A New Class of Indirect Estimators: Higher Order Asymptotics and Approximate Bias Correction

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

January 10, 2012

Abstract

In this paper we define a set of indirect estimators based on moment approximations of the auxiliary ones. Their introduction is motivated by reasons of analytical and computational facilitation. We provide results that describe higher order asymptotic properties of these estimators. We extend this to a class of multistep indirect estimators that have zero higher order bias without increasing the approximate Mean Squared Error, up to the same order.

KEYWORDS: Indirect Estimator, Recursive Indirect Estimator, Binding Function, Asymptotic Approximation, Edgeworth Expansion, Moment Approximation, Higher Order Bias, Higher Order Mean Squared Error, Approximate Bias Correction.

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 (IE) are (semi-) parametric ones, usually emerging from two-step optimization procedures. They were formally introduced by Gourieroux Monfort and Renault [27]. 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. The inversion criterion depends on a function connecting the underlying statistical models, termed *the binding function*. Minimization of the inversion criterion, which usually has the form of a stochastic norm, essentially inverts the binding function.

Given an auxiliary estimator, IE differ due to differences in the inversion criteria that hinge on differences between the binding functions that each one involves. Among the IE involving the same auxiliary estimator, the consistent ones depend on sequences of binding functions that converge appropriately to a common limit binding function 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 of 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 casts them as unattractive among practitioners. The same IE under a more involved assumption framework also have attractive high order asymptotic properties,[1] that are not exploited due to the aforementioned numerical burden.

Part of the scope of the present paper, is the introduction of a class of (potentially multistep) IE, where 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 employed 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 with-

[1] See e.g. Gourieroux and Monfort [26], and Gourieroux, Renault and N. Touzi [28].

out sacrificing useful properties. Furthermore, 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 approximate bias correction. In fact, by providing to these procedures a framework, this of IE, we are able to provide sufficient conditions under which these approximations are valid ones.

Before the discussion of the framework on which the current results are based upon, in section 2, notice that indirect inference algorithms were initially employed by Smith [46], were formally introduced by Gourieroux et al. [27], complemented by Gallant and Tauchen [21] and extended by Calzolari, Fiorentini and Sentana [8]. Properties similar to those studied here were studied in Gourieroux et al. [28] and were validated and extended 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. Throughout all sections we employ two examples, concerning the MA(1) and GARCH$(1,1)$ processes, to facilitate the exposition of our results. In the first case almost all needed formulae are known analytically, while in the second one they are numerically approximated. In section 6 we present an additional example, concerning the fractional Gaussian process, and Monte Carlo experiments. Conclusions are gathered in section 7 and in the appendices we collect our proofs along with some useful tools concerning the derivation of our results.

2 General Framework

In this section a general assumption framework is described, that facilitates the presentation of the already established IE. In order to do so we establish first some initial notation.

Given a metric space (X, d_X) the symbol $\mathcal{O}_{\varepsilon}(x)$ will denote the ε-ball around the point x and $\overline{\mathcal{O}}_{\varepsilon}(x)$ its closure. For a matrix W, $\|W\|$ will denote a *submultiplicative* matrix norm,[2] 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.

Assumption A.1 Let the following hold:

1. Θ denotes a compact subset of the p-dimensional Euclidean space. Given a measurable space $(\Omega, \mathcal{F})$, the statistical model at hand is de-

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

fined by a correspondence par : $\Theta \twoheadrightarrow \mathcal{P}$ the set of probability measures on $\mathcal{F}$ such that par $(\theta) \cap$ par $(\theta') \neq \varnothing$ iff $\theta = \theta'$. Let P_θ denote any member of par (θ).

2. The limit binding function (lbf) $b : \Theta \to B$, for B a compact subset of $\mathbb{R}^q$, $b(\Theta) \subset \operatorname{Int} B$. Moreover it is continuous, injective and, for a natural number s specified in the sequel, it is $s+2$ times continuously differentiable when restricted to $\operatorname{Int}\Theta$.[3] Also, 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.[4]

3. Let $W_n(\cdot, \theta')$ be $\mathcal{B}_{M(\mathbb{R},q)} / (\mathcal{B}_{\mathcal{F}_n} \otimes \mathcal{B}_\Theta)-$measurable and P_θ-almost surely positive definite, for every $\theta', \theta \in \Theta$. Also, let θ_n^+ denote a random element on Ω with values in Θ.

For an appropriate sequence of measurable spaces $(\Omega_n, \mathcal{F}_n)_{n=1}^{\infty}$, we usually have that $\Omega = \times_n \Omega_n$, i.e. the Cartesian product of the $\Omega_n' s$, $\mathcal{F} = \bigotimes_n \mathcal{F}_n$, i.e. the $\sigma - field$ of the product sets, and that any $P^* \in$ par (Θ) 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.

In the following we suppress the dependence of the aforementioned binding functions on Ω where unnecessary. 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 $W \in \mathcal{PD}(\mathbb{R}, r)$.

[3] Due to the obvious separability 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 $\overline{\mathcal{O}}_\varepsilon(\theta)$ for any θ and appropriate ε.

[4] Componentwise lower semi-continuity of the lbf would follow from the continuity of par (i.e. for any θ, $\theta_n \to \theta$ the sequence comprised by any member of par (θ_n) converges inside par (θ) with respect to the weak topology), the uniform w.r.t. P_θ $\varsigma_n(\omega, \beta) \rightsquigarrow \varsigma(\theta, \beta)$ uniformly over $b(\Theta)$ ($\rightsquigarrow$ denotes convergence in distribution.), for $\varsigma : \Theta \times b(\Theta) \to \mathbb{R}$ such that $\varsigma(\theta, b(\theta)) < \varsigma(\theta, \beta)$ $\forall \beta \in b(\Theta)$ and $\forall \theta \in \Theta$, and the 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.

Definitions of Already Known IE We can now define the auxiliary and the already established, GMR1, and GMR2 estimators. They were initially formalized by Gourieroux et al. [27].

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

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

The GMR1 estimator is defined by

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

Finally, let $b_n(\theta) = E_\theta \beta_n$, then the GMR2 estimator is defined by

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

Remark R.1 By A.1.2-3 the above estimators are well defined. The IE computation relies on the analytical availability of b, and they are in most cases numerically approximated. Thereby, they are associated with *nested numerical optimizations* producing large numerical costs especially in the case of GMR2.

Let us now introduce the two examples.

