ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF ECONOMICS

WORKING PAPER SERIES **05-2011**

A New Class of Indirect Estimators and Bias Correction
(Preliminary and Incomplete)

Stelios Arvanitis and Antonis Demos

76 Patission Str., Athens 104 34, Greece
Tel. (++30) 210-8203911 - Fax: (++30) 210-8203301
www.econ.aueb.gr

A New Class of Indirect Estimators and Bias Correction
(Preliminary and Incomplete)

Stelios Arvanitis and Antonis Demos*
Athens University of Economics and Business

June 4, 2011

Abstract

In this paper we define a set of indirect estimators based on moment approximations of the auxiliary 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 optimization step approximation of the suggested one.

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

JEL: C10, C13

*Corresponding author A. Demos, Athens University of Economics and Business, Patission 76, Athens 10434, Greece. Tel: +302108203451, mail: demos@aueb.gr

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 [[13]]. 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 (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

[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.

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

[3] See e.g. Gourieroux and Monfort [[12]], and Gourieroux, Renault and N. Touzi [[14]].

(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.

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 [[14]] that the indirect estimator proposed by Gourieroux et al. [[13]], 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. [[14]] 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. [[13]]. 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 employing 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. [[13]] 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 [[24]], were formally introduced by Gourieroux et al. [[13]], complemented by Gallant and Tauchen [[10]] and extended by Calzolari, Fiorentini and E. Sentana [[8]]. Properties similar to those studied here were more or less algebraically studied in Gourieroux et al. [[14]] and more formally in Arvanitis and Demos [[4]]. 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 known IE are presented along with some of their properties and relations, that rely upon the particular assumption canvas.

Given a metric space (X, d_X) and $\mathbb{R}^q$ equiped with the usual metric we denote with $\mathrm{LS}(X, \mathbb{R}^q)$ the set of functions $X \to \mathbb{R}^q$ with lower semi-continuous components, suppressing the dependence on the metrics. The symbol $\mathcal{O}_\varepsilon(\theta)$ will denote the ε-ball around the point θ in a relevant metric space and $\overline{\mathcal{O}}_\varepsilon(\theta)$ its closure. 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] 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. For $a \in A \doteqdot \left\{\frac{i}{2}, i \in \mathbb{N}\right\}$, $d = 2a + 2$ and $\rightsquigarrow$ denote convergence in distribution.

Assumption A.1 The following characterize the basic framework:

1. Θ denotes a compact subset of the p-dimensional Euclidean space for $p \in \mathbb{N}$, equipped with the relevant subspace topology. Let $\theta_0 \in \mathrm{Int}(\Theta)$. Given a measurable space $(\Omega, \mathcal{F})$, the statistical model at hand is defined by a correspondence $\mathrm{par} : \Theta \twoheadrightarrow \mathcal{P}$ the set of probability measures on $\mathcal{F}$ such that $\mathrm{par}(\theta) \cap \mathrm{par}(\theta') \neq \varnothing$ iff $\theta = \theta'$. The (*arbitrary*) unknown probability measure (say P) at which the inferential procedures defined later aim, belongs to $\mathrm{par}(\theta_0)$. Also let P_θ denote any member of $\mathrm{par}(\theta)$.

2. The limit binding function (*lbf*) $b \in \mathrm{LS}(X, \mathbb{R}^q)$, for B be a compact subset of $\mathbb{R}^q$ for $q \geq p$, such that $b(\Theta) \subset \mathrm{Int}(B)$ and suppose that $b(\theta_0) = b(\theta)$ iff $\theta = \theta_0$. b is $d + 1$ times continuously differentiable on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ for some $\varepsilon > 0$, and $\mathrm{rank}\left(\frac{\partial b}{\partial \theta'}(\theta_0)\right) = p$.

3. There exists a function $\varsigma_n : \Omega \times B \to \mathbb{R}$ that is $\mathcal{B}_{\mathbb{R}} / (\mathcal{F} \otimes \mathcal{B}_B)$ $-$measurable and $\varsigma_n(\omega, \beta)$ is (lower semi) continuous on $b(\Theta)$ for $P_\theta-$almost all ω, for any θ, and there exists a function $\varsigma : \Theta \times b(\Theta) \to \mathbb{R}$ such that $\varsigma_n(\omega, \beta) \rightsquigarrow \varsigma(\theta, \beta)$ uniformly over $b(\Theta)$, uniformly over P_θ for any $\theta \in \Theta$. Also $\varsigma(\theta, b(\theta)) < \varsigma(\theta, \beta)$ $\forall \beta \in b(\Theta)$ and $\forall \theta \in \Theta$.[5]

[4] Notice that due to the fact that finite dimensional matrix spaces are identified with finite dimensional Euclidean spaces, the norm equivalence theorem applies.

[5] Componentwise lower semi-continuity of the *lbf* would follow from the continuity of

4. Let $W_n^*(\cdot,\theta)$ be $\mathcal{B}_{M(\mathbb{R},q)}/\left(\mathcal{B}_{\mathcal{F}_n}\otimes\mathcal{B}_\Theta\right)$ −measurable and P_{θ_0}-almost surely positive definite, for every $\theta\in\Theta$.

We denote with $E_\theta f(\theta')=\displaystyle\int f(\omega,\theta')\,dP_\theta(\omega)$ for any appropriate f and $\theta,\theta'\in\Theta$.

Remark R.1 For an appropriate sequence of measurable spaces $((\Omega_n,\mathcal{F}_n))_{n=1}^\infty$, we usually have that $\Omega=\prod_n\Omega_n$, $\mathcal{F}=\bigotimes_n\mathcal{F}_n$ and that any $P^*\in\operatorname{par}(\Theta)$ is the unique extension on $\mathcal{F}$, of a sequence of probability measures $(P_n^*)_{n=1}^\infty$- with P_n^* defined on $\bigotimes_{i=1}^n\mathcal{F}_i$- that is Kolmogorov consistent. Given the Kolmogorov consistency, the existence of P^* is guaranteed when Ω_n is a Hausdorff topological space, $\mathcal{F}_n$ is the relevant Borel algebra, and P_n^* is tight for any n (see corollary 15.28 of Aliprantis and Border [[1]]).Usually Ω_n is homeomorphic to $\mathbb{R}^m$ for some m in $\mathbb{N}$ and $\mathcal{F}_n$ is the Borel algebra with respect to the Euclidean topology.

Remark R.2 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 something that is also true for its derivatives on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$.

In the following we suppress the dependence of the aforementioned binding functions on Ω where unnecessary. We also let θ_n^+ denote a random element with values in Θ. 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).

par (i.e. for any θ, $\theta_n\to\theta$ the sequence comprised by any member of $\operatorname{par}(\theta_n)$ converges inside $\operatorname{par}(\theta)$ with respect to the weak topology), the uniform w.r.t. P_θ $\varsigma_n(\omega,\beta)\rightsquigarrow\varsigma(\theta,\beta)$ uniformly over $b(\Theta)$, a furher condition of injectivity for b and he identification of $b(\theta)$ with the degenerate probability measure at $b(\theta)$. The differeniability assumptions could follow from analogous assumptions for $\varsigma(\theta,\beta)$ and the implicit function theorem.

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

$$\varsigma_n(\beta_n) = \inf_{\beta \in B} \varsigma_n(\beta)$$

Remark R.3 In view of assumption A.1.3 and by remark AR.4 (in Appendix B) the above estimator is well defined.

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.4 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

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

and

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

Their existence is justified by remark AR.4 in view of assumption A.1. The computation of the estimators relies on the knowledge of b and b_n which is in most cases unavailable. Hence the estimators are usually approximated by the use of resampling techniques such as Monte Carlo simulations, which itself involves *nested numerical optimizations* that is of potentially large computational cost especially in the case of the second estimator.

Assumptions Specific to a New Class of IE Let $a^* = \frac{s-1}{2}$ for $s \geq 2a + 1$. Let also $\mathcal{EDG}_{n,\theta,a^*}$ denote the n^{th} term of an Edgeworth measure of order s with respect to $N(0, V_\theta)$, where V_θ is a positive definite $q \times q$ matrix for any θ, and $\mathcal{B}_C$ the collection of measurable convex subsets of the Euclidean $\mathbb{R}^{q+p}$ (see Appendix B).

Assumption A.2

$$\sup_{A \in \mathcal{B}_C} \left| P_\theta\left(\sqrt{n}\left(\beta_n - b(\theta)\right) \in A\right) - \mathcal{EDG}_{n,\theta,a^*}(A) \right| = o\left(n^{-a^*}\right)$$

for any $\theta \in \Theta$. Then, there exist $k_{i+1}(z, \theta)$, $i = 0, \ldots, 2a$ that are polynomial functions in z, with $O(1)$ coefficients such that

$$\int_{\mathbb{R}^q} z d\mathcal{EDG}_{n,\theta,a} = \sum_{i=0}^{2a} \frac{1}{n^{i/2}} Ek_{i+1}(z, \theta) \qquad (2)$$

where $Ek_1(z, \theta) = \mathbf{0}_{q \times 1}$ on Θ.

Lemma 2.1 If $s > 2a + 1$

$$\left\| E_\theta \beta_n - b(\theta) - \sum_{i=0}^{2a} \frac{1}{n^{(i+1)/2}} E k_{i+1}(z, \theta) \right\| = o\left(n^{-a-\frac{1}{2}}\right) \tag{3}$$

Remark R.5 The assumption on the $\sqrt{n}$ rate of convergence of the auxiliary estimator could be extended so as to allow different rates as long as these do not depend on θ.

The remaining assumptions are *local*. The next one concerns the asymptotic behavior of the sequence of stochastic weighting matrices described in A.1.5,6.

Assumption A.3 $W_n^*(\theta)$ is d-continuously differentiable P_{θ_0}-almost surely $\forall \theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$. There exists a $M(\mathbb{R}, l)$ valued function denoted by $W^*(\theta)$, defined on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$, such that

$$P(\|W_n^*(\theta) - W^*(\theta)\| > \delta) = o\left(n^{-a^*}\right), \forall \delta > 0, \forall \theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$$

Moreover

$$P\left(\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} \left\|D^d W_n^*(\theta)\right\| > C_{W^*}\right) = o\left(n^{-a^*}\right)$$

for some $C_{W^*} > 0$.

The following assumption, enables the *stochastic approximation* of the Edgeworth mean in (2). This can facilitate the definition of the IE that depend on the latter, in case where $Ek_{i+1}(z, \theta)$ are analytically unknown for some i, due to the structure of statistical model 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 sampling methods like Monte Carlo simulations, bootstrap e.t.c.

Assumption A.4 The following characterize the basic framework:

1. For a probability space $(\Omega', \mathcal{F}', P')$ and each $i = 1, \ldots, 2a$, there exist $\zeta_{i+1_n} : \Omega \times \Omega' \times \Theta \to \mathbb{R}^q$, that is $\mathcal{B}_{\mathbb{R}^q} / (\mathcal{F} \otimes \mathcal{F}' \otimes \mathcal{B}_\Theta)$-measurable, Q-almost everywhere continuous on Θ and Q-almost everywhere $d+1$ continuously differentiable on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$, where $Q = P \times P'$.

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

3. $Q\left(\sup_{\theta\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)}\left\|D^r\zeta_{i+1_n}(\omega,\omega',\theta)\right\|>M_{i,r}\right)=o\left(n^{-a^*}\right)$, for $M_{i,r}>0$, $\forall i=1,\ldots,2a$, for $r=1,2$.

ω' can be thought of as a simulated random element, which along with the "observed" sample ω constitutes a generalized sample that can be employed to approximate the relevant expectations. The space Ω' 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.

Remark R.6 This setup 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 it 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 ω' and θ, or their value at θ_0 is approximated by measurable functions of ω. This allows also for approximations of $Ek_{i+1}(z,\theta)$ when the latter is *partially computed* at stochastic point close to θ_0, enabling the derivation of estimators that emerge from *partial optimization.*

We enrich our assumption framework by a partial 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 $\operatorname{vec}D^iW_n^*(\theta)-T_i(\theta)$, for $i=1,\ldots,d$ for $T_i(\theta):\overline{\mathcal{O}}_{\varepsilon}(\theta_0)\to\mathbb{R}^{\dim\operatorname{vec}D^iW_n^*(\theta)}$, for any $\theta\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$. Let $q_n(\theta)$ be the vector containing the elements of $\zeta_{i+1_n}(\omega,\omega',\theta)-Ek_{i+1}(z,\theta)$ and $\operatorname{vec}D^i\zeta_{i+1_n}(\omega,\omega',\theta)-Z_i(\theta)$, for $i=1,\ldots,d+1$ for $Z_i(\theta):\overline{\mathcal{O}}_{\varepsilon}(\theta_0)\to\mathbb{R}$ for any $\theta\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$. Let also $m_n^*(\theta_0)$ be $\left((\beta_n-b(\theta_0))',\sqrt{n}\left(\theta_n^+-\theta_0\right),f_n'(\theta_0),q_n'(\theta_0)\right)'$ and $\mathcal{EDG}^*_{n,\theta,a^*}$ denote the n^{th} term of an Edgeworth sequence of order s, for $\sqrt{n}m_n^*(\theta_0)$ under Q.

Assumption A.5

$$\sup_{A\in\mathcal{B}_C}\left|Q\left(\sqrt{n}m_n^*(\theta_0)\in A\right)-\mathcal{EDG}^*_{n,\theta_0,a^*}(A)\right|=o\left(n^{-a^*}\right)$$

Remark R.7 Due to assumption A.2 and lemma 2 of [18] we have that,

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

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

A discussion on the validation of assumptions A.2, A.3, A.5 is provided in Appendix B. The following lemma is useful, especially for lemma 4.5.

Lemma 2.2 Assumptions A.3, A.2,A.5 imply that

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

Assumption A.5 along with the postulated behavior of the binding function and the definition of the GMR1 estimator implies that the latter admits also a valid Edgeworth expansion of the same order. Due to reasons that will become apparent in the next paragraph we also denote with $\theta_n(0)$ the GMR1 estimator.

Lemma 2.3 Under assumptions A.1, A.3, A.4, and A.5 $\sqrt{n}\left(\theta_n(0)-\theta_0\right)$ admits a valid Edgeworth expansion of order $2a^*+1$.

