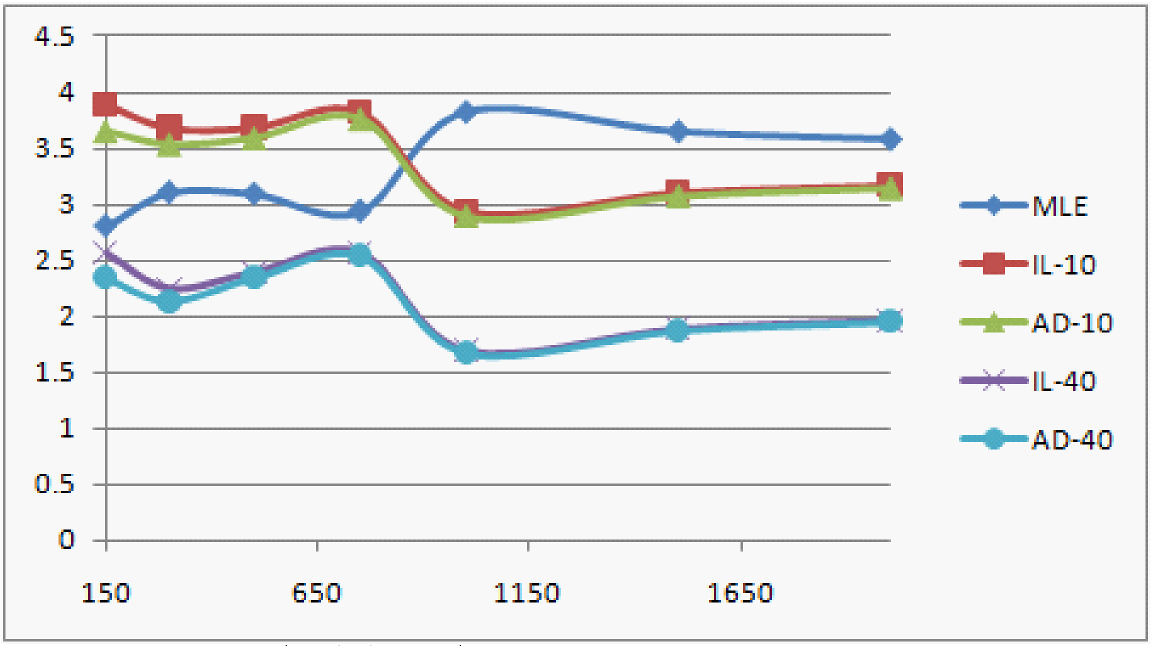

Figure 4: $n\left|E\left(\widehat{\theta_1}\right)-\theta_1\right|$ $ARCH(1)$ model, $\theta_1 = 1.0$ and $\theta_2 = 0.5$.

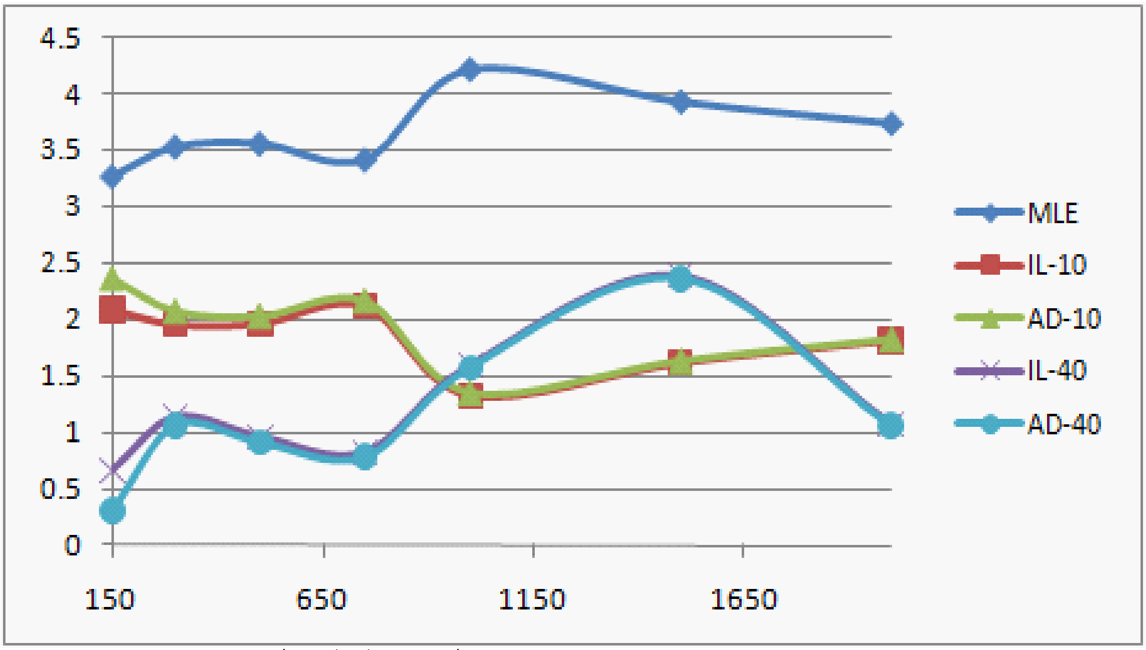

Figure 5: $n\left|E\left(\widehat{\theta_2}\right)-\theta_2\right|$ $ARCH(1)$ model, $\theta_1 = 1.0$ and $\theta_2 = 0.5$.