MA(1) **Example** Consider the invertible MA(1) process

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

Now Θ is a compact subset of $(-1, 1)$. The usual estimation method of θ is QML, however one can find the sample first order autocorrelation of the process, say β_n, and find the θ that equates the sample and the theoretical autocorrelations. In fact, this is the GMR1 estimator of θ, and as the binding function is $b(\theta) = \frac{\theta}{1+\theta^2}$ the GMR1 is equal to $\frac{1-\sqrt{1-4\beta_n^2}}{2\beta_n}$. Notice that β_n is also the QMLE of the coefficient of an AR(1) auxiliary model (see Gourieroux et al. [27], and Demos and Kyriakopoulou [14]). Let B be a compact subset of $\left(-\frac{1}{2}, \frac{1}{2}\right)$ compliant to assumption A.1. To evaluate the GMR2 estimator the $E_\theta \beta_n$ is needed, which is analytically intractable. This expectation is usually approximated by Monte Carlo integration (see e.g. Gourieroux et al. [27] or Arvanitis and Demos [4]).

GARCH$(1,1)$ **Example** Consider the second order stationary GARCH $(1,1)$ model (Bollerslev [7])

$$\begin{aligned} y_t &= u_t^{1/2} z_t, \quad u_t = \theta_1 + \theta_2 y_{t-1}^2 + \theta_3 u_{t-1}, \quad t = ..., -1, 0, 1, ..., \qquad (2) \\ \theta_1, \theta_2 &> 0, \quad \theta_3 \geq 0, \quad \theta_2 + \theta_3 < 1 \quad z_t \overset{iid}{\backsim} N(0,1). \end{aligned}$$

Now Θ is a compact subset of $\mathbb{R}^3$ defined by the above inequalities. Let $\beta_n = \left(\beta_{1,n}, \beta_{2,n}, \beta_{3,n}\right)'$ represent the MLE of $\theta = (\theta_1, \theta_2, \theta_3)'$. It is well known that β_n is consistent and asymptotically normal (see e.g. Lee and Hansen [33] or Lumsdaine [38]). However, it is also known that in small samples it can be severely biased (see e.g. Linton [37] or Lumsdaine [39]). Now treating β_n as the auxiliary estimator we have that, as the binding function is the identity, the GMR1 $= \beta_n$. Again as the $E_\theta \beta_n$ is analytically intractable and, for the evaluation of the GMR2 estimator, it is approximated via Monte Carlo integration. It is in this case that the evaluation of the GMR2 estimator involves nested optimizations.

Assumptions Specific to a New Class of IE The following assumptions enable the definition and the derivation of properties of a new class of IE. We need some further notation. For $s^*, s_*, s \in \mathbb{N}$ with $s^*, s_* \geq s$, let $a^* = \frac{s^*-1}{2}$, $a_* = \frac{s_*-1}{2}$ and $a = \frac{s-1}{2}$. For the Edgeworth measure of order s^* with density $\left(1 + \sum_{i=1}^{s^*} \frac{\pi_i(z,\theta)}{n^{\frac{i}{2}}}\right) \varphi_{V_\theta}(z)$, where φ_{V_θ} denotes the density of $N(0, V_\theta)$ for V_θ is a positive definite $q \times q$ matrix (see for example Magdalinos [41] eq. 3.7-8, p. 348). The three expansion orders are necessary in order to make the set-up as general as possible. In general, s^* will denote the order of the Edgeworth expansion of the auxiliary estimator, s_* will be employed for the definition of the newly established estimator, whereas s will refer to the order of the moment approximations. Let $k_i(z,\theta) = z\pi_{i-1}(z,\theta)$ for $i = 1, \ldots, s^*$, and with $\mathcal{I}_{V_\theta}(k_i(z,\theta)) = \int_{\mathbb{R}^q} k_i(z,\theta) \varphi_{V_\theta}(z) dz$. D^r denotes the r-derivative operator and $D^r(f(x_0))(x^r)$ the r^{th}-linear function defined by the evaluation of $D^r f$ at x_0 evaluated at $\underbrace{(x, ..., x)}_{r \ \ times}$. The first assumption essentially builds the binding function upon which the definitions of IE that follow will be based.

Assumption A.2 There exist $\xi_i : \Theta \to \mathbb{R}^q$ for which

$$\left\| E_\theta \beta_n - b(\theta) - \sum_{i=1}^{s_*} \frac{\xi_i(\theta)}{n^{\frac{i}{2}}} \right\| = o\left(n^{-\frac{s}{2}}\right) \qquad (3)$$

for any $\theta \in \Theta$ with $\xi_1 = 0_q$.

Remark R.2 This holds when $\sqrt{n}\left(\beta_n - b\left(\theta\right)\right)$ admits an Edgeworth expansion of order $s^* > s$ in the light of lemma AL.2, in appendix B, with $\xi_i\left(\theta\right) = \mathcal{I}_{V_\theta}\left(k_i\left(z,\theta\right)\right)$ for all $\theta \in \Theta$. Due to assumption A.1.2, $b\left(\theta\right)$ lies in $\operatorname{Int} B$ hence $\xi_1 = 0_q$. The assumption on the $\sqrt{n}$ rate of convergence of the auxiliary estimator could be extended so as to allow for different rates as long as these do not depend on θ.

The next assumption, enables the *stochastic approximations* of the $\xi_i\left(\theta\right)' s$ in (3). This can facilitate the definition of IE in cases where (some of) the $\xi_i' s$ are analytically unknown. This could be the case if the structure of statistical model involves nuisance parameters, analytically unknown moments etc. We assume the existence of another probability space that enables the possibility of stochastic approximations via numerical methods like Monte Carlo simulations, bootstrap etc.

Assumption A.3 For some $\delta > 0$ small enough that could depend on θ:

1. For any $\theta \in \Theta$ and a probability space $(\Omega', \mathcal{F}', P_\theta')$ and each $i = 2, \ldots, s_*$, there exist $\zeta_{i_n} : \Omega \times \Omega' \times \Theta \to \mathbb{R}^q$, that is $\mathcal{B}_{\mathbb{R}^q} / \left(\mathcal{F} \otimes \mathcal{F}' \otimes \mathcal{B}_\Theta\right)$-measurable, Q_θ-almost everywhere continuous on Θ where $Q_\theta = P_\theta \times P_\theta'$. If $\xi_i = 0_q$ then set $\zeta_{i_n} = 0_q$.

2. $Q_\theta\left(\sup_{\theta' \in \Theta}\left\|\zeta_{i_n}\left(\omega, \omega', \theta'\right)\right\| > o\left(n^{\frac{i-1}{2}-\delta}\right)\right) = o\left(n^{-a^*}\right)$, for all $i = 2, \ldots, s_*$.