3 Definition of the GMR2$^*(a)$ Estimators

We are now ready to define a new class of IE based on these moment approximations. In what follows we suppress the dependence of the approximating functions ζ_{i+1_n} on the generalized sample space for notational convenience and denote $\zeta_n(\theta,a)=\left(\zeta_{2_n}(\theta),\ldots,\zeta_{2a+1_n}(\theta)\right)$ 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

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

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

Remark R.8 The existence of $GMR2^*(a)$ is facilitated by assumptions A.4 and A.3, and remark AR.4 in the appendix.

Remark R.9 Due to $Ek_1(z,\theta)=\mathbf{0}_{q\times1}$ we identify the $GMR1$ estimator with the $GMR2^*(0)$ one, and this justifies the relevant choice of notation in the previous paragraph.

Remark R.10 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.11 In the case that $\beta_n = \theta_n(0)$, and $b(\theta) = \theta$, we consider a variant of the GMR2$^*(a)$, defined as

$$\theta_n^*(a) = \theta_n(0) - \sum_{i=1}^{2a} \frac{1}{n^{(i+1)/2}} \zeta_{i+1_n}(\theta_n(0))$$

Q-almost everywhere, 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 literature in the case where $a = \frac{1}{2}$, and in this case it is called "feasible bias corrector" 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 $\forall \varepsilon > 0$,

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

and therefore

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

Remark R.12 In the light of lemma 4.1 it is evident that for example, θ_n^+ could be defined as $\theta_n^*\left(a\right)$ 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

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

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

Proposition 4.2 If $\sup_{\theta\in\Theta}\left\|E_\theta\beta_n - b\left(\theta\right)\right\| = o\left(1\right)$, then

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

Remark R.13 The assumption $\sup_{\theta\in\Theta}\left\|E_\theta\beta_n - b\left(\theta\right)\right\| = o\left(1\right)$ would follow from the uniform pseudo consistency of the auxiliary estimator given the compactness of B. If the lbf is a bijection $\sup_{\theta\in\Theta} P_\theta\left(\sup_{\beta\in B}\left\|\varsigma_n\left(\beta\right) - \varsigma(\theta,\beta)\right\| > \delta\right) = o\left(1\right)$ $\forall\delta > 0$ would be sufficient for this, given assumption A.1.3.

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

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

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

Remark R.14 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.11.2).

Remark R.15 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

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

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 $\zeta_{i+1_n}(\theta) = E k_{i+1}(z,\theta)$, 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.16 Again, any $GMR2^*(a')$ is an approximate $GMR2^*(a)$ for any $a > a'$. This is particularly valid when $a' = 0$, since $GMR1 = GMR2^*(0)$.

Remark R.17 Obviously lemma R.7 could be readily deduced from lemma 4.4 and an easily established uniform convergence of $J(\beta_n, b(\theta), W_n^*(\theta_n^*))$ to $J(b(\theta_0), b(\theta), W^*(\theta_0))$ in P_{θ_0}-probability $1 - o(n^{-a})$.

4.2 Asymptotic Tightness and Validity of Edgeworth Approximation

In this paragraph, we are concerned with the higher order approximation of the distribution of $GMR2^*(a)$ for $a > 0$. We essentially rely on the previous results, the local differentiability of the criterion from which it emerges and lemma AL.1 presented at the appendix.

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

The validity of the Edgeworth expansion of $\sqrt{n}(\theta_n(a) - \theta_0)$ 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 $GMR2^*(a)$ 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 $Q\left(\sqrt{n}\|\theta_n(a) - \theta_0\| > C^* \ln^{1/2} n\right) = o\left(n^{-a^*}\right)$.

Lemma 4.6 does not provide any further insight on the form of the Edgeworth approximation for $GMR2^*(a)$. However, lemma 4.7 along with the first part of lemma AL.3 can validate an Edgeworth approximation, the polynomials of the density of which, are obtained as in the proof of the first part of latter lemma.

Lemma 4.8 Under the assumptions of lemma 4.5, the $GMR2^*(a)$ admits an Edgeworth expansion of order $s = 2a^* + 1$, the density of which is obtained as in the proof of the first part of lemma AL.3.

4.3 Moment Approximations

Lemma 4.8 in the light of lemmas 7.1 and AL.3 if $a^* > a$, provide with an approximation of the sequence of (any order) moments of the defined estimator. In the following we expicitily provide this type of approximation in the case where $a = \frac{1}{2}$ for the mean.

Lemma 4.9 If $a^* \geq a + \frac{m}{2}$, then

$$\left| E_{\theta_0}\left(K\left(\sqrt{n}\left(\theta_n\left(a\right) - \theta_0\right)\right)^m\right) - \int_{\mathbb{R}^q} K\left(\left(\psi_n\left(z\right)\right)^m\right)\left(1 + \sum\nolimits_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i\left(z\right)\right)\varphi_V\left(z\right)dz\right|$$

is $o\left(n^{-a}\right)$ where $\psi_n\left(z\right) = \Pr_p\left(\sum_{i=1}^{2a+1} D^i \kappa_n^{*-1}\left(0\right)\left(z^j\right)\right)$, $\kappa_n^* = \left(\kappa_n, \varpi_n\right)$, κ_n is as in the proof of lemma 4.8 $\varpi_n : \mathbb{R}^{\dim(m_n^*)} \to \mathbb{R}^{\dim(m_n^*)-p}$ is an orthogonal projection of $\mathbb{R}^{\dim(m_n^*)}$ to a $\dim\left(m_n^*\right) - p$ dimensional subspace composed with a linear isometry with $\mathbb{R}^{\dim(m_n^*)-p}$, such $\left(Dg_n\left(0\right)\right)^{-1} \cap \left(\varpi_n\left(0\right)\right)^{-1} = \{0\}$, K is any m-linear function on $\mathbb{R}^p$ and $x^m = \underbrace{\left(x, x, \ldots x\right)}_{m}$ and $\left(1 + \sum\nolimits_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i\left(z\right)\right)\varphi_V\left(z\right)$ denotes the density of the Edgeworth approximation of order $2a + 1$ of $\sqrt{n}m_n^*\left(\theta_0\right)$ and $\Pr_p$ is the projection on the first p coordinates.

We next provide the analogous approximation for $a = \frac{1}{2}$. We obtain the following lemma.

Lemma 4.10 If the previous assumptions hold for $a = \frac{1}{2}$ $a^* > \frac{1}{2}$, then

$$\left\| E_{\theta_0}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right) - \theta_0\right) - E\left(q_1\left(z, \theta_0\right)\right) - \frac{1}{\sqrt{n}}E\left(q_2\left(z, \theta_0\right)\right)\right\| = o\left(n^{-a}\right)$$

where

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

and

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

$A = \left(Id_q - \frac{\partial b(\theta_0)}{\partial \theta'} \left(\frac{\partial b'(\theta_0)}{\partial \theta} W^* (\theta_0) \frac{\partial b(\theta_0)}{\partial \theta'} \right)^{-1} \frac{\partial b'(\theta_0)}{\partial \theta} W^* (\theta_0) \right)$,
$B = \left(\frac{\partial b'(\theta_0)}{\partial \theta} W^* (\theta_0) \frac{\partial b(\theta_0)}{\partial \theta'} \right)^{-1} \frac{\partial b'(\theta_0)}{\partial \theta} \; w^* (z, \theta_0)$ and q_1^* are the $O\left(\frac{1}{\sqrt{n}}\right)$ terms of the mean of the Edgeworth distribution of $\sqrt{n} m_n^* (\theta_0)$ corresponding to the weightening matrix and the initial estimator respectively, and z is a zero mean normal random vector of dimension equal to $\dim (m_n^* (\theta_0))$. Moreover

$$\left\| E_{\theta_0} \sqrt{n} (\theta_n (0) - \theta_0) - E (q_1^* (z, \theta_0)) - \frac{1}{\sqrt{n}} E (q_2^* (z, \theta_0)) \right\| = o \left(n^{-a} \right)$$

where

$$q_1^* = q_1$$

and

$$q_2^* = q_2 + BW^* (\theta_0) E k_2$$

Remark R.18 Lemma 4.10 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 Newey and Smith 2001).

Remark R.19 Notice that neither q_1 nor q_2 depend on the analogous terms in the approxiamtions of the remaining elements of $\sqrt{n} m_n^* (\theta_0)$. This would not hold in higher order expansions concerning $\theta_n \left(\frac{1}{2}\right)$.

Remark R.20 It is easy to see that q_1 and q_2 would provide the analogous approximation to bias of $\theta_n (\widetilde{a})$ for $\widetilde{a} > \frac{1}{2}$.

Remark R.21 Even if the aforementioned moment approximations are not valid (for example in cases where $a^* = \frac{1}{2}$), the relevant moments of the Edgeworth measures could be used for comparisons between the employed estimators in the spirit of Magdalinos [[18]] (see the second paragraph immediately after Theorem 2).

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 in lemma 4.10 and for $p = q$ we obtain

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

and

$$q_2 = \left(\frac{\partial b(\theta_0)}{\partial \theta'}\right)^{-1} \left(k_2 - Ek_2 - \frac{1}{2}\left[trq_1 q_1' \frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,q}\right)$$

Use the following definition.

Definition D.5 An estimator admitting a moment expansion such as the ones considered in the previous sections, will be termed approximately unbiased of order $s = 2a + 1$ if the relevant expansion is valid, and

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

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

Hence 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 $\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'} = \mathbf{0}_{p\times p}\ \forall j = 1, \ldots, q$, $Eq_2 = \mathbf{0}_p$, hence the estimator is approximately unbiased of order 2. This is not true for the GMR1 estimator unless the auxiliary estimator is is approximately unbiased of order 2 since $Eq_2 = Ek_2$.

Remark R.22 As it will become apparent next, 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.

Remark R.23 Under conditions that ensure that the derivatives of $E_\theta \beta_n$ are uniformly bounded around θ_0, an analogous result would hold for the GMR2 estimator (see Arvanitis and Demos [[4]]).

Remark R.24 If the previous assumptions hold for $a = \frac{1}{2}\ a^* > 1$, in the light of lemma 4.9 q_1 and q_2 can be used in order to provide the second order MSE of the estimator. In the particular case and due to the fact that Ek_1 is zero, it can be easily seen that

$$\left\| E_{\theta_0} n\left((\theta_n(\beta) - \theta_0)\right)(\theta_n(\beta) - \theta_0)' - Eq_1 q_1' - \frac{1}{\sqrt{n}} E\left(q_1 q_2' + q_2 q_1'\right)\right\| = o\left(n^{-a}\right)$$

for any $\beta \in A$, thereby establishing the superiority of $\theta_n^*(a)$ for any $a \geq \frac{1}{2}$, over $\theta_n^*(0)$ w.r.t. 2^{nd} order mean-MSE comparisons.

Returning to the issue of the bias approximation we extend the previous result when the binding function is in local canonical form. For the definition of the local canonical form, see Arvanitis and Demos [[4]], section 4.2, which is derived by theorem 10.2 of Spivak [[25]]) (p. 44). This notion concerns the choice of the auxiliary parameterization 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 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.25 The $GMR2^*\left(\frac{1}{2}\right)$ estimator is second order unbiased even in cases where $q > p$, when there is non stochastic weighting given that the binding function is in local canonical form. However, *given an admissible auxiliary statistical model, there always exists an auxiliary parameterization such that the previous result is valid*, given the relevant weighting structure.

Remark R.26 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.11. It can be verified using our assumption framework that

$$\left\| E_{\theta_0}\left(\sqrt{n}\left(\theta_n^*\left(\frac{1}{2}\right) - \theta_0\right)\right) - E(k_1) + \frac{1}{\sqrt{n}} E(k_2 - Ek_2) \right\| = o\left(n^{-1/2}\right)$$

thereby it is in this respect equivalent to $\sqrt{n}\left(GMR2^*\left(\frac{1}{2}\right) - \theta_0\right)$, while it is of minimal computational burden. However, we make the following observations:[6]

1. Due to remark R.11.1 $\theta_n^*\left(\frac{1}{2}\right)$ could be non-definable for fixed 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.

[6] 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.

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 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.[7] We make the following assumption.

Assumption A.6 ,The binding function is the identity, and the auxiliary estimator satisfies assumption A.2 for some $s^* \geq 2a+1$ given a. For any $a^+ \leq a$ consider any function of the form $\zeta_n(\theta, a^+) = \left(\zeta_{2_n}(\theta), \ldots, \zeta_{(2a^+)_n}(\theta)\right)$ where for each $i = 1, \ldots, 2a^+$, assumptions A.4 and A.5 hold, under the convention that if β_n is s_*^{th}-order approximately 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. Denote by $GMR^*(\zeta_n(\theta, a^+))$ the GMR2$^*(a^+)$ derived via the use of $\zeta_n(\theta, a^+)$.

The first part of the previous assumption does not pose any loss of generality compared to the previous sections, since β_n could itself be an IE of θ_0. In this respect any concern about the asymptotic behavior of sequences of weighting matrices becomes asymptotically irrelevant. Denote the set of functions satisfying the second part by $Z_n(a^+, \beta_n)$. This convention is motivated by the fact that $Z_n(0, \beta_n) = \{\mathbf{0}_\Theta\}$. We denote with $GMR^*(\zeta_n(\theta, a^+))$ $GMR2^*(a^+)$ w.r.t. $\zeta_n(\theta, a^+) \in Z_n(a^+, \beta_n)$.[8]

[7] It will also provide an algorithm for the computation of the local canonical form of the binding function, discussed in the previous section.

[8] 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.

Definition D.6 Given $a_1, a_2 \leq a$, 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 the derivation of 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 $a_f : \{1, 2, \ldots, K\} \to \{0, \ldots, a\}$. 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 prevents this set of estimators from obtaining a rich enough algebraic structure.

Remark R.27 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.

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

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

Given the following question:

Question Does there exist an IE procedure providing an estimator that is is approximately unbiased of order $2a + 1$?

An answer is provided by the next algorithmic procedure.

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 $\theta_n^{(i)}$ =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 unbiased of order $\left(2a^{(i)}+1\right)$.

Then $\theta_n^{(2(a-a_1))}$ is approximately 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.[9]

Remark R.28 In the case the $\theta_n^*\left(a^{(i)}\right)$ is needed for some i, remarks R.11 and R.26 would also hold.

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 corresponding to the $MA(1)$ model and the MLE for an $ARCH(1)$ one under conditional normality. To provide with additional evidence for our theoretical results we perform a small simulation exercise.