3. ζ_{i_n} is s^*+2 continuously differentiable on $\operatorname{Int}\Theta$, Q_θ-almost everywhere, and for any $\theta' \in \operatorname{Int}\Theta$ there exists an $\varepsilon > 0$ (independent of θ) such that $Q_\theta\left(\sup_{\theta'' \in \overline{\mathcal{O}}_\varepsilon(\theta')}\left\|D^j\zeta_{i_n}\left(\omega, \omega', \theta''\right)\right\| > o\left(n^{\frac{i-1}{2}-\delta}\right)\right) = o\left(n^{-a^*}\right)$, for all $i = 2, \ldots, s_*$, $j = 1, \ldots, s^* + 2$.

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 $\xi_i' s$. The space Ω' can also depend on some index that indicates the number of Monte Carlo and/or bootstrap samples which is suppressed. This setup is general enough to allow for cases in which ζ_{i_n} is evaluated on initial estimators of θ, and/or on estimators of nuisance parameters. Similarly it allows for cases in which the $\xi_i' s$ depend 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, possibly composed with measurable functions $\Omega \to \Theta$. This allows also for approximations of $\xi_i's$ when they are *partially computed* at a stochastic point, enabling the derivation of estimators that emerge from *partial optimization.*

Remark R.3 Assumption A.3.1 establishes essentially the almost everywhere continuity of the relevant simulators. If $\zeta_{i_n} = \xi_i = \mathcal{I}_{V_\theta}(k_i(z,\theta))$ this could follow from the continuity of moments appearing in the Edgeworth polynomials. A.3.2 and A.3.3 would follow from Lipschitz conditions with probability $1-o\left(n^{-a^*}\right)$ on $\zeta_{i_n}(\omega,\omega',\theta)$ (on Θ), and $D^j\zeta_{i_n}(\omega,\omega',\theta)$ (on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$) and the $o\left(n^{\frac{i-1}{2}-\delta}\right)$ bound for the analogous Lipschitz coefficients and $\zeta_{i_n}(\omega,\omega',\theta)$, $D^j\zeta_{i_n}(\omega,\omega',\theta)$ evaluated at some arbitrary point. In the case that $\zeta_{i_n}(\omega,\omega',\theta)=\frac{1}{n}\sum\zeta_i(y_j,\theta)$ for ζ_i and y_j appropriate measurable functions that can be defined on some open superset of Θ, then a condition of the form

$$E_\theta\left\|\sup_{\theta'}D^{s^*+2}\zeta_i(y_1,\theta')-E_\theta\sup_{\theta'}D^{s^*+2}\zeta_i(y_1,\theta')\right\|^p<+\infty$$

for $p>2a^*$ along with stationarity and mixing conditions would validate A.3.2 and A.3.3 with $o\left(n^{\frac{i-1}{2}-\delta}\right)$ replaced by constants via results such as the Yokoyama moment inequality (see Andrews [2], proof of lemma 3).

The next assumption concerns the asymptotic behavior of the weighting matrix.

Assumption A.4 Suppose that there exists a sequence of random elements $x_n:\Omega\to\mathbb{R}^m$, such that $W_n(\theta)=\frac{1}{n}\sum W(x_i(\omega),\theta)$ for $W:\mathbb{R}^m\times\Theta\to\mathcal{PD}(q,\mathbb{R})$ integrable with respect to P_θ for any $\theta\in\operatorname{Int}\Theta$, such that a)

$$P_\theta\left(\|W_n(\theta')-E_\theta W(\theta')\|>\varepsilon\right)=o\left(n^{-a^*}\right),\forall\varepsilon>0$$

$W(\theta')$ is Lipschitz on $\overline{\mathcal{O}}_\varepsilon(\theta)$ with coefficient κ_{W^*} integrable w.r.t. P_θ and $P_\theta\left(\frac{1}{n}\sum\kappa_W(x_i)>M\right)=o\left(n^{-a^*}\right)$ for any $\theta\in\operatorname{Int}\Theta$. b) W_n is s^*+1-continuously differentiable P_θ-almost surely on $\operatorname{Int}\Theta$ and for any $\theta'\in\operatorname{Int}\Theta$

$$P_\theta\left(\sup_{\theta''\in\overline{\mathcal{O}}_\varepsilon(\theta')}\left\|D^{s^*+1}W_n^j(\theta'')\right\|>M\right)=o\left(n^{-a^*}\right)$$

Remark R.4 The first part of assumption A.4.a can be justified by conditions on the asymptotic behavior of $E_\theta\left(\|W_n^j(\theta')-E_\theta W^j(\theta')\|^q\right)$. The second part can be justified by

$$E_\theta\sup_{\theta'\in\overline{\mathcal{O}}_\varepsilon(\theta)}\left\|DW^j(x_i(\omega),\theta')\right\|<+\infty$$

Part b) can be justified analogously.

Obviously when $W^j(x,\theta)$ is independent of x and θ the above is trivially satisfied.

The next assumption makes evident the sense of the approximation of ξ_i from ζ_{i_n}. For $\theta \in \operatorname{Int}\Theta$, let $f_n(\theta)$ be the vector containing the elements of $W_n(\theta) - E_\theta W(\theta)$. Let $q_n(\theta)$ be the vector containing the elements of $\zeta_{i_n}(\omega,\omega',\theta) - \xi_i(\theta)$ for all i. Let also $m_n^*(\theta)$ be $\left((\beta_n - b(\theta))', (\theta_n^+ - \theta)'\right)'$.

Assumption A.5 For $\theta \in \operatorname{Int}\Theta$:

1. the distribution of $\sqrt{n}m_n^*(\theta)$ under Q_θ admits an Edgeworth expansion of order s^*.
2. $Q_\theta(\|q_n(\theta)\| > o(1)) = o\left(n^{-a^*}\right)$, $P_\theta(\|f_n(\theta)\| > o(1)) = o\left(n^{-a^*}\right)$.

Remark R.5 This first part can be established 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

$$S_n(\omega,\omega',\theta) = n^{-1/2}\left[X_n(\omega,\omega',\theta) - E_\theta X_n(\omega,\omega',\theta)\right]$$

and assumptions 2-4 of Durbin [16] 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 X_n is in a form of a sum. They were employed, 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 MLE and WMLE, respectively, uniformly on their parameter space for long memory Gaussian processes.