6.1 MA(1)

Consider the invertible $MA(1)$ process

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

Suppose that Θ is a compact subset of $(-1, 1)$, 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

[9] Due to the need to avoid the relevant moment approximations at boundary points we could consider a finite sequence of parameter spaces, say $\Theta^{(i)}$ for $i = 0, \ldots, 2(a - a_1)$ such that $\Theta^{(i)} \subset \Theta^{(i-1)}$, and $\Theta^{(i)}$ is compact with non empry interior, and the relevant results hold for any θ in the interior of $\Theta^{(i)}$.

of an $AR(1)$ auxiliary model (see Gourieroux et al. [[13]], and Demos and Kyriakopoulou [[6]]). Notice also that $b(\theta) = \frac{\theta}{1+\theta^2}$, and let B be a compact subset of $\left(-\frac{1}{2}, \frac{1}{2}\right)$ compliant to assumption A.1.

From the calculations in appendix C we have that

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

as in Demos and Kyriakopoulou [[6]]. 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 $GMR2^*\left(\frac{1}{2}\right) \otimes GMR1$ (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 $\theta = 0$. Hence

$$E\left[\sqrt{n}\left(GMR2R-\theta\right)\right] = \frac{1}{\sqrt{n}}\frac{\theta^2+4\theta^4+\theta^6+\theta^8+1}{\left(\theta^2+1\right)}\frac{\theta\left(3-\theta^2\right)}{\left(1-\theta^2\right)^3}. \tag{5}$$

However, considering $GMR2RS \doteqdot GMR2^*\left(\frac{1}{2}\right) \otimes GMR2R$, 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

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

i.e. the $GMR2$ is not 2^{nd} order *unbiased*, as expected due to nonlinearities in the binding function. Comparing equation (5) with (6) 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 *unbiased* at 0.

Notice that our assumption framework is easily validated for any of the IE described in the previous paragraph. More specifically, assumption A.1.1-3 follows easily while A.1.4 is irrelevant. Assumption A.2 follows from the results of Demos and Kyriakopoulou [[6]] for $a = \frac{1}{2}$ and $a^* = 1$. Assumption A.3 is again irrelevant, while due to the fact that $\zeta_2 = Ek_2$ assumptions A.4-A.5 follow trivially from the continuity of the relevant expressions and their derivatives and the compactness of Θ. For the aforementioned IE in the realm of section 5, assumption A.6 follows trivialy ($\Theta^{(2)}$ is chosen as a compact subset of the initial Θ).

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 4 for bias correction, i.e.

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

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. [[13]]). Of course under the assumptions in Gourieroux et al. [[13]] 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 6). 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.[10]

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 the Edgeworth expansions validated in Iglesias and Linton [[16]], and Iglesias and Phillips [[17]], that $Ek_2 = \begin{pmatrix} G \\ G^* \end{pmatrix}$ for the MLE $\theta_n = \begin{pmatrix} \widehat{\theta}_{1_n} \\ \widehat{\theta}_{2_n} \end{pmatrix}$ where G and G^* are given in the appendix C.

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 [[16]], named IL, and the indirect estimator suggested

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

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 [[16]], and Iglesias and Phillips [[17]]).

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 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 [[9]]). This experiment elucidates remark R.6. 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.

Our assumption framework can be validated for the AD type of estimators as follows. Θ is chosen compact and the binding function is obviously the identity. Hence assumption A.1 (again due to the coincidence of the underlying dimensions assumption A.1.4 is irrelevant) follows trivially. A.2 follows for the results of the aforementioned papers for $a = \frac{1}{2}$, $a^* = 1$. A.3 is again irrelevant while assumption A.4 follows from the continuity of G, G^* and their derivatives with respect to $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$, the compactness of Θ, the conditional normality and the fact that $u_t(\theta)$ has a uniform strictly positive lower bound. For A.5 consider the random element represented by $\frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \lambda_{1_i}(\theta) \\ \lambda_{2_i}(\theta) \end{pmatrix}$ where $\lambda_{1_i}(\theta)$ containes the elements of the derivatives of the of the i^{th} factor of the logarithmic likelihood function up to order $d = 3$. $\lambda_{2_i}(\theta)$ contains the elements that are used in the description above to approximate, via simulations, the analytically unknown moments that emerge in G and G^*. These are in the form of arithmetic means with respect to independent random elements that depend on θ (for example an element of $\lambda_{2_i}(\theta)$ would be $\frac{1}{h}\sum_{j=1}^{h} \frac{1}{u_j(\theta)}$, where $h = \frac{n^*}{n}$). Due to the compactness of Θ, the conditional normality and the fact that $u_t(\theta)$ has a uniform strictly positive lower bound and the results of the aforementioned papers, the conditions 2-4 of Gotze and Hipp [[11]] for $\begin{pmatrix} \lambda_{1_i}(\theta) \\ \lambda_{2_i}(\theta) \end{pmatrix}$ are easily (yet tediously) established, a fact that validates an Edgeworth expansion of order 3 for $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \lambda_{1_i}(\theta_0) - E\lambda_{1_i}(\theta_0) \\ \lambda_{2_i}(\theta_0) - E\lambda_{2_i}(\theta_0) \end{pmatrix}$. Notice that $\sqrt{n}(\theta_n - \theta_0) = \pi_n\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \lambda_{1_i}(\theta_0) - E\lambda_{1_i}(\theta_0) \\ \lambda_{2_i}(\theta_0) - E\lambda_{2_i}(\theta_0) \end{pmatrix}\right) + R_n$ with $P(\|R_n\| > \gamma_n) = o(n^{-1})$ for $\gamma_n = o(n^{-1})$, and π_n satisfying the assumptions of lemma AL.3 (see for example the proof of lemma 8 of Andrews [[2]]). Hence $\begin{pmatrix} \sqrt{n}(\theta_n - \theta_0) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\lambda_{1_i}(\theta_0) - E\lambda_{1_i}(\theta_0)) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\lambda_{2_i}(\theta_0) - E\lambda_{2_i}(\theta_0)) \end{pmatrix}$ admits a valid Egdeworth expansion of order 3. Moreover the elements of $\lambda_{2_i}(\theta)$ are smooth functions of θ and moreover $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\lambda_{2_i}(\theta_n) - E\lambda_{2_i}(\theta_0)) = \pi_n^*\begin{pmatrix} \sqrt{n}(\theta_n - \theta_0) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\lambda_{1_i}(\theta_0) - E\lambda_{1_i}(\theta_0)) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\lambda_{2_i}(\theta_0) - E\lambda_{2_i}(\theta_0)) \end{pmatrix} +$ R_n^* with $P(\|R_n^*\| > \gamma_n^*) = o(n^{-1})$ for $\gamma_n^* = o(n^{-1})$, and π_n^* in the premises of remark R.29. Hence $\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\lambda_{2_i}(\theta_n) - E\lambda_{2_i}(\theta_0)) \\ \sqrt{n}(\theta_n - \theta_0) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\lambda_{1_i}(\theta_0) - E\lambda_{1_i}(\theta_0)) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\lambda_{2_i}(\theta_0) - E\lambda_{2_i}(\theta_0)) \end{pmatrix}$ admits a valid Egdeworth expansion of order 3. Now,

$$\sqrt{n}m_n(\theta_0) = \begin{pmatrix} \sqrt{n}(\theta_n - \theta_0) \\ \sqrt{n}G\left(\left(\frac{1}{n}\sum_{i=1}^{n}\lambda_{2_i}(\theta_n)\right), \theta_0\right) - G(E\lambda_{2_i}(\theta_0), \theta_0) \\ \sqrt{n}G^*\left(\left(\frac{1}{n}\sum_{i=1}^{n}\lambda_{2_i}(\theta_n)\right), \theta_0\right) - G^*(E\lambda_{2_i}(\theta_0), \theta_0) \end{pmatrix}$$

can be easily seen to satisfy an analogous expression with respect to a sequence of smooth functions of $\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\lambda_{2_i}(\theta_n) - E\lambda_{2_i}(\theta_0)) \\ \sqrt{n}(\theta_n - \theta_0) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\lambda_{1_i}(\theta_0) - E\lambda_{1_i}(\theta_0)) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\lambda_{2_i}(\theta_0) - E\lambda_{2_i}(\theta_0)) \end{pmatrix}$ and an appropriate remainder. Hence assumption A.5 follows. An analogous reasoning would justify the Edgeworth expansions for the IL estimators.

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.11). In these cases we throw away the particular Monte Carlo samples and draw new ones.[11]

[11] 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.

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 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] Aliprantis Ch. D., and K. C. Border (2005), "*Infinite Dimensional Analysis*", Third Edition, Springer.

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

[3] Andrews D.W.K., Of. Lieberman (2005), "Valid Edgeworth Expansions for the Whittle Maximum Likelihood Estimator for Stationary Long-Memory Gaussian Time Series", *Econometric Theory* 21, pp. 710-734.

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

[5] Bhattacarya R.N. and R.R. Rao (1975), "*Normal Approximations and Asymptotic Expansions*", Wiley Series in Probability and Mathematical Statistics.

[6] 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.

[7] Durbin J. (1980), "Approximations for Densities of Sufficient Estimators", *Biometrica* 67-2, pp. 311–333.

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

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

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

[11] Gotze F., and C. Hipp (1994), "Asymptotic distribution of statistics in time series", *The Annals of Statistics*, 22, pp. 2062-2088.

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

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

[14] 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.

[15] Gradshteyn I. S., and I.M. Ryzhik (1994), "Table of Integrals, Series, and Products" A. Jeffrey Editor, Academic Press.

[16] 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.

[17] 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.

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

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

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

[21] R.R. Rao (1962), "Relations Between Weak and Uniform Convergence of Measures with Applications", *The Annals of Mathematical Statistics*, Vol. 33, 2, pp. 659–680.

[22] Skovgaard IB M. (1981), "Transformation of an Edgeworth Expansion by a Sequence of Smooth Functions", *Scandinavian Journal of Statistics 8*, pp. 207-217.

[23] Stokey N.L. and R.E. Lucas (1989), "*Recursive Methods in Economic Dynamics*", Harvard University Press.

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

[25] 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 2.1. The result follows from the compactness of B and lemma 7.1. ■

Proof of Lemma 2.2. Due to assumpion AL.1 θ_n^+ lies in $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ with P-probability $1-o(n^{-a})$. Then a Taylor expansion of order d of $W_n\left(\theta_n^+\right)$ around θ_0 along with assumptions A.3, AL.1, implies that for any $\delta>0$, exist $\delta_i>0$, $i=0,\ldots,d$ such that

$$\begin{aligned}
& P\left(\left\|W_n^*\left(\theta_n^+\right)-W\left(\theta_0\right)\right\|>\delta\right) \\
\leq\;& P\left(\left\|W_n^*\left(\theta_0\right)-W\left(\theta_0\right)\right\|>\delta_0\right)+\textstyle\sum_{i=1}^d P\left(\left\|\theta_n^+-\theta_0\right\|>\delta_i\right)=o\left(n^{-a}\right)
\end{aligned}$$

■

Proof of Lemma 2.3. First notice that due to the definition of the GMR1 and remark R.7 the estimator lies in $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ with probability $1-o\left(n^{-a^*}\right)$. Hence it satisfies first order conditions with the same probability due to A.1.2. A mean value expansion of the first order conditions around θ_0 along with A.2, A.4, A.5 implies that $P\left(\sqrt{n}\left\|\theta_n(0)-\theta_0\right\|>C\sqrt{\ln n}\right)=o\left(n^{-a^*}\right)$, for some $C>0$. A Taylor expansion of order d of the first order conditions implies that with probability $1-o\left(n^{-a^*}\right)$ $\sqrt{n}\left(\theta_n(0)-\theta_0\right)=L\sqrt{n}m_n^*(\theta_0)+\frac{1}{\sqrt{n}}\rho_n\left(\sqrt{n}m_n^*(\theta_0)\right)+R_n$ where L is an $p\times\dim\left(m_n(\theta)\right)$ matrix of rank p due to A.1, A.3, ρ_n is a polynomial function with absolutely bounded coefficients due to A.1, and $P\left(\left\|R_n\right\|>\gamma_n\right)=o\left(n^{-a^*}\right)$ for some $\gamma_n=o\left(n^{-a^*}\right)$ due to $P\left(\sqrt{n}\left\|\theta_n(0)-\theta_0\right\|>C\sqrt{\ln n}\right)=o\left(n^{-a^*}\right)$, A.2, A.4, A.5. Hence from lemma AL.1 the result would follow if $L\sqrt{n}m_n^*(\theta_0)+\frac{1}{\sqrt{n}}\rho_n\left(\sqrt{n}m_n^*(\theta_0)\right)$ has a valid Edgeworth expansion of the respective order. This is established by lemma 3 of [18] and assumption A.5. ■