Remark R.6 In a similar fashion $S_n(\omega,\omega',\theta)$ could be of the form

$$n^{-1/2}\left(\textstyle\sum_{i=1}^n \left(X_i(\omega,\omega',\theta) - EX_i(\omega,\omega',\theta)\right)\right)$$

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 [25] (Lemma 2.3, Assumptions 2-4) if for any θ, g satisfies some Lipschitz conditions, has almost everywhere continuous derivatives w.r.t. to ε, which are appropriately non-degenerate in a set of positive probability, $E\left\|X_1(\omega,\omega',\theta)\right\|^{2a^*+3}<\infty$, and $g(\varepsilon_{n-i}(\omega,\omega'),\theta; i\in\mathbb{N})$ satisfies a weak dependence condition, then $S_n(\omega,\omega',\theta)$ again admits a valid formal Edgeworth expansion of order s^* for any θ. Given the previous if $\sqrt{n}m_n^*(\theta)=\pi_n(S_n(\omega,\omega',\theta))+R_n(\theta)$ with Q_θ-probability $1-o\left(n^{-a^*}\right)$, where for any θ, π_n satisfies the provisions of lemma 4.6 of Skovgaard [45] (or in a more specific form AL.4 in appendix B) and

$$Q_\theta\left(\|R_n(\theta)\|>\gamma_n(\theta)\right)=o\left(n^{-a^*}\right)$$

for $\gamma_n(\theta)=o\left(n^{-\epsilon(\theta)}\right)$ with $\epsilon(\theta)>0$, then the assumption follows from aforementioned lemmas and AL.1 (in appendix B). In the case that $\epsilon(\theta)=a^*$ then the relevant expansion coincides with the one of $\pi_n(S_n(\omega,\omega',\theta))$. A.5.2 would follow if $\sqrt{n}q_n(\theta)$ admits under Q_θ, an Edgeworth expansion of the analogous order due to lemma 2 of Magdalinos [41]. Analogously for the second part of A.5.2

The following lemmas are useful and immediate.

Lemma 2.1 Assumptions A.1, A.4, and A.5.1 imply that for any $\theta\in\operatorname{Int}\Theta$

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

Lemma 2.2 Assumptions A.1, A.4, and A.5.1 imply that for any $\theta\in\operatorname{Int}\Theta$ the distribution of $\sqrt{n}\,(\text{GMR1}-\theta)$ under P_θ admits an Edgeworth expansion of order s^*.

MA(1) **Example Cont.** Notice that for the MA(1) process in equation 1, assumption A.4 is irrelevant as the dimensions of $b(\theta)$ and θ are equal, i.e. $p=q$. Furthermore, from the results of Arvanitis and Demos [4], and Demos and Kyriakopoulou [14] we have that if $Eu_i^{14}<\infty$ and if $D(0,\sigma^2)$ is a smooth continuous density, then the β_n estimator admits a 5^{th} order valid Edgeworth expansion, uniformly over Θ, and by lemma AL.2 of appendix B we have that assumption A.2 applies for $s_*=4$ and the $\xi_i(\theta)$ are known functions of θ only (see Demos and Kyriakopoulou [14]). Consequently, assumption A.5 applies, with $m_n^*(\theta)=\beta_n-\frac{\theta}{1+\theta^2}$ and $s^*=4$, validating the 4^{rd} order expansion of the GMR1 estimator, by lemma 2.2.

GARCH$(1,1)$ **Example Cont.** For the GARCH$(1,1)$ process in equation 2, again, assumption A.4 is irrelevant as the binding function is the identity. Further, by Corradi and Iglesias [11] we have that, under normality, the Edgeworth expansion of the MLE, β_n, is valid for any order s^*. Employing the analytic formulae for the Edgeworth expansion, given in Linton [37] and

Iglesias and Linton [31], we have that $E(\beta_n - \theta) = -\frac{1}{n}(\lambda_0 + \lambda_2) + o(n^{-1})$, where λ_0 and λ_2 depends on the moments of the, up to 3^{rd} order, likelihood derivatives, which in turn they depend on the derivatives of the conditional variance process in 2, up to 2^{nd} order. These do not have an explicit form. For example, λ_0 is a smooth function of $u_t^{-1}(u_{t;1}, u_{t;2}, u_{t;3})'$, where $u_{t;i} = \frac{\partial u_t}{\partial \theta_i}$, and of the elements of a matrix of the form $\left\{\frac{1}{n}\sum_{t=1}^{n} E_\theta\left(u_t^{-2}u_{t;i}u_{t;j}\right)\right\}_{i,j=1,\ldots,3}$. In fact these expectations are evaluated via Monte Carlo integrations or any other type of numerical methods such as boostrap. Now the validity of assumption A.3 follows from the smoothness of λ_0 and λ_2, the compactness of Θ, the conditional normality and the fact that $u_t(\theta)$ has a uniform strictly positive lower bound. Further, the validity of the Edgeworth expansions of the derivatives, up to 3^{rd} order, of the likelihood function, say $l_{1_i}(\theta)$, is presented in Lemma A.1 of Corradi and Iglesias [11]. Hence we know that $\sqrt{n}(\beta_n - \theta)$ and $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(l_{1_t}(\theta) - El_{1_t}(\theta))$ have valid Edgeworth expansions. Now let $l_{2_t}(\theta)$ represent the elements involved in λ_0 and λ_2 with analytically unknown expectations, e.g. $u_t^{-1}u_{t;1}$ or $u_t^{-2}u_{t;i}u_{t;j}$ etc. In the same way as in lemma A.1 of Corradi and Iglesias [11] it is possible to prove that the vector $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix} l_{1_t}(\theta) - El_{1_t}(\theta) \\ l_{2_t}(\theta) - El_{2_t}(\theta)\end{pmatrix}$ has a valid Edgeworth expansion (conditions 2-4 of Gotze and Hipp [25] are easily, yet tediously, established). Notice that $\sqrt{n}(\beta_n - \theta) = \pi_n\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix} l_{1_t}(\theta) - El_{1_t}(\theta) \\ l_{2_t}(\theta) - El_{2_t}(\theta)\end{pmatrix}\right) + R_n$ with $P(\|R_n\| > \gamma_n) = o(n^{-1})$ for $\gamma_n = o(n^{-1})$, and π_n satisfying the assumptions of lemma AL.4 (see also the proof of lemma 8 of Andrews [2]). Hence $\begin{pmatrix} \sqrt{n}(\beta_n - \theta) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(l_{1_t}(\theta) - El_{1_t}(\theta)) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(l_{2_t}(\theta) - El_{2_t}(\theta))\end{pmatrix}$ admits a valid Egdeworth expansion. Moreover the elements of $l_{2_t}(\theta)$ are smooth functions of θ and additionally $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(l_{2_t}(\theta_n) - El_{2_t}(\theta)) = \pi_n^*\begin{pmatrix} \sqrt{n}(\beta_n - \theta) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(l_{1_t}(\theta) - El_{1_t}(\theta)) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(l_{2_t}(\theta) - El_{2_t}(\theta))\end{pmatrix} + R_n^*$ with $P(\|R_n^*\| > \gamma_n^*) = o(n^{-1})$ for $\gamma_n^* = o(n^{-1})$, and π_n^* in the premises of remark R.6. Hence $\begin{pmatrix} \sqrt{n}(\beta_n - \theta) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(l_{2_t}(\theta_n) - El_{2_t}(\theta)) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(l_{1_t}(\theta) - El_{1_t}(\theta)) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(l_{2_t}(\theta) - El_{2_t}(\theta))\end{pmatrix}$ admits a valid Egdeworth expansion. Now,