Proof of Lemma 4.1. We have that

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

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

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

Now, due to the triangle inequality, submultiplicativity A.4.3, A.2, R.7 and 2.2 we have for the first term of the last sum that it is less than or equal to $Q\left(\sup_{\theta\in\Theta}\left\|\beta_n-b(\theta)-\sum_{i=0}^{2a}\frac{1}{n^{(i+1)/2}}\zeta_{i+1_n}(\theta)\right\|^2\left\|W_n^*\left(\theta_n^+\right)-W^*(\theta_0)\right\|>\frac{\varepsilon}{2}\right)\leq$

$$Q\left(\begin{array}{c}\sup_{\beta_n\in\overline{\mathcal{O}}_{\varepsilon_1^*}(b(\theta_0))}\sup_{\theta\in\Theta}\left\|\beta_n-b(\theta)-\sum_{i=0}^{2a}\frac{1}{n^{(i+1)/2}}\zeta_{i+1_n}(\theta)\right\|\\ \left\|W_n^*\left(\theta_n^+\right)-W^*(\theta_0)\right\|^{1/2}>\frac{\varepsilon}{2}\end{array}\right)+$$
$Q\left(\beta_n\notin\mathcal{O}_{\varepsilon_1^*}(b(\theta_0))\right)\leq$
$$Q\left(\begin{array}{c}\left(\|\beta_n\|+\sup_\theta\|b(\theta)\|+\sum_{i=0}^{2a}\frac{1}{n^{(i+1)/2}}\sup_{\theta\in\Theta}\left\|\zeta_{i+1_n}(\theta)\right\|\right)\\ \left\|W_n^*\left(\theta_n^+\right)-W^*(\theta_0)\right\|>\frac{\varepsilon}{2}\end{array}\right)+$$
$P\left(\beta_n\notin\overline{\mathcal{O}}_{\varepsilon_1^*}(b(\theta_0))\right)\leq$
$$P\left(\left\|W_n^*\left(\theta_n^+\right)-W^*(\theta_0)\right\|>\frac{\varepsilon}{2\left(c_1+c_2+\sum_{i=0}^{2a}\frac{M_i}{n^{(i+1)/2}}\right)}\right)+$$
$P\left(\beta_n\notin\overline{\mathcal{O}}_{\varepsilon_1^*}(b(\theta_0))\right)+\sum_{i=0}^{2a}Q\left(\sup_{\theta\in\Theta}\left\|\zeta_{i+1_n}(\theta)\right\|>M_i\right)=o\left(n^{-a^*}\right)$ due to remark R.7. For the second term we have that due to the continuous mapping theorem $\exists\epsilon>0:$
$Q\left(\sup_\theta\left|J^2(\beta_n,b_n(\theta,\delta_n,a),W^*(\theta_0))-J^2(b(\theta_0),b(\theta),W^*(\theta_0))\right|>\frac{\varepsilon}{2}\right)\leq$
$Q\left(\sup_\theta\left|J(\beta_n,b_n(\theta,\delta_n,a),W^*(\theta_0))-J(b(\theta_0),b(\theta),W^*(\theta_0))\right|>\epsilon\right)$ and due to the triangle inequality, this is less than or equal to
$$Q\left(\sup_\theta\left\|\left(\beta_n-b(\theta)-\sum_{i=0}^{2a}\frac{1}{n^{(i+1)/2}}\zeta_{i+1_n}(\theta)\right)-(b(\theta_0)-b(\theta))\right\|_{W^*(\theta_0)}>\epsilon\right)\leq$$
$Q\left(\|\beta_n-b(\theta_0)\|+\sum_{i=0}^{2a}\frac{1}{n^{(i+1)/2}}\sup_{\theta\in\Theta}\left\|\zeta_{i+1_n}(\theta)\right\|>\epsilon\right)$. The last term is less than or equal to
$Q\left(\|\beta_n-b(\theta_0)\|>\frac{\epsilon}{\sum_{i=0}^{2a}\frac{M_i}{n^{(i+1)/2}}}\right)+\sum_{i=0}^{2a}Q\left(\sup_{\theta\in\Theta}\left\|\zeta_{i+1_n}(\theta)\right\|>M_i\right)$ which is obviously $o(n^{-a})$. The result follows from the continuous mapping theorem and assumption A.1.3 which implies that $J(b(\theta_0),b(\theta),W^*(\theta_0))$ is uniquely minimized at θ_0. ■

Proof of Proposition 4.2. We have that $Q\left(\sup_{\theta\in\Theta}\|b_n(\theta)-b_n(\theta,\zeta_n(\theta,a))\|>\varepsilon\right)\leq$
$P\left(\sup_{\theta\in\Theta}\|b_n(\theta)-b(\theta)\|>\frac{\varepsilon}{2}\right)+Q\left(\sum_{i=1}^{2a}\frac{1}{n^{i+1/2}}\sup_{\theta\in\Theta}\left\|\zeta_{i+1_n}(\theta)\right\|>\frac{\varepsilon}{2}\right)\leq$
$P\left(\sup_{\theta\in\Theta}\|b_n(\theta)-b(\theta)\|>\frac{\varepsilon}{2}\right)+Q\left(\sum_{i=1}^{2a}\frac{M_i}{n^{i+1/2}}>\frac{\varepsilon}{2}\right)+\sum_{i=0}^{2a}Q\left(\sup_{\theta\in\Theta}\left\|\zeta_{i+1_n}(\theta)\right\|>M_i\right)$, for some $\varepsilon>0$. Now due to assumption A.1.3 and due to the hypothesis that $\sup_{\theta\in\Theta}\|b_n(\theta)-b(\theta)\|$ converges to zero uniformly on Θ, the first two probabilities are exactly zero for large enough n, and therefore the result follows from remark R.7. ■

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

$P\left(\sup_{\theta\in\Theta}\|(b_n(\theta) - b_n(\theta,\zeta_n(\theta_n(a),a)))\|_{W_n^*\left(\theta_n^+\right)} > \varepsilon\right) \leq$
$P\begin{pmatrix} \sup_{\theta\in\Theta}\|b_n(\theta) - b_n(\theta,\zeta_n(\theta_n(a),a))\|\left\|W_n^*\left(\theta_n^+\right) - W^*(\theta_0)\right\|^{1/2} \\ +\sup_{\theta\in\Theta}\|b_n(\theta) - b_n(\theta,\zeta_n(\theta_n(a),a))\|\|W^*(\theta_0)\|^{1/2} > \varepsilon \end{pmatrix} \leq$
$2P(\sup_{\theta\in\Theta}\|b_n(\theta) - b_n(\theta,\zeta_n(\theta_n(a),a))\| > \varepsilon_*) + P\left(\left\|W_n^*\left(\theta_n^+\right) - W^*(\theta_0)\right\| > K\right) =$
$o\left(n^{-a^*}\right)$ 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
$P\left(\sup_{\theta\in\Theta}\left|J\left(\beta_n, b_n(\theta,\zeta_n(\theta,a)), W_n^*\left(\theta_n^+\right)\right) - J\left(\beta_n, b_n(\theta,\zeta_n(\theta,a')), W_n^*\left(\theta_n^+\right)\right)\right| > \frac{\gamma_n}{2}\right) \leq$

$P\left(\sup_{\theta\in\Theta}\|(b_n(\theta,\zeta_n(\theta,a)) - b_n(\theta,\zeta_n(\theta,a')))\|_{W_n^*\left(\theta_n^+\right)} > \frac{\gamma_n}{2}\right) \leq$
$P\begin{pmatrix} \sup_{\theta\in\Theta}\|b_n(\theta,\zeta_n(\theta,a)) - b_n(\theta,\zeta_n(\theta,a'))\|\left\|W_n^*\left(\theta_n^+\right) - W^*(\theta_0)\right\|^{1/2} \\ +\sup_{\theta\in\Theta}\|b_n(\theta,\delta_n,a) - b_n(\theta,\zeta_n(\theta,a'))\|\|W^*(\theta_0)\|^{1/2} > \frac{\gamma_n}{2} \end{pmatrix} \leq$
$2P\left(\sup_\theta\|b_n(\theta,\delta_n,a) - b_n(\theta,\delta_n,a)\| > c_*\frac{\gamma_n}{2}\right) + P\left(\left\|W_n^*\left(\theta_n^+\right) - W^*(\theta_0)\right\| > K\right)$,
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 A.4, 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.7) we have that $\theta_n(a)$ and $\theta_n(a')$ are in $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ 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 $\sqrt{n}(\theta_n(0) - \theta_n(a)) = \left(D^2J^2\left(\beta_n, b_n\left(\theta_n^{++},\zeta_n\left(\theta_n^{++},a\right)\right), W_n^*\left(\theta_n^+\right)\right)\right)^{-1}\times$

$\sqrt{n}DJ^2\left(\beta_n, b_n(\theta_n(0),\zeta_n(\theta_n(0),a)), W_n^*\left(\theta_n^+\right)\right)$.
It suffices to prove that $Q\left(\sqrt{n}\|\theta_n(0) - \theta_n(a)\| > \gamma_n'\right) = o\left(n^{-a^*}\right)$, 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 $Q\left(\sqrt{n}\|\theta_n(0) - \theta_n(a)\| > \gamma_n'\right) \leq$

$Q\left(\begin{array}{c}\left\|\left(D^{2}J\left(\beta_{n},b_{n}\left(\theta_{n}^{++},\zeta_{n}\left(\theta_{n}^{++},a\right)\right),W_{n}^{*}\left(\theta_{n}^{+}\right)\right)\right)^{-1}\right\|\times\\ \left\|\sqrt{n}DJ\left(\beta_{n},b_{n}\left(\theta_{n}\left(0\right),\zeta_{n}\left(\theta_{n}\left(0\right),a\right)\right),W_{n}^{*}\left(\theta_{n}^{+}\right)\right)\right\|>\gamma_{n}^{\prime}\end{array}\right)$. Now, due to the definition of GMR1, the triangle inequality, norm submultiplicativity, assumptions A.1-A.4 and the subsequent R.2, 2.2 and R.7, and by choosing $\delta_1, \delta_2 > 0$ and positive constants K_b, M_i, M_i' for $i = 1, \ldots, 2a$ we have that $Q\left(\left\|\sqrt{n}DJ^{2}\left(\beta_{n},b_{n}\left(\theta_{n}\left(0\right),\delta_{n},a\right),W_{n}^{*}\left(\theta_{n}^{+}\right)\right)\right\|>2\rho_{n}\right)\leq$

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

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

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

$$+P\left(\left\|W_{n}^{*}\left(\theta_{n}^{*}\right)\right\|>M_{W}\right)\leq$$

$$Q\left(\begin{array}{c}\sup_{\theta\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)}\left\|\frac{\partial b'(\theta)}{\partial\theta}\right\|\times\\ \sup_{\theta\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)}\sum_{i=1}^{2a}\frac{1}{n^{i/2}}\left\|\zeta_{i+1_n}\left(\theta\right)\right\|\\ +\sup_{\theta\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)}\sum_{i=1}^{2a}\frac{1}{n^{i//2}}\left\|\frac{\partial\zeta_{i+1_n}'(\theta)}{\partial\theta}\right\|\times\\ \left(\left\|\beta_{n}-b\left(\theta_{0}\right)\right\|+\left\|b\left(\theta_{n}\left(0\right)\right)-b\left(\theta_{0}\right)\right\|\right)\\ +\sup_{\theta\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)}\sum_{i=1}^{2a}\frac{1}{n^{(i+1)/2}}\left\|\frac{\partial\zeta_{i+1_n}'(\theta)}{\partial\theta}\right\|\times\\ \sup_{\theta\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)}\sum_{i=1}^{2a}\frac{1}{n^{i/2}}\left\|\zeta_{i+1_n}\left(\theta\right)\right\|>\frac{\rho_{n}}{M_W}\end{array}\right)$$

$$+P\left(\theta_{n}\left(0\right)\in\overline{\mathcal{O}}_{\varepsilon}\left(\theta_{0}\right)\right)+P\left(\left\|W_{n}^{*}\left(\theta_{n}^{+}\right)\right\|>M_{W}\right)\leq$$

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

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

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

$+P\left(\theta_n(0)\in\overline{B}_{\varepsilon}(\theta_0)\right)+P\left(\left\|W_n^*\left(\theta_n^+\right)\right\|>M_W\right)$
$+\sum_{i=1}^{2a}Q\left(\left\|\zeta_{i+1_n}(\theta)\right\|>M_i\right)+\sum_{i=1}^{2a}Q\left(\left\|\frac{\partial\zeta_{i+1_n}'(\theta)}{\partial\theta}\right\|>M_i'\right)\leq$

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

$+P\left(\theta_n(0)\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)\right)+P\left(\left\|W_n^*\left(\theta_n^+\right)\right\|>M_W\right)$
$+\sum_{i=1}^{2a}Q\left(\left\|\zeta_{i+1_n}(\theta)\right\|>M_i\right)+\sum_{i=1}^{2a}Q\left(\left\|\frac{\partial\zeta_{i+1_n}'(\theta)}{\partial\theta}\right\|>M_i'\right)$
$+P\left(\theta_n(0)\in\overline{\mathcal{O}}_{\delta_2}(\theta_0)\right)+P\left(\beta_n\in\overline{\mathcal{O}}_{\delta_1}(b(\theta_0))\right)\leq$
$Q\left(K_b\sum_{i=1}^{2a}\frac{M_i}{n^{i/2}}+(\delta_1+L_b\delta_2)\sum_{i=1}^{2a}\frac{M_i'}{n^{i//2}}+\sum_{i=1}^{2a}\frac{M_i'}{n^{(i+1)/2}}\sum_{i=1}^{2a}\frac{M_i}{n^{i/2}}>\frac{\rho_n}{M_W}\right)$
$+P\left(\theta_n(0)\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)\right)+P\left(\left\|W_n^*\left(\theta_n^+\right)\right\|>M_W\right)$
$+\sum_{i=1}^{2a}Q\left(\left\|\zeta_{i+1_n}(\theta)\right\|>M_i\right)+\sum_{i=1}^{2a}Q\left(\left\|\frac{\partial\zeta_{i+1_n}'(\theta)}{\partial\theta}\right\|>M_i'\right)$
$+P\left(\theta_n(0)\in\overline{\mathcal{O}}_{\delta_2}(\theta_0)\right)+P\left(\beta_n\in\overline{\mathcal{O}}_{\delta_1}(b(\theta_0))\right)\leq o\left(n^{-a^*}\right)$ 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 $Q\left(\left\|\left(D^2J\left(\beta_n,b_n\left(\theta_n^{++},\zeta_n\left(\theta_n^{++},a\right)\right),W_n^*\left(\theta_n^+\right)\right)\right)^{-1}\right\|>C^*\right)=$ $o\left(n^{-a}\right)$ 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. First notice that due to lemma 4.7 the estimator lies in $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$ with probability $1-o\left(n^{-a^*}\right)$. Hence it satisfies first order conditions with the same probability due to A.1.2. A Taylor expansion of order d of the first order conditions implies that with probability $1-o\left(n^{-a^*}\right)$ $\sqrt{n}\left(\theta_n(0)-\theta_0\right)=\kappa_n\left(\sqrt{n}m_n^*(\theta_0)\right)+R_n$ where κ_n is a polynomial function that satisfies the conditions of lemma AL.3 due to assumptions A.2, A.4, A.5 and lemma 3.5 of Skovgaard (1981), while $P\left(\left\|R_n\right\|>\gamma_n\right)=o\left(n^{-a^*}\right)$ for some $\gamma_n=o\left(n^{-a^*}\right)$ due to the previous and lemma 4.7. Hence the result follows from the first part of lemma AL.3. ■

Proof of Lemma 4.9. It follows from lemmas 4.8, 7.1, AL.3 (second part), and the fact that Θ is compact. ■