$$\sqrt{n}m_n^*(\theta) = \begin{pmatrix} \sqrt{n}\left(\beta_n - \theta\right) \\ \sqrt{n}\left[\lambda_0\left(\frac{1}{n}\sum_{t=1}^{n} l_{2_t}(\theta_n)\right) - \lambda_0\left(El_{2_t}(\theta)\right)\right] \\ \sqrt{n}\left[\lambda_2\left(\frac{1}{n}\sum_{t=1}^{n} l_{2_t}(\theta_n)\right) - \lambda_2\left(El_{2_t}(\theta)\right)\right] \end{pmatrix}$$

can be easily seen to satisfy an analogous expression with respect to a sequence of smooth functions of $\begin{pmatrix} \sqrt{n}\left(\beta_n - \theta\right) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(l_{2_t}(\theta_n) - El_{2_t}(\theta)\right) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(l_{1_t}(\theta) - El_{1_t}(\theta)\right) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(l_{2_t}(\theta) - El_{2_t}(\theta)\right) \end{pmatrix}$ and an appropriate remainder. Hence, by remark R.6, assumption A.5 follows.

Let us now turn our attention to the newly established estimator.

3 Definition of the $\mathrm{GMR2}(a_*)$ 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_n} on the generalized sample space for notational convenience and denote $\zeta_n(\theta, a) = \left(\zeta_{2_n}(\theta), \ldots, \zeta_{s_n}(\theta)\right)$ and $b_n(\theta, \zeta_n(\theta, a_*)) = b(\theta) + \sum_{i=2}^{s_*} \frac{1}{n^{i/2}}\zeta_{i_n}(\theta)$.

Definition D.2 Given Assumptions A.3 and A.4, the $\mathrm{GMR2}(a_*)$ estimator is defined by

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

Remark R.7 The definition depends on i) the analytical knowledge of b, and ii) a numerical approximation of the remaining mean approximation. Due to the fact that for a large class of models the form of the analogous formal Edgeworth approximations is known, up to their dependence of analytically unknown yet simulable moments, i) is the hardest part to establish. However this constitutes of a fixed cost. Given this, the $\mathrm{GMR2}(a_*)$ estimator can surpass the nested optimization burden associated with the GMR2 estimator. Furthermore, due to A.1 and A.2 the GMR1 estimator can be identified as the $\mathrm{GMR2}(0)$ one.

Remark R.8 Suppose that $\beta_n = \theta_n(0)$, and $b(\theta) = \theta$, $\zeta_{i_n}^* = \zeta_{i_n}(\theta_n(0))$ and consider the $\mathrm{GMR2}(a_*)$, defined by

$$\theta_n^*(a_*) = \theta_n(0) - \sum_{i=2}^{s_*} \frac{1}{n^{i/2}}\zeta_{i_n}^*$$

Q_θ-almost everywhere, the computation of which is of minimal numerical burden. $\theta_n^*(a_*)$ admits another interesting characterization. It is the first term of a sequence defined by $\theta_n^{(i)} = \theta_n(0) - \sum_{i=2}^{s_*} \frac{1}{n^{i/2}} \zeta_{i_n}\left(\theta_n^{(i-1)}\right)$, for $i = 0, 1, 2, \ldots$, and $\theta_n^{(0)} = \theta_n(0)$. It is easy to prove that under assumptions A.3 and A.5 for any $\theta \in \operatorname{Int}\Theta$ and $\delta > 0$

$$P_\theta\left(\sup_{\theta' \in \overline{\mathcal{O}}_\varepsilon(\theta)} \left\| \sum_{i=2}^{s_*} \frac{1}{n^{i/2}} D\zeta_{i_n}(\theta) \right\| \sup_{\theta' \in \overline{\mathcal{O}}_\varepsilon(\theta)} \left\| \sum_{i=1}^{s_*} \frac{1}{n^{i/2}} \zeta_{i_n}(\theta) \right\| > o\left(n^{-\delta}\right)\right) = o\left(n^{-a^*}\right).$$

Hence $\theta_n^*(a_*)$ is in fact the first term of an asymptotically approximate Newton-Ralphson sequence (that need not converge to $\theta_n(a_*)$ as $i \to \infty$). It is widely used in the literature of bias correction especially in the case where $a_* = \frac{1}{2}$ (see e.g. Bao and Ullah [5], Cox and Hinkley [12], Linton [37], MacKinnon and Smith [40] etc.). Notice that it is possible that for some n, $\theta_n^*(a_*) \notin \Theta$ or it will be in the boundary of Θ with positive Q_θ probability, as it will be the case in some of the examples considered later. Finally, our framework enables the possibility that intermediate cases can be also characterized as GMR2(a) estimators, i.e. cases $\zeta_{i_n}^*$ depend on θ for some i.

MA(1) **Example Cont.** From Demos and Kyriakopoulou [14] we have that

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

As expected the GMR1 is not 3^{rd} order unbiased as the binding function is not linear. Treating the GMR1 ($\theta_n(0)$ in our terminology, see remark R.8) estimator as the auxiliary one we can define the GMR2(1) as:

$$\theta_n(1) = \text{GMR2}(1) = \arg\min_{\theta \in \Theta} \left\| \theta_n(0) - \theta - \frac{1}{n} \theta \frac{1 + 5\theta^2 + 2\theta^4 + \theta^6 - \theta^8}{\left(1 - \theta^2\right)^3} \right\|.$$

Notice that in this case the binding function is the identity, i.e. $b(\theta) = \theta$. Further, notice that, what is commonly known as, the biased corrected estimator is given by $\theta_n(0) - \frac{1}{n}\theta_n(0) \frac{1+5\theta_n^2(0)+2\theta_n^4(0)+\theta_n^6(0)-\theta_n^8(0)}{\left(1-\theta_n^2(0)\right)^3}$ (see e.g. Bao and Ullah [5] or Demos and Kyriakopoulou [14]). We refer to this estimator as $\theta_n^*(1)$ (see remark R.8) and it is defined as

$$\theta_n^*(1) = \arg\min_{\theta \in \Theta} \left\| \theta_n(0) - \theta - \frac{1}{n} \theta_n(0) \frac{1 + 5\theta_n^2(0) + 2\theta_n^4(0) + \theta_n^6(0) - \theta_n^8(0)}{\left(1 - \theta_n^2(0)\right)^3} \right\|.$$

GARCH$(1,1)$ **Example Cont.** As the binding function is the identity then the GMR1 estimator equals toβ_n, i.e. the MLE one. Hence the GMR2 $\left(\frac{1}{2}\right)$ estimator is defined as

$$\theta_n\left(\frac{1}{2}\right) = \text{GMR2}\left(\frac{1}{2}\right) = \arg\min_{\theta\in\Theta}\left\|\beta_n - \theta + \frac{\lambda_0(\theta) + \lambda_2(\theta)}{n}\right\|.$$

The commonly employed biased corrected estimator (see e.g. Bao and Ullah [5] Linton [37]), $\theta_n^*\left(\frac{1}{2}\right)$ (see remark R.8), is defined as

$$\theta_n^*\left(\frac{1}{2}\right) = \arg\min_{\theta\in\Theta}\left\|\beta_n - \theta + \frac{\lambda_0(\beta_n) + \lambda_2(\beta_n)}{n}\right\|.$$

Notice that numerically approximated terms, λ_0 and λ_2, are evaluated at β_n, as opposed to the $\theta_n\left(\frac{1}{2}\right)$ where they are functions of the minimization parameters.