Proof of Lemma 4.10. Lemma 4.9 is validated and therefore we essentially compute $K_j\left(\left(\psi_n(z)\right)\right)\left(1+\sum_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i(z)\right)$ for $m=1$ and $K_j(x)=x_j$, $j=1,\ldots,p$. Holding terms of the relevant order, we thus obtain

$\frac{\partial}{\partial\theta}b_n'\left(\theta_n\left(\frac{1}{2}\right),\zeta_n\left(\theta_n\left(\frac{1}{2}\right),a\right)\right)W_n^*\left(\theta_n^+\right)\sqrt{n}\left(\beta_n-b_n\left(\theta_n\left(\frac{1}{2}\right),\zeta_n\left(\theta_n\left(\frac{1}{2}\right),a\right)\right)\right)=$ $\mathbf{0}_p\Rightarrow$

$$\left(\frac{\partial b'(\theta_0)}{\partial\theta}+\frac{1}{\sqrt{n}}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n-\theta_0\right)\right]_{j=1,\ldots,q}\right)\times$$
$$\left(\begin{array}{c}W_n^*(\theta_0)\\+\frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\theta'}W_{rj_n}^*(\theta_0)\sqrt{n}\left(\theta_n^+-\theta_0\right)\right]_{r,j=1,\ldots,q}\end{array}\right)\times$$
$$\left(\begin{array}{c}\left(\sqrt{n}\left(\beta_n-b(\theta_0)\right)-\frac{1}{\sqrt{n}}\zeta_{2_n}(\theta_0)\right)\\-\left(\frac{\partial b(\theta_0)}{\partial\theta'}+\frac{1}{n}\frac{\partial\zeta_{2_n}(\theta_0)}{\partial\theta'}\right)\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\\-\frac{1}{2\sqrt{n}}\left[\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)^{/}\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]_{j=1,\ldots,q}\end{array}\right)=\mathbf{0}_p\Rightarrow$$

$$\left(\frac{\partial b'(\theta_0)}{\partial\theta}+\frac{1}{\sqrt{n}}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]_{j=1,\ldots,q}\right)$$
$$\times\left(\begin{array}{c}W^*(\theta_0)\\+\frac{1}{\sqrt{n}}w^*(z,\theta_0)+\frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\theta'}W_{rj}^*(\theta_0)q_1^*\right]_{r,j=1,\ldots,q}\end{array}\right)$$
$$\left(\begin{array}{c}\left(\sqrt{n}\left(\beta_n-b(\theta_0)\right)-\frac{1}{\sqrt{n}}Ek_2(\theta_0,z)\right)\\-\frac{\partial b(\theta_0)}{\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\\-\frac{1}{2\sqrt{n}}\left[\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)^{/}\frac{\partial^2 b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]_{j=1,\ldots,q}\end{array}\right)=\mathbf{0}_p\Rightarrow$$

$$\left(\begin{array}{c}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)+\frac{1}{\sqrt{n}}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]W^*(\theta_0)\\+\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}w^*(z,\theta_0)+\frac{1}{\sqrt{n}}\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta'}W_{rj}^*(\theta_0)q_1^*\right]_{r,j=1,\ldots,q}\end{array}\right)$$
$$\left(\begin{array}{c}\left(k_1(z,\theta_0,\delta_0)+\frac{1}{\sqrt{n}}\left(k_2(z,\theta_0,\delta_0)-Ek_2(z,\theta_0,\delta_0)\right)\right)\\-\left(\frac{\partial b(\theta_0)}{\partial\theta'}+\frac{1}{n}\frac{\partial Ek_2(\theta_0,z)}{\partial\theta'}\right)\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\\-\frac{1}{n}\frac{\partial Ek_2(z,\theta_0,\delta_0)}{\partial\delta'}\sqrt{n}\left(\delta_n-\delta_0\right)\\-\frac{1}{2\sqrt{n}}\left[\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)_j^{/}\frac{\partial^2 b(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right)\right]_{j=1,\ldots,q}\end{array}\right)=\mathbf{0}_p\Rightarrow$$

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

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

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

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

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

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

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

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

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

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

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

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

$$\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta_0\right) = \left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(\theta_0\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(\theta_0\right)k_1\left(z,\theta_0,\delta_0\right)$$

$$+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*\left(\theta_0\right)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n-\theta_0\right)\right]W^*\left(\theta_0\right)k_1\left(z,\theta_0,\delta_0\right)$$

$$
\begin{aligned}
&+\tfrac{1}{\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}w^*(z,\theta_0)\,k_1(z,\theta_0,\delta_0)\\
&+\tfrac{1}{\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}\left[\tfrac{\partial}{\partial\theta'}W^*_{rj}(\theta_0)\,q_1^*\right]_{r,j=1,\dots,q}k_1(z,\theta_0,\delta_0)\\
&+\tfrac{1}{\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\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)\\
&-\tfrac{1}{2\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\left[\sqrt{n}\left(\theta_n\left(\tfrac12\right)-\theta_0\right)_j^{/}\tfrac{\partial^2 b(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\tfrac12\right)-\theta_0\right)\right]_{j=1,\dots,q}\\
&-\tfrac{1}{\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\\
&\times\left(\begin{array}{c}\left[\tfrac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\tfrac12\right)-\theta_0\right)\right]W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\\ +\tfrac{\partial b'(\theta_0)}{\partial\theta}w^*(z,\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\\ +\tfrac{\partial b'(\theta_0)}{\partial\theta}\left[\tfrac{\partial}{\partial\theta'}W^*_{rj}(\theta_0)\,q_1^*(z,\theta_0,\delta_0)\right]_{r,j=1,\dots,q}\tfrac{\partial b(\theta_0)}{\partial\theta'}\end{array}\right)\\
&\times\qquad\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\,k_1\qquad\Rightarrow\\
&\sqrt{n}\left(\theta_n\left(\tfrac12\right)-\theta_0\right)=\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\,k_1\\
&+\quad\tfrac{1}{\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left[\tfrac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n-\theta_0\right)\right]W^*(\theta_0)\,k_1(z,\theta_0,\delta_0)\\
&+\tfrac{1}{\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}w^*(z,\theta_0)\,k_1(z,\theta_0,\delta_0)\\
&+\tfrac{1}{\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}\left[\tfrac{\partial}{\partial\theta'}W^*_{rj}(\theta_0)\,q_1^*\right]_{r,j=1,\dots,q}k_1(z,\theta_0,\delta_0)\\
&+\tfrac{1}{\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\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)\\
&-\tfrac{1}{2\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\left[\sqrt{n}\left(\theta_n-\theta_0\right)_j^{/}\tfrac{\partial^2 b(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n-\theta_0\right)\right]_{j=1,\dots,q}\\
&-\tfrac{1}{\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left[\tfrac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\tfrac12\right)-\theta_0\right)\right]\\
&\times\qquad W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\,k_1\\
&-\tfrac{1}{\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}w^*(z,\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\\
&\times\qquad\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\,k_1(z,\theta_0,\delta_0)\\
&-\tfrac{1}{\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}\left[\tfrac{\partial}{\partial\theta'}W^*_{rj}(\theta_0)\,q_1^*(z,\theta_0,\delta_0)\right]_{r,j=1,\dots,q}\\
&\times\qquad\tfrac{\partial b(\theta_0)}{\partial\theta'}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\,k_1(z,\theta_0,\delta_0)\qquad\Rightarrow\\
&\sqrt{n}\left(\theta_n\left(\tfrac12\right)-\theta_0\right)=\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\,k_1(z,\theta_0,\delta_0)\\
&+\tfrac{1}{\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\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)\\
&-\tfrac{1}{2\sqrt{n}}\left(\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\tfrac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\tfrac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)
\end{aligned}
$$

$$\times \qquad \left[\sqrt{n}\left(\theta_n\left(\tfrac{1}{2}\right)-\theta_0\right)^{/}_{j}\frac{\partial^2 b(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n-\theta_0\right)\right]_{j=1,\ldots,q}$$

$$+\frac{1}{\sqrt{n}}\left(\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\left[\frac{\partial b_j(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n\left(\tfrac{1}{2}\right)-\theta_0\right)\right]W^*(\theta_0)\right)$$

$$\times \qquad \left(Id_q-\frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\right)k_1(z,\theta_0,\delta_0)$$

$$+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}w^*(z,\theta_0)$$

$$\times \qquad \left(Id_q-\frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\right)k_1(z,\theta_0,\delta_0)$$

$$+\frac{1}{\sqrt{n}}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}\left[\frac{\partial}{\partial\theta'}W^*_{rj}(\theta_0)\,q^*_1(z,\theta_0,\delta_0)\right]_{r,j=1,\ldots,q}$$

$$\times \qquad \left(Id_q-\frac{\partial b(\theta_0)}{\partial\theta'}\left(\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\frac{\partial b(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial b'(\theta_0)}{\partial\theta}W^*(\theta_0)\right)k_1(z,\theta_0,\delta_0)$$

which is the required result. The result for the GMR estimator follows analogously. ■

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.27. 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

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

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 A.6, and therefore we obtain that $\sqrt{n}\left(\theta_n(a_2)-\theta_0\right)\underset{a_1}{\backsim}\sqrt{n}\left(\beta_n-\theta_0\right)$ 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 $\theta_n(a^*)$, by expanding analogously

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

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 A.6, and therefore by keeping terms up to $O\left(n^{-a_1-\frac{1}{2}}\right)$ we obtain

$$\sqrt{n}\left(\theta_n\left(a_2\right)-\theta_0\right) \underset{a_1+\frac{1}{2}}{\sim} \sqrt{n}\left(\beta_n-\theta_0\right)-\frac{1}{n^{a_1+\frac{1}{2}}}\zeta_{2a_1+2_n}\left(\theta_0\right)$$

$$\underset{a_1+\frac{1}{2}}{\sim} \sum_{i=0}^{2a_1}\frac{1}{n^{\frac{i}{2}}}\left(k_{i+1}\left(z,\theta_0\right)-Ek_{i+1}\left(z,\theta_0\right)\right)+\frac{1}{n^{a_1+\frac{1}{2}}}\left(k_{2a_1+2_n}\left(z,\theta_0\right)-Ek_{2a_1+2_n}\left(z,\theta_0\right)\right)$$

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. Let $\{\zeta_n\}$ denote a generic sequence of random vectors. In the following π_i denote polynomial real functions on $\mathbb{R}^q$ for i in some index set, with $O(1)$ coefficients. Finally φ_V denotes the standard density function of the q-dimensional Normal distribution with zero mean and covariance matrix V. V may also depend on n, hence we suppose that it converges to a positive definite matrix which we also denote with V.

Definition D.7 Suppose that A is a borel set. We say that A has the ε-neighborhood property with respect to a probability measure P if P attributes probability of order $O(\varepsilon)$ as $\varepsilon \downarrow 0$ to the ε-neighborhood of the boundary of A.

Definition AD.1 ζ_n admits an Edgeworth expansion of order s around the normal distribution with zero mean and covariance matrix V if there exist polynomial functions π_i for $i = 1, \ldots, 2a$ where $a = \frac{s-1}{2}$ such that

$$\sup_{A\in\mathcal{B}_C}\left|P\left(\zeta_n \in A\right)-\int_A\left(1+\sum\nolimits_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i\left(x\right)\right)\varphi_V\left(x\right)dx\right| = o\left(n^{-a}\right)$$

where $\mathcal{B}_C$ denotes any class of Borel sets with the ε-neighborhood property with respect to the Normal distribution with mean zero and covariance matrix V. The resulting signed measure whose density is $\left(1+\sum\nolimits_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i\left(x\right)\right)\varphi_V\left(x\right)$

is denoted as $\mathcal{EDG}_{n,a}$. In the case that the π_i are derived from the inversion of a Taylor approximation of the Fourier-Stieljes transform of P, then the Edgeworth measure is known as the formal one (see among others Magdalinos [[18]]) section 4, or Gotze and Hipp [[11]] page 2063, for the definition of the polynomials in this case).

Remark AR.1 When the Edgeworth expansion exists it is not unique. This is due to the fact if π_i^* is another polynomial such that $|\pi_i(x) - \pi_i^*(x)| = o\left(n^{-a+\frac{i}{2}}\right)$ for any x and some i then the relevant Edgeworth measure constructed by the new polynomials provides also a valid Egdeworth expansion of the same length to ζ_n. It is also trivial to establish that when an Edgeworth expansion of order s is valid then, the Edgeworth expansion of order s^* is also valid for any $s^* \leq s$, where the latter is the one with density $\left(1 + \sum_{i=1}^{s^*} \frac{1}{n^{\frac{i}{2}}}\pi_i(x)\right)\varphi_V(x)$ given the former.

Remark AR.2 When $\mathcal{B}_C$ denotes the largest class of sets with the ε-neighborhood property, then it can be easily seen that it also satisfies the provisions of theorem 12.5 of Stokey and Lucas [[23]], hence it comprises of a weak convergence determining class. The definition hence implies that the sequence of $\mathcal{EDG}_{n,a}$ comprises of an $o(n^{-a})$ approximation to the sequence of distributions of ζ_n with respect to the weak topology. This in turn implies that the definition uniformly holds for any class of Borel sets with the relevant property (see for example Rao [[21]] or Bhattacharya and Rao [[21]]. Hence the definition is equivalent to the aforementioned weak approximation. The same conclusion could also be drawn if $\mathcal{B}_C$ is chosen smaller than the previous. A typical example consists of the class of convex Borel sets which is used in the main body of the present paper.

Remark AR.3 It can be easily seen that given $\mathcal{EDG}_{n,a}$ then

$$\int_A x\left(1 + \sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i(x)\right)\varphi_V(x)\,dx = \sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}k_i$$

where $k_i = \int_A x\pi_i(x)\varphi_V(x)\,dx$ which always exists, could depend on n and converges to zero. This is essentialy used for the definition of the proposed estimators in the present paper.

Lemma AL.1 Suppose that ζ_n 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 $P(\|x_n\| > a_n) = o(n^{-a})$. Then any η_n, such that $P(\zeta_n + x_n = \eta_n) = 1 - o(n^{-a})$, admits a valid Edgeworth expansion of the same order.