When the binding function is the identity (or the inclusion) function, then the GMR2 (a_*) estimator lies in the class of estimators considered by MacKinnon and Smith [40] where the bias function is approximated as our assumptions A.2-A.4 indicate. The form of the objective functions from which they emerge and the derivation of their higher order asymptotic properties imply that under our assumption framework, they constitute a subclass MacKinnon-Smith estimators that perform second order bias correction while retaining the analogous order approximate mean squared error (see section 4.3). Moreover they facilitate the definition of multistep estimators that approximate the bias function with increasing accuracy and thereby can perform approximate bias correction of any order (see section 5). We establish these properties in what follows.

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 estimators. Consistency, Edgeworth and moment approximations are established in that order.

4.1 Consistency

It is proven that the GMR2 (a_*) under Q_θ is contained in an arbitrary neighborhood of $\theta \in \operatorname{Int}\Theta$ 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 GMR2 criterion. Analogous relations are established between GMR2 (a_*) and GMR2 (a'_*), for potentially different a_*, a'_* including the case of the GMR1. For notational simplicity denote $E_\theta W(\theta)$ with $W(\theta)$.

Lemma 4.1 Under assumptions A.1, A.4, A.2, A.3.1 and A.5.1 for any $\theta \in \operatorname{Int}\Theta$, $\varepsilon > 0$

$$Q_\theta\left(\sup_{\theta'\in\Theta}\left|\|\beta_n - b_n(\theta', \zeta_n(\theta', a_*))\|_{W_n\left(\theta_n^+\right)} - \|b(\theta) - b(\theta')\|_{W(\theta')}\right| > \varepsilon\right) = o\left(n^{-a^*}\right)$$

and therefore

$$Q_\theta\left(\|\theta_n(a_*) - \theta\| > \varepsilon\right) = o\left(n^{-a^*}\right)$$

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

Lemma 4.2 Under assumptions A.1, A.4, A.2, A.3.1 and if $\sup_{\theta\in\Theta}\|E_\theta\beta_n - b(\theta)\| = o(1)$ then for any $\theta \in \operatorname{Int}\Theta$

$$\left\|\beta_n - E_{\theta_n(a_*)}\beta_n\right\|_{W_n\left(\theta_n^+\right)} \leq \|\beta_n - E_{\mathrm{GMR2}}\beta_n\|_{W_n\left(\theta_n^+\right)} + \eta_n$$

with $Q_\theta(\eta_n > \varepsilon) = o\left(n^{-a^*}\right)$, for any $\varepsilon > 0$ and η_n is Q_θ-almost surely non negative.

Remark R.9 The condition $\sup_{\theta\in\Theta}\|E_\theta\beta_n - b(\theta)\| = o(1)$ would follow from the uniform (pseudo) consistency of the auxiliary estimator given the compactness of B. If the lbf is a bijection $\sup_{\theta\in\Theta} P_\theta\left(\sup_{\beta\in B}\|\varsigma_n(\beta) - \varsigma(\theta,\beta)\| > \delta\right) = o(1)\ \forall\delta > 0$ would be sufficient for this, given assumption A.1.2 (see also the analogous footnote). The examined estimator is essentially an η_n-GMR2 estimator. Finally notice that 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 $b(\theta)$. Obviously, the GMR1 estimator is an approximate GMR2 estimator.

The previous reasoning can also establish analogous relations between GMR2 (a) and GMR2 (a') estimators, with a and a' not necessarily the same that could also be defined by different ζ_{i_n}.

Lemma 4.3 Suppose that assumptions A.1, A.4, A.2, A.3.1 hold for both a_*, a'_* ζ_{i_n} and ζ'_{i_n} defined analogously, then for any $\theta \in \operatorname{Int}\Theta$, there exists a real sequence $\gamma_n = o\left(n^{-\left(\delta+\frac{1}{2}\right)}\right)$ such that

$$\|\beta_n - b_n(\theta_n(a'_*), \zeta_n(\theta_n(a'_*), a_*))\|_{W_n\left(\theta_n^+\right)} \leq \|\beta_n - b_n(\theta_n(a_*), \zeta_n(\theta_n(a_*), a_*))\|_{W_n\left(\theta_n^+\right)} + \eta_n$$

with $Q_\theta(\eta_n > \gamma_n) = o\left(n^{-a^*}\right)$, where δ might depend on θ.

Therefore any GMR2(a'_*) is an approximate GMR2(a_*) estimator in this sense given our assumption framework. Obviously this relation holds between a given GMR2(a_*) and the $\theta^*_n(a_*)$ or its variants discussed in R.8. The following lemma provides with a more special case.

Lemma 4.4 Suppose that assumptions A.1, A.4, A.2, A.3 hold for both a_* and a'_* When $a_* > a'_*$, ζ_{i_n} coincide for any i up to $2a'_*$ and in assumption A.3.2 the $o\left(n^{\frac{i-1}{2}-\delta}\right)$ sequences are replaced by a constant, then for any $\theta \in \operatorname{Int}\Theta$, there exists a real sequence $\gamma_n = o\left(n^{-\rho-\frac{1}{2}}\right)$ such that

$$\|\beta_n - b_n(\theta_n(a'_*), \zeta_n(\theta_n(a'_*), a_*))\|_{W_n(\theta_n^+)} \leq \|\beta_n - b_n(\theta_n(a_*), \zeta_n(\theta_n(a_*), a_*))\|_{W_n(\theta_n^+)} + \eta_n$$

with $Q_\theta(\eta_n > \gamma_n) = o\left(n^{-a^*}\right)$, where $\rho = \begin{cases} \frac{1}{2}+\varepsilon \text{ if } a_* = \frac{1}{2} \\ a'_* \text{ if } a_* > \frac{1}{2} \end{cases}$ with $0 < \varepsilon < \frac{1}{2}$ and ε might depend on θ.