Proof. We have that

$$\begin{aligned}\sup_{A\in\mathcal{B}_C}\left|P\left(\eta_n\in A\right)-P\left(\zeta_n+x_n\in A\right)\right| &\leq \sup_{A\in\mathcal{B}_C}\left|P\left(\eta_n\in A,\zeta_n+x_n=\eta_n\right)-P\left(\zeta_n+x_n\in A\right)\right|\\ &\quad +P\left(\zeta_n+x_n\neq\eta_n\right)\end{aligned}$$

the last term being $o\left(n^{-a}\right)$. Now by assumption

$$\sup_{A\in\mathcal{B}_C}\left|P\left(\zeta_n\in A\right)-\int_A\left(1+\sum_{i=1}^{2a+1}n^{-\frac{i}{2}}\pi_i\left(z\right)\right)\phi\left(z\right)dz\right|=o\left(n^{-a}\right)$$

where $\mathcal{B}_C$ denotes the collection of convex Borel sets of $\mathbb{R}^q$ and $\pi_i\left(z\right)=O\left(1\right)$.
Then,

$$P\left(\zeta_n+x_n\in A\right)\leq P\left(\zeta_n\in A-a_n\right)+o\left(n^{-a}\right)$$

uniformly over $\mathcal{B}_C$. Therefore

$$\begin{aligned}&\sup_{A\in\mathcal{B}_C}\left|P\left(\zeta_n+x_n\in A\right)-\int_{A-a_n}\left(1+\sum_{i=1}^{2a+1}n^{-\frac{i}{2}}\pi_i\left(y\right)\right)\phi\left(y\right)dy\right|\\ \leq\ &\sup_{A\in\mathcal{B}_C}\left|P\left(\zeta_n+x_n\in A-a_n\right)-\int_{A-a_n}\left(1+\sum_{i=1}^{2a+1}n^{-\frac{i}{2}}\pi_i\left(y\right)\right)\phi\left(y\right)dy\right|+o\left(n^{-a}\right)=o\left(n^{-a}\right)\end{aligned}$$

as $A-a_n$ is convex. Now,

$$\int_{A-a_n}\left(1+\sum_{i=1}^{2a+1}n^{-\frac{i}{2}}\pi_i\left(y\right)\right)\phi\left(y\right)dy=\int_A\left(1+\sum_{i=1}^{2a+1}n^{-\frac{i}{2}}\pi_i\left(z-a_n\right)\right)\phi\left(z-a_n\right)dz$$

Hence, if $H_k\left(z\right)$ denotes the k^{th} order Hermite multivariate polynomial, $L\left(H_k\left(z\right),a_n,i\right)$ and i-linear function of a_n with coefficients from $H_k\left(z\right)$, and

$$\phi\left(z-a_n\right)=\phi\left(z\right)\sum_{k=0}^{K}\frac{1}{k!}L\left(H_k\left(z\right),a_n,k\right)+\rho_n\left(z\right)$$

where

$$\rho_n\left(z\right)=\frac{1}{\left(2K+1\right)!}\left(-1\right)^{K+1}L\left(H_k\left(z-a_n^*\right),a_n,K+1\right)\phi\left(z-a_n\right)$$

and a_n^* lies between a_n and zero. If $a\leq\varepsilon$ set $K=0$, else, choose some natural $K\geq\frac{a}{\varepsilon}-1$.
Then,

$$\left(1+\sum_{i=1}^{2a+1}n^{-\frac{i}{2}}\pi_i\left(z-a_n\right)\right)\phi\left(z-a_n\right)=\phi\left(z\right)\left(1+\sum_{i=1}^{2a+1}n^{-\frac{i}{2}}\pi_i^*\left(z\right)\right)+q_n\left(z\right)$$

where the $\pi_i^*(z)' s$ are $O(1)$ polynomials in z and $q_n(z) = \left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}} \pi_i(z - a_n)\right) \rho_n(z)$. Hence

$$\begin{aligned}
& \int_A \left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}} \pi_i(z - a_n)\right) \phi(z - a_n)\, dz \\
= & \int_A \phi(z) \left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}} \pi_i^*(z)\right) dz + \int_A q_n(z)\, dz
\end{aligned}$$

and

$$\begin{aligned}
\sup_{A \in \mathcal{B}_C} \left| \int_A q_n(z)\, dz \right| & \leq \sup_{A \in \mathcal{B}_C} \int_A \left| \left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}} \pi_i(z - a_n)\right) \rho_n(z) \right| dz \\
& \leq \int_{\mathbb{R}^q} \left| \left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}} \pi_i(z - a_n)\right) \rho_n(z) \right| dz \leq \frac{C}{n^{a+\delta}} = o\left(n^{-a}\right)
\end{aligned}$$

for some $C, \delta > 0$. Hence, since $\sup_{A \in \mathcal{B}_C} \left| R_n - \int_A q_n(z)\, dz \right| = o(n^{-a})$, and therefore

$$\begin{aligned}
\sup_{A \in \mathcal{B}_C} \left| R_n - \int_A q_n(z)\, dz \right| & \geq \sup_{A \in \mathcal{B}_C} \left| R_n - \left| \int_A \phi(z)\, q_n(z)\, dz \right| \right| \\
& \geq \left| \sup_{A \in \mathcal{B}_C} |R_n| - \sup_{A \in \mathcal{B}_C} \left| \int_A \phi(z)\, q_n(z)\, dz \right| \right| = o\left(n^{-a}\right)
\end{aligned}$$

and $\sup_{A \in \mathcal{B}_C} \left| P(\zeta_n + x_n \in A) - \int_A \phi(z) \left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}} \pi_i^*(z)\right) dz \right| = o(n^{-a})$ due to the fact that the transformation from $\pi_i(z)$ to $\pi_i^*(z)$ does not depend on A but only on a_n with $R_n = P(\zeta_n + x_n \in A) - \int_A \phi(z) \left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}} \pi_i^*(z)\right) dz$. ■

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.

Now, denote by P_n the measure $P \circ \zeta_n^{-1}$. 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 7.1 Suppose that K is a m-linear real function on $\mathbb{R}^p$, if the support of ζ_n is bounded by $\mathcal{O}_{\sqrt{n}\rho}(0)$ for some $\rho > 0$ and ζ_n admits an Edgeworth expansion of order $2a + m + 1$ then

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

where Q_n denotes the analogous Edgeworth measure of order $2a + 1$ and $x^m = \underbrace{(x, x, \ldots x)}_{m}$.

Proof. Since $2a+m+1 > 2a+1$, we have that $\sup_{A\in\mathcal{B}_C} |P_n(A) - Q_n(A)| = O(n^{-a-\eta})$, where $\eta > 0$. Hence

$$\begin{aligned}
& n^a \left| \int_{\mathbb{R}^q} K(x^m)(dP_n - dQ_n) \right| \leq n^a \left| \int_{\mathcal{O}_{c(\ln n)^\epsilon}(0)} K(x^m)(dP_n - dQ_n) \right| \\
& + n^a \left| \int_{\mathbb{R}^q \setminus \mathcal{O}_{c(\ln n)^\epsilon}(0)} K(x^m)\, dP_n \right| + n^a \left| \int_{\mathbb{R}^q \setminus \mathcal{O}_{c(\ln n)^\epsilon}(0)} K(x^m)\, dQ_n \right| \\
\leq\; & n^a c^m (\ln n)^{m\epsilon} \int_{\mathcal{O}_{c(\ln n)^\epsilon}(0)} |dP_n - dQ_n| + n^a \int_{\mathbb{R}^q \setminus \mathcal{O}_{c(\ln n)^\epsilon}(0)} |K(x^m)| (dP_n + |dQ_n|) \\
\leq\; & c^m (\ln n)^{m\epsilon} \sup_{A\in\mathcal{B}_C} n^a |P_n(A) - Q_n(A)| + n^a \int_{\mathbb{R}^q \setminus \mathcal{O}_{c(\ln n)^\epsilon}(0)} |K(x^m)| (dP_n + |dQ_n|)
\end{aligned}$$

Due to the hypothesis for the support of P_n

$$\begin{aligned}
& n^a \int_{\mathbb{R}^q \setminus \mathcal{O}_{c(\ln n)^\epsilon}(0)} |K(x^m)|\, dP_n \\
=\; & n^a \int_{\left[\mathbb{R}^q \setminus \mathcal{O}_{c(\ln n)^\epsilon}(0)\right] \cap \mathcal{O}_{\sqrt{n}\rho}(0)} |K(x^m)|\, dP_n + n^a \int_{\left[\mathbb{R}^q \setminus \mathcal{O}_{c(\ln n)^\epsilon}(0)\right] \cap \left(\mathcal{O}_{\sqrt{n}\rho}(0)\right)^c} |K(x^m)|\, dP_n \\
=\; & n^a \int_{\left[\mathbb{R}^q \setminus \mathcal{O}_{c(\ln n)^\epsilon}(0)\right] \cap \mathcal{O}_{\sqrt{n}\rho}(0)} |K(x^m)|\, dP_n = n^a \int_{\mathcal{O}_{\sqrt{n}\rho}(0) \setminus \mathcal{O}_{c(\ln n)^\epsilon}(0)} |K(x^m)|\, dP_n \\
\leq\; & n^{a+m\beta} \rho^m q^m \int_{\mathbb{R}^q} 1_{\|x\| > c(\ln n)^\epsilon}\, dP_n
\end{aligned}$$

Hence

$$\begin{aligned}
& n^a \left| \int_{\mathbb{R}^q} x^m (dP_n - dQ_n) \right| \\
\leq\; & c^{\overline{m}} (\ln n)^{m\epsilon} \sup_{A\in\mathcal{B}_C} n^a |P_n(A) - Q_n(A)| \\
& + n^{a+m\beta} \rho^m q^m P\left(\|\zeta_n\| > c(\ln n)^\epsilon\right) \\
& + n^a \int_{\mathbb{R}^q \setminus \mathcal{O}_{c(\ln n)^\epsilon}(0)} |K(x^m)|\, |dQ_n| .
\end{aligned}$$

As $\sup_{A\in\mathcal{B}_C} n^a \left|P_n(A) - Q_n(A)\right| = O\left(n^{-\eta}\right)$ for $\eta > 0$, we have that

$$(\ln n)^{2\epsilon} \sup_{A\in\mathcal{B}_C} n^a \left|P_n(A) - Q_n(A)\right| = o(1)$$

and $n^{a+\frac{m}{2}}\rho^m q^m P\left(\|\zeta_n\| > c(\ln n)^{\epsilon}\right) = o(1)$ if $\epsilon \geq \frac{1}{2}$ and $c \geq \sqrt{2a+\overline{m}+1}$ by lemma 2 of Magdalinos [[18]]. Finally $n^a \int_{\mathbb{R}^q\setminus\mathcal{O}_{c(\ln n)^{\epsilon}}(0)} |K(x^m)|\,|dQ_n| = o(1)$ due to Gradshteyn and Ryzhik [[15]] formula 8.357. ■

We extend the previous result in the case where ζ_n has unbounded support in exchage for the boundeness of the relevant moments of this random element.

In the next lemma let $\Pr_p : \mathbb{R}^q \to \mathbb{R}^p$ $(p < q)$ denote projection to the first p coordinates.

Lemma AL.2 Suppose that ζ_n admits a valid Edgeworth expansion of order s. Then $\Pr_p(\zeta_n)$ admits an analogous expansion of the same order.

Proof. Let $\mathcal{B}_C$ denote the class of convex Borel sets on $\mathbb{R}^p$. Then for $A \in \mathcal{B}_C$ we have that since $\Pr_p^{-1}(A)$ has the ε-neighborhood property with respect to the relevant normal distribution, hence

$$\begin{aligned} P\left(\Pr_p(\zeta_n) \in A\right) &= P\left(\zeta_n \in \Pr_p^{-1}(A)\right) \\ &= \int_{A\times\mathbb{R}^{q-p}} \left(1 + \sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i(x)\right)\varphi_V(x)\,dx + o\left(n^{-a}\right) \\ &= \int_A \left(1 + \sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i^*(v)\right)\varphi_{IVI'}(v)\,dv + o\left(n^{-a}\right) \end{aligned}$$

where $v = \Pr_p(x)$, $x = (v, v^*)$, $I = (\mathrm{Id}_{p\times p}, 0_{p\times q-p})$, and $\pi_i^*(v) = \int_{\mathbb{R}^{q-p}} \pi_i(v, v^*)\,\varphi_V(v, v^*)\,dv^*$. The $o(n^{-a})$ is independent of A hence the result holds uniformly over $\mathcal{B}_C$ and therefore uniformly over any class of Borel sets with the ε-neighborhood property with respect to the normal distribution with zero mean and covariance matrix IVI' due to remark AR.2. ■

Lemma AL.3 Suppose that ζ_n admits a valid Edgeworth expansion of order s. Let also $H_n(C) = \left\{x \in \mathbb{R}^q : \|x\| < C\ln^{1/2} n\right\}$ for $C > 4a+2$ and $g_n : \mathbb{R}^q \to \mathbb{R}^p$ $(p \leq q)$ be measurable and continuously differentiable of order $s+1$ on H_n, for large enough n, with $\lim_{n\to\infty} \operatorname{rank} Dg_n(0) = p$, $\|D^i g_n(0)\|^{\frac{s}{i-1}} = o(n^{-a})$ for $1 < i < s+1$ and $\|D^m g_n(x)\| = o(n^{-a})$ uniformly in H_n. Then $g_n(\zeta_n)$ admits an analogous expansion of the same order, i.e. there exist polynomials $\pi_i^* : \mathbb{R}^p \to \mathbb{R}$, $i = 1, \ldots, s$ such that

$$\sup_{A\in\mathcal{B}_C} \left|P\left(g_n(\zeta_n) \in A\right) - \int_A \left(1 + \sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i^*(x)\right)\varphi_{LVL'}(x)\,dx\right| = o\left(n^{-a}\right)$$

Furthermore, if K is a m-linear real function on $\mathbb{R}^p$

$$\begin{aligned}&\int_{\mathbb{R}^p} K\left(x^m\right)\left(1+\sum_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i^*\left(x\right)\right)\varphi_{LVL'}\left(x\right)dx\\ =\;&\int_{\mathbb{R}^q} K\left(\left(\psi_n\left(z\right)\right)^m\right)\left(1+\sum_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i\left(z\right)\right)\varphi_V\left(z\right)dz+o\left(n^{-a}\right)\end{aligned}$$

where $\mathcal{B}_C$ denotes the class of convex Borel sets on $\mathbb{R}^p$, $L=\lim_{n\to\infty}Dg_n\left(0\right)$, $\psi_n\left(z\right)=\mathrm{Pr}_p\left(\frac{1}{i!}\sum_{i=0}^{s}D^i g_n^*\left(0\right)\left(z^j\right)\right)$, $g_n^*=\left(g_n,\varpi_n\right)$, $\varpi_n:\mathbb{R}^q\to\mathbb{R}^{q-p}$ is an orthogonal projection of $\mathbb{R}^q$ to a $q-p$ dimensional subspace composed with a linear isometry with $\mathbb{R}^{q-p}$, such $\left(Dg_n\left(0\right)\right)^{-1}\cap\left(\varpi_n\left(0\right)\right)^{-1}=\{0\}$, g_n is restricted to H_n and $x^m=\underbrace{\left(x,x,\ldots x\right)}_{m}$.