Remark R.10 The GMR1 estimator is an approximate GMR2(a_*) for any a_*.

4.2 Validity of Edgeworth Approximation

In this subsection, 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 B.

Lemma 4.5 Under assumptions A.1, A.4, A.2, A.3 and A.5.1 for any $\theta \in \operatorname{Int}\Theta$, there exists an $\{\eta''_n\}_n$, with $Q_\theta\left(\sqrt{n}\,\|\eta''_n\| > \gamma'_n\right) = o\left(n^{-a^*}\right)$, and $\gamma'_n = o\left(n^{-\varepsilon}\right)$ for some $\varepsilon > 0$ that could depend on θ, and $\sqrt{n}\left(\theta_n(a_*) - \theta_n(0)\right) = \eta''_n$ with Q_θ-probability $1 - o\left(n^{-a^*}\right)$.

The validity of the Edgeworth expansion of the distribution of $\sqrt{n}\left(\theta_n(a_*) - \theta\right)$ under Q_θ of order s^* can now be established by assumption A.5, lemma 2.2 and corollary AC.1 presented in the appendix.

Lemma 4.6 Under assumptions A.1, A.4, A.2, A.3 and A.5 for any $\theta \in \operatorname{Int}\Theta$ the distribution of $\sqrt{n}\left(\theta_n(a_*) - \theta\right)$ under Q_θ has an Edgeworth expansion of order s^*.

Lemma 4.6 does not provide any further insight on the form of the Edgeworth approximation for GMR2 (a_*). However, this, along with the first part of lemma AL.4 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 4.7 Under assumptions A.1, A.4, A.2, A.3 and A.5 for any $\theta \in \operatorname{Int}\Theta$ the distribution of $\sqrt{n}\left(\theta_n\left(a_*\right)-\theta\right)$ under Q_θ, as described in the first part of the proof of lemma AL.4, is a valid Edgeworth expansion of order s^*.[5]

4.3 Valid Moment Approximations

Lemma 4.7 in the light of lemmas AL.2 and AL.4 under the correct relation between a^* and a, provides with an approximation of the sequence of moments (of any order) of the defined estimator. The next lemma clarifies this relation.

Lemma 4.8 Under assumptions A.1, A.4, A.2, A.3 and A.5 If $a^* \geq a + \frac{m}{2}$, then for any $\theta \in \operatorname{Int}\Theta$

$$\left\| E_\theta\left(K\left(\sqrt{n}\left(\theta_n\left(a_*\right)-\theta\right)\right)^m\right) - \int_{\mathbb{R}^q} K\left(\left(g_n\left(z\right)\right)^m\right)\left(1+\sum\nolimits_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i\left(z,\theta\right)\right)\varphi_{V_\theta}\left(z\right)dz\right\| = o\left(n^{-a}\right)$$

where g_n as in the proof of lemma 4.7 and $\left(1+\sum\nolimits_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i\left(z,\theta\right)\right)$ is the density of the Edgeworth approximation of order s of $\sqrt{n}m_n^*\left(\theta\right)$.

In the following we explicitly provide this type of approximation for the mean and the mean squared error for the GMR2 (a_*) for any a_* when $a=\frac{1}{2}$. We suppress the dependence on θ and z where possible for notational convenience. For the rest of this section we denote by $b_{,j}$ the j^{th} element of b, $W_{j,j'}$ the (j,j') element of W, $\mathcal{C}=\frac{\partial b'}{\partial\theta}W\frac{\partial b}{\partial\theta'}$, $k_{i_\beta}\left(z,\theta\right)=\operatorname{pr}_{1,q}\left(z\right)\pi_{i-1}\left(z,\theta\right)$, $k_{i_{\theta+}}\left(z,\theta\right)=\operatorname{pr}_{q+1,p+q}\left(z\right)\pi_{i-1}\left(z,\theta\right)$, and $k_{i_{w^*}}\left(z,\theta\right)$ is the matrix containing the elements of $\operatorname{pr}_{p+q+1,q^2}\left(z\right)\pi_{i-1}\left(z,\theta\right)\ k_{i_w}\left(z,\theta\right)$ ordered appropriately and $k_{1_\theta}=\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}$, where $\operatorname{pr}_{i,j}\left(x\right)$ denotes the transformation of an r^{th} dimensional vector, say $x=\left(x_1,x_2,...,x_r\right)'$, to a vector containing only the elements of x from the i^{th} to the j^{th} coordinate, i.e. $\operatorname{pr}_{i,j}\left(x\right)=\left(x_i,x_{i+1},...,x_j\right)'$, where naturally $1\leq i\leq j\leq r$.

[5] The Edgeworth measures described in lemmas 4.6 and 4.7 need not coincide. However the validity of the lemmas forces the distance between their evaluations on the same Borel set to be $o\left(n^{-a^*}\right)$ uniformly over the Borel sets on $\mathbb{R}^p$.

4.4 Valid 2nd order Bias approximation for the $\mathrm{GMR2}\,(a_*)$

We are ready to provide the results for the second order bias approximation for the $\mathrm{GMR2}\,(a_*)$. Notice that due to its form, the results in Newey and Smith [[42]] imply that the bias will depend on the relation between p and q, the non linearities of the relevant estimating vectors and the stochastic weighting.

We obtain the following lemma.[6]

Lemma 4.9 Under assumptions A.1, A.4, A.2, A.3 and A.5, If $a^* \geq 1$, then for any $\theta \in \operatorname{Int}\Theta$

$$\left\| E_\theta \sqrt{n}\left(\theta_n\left(a_*\right)-\theta\right)-\frac{\xi\left(\theta\right)}{\sqrt{n}}\right\| = o\left(n^{-\frac{1}{2}}\right)$$

where

$$\begin{aligned}\xi\left(\theta\right) &= \mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\mathcal{I}_V\left(k_{2_\beta}\right)-\mathcal{C}^{-1}\mathcal{I}_V\left(\left[k'_{1_\theta}\frac{\partial^2 b_j}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,l}W\frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\right)\\ &-\frac{1}{2}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\mathcal{I}_V\left(\left[\frac{\partial b'}{\partial\theta}k'_{1_\beta}W\mathcal{C}^{-1}\frac{\partial b_j}{\partial\theta\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\right]_{j=1,\ldots,q}\right)\\ &-\mathcal{C}^{-1}\mathcal{I}_V\left(\left(\frac{\partial b'}{\partial\theta}k_{1_w}+\left[\frac{\partial}{\partial\theta'}W_{j,j'}k_{1_{\theta+}}\right]_{j,j'=1,\ldots,l}\right)\frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\right)\end{aligned}$$

if $a_* = 0$ and

$$\begin{aligned}\xi\left(\theta\right) &= -\mathcal{C}^{-1}\mathcal{I}_V\left(\left[k'_{1_\theta}\frac{\partial^2 b_j}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,l}W\frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\right)\\ &-\frac{1}{2}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\mathcal{I}_V\left(\left[\frac{\partial b'}{\partial\theta}k'_{1_\beta}W\mathcal{C}^{-1}\frac{\partial b_j}{\partial\theta\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\right]_{j=1,\ldots,q}\right)\\ &-\mathcal{C}^{-1}\mathcal{I}_V\left(\left(\frac{\partial b'}{\partial\theta}k_{1_w}+\left[\frac{\partial}{\partial\theta'}W_{j,j'}k_{1_{\theta+}}\right]_{j,j'=1,\ldots,l}\right)\frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\right)\end{aligned}$$

if $a_* > 0$.