Proof. For the first part assume $p=q$ without loss of generality. For if $p<q$ then consider g_n^* satisfies the assumptions of the lemma if g_n does. Then the result would follow from lemma AL.2. Due to the assumption for ζ_n the choice of $C>4a+2$ stems from lemma 2 of Magdalinos [[18]]. Notice that when g_n is restricted to $H_n\left(C\right)$ then due to a slight modification of lemmas 3.5 and 4.4 of Skovgaard [[22]], which consists in replacing $1_{\mathbb{R}^k}$ with L in equations (4.4) and (4.5) and in the proof of 3.5, it has an inverse g_n^{-1} satisfying the same assumptions, defined on $H_n\left(C^*\right)$ for some $C^*<C<4a+2$. Then ψ_n is essentially the s-th order McLaurin polynomial of g_n. Then for $A\in\mathcal{B}_C$ we have that $g_n^{-1}\left(A\right)$ and $g_n^{-1}\left(A\right)\cap H_n\left(C\right)$ possess the ε-neighborhood property with respect to the normal distribution with mean zero and covariance matrix LVL'. Hence

$$\begin{aligned}P\left(g_n\left(\zeta_n\right)\in A\right)&=P\left(\zeta_n\in g_n^{-1}\left(A\right)\right)\\&=\int_{g_n^{-1}(A)}\left(1+\sum_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i\left(x\right)\right)\varphi_V\left(x\right)dx+o\left(n^{-a}\right)\\&=\int_{g_n^{-1}(A)\cap H_n(C)}\left(1+\sum_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i\left(x\right)\right)\varphi_V\left(x\right)dx\\&\quad+\int_{g_n^{-1}(A)\cap H_n(C)}\left(1+\sum_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i\left(x\right)\right)\varphi_V\left(x\right)dx+o\left(n^{-a}\right)\\&=\int_{g_n^{-1}(A)\cap H_n(C)}\left(1+\sum_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i\left(x\right)\right)\varphi_V\left(x\right)dx+o\left(n^{-a}\right)\end{aligned}$$

due to lemma 2 of Magdalinos [[18]]. Now since g_n is inverible on $H_n\left(C\right)$ the last integral equals

$$\int_{A\cap g_n(H_n(C))}\left(1+\sum_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i\left(g_n^{-1}\left(z\right)\right)\right)\varphi_V\left(g_n^{-1}\left(z\right)\right)\det\left(Dg_n^{-1}\left(z\right)\right)dz$$

and due to the proof of lemma 3.5 of Skovgaard [[22]] $H_n(C_*) \subseteq g_n(H_n(C))$ hence this equals

$$\int_{A\cap H_n(C_*)} \left(1+\sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i\left(g_n^{-1}(z)\right)\right)\varphi_V\left(g_n^{-1}(z)\right)\det\left(Dg_n^{-1}(z)\right)dz$$
$$+\int_{A\cap(g_n(H_n(C))/H_n(C_*))} \left(1+\sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i\left(g_n^{-1}(z)\right)\right)\varphi_V\left(Dg_n^{-1}(z)\right)\det\left(g_n^{-1}(z)\right)dz$$

the latter is bounded from

$$\int_{H_n^c(C_*)} \left(1+\sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i\left(g_n^{-1}(z)\right)\right)\varphi_V\left(g_n^{-1}(z)\right)\det\left(Dg_n^{-1}(z)\right)dz$$

which is $o(n^{-a})$.Then the needed polynomials are obtained from

$$\int_{A\cap H_n(C_*)} \left(1+\sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i\left(g_n^{-1}(z)\right)\right)\varphi_V\left(g_n^{-1}(z)\right)\det\left(Dg_n^{-1}(z)\right)dz$$

as in the proof of the first part of lemma 4.6 of Skovgaard [[22]] using repeated Taylor expansions and the fact that $\det\left(Dg_n^{-1}(z)\right) = \det^{-1}(L)+o(1)$ uniformly on $H_n(C_*)$, holding terms of the relevant order and esimate the remainders as $o(n^{-a})$ terms. The previous holds uniformly on $\mathcal{B}_C$ due to the fact that none of the employed consrucions depends on A. For the second part, when $p=q$ the result follows directly from the second part of lemma 4.6 of Skovgaard [[22]]. When $p<q$ use the construction described in the begining of the proof so that

$$\begin{aligned}
&\int_{\mathbb{R}^q} K\left((\psi_n(z))^m\right)\left(1+\sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i(z)\right)\varphi_V(z)\,dz \\
=\;& \int_{H_n(C_*)} K\left((\psi_n(z))^m\right)\left(1+\sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i(z)\right)\varphi_V(z)\,dz \\
&+\int_{H_n^c(C_*)} K\left((\psi_n(z))^m\right)\left(1+\sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i(z)\right)\varphi_V(z)\,dz
\end{aligned}$$

due to the previous. The last integral is $o(n^{-a})$ due to equation (A.8) in the proof of lemma 2 of Magdalinos [[18]]. Hence the previous equals

$$\begin{aligned}
&\int_{H_n(C)} K\left(\left(\mathrm{Pr}_p\left(\frac{1}{i!}\sum_{i=0}^{s} D^i g_n^*(0)\left(z^j\right)\right)\right)^m\right)\left(1+\sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i(z)\right)\varphi_V(z)\,dz+o\left(n^{-a}\right) \\
=\;& \int_{g_n^*(H_n(C))} K\left(\left(\mathrm{Pr}_p\left(\frac{1}{i!}\sum_{i=0}^{s} D^i g_n^*(0)\left(g_n^{*-1}(z)\right)^j\right)\right)^m\right)\left(1+\sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i\left(g_n(z)\right)\right)\varphi_V\left(g_n\left(\right.\right. \\
=\;& \int_{g_n^*(H_n(C))} K\left(x^m\right)\left(1+\sum_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i^*(x)\right)\varphi_{LVL'}(x)\,dx+o\left(n^{-a}\right)
\end{aligned}$$

$$= \int_{\mathbb{R}^p} K(x^m) \left(1 + \sum\nolimits_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}} \pi_i^* (x)\right) \varphi_{LVL'}(x)\, dx$$
$$- \int_{\mathbb{R}^p / g_n^* (H_n(C))} K(x^m) \left(1 + \sum\nolimits_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}} \pi_i^* (x)\right) \varphi_{LVL'}(x)\, dx + o\left(n^{-a}\right)$$

and

$$\int_{\mathbb{R}^p / g_n^* (H_n(C))} \|K(x^m)\| \left(1 + \sum\nolimits_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}} \pi_i^* (x)\right) \varphi_{LVL'}(x)\, dx = o\left(n^{-a}\right)$$

since $g_n^*(x) = Lx + o(1)$ uniformly in $H_n(C)$, the fact that ζ_n admits a valid Edgeworth expansion of order $s+1$ and due to equation (A.8) in the proof of lemma 2 of Magdalinos [[18]]. ■

Remark R.29 Notice that the previous result can be easily established in the case where g_n is stochastic and it satisfies the needed conditions with probability $1 - o(n^{-a})$.

Further Comments on the Assumption Framework

In this section we provide a brief commentary on our assumption framework. This is in the form of further remarks with respect to several groupings of the assumptions that appear in the main text.

Remark R.30 (Assumptions A.3, A.4 A.5) The first part of assumption A.3 can be established for example when $W_n^*(\theta) = \frac{1}{n}\sum W(x_j(\omega), \theta)$ with x_j and W appropriate random elements and function respectively. If for any $\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$ the sequence $(W(x_n(\omega), \theta))$ is stationary and strong mixing (see Assumption 3 of Gotze and Hipp [[11]]), and

$$E_{\theta_0} \|W(x_1(\omega), \theta) - E_{\theta_0} W(x_1(\omega), \theta)\|^p < +\infty$$

for $p > 2a^*$ then the first part of the assumption follows from the Yokoyama moment inequality (see Andrews [[2]], proof of lemma 3) with $W^*(\theta) = E_{\theta_0} W(x_1(\omega), \theta)$. In an analogous manner the existence of the T_i functions can be established. Hence if

$$E_{\theta_0} \left\|D^i W(x_1(\omega), \theta) - E_{\theta_0} D^i W(x_1(\omega), \theta)\right\|^p < +\infty$$

then the T_i can be identified with $\operatorname{vec} E_{\theta_0} D^i W(x_1(\omega), \theta)$ for all $i = 1, \ldots, d$. The second part of the assumption could be obtained from an asymptotic equi-Lipschitz condition with probability $1 - o(n^{-a})$ for $D^d W_n^*(\theta)$ and the compactness of $\overline{\mathcal{O}}_\varepsilon(\theta_0)$. In the case that $W_n^*(\theta) = \frac{1}{n}\sum W(x_j(\omega), \theta)$ then a

Lipschitz coefficient of $D^dW_n^*(\theta) - E_{\theta_0}D^dW(x_1(\omega), \theta)$ would be $\frac{1}{n}\sum(l(x_j(\omega)) + El(x_j(\omega)))$ where l_j is the analogous coefficient for $D^dW(x_j(\omega), \theta)$ and if the sequence $(l(x_j(\omega)))$ is stationary and strong mixing, and

$$E_{\theta_0}\|l(x_1(\omega)) - E_{\theta_0}l(x_1(\omega))\|^p < +\infty$$

then the equi-Lipschitz property would follow from lemma 3.b of Andrews [[2]]. Assumption A.4.2,3 would, analogously to the previous, follow from equi-Lipschitz conditions with probability $1 - o(n^{-a})$ on $\zeta_{i+1_n}(\omega, \omega', \theta)$ (on Θ), and $D^2\zeta_{i+1_n}(\omega, \omega', \theta)$ (on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$). In the case that $\zeta_{i+1_n}(\omega, \omega', \theta) = \frac{1}{n}\sum\xi_{i+1}(y_j(\omega, \omega', \theta), \theta)$ for ξ_{i+1} appropriate functions and $y_j(\omega, \omega', \theta)$ appropriate random elements then along the same lines of the previous discussion the needed conditions would follow from strong mixing, stationarity and moment conditions on the relevant Lipschitz coefficients. In the same respect the functions $Z_i(\theta)$, used in assumption A.5, can be identified with $\mathrm{vec}\, E_Q D^i\xi_{i+1}(y_1(\omega, \omega', \theta_0), \theta)$.

Remark R.31 (Assumptions A.2, A.5, A.6) This particular set of assumptions requires the validity of Edgeworth approximations of the distributions of random elements smooth transformations of which are the IE at hand. Obviously this can be established with the help of the general results that appear in the previous section along with more primitive assumptions that guarantee analogous expansions for similar random elements. Assume the existence of a sequence of random elements $S_n(\omega, \omega', \theta)$ with values in $\mathbb{R}^l$ ($l \geq \dim(m_n^*)$) of the form $S_n(\omega, \omega', \theta) = n^{-1/2}(X_n(\omega, \omega', \theta) - E_\theta X_n(\omega, \omega', \theta))$, Assumptions 2-4 of Durbin [[7]] hold. These assumptions essentially concern the rate of uniform integrability of the characteristic function of X_n, the asymptotic behavior of its derivatives and of the cumulants of X_n of order $2a + 2$ uniformly in a neighborhood of θ_0. These assumptions guarantee the validity of the formal Edgeworth expansion *uniformly* in the aforementioned neighborhood. These do not require independence between the random variables comprising X_n, nor that the latter is in a form of a sum. They were used for example by Andrews and Lieberman [[3]] for the validation of the formal Edgeworth expansions of S_n when X_n is comprised by the elements of the derivatives of the aforementioned order of the likelihood function or the Whittle likelihood function and the consequent validation of the Edgeworth expansion of the analogous MLE and WMLE respectively uniformly on their parameter space for long memory Gaussian processes. In a similar fashion $S_n(\omega, \omega', \theta)$ could be of the form $n^{-1/2}\left(\sum_{i=1}^n(X_i(\omega, \omega', \theta) - EX_i(\omega, \omega', \theta))\right)$ where $X_n(\omega, \omega', \theta) = g(\varepsilon_{n-i}(\omega, \omega'), \theta; i \in \mathbb{N})$ and the ε_n comprise an i.i.d. process. Using the results of Gotze and Hipp [[11]] (Lemma 2.3, Assumptions 2-4) if for any θ, g satisfies some Lipschitz conditions, has almost everywhere continuous derivativw w.r.t. to ε, these are appropriately non-degenerate in a set of positive probability, $E\left\|X_1(\omega,\omega',\theta)\right\|^{2a+3}<\infty$, and $g(\varepsilon_{n-i}(\omega,\omega'),\theta;i\in\mathbb{N})$ satifies a weak dependence condition, then $S_n(\omega,\omega',\theta)$ again admits a valid formal Edgeworth expansion of order $2a+2$ for any θ. Given the previous if $\sqrt{n}m_n^*(\theta)=\pi_n(S_n(\omega,\omega',\theta))+R_n(\theta)$ with P_θ-probability $1-o(n^{-a})$, where for any θ, π_n satisfies the provisions of lemma AL.3 and $P_\theta(\|R_n(\theta)\|>\gamma_n(\theta))=o(n^{-a})$ for $\gamma_n(\theta)=o\left(n^{-\epsilon(\theta)}\right)$ with $\epsilon(\theta)>0$, then due to lemmas AL.3 and AL.1 the discussed assumptions hold (in fact assumption A.5 holds for any θ). In the case that $\epsilon(\theta)=a$ then the relevant expansion can be identified as the formal one. Notice finally, that due to the fact that $m_n^*(\theta)$ contains the elements of $\beta_n-b(\theta)$ and $\theta_n^+-\theta$, which can be assumed to satisfy smooth first order conditions with P_θ-probability $1-o(n^{-a})$ which can be reduced with P_θ-probability $1-o(n^{-a})$ to be formed by elements of $S_n(\omega,\omega',\theta)$, the elements of π_n and R_n corresponding to the elements of $\beta_n-b(\theta)$ and $\theta_n^+-\theta$, say π_n^* and R_n^*, can be obtained by recursive applications of the implicit function theorem. The conditions imposed by lemma AL.3 on π_n^* would then be justified by conditions-imposed on the function that defines these first order conditions-enabling that the jacobian at θ is asymptotically invertible and the higher order derivatives satisfy conditions similar to the ones imposed by this lemma . Analogous Lipschitz continuity conditions on the highest order derivative, would ensure that $\|R_n^*(\theta)\|$ is bounded with P_θ-probability $1-o(n^{-a})$ by $\|z_n(\theta)\|\left\|\frac{S_n(\omega,\omega',\theta)}{\sqrt{n}}\right\|^{2a+2}$ with $P_\theta(\|z_n(\theta)\|>M(\theta))=o(n^{-a})$ for $M(\theta)>0$ (see for example the proof of lemma 8 of Andrews [[2]]).

arg min **Properties**

In the following, let Θ be a compact metric space, and $(\Omega,\mathcal{F},P)$ a complete probability space. Let $(\mathcal{K}(\Theta),\mathcal{H})$ 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.4 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 [[19]] $\arg\min_\theta\circ J$ is non empty, measurable and almost surely compact valued. By theorem 2.13 of Molchanov [[19]], $\arg\min_\theta\circ J$ has a measurable selection.

C Examples' Expansions

MA(1) calculations

Given the expansion results in [6], and employing the notation of lemma 4.2 in Arvanitis and Demos [[4]] 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

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

and

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

where

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

Now taking into account that

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

the above formulae can be simplified to

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

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

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.

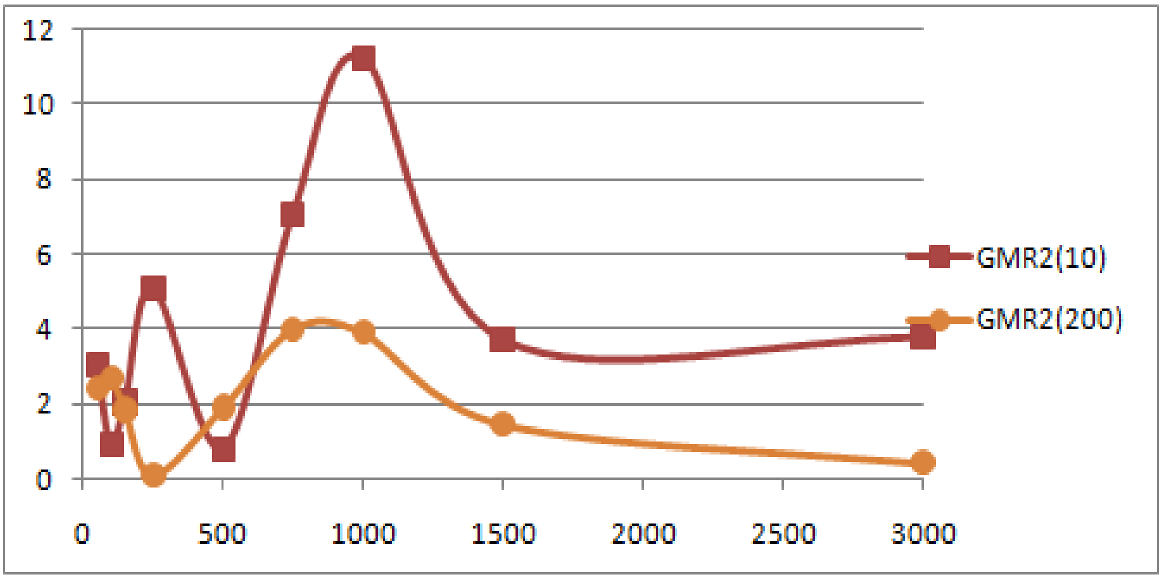

Figure 1: $n\,|E\,(GMR2(i) - \theta)|$, $i = 10, 200$, $MA(1)$ model, $\theta = -0.4$

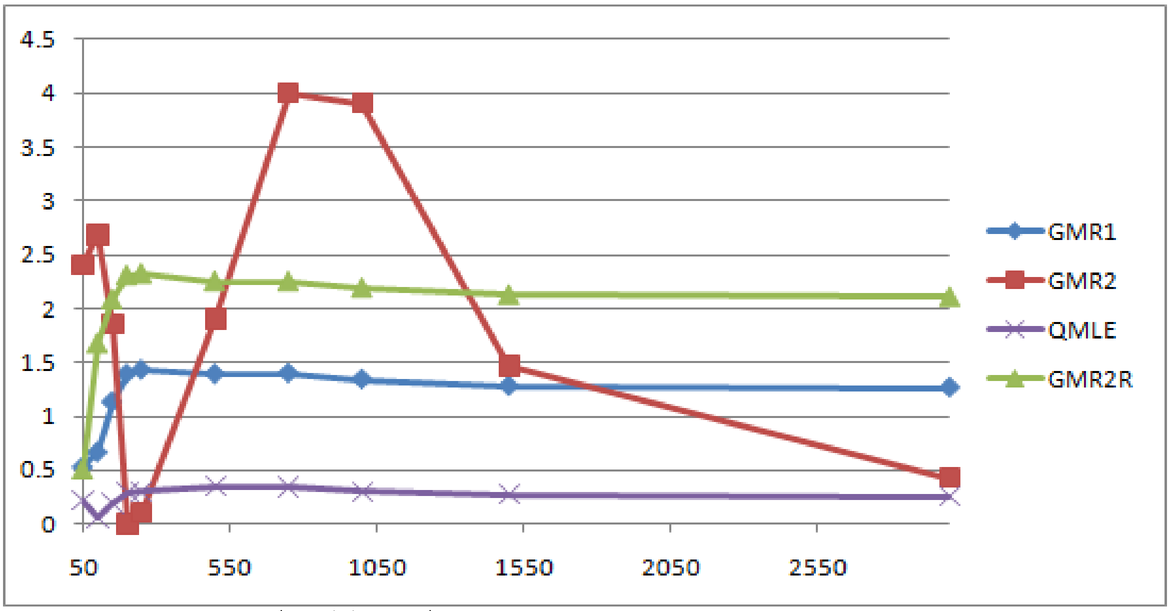

Figure 2: $n\left|E\left(\widehat{\theta}\right)-\theta\right|$ Biased Estimators, $MA(1)$ model, $\theta=-0.4$.

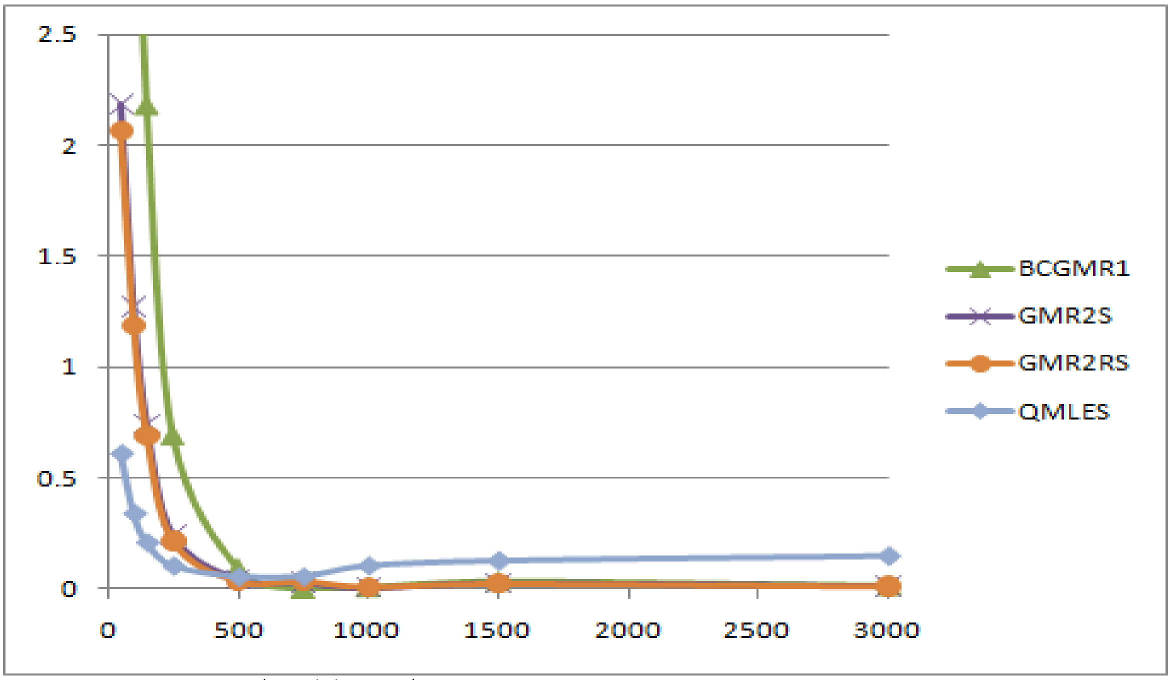

Figure 3: $n\left|E\left(\widehat{\theta}\right)-\theta\right|$ Unbiased Estimators, $MA(1)$ model, $\theta=-0.4$.

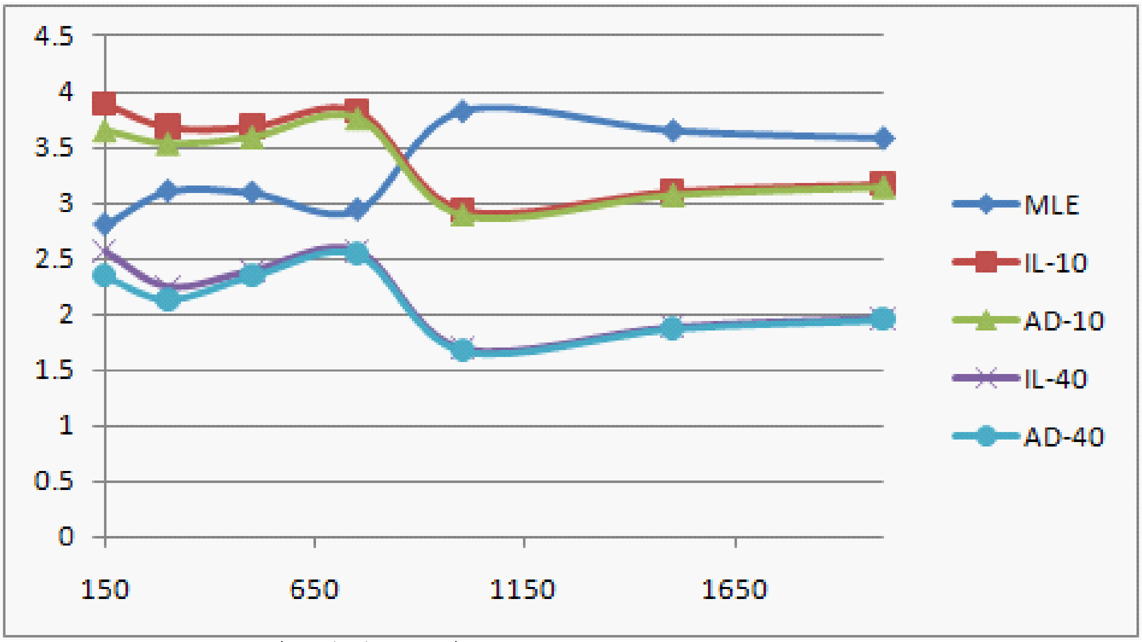

Figure 4: $n\left|E\left(\widehat{\theta_1}\right)-\theta_1\right|$ $ARCH(1)$ model, $\theta_1=1.0$ and $\theta_2=0.5$.

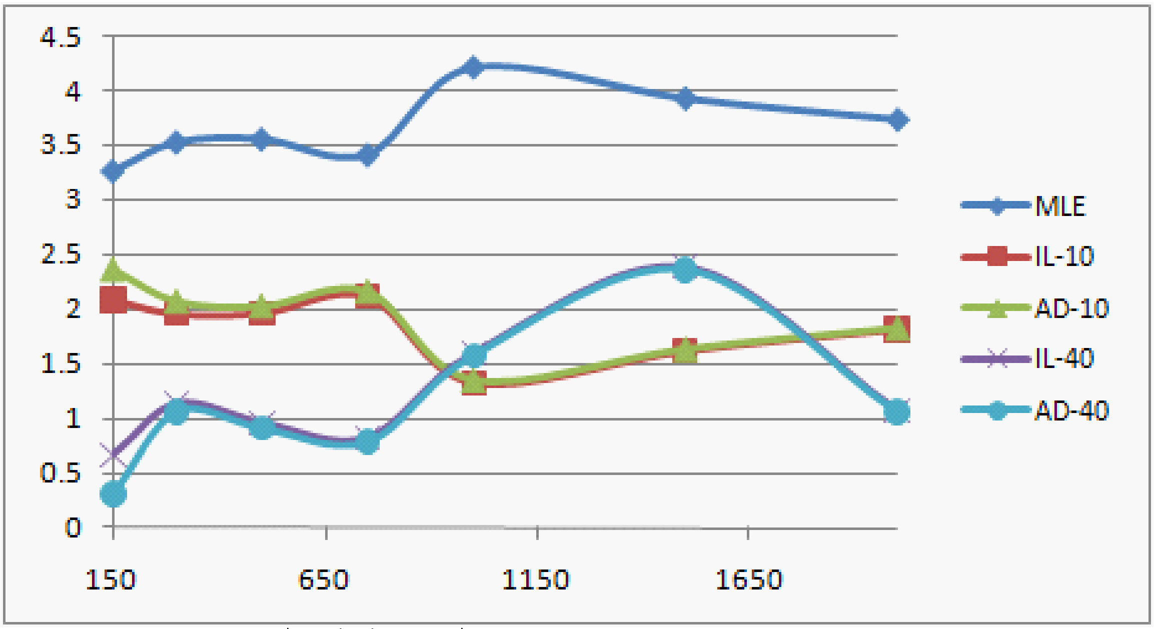

Figure 5: $n\left|E\left(\widehat{\theta_2}\right)-\theta_2\right|$ $ARCH(1)$ model, $\theta_1=1.0$ and $\theta_2=0.5$.