[6] 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 [41] (see the second paragraph immediately after Theorem 2).

The following corollary is trivial.

Corollary 1 When W is independent of x and θ and $b(\theta)$ is affine then for any $\theta \in \operatorname{Int}\Theta$

$$\xi(\theta) = \begin{cases} \mathcal{C}^{-1}\frac{\partial b'}{\partial \theta} W \mathcal{I}_V\left(k_{2_\beta}\right) \text{ if } a_* = 0 \\ 0_p \text{ if } a_* > 0 \end{cases}$$

Thus when W is independent of x and θ and $b(\theta)$ is affine the GMR2(a) is second order unbiased when $a > 0$ and our assumption framework is valid. This result holds as well for the GMR2 estimator (see Arvanitis and Demos [4]-Corollary 2) under a more restrictive assumption framework that validates the analogous Edgeworth approximations. The framework of non stochastic weighting and of affinity of the lbf is the most general known for this kind of results to hold.

MA(1)**,** GARCH(1,1) **Examples Cont.** From the above corollary, and for the MA(1) example, it is obvious $E_\theta \theta_n(1) = E_\theta \theta_n^*(1) = \theta + o(n^{-1})$. Furthermore, for the GARCH(1,1) one, we have that $E_\theta \theta_n\left(\frac{1}{2}\right) = E_\theta \theta_n^*\left(\frac{1}{2}\right) = \theta + o(n^{-1})$.

4.5 MSE 2^{nd} order Approximations for the GMR2(a_*)

Given the results of the previous subsection the question arising concerns the comparison between the analogous MSE approximations of the GMR2(a_*) for any a_*. We obtain the following lemmas.

Lemma 4.10 If $W(x,\theta)$ is independent of x and θ, b is affine, assumptions A.1, A.4, A.2, A.3 and A.5 hold and $a^* \geq \frac{3}{2}$ then, for any $\theta \in \operatorname{Int}\Theta$

$$\left\| E_\theta\left(n\left(\theta_n(a_*) - \theta\right)\left(\theta_n(a_*) - \theta\right)'\right) - H_1(\theta) - \frac{H_2(\theta)}{\sqrt{n}} \right\| = o\left(n^{-1/2}\right)$$

where

$$\begin{aligned} H_1(\theta) &= \mathcal{C}^{-1}\frac{\partial b'}{\partial \theta} W V(\theta) W \frac{\partial b}{\partial \theta'} \mathcal{C}^{-1} \\ H_2(\theta) &= \mathcal{C}^{-1}\frac{\partial b'}{\partial \theta} W \mathcal{I}_V\left(k_{2_\beta} k'_{1_\beta}\right) W \frac{\partial b}{\partial \theta'} \mathcal{C}^{-1} \end{aligned}$$

for all a_*.

This along with corollary 4.9 establishes the second order equivalence of the GMR2(a_*) estimators when $a_* > 0$ and their superiority to the GMR1 estimator.[7]

[7] The same result holds the GMR2 estimator (see Arvanitis and Demos [4]-Lemma 3.6)

MA(1) **Example Cont.** Recall that $\theta_n(1)$ is the GMR2 (1) with auxiliary estimator the GMR1 $= \theta_n(0)$ one. Now from Demos and Kyriakopoulou [14] we have that $\mathcal{I}_V\left(k_{2_{\mathrm{GMR1}}}k'_{1_{\mathrm{GMR1}}}\right) = 0$ and the binding function is the identity, i.e. $b(\theta) = \theta$, we have by the above lemma that

$$E\left[\sqrt{n}\left(\theta_n(1) - \theta\right)\right]^2 = E\left[\sqrt{n}\left(\theta_n(0) - \theta\right)\right]^2 = \frac{1 + \theta^2 + 4\theta^4 + \theta^6 + \theta^8}{\left(1 - \theta^2\right)^2} + o\left(n^{-\frac{1}{2}}\right)$$

which is the asymptotic variance of $\theta_n(0)$ (see Fuller [20]).

GARCH(1, 1) **Example Cont.** Again, as the binding function is the identity and applying once more the lemma above we have that the $\theta_n\left(\frac{1}{2}\right)$ has the same asymptotic variance as the GMR1 one, up to $o\left(n^{-\frac{1}{2}}\right)$, which is equal to β_n, the MLE. Hence,

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

Due to the structure of the mean approximations the approximate MSE of the unbiased GMR2 (a_*) presented in this paper, even when these are derived via the use of Monte Carlo and/or Bootstrap sampling techniques do not depend on the number of these samples. This is not obviously the case with other simulation based bias correctors whence the corresponding MSEs are inflated by a factor depending on this number. Under suitable (and stronger than the ones employed here) conditions the latter perform the analogous correction even with minimal number of simulated samples at the expense of large MSE. This can be reduced when this number is augmented at the expense of a large numerical cost. This trade off is not obviously faced by the GMR2 (a_*) estimators under our assumption framework.

5 Recursive GMR2(a_*) Estimators

The previous results imply that the second order asymptotic properties of the GMR2 (a_*) depend among others on the local behavior of the lbf. Due to assumption A.1.2, and theorem 10.2 of Spivak [49] (p. 44) since $p \leq q$, B **can always be chosen** so that $b(\theta)$ is of the form $\left(\theta', 0'_{q-p}\right)'$ at least in a small enough neighborhood of θ. This along with non stochastic weighting and corollary 1 imply that there always exists an auxiliary parametrization such that the the GMR2 (a_*) estimators for $a_* > 0$ are second order unbiased. Usually, this reparametrization is analytically intractable.

Notice though, that there exists at least one indirect estimation procedure that can be employed in order to approximate this "canonical" parameterization. Given the GMR1, let $\beta_n = \left(\text{GMR1}', 0'_{q-p}\right)'$. Given the validity of our assumption framework, lemma 4.9 implies the validation of assumption **??** for β'_n. Notice that the corresponding terms of the new approximation would be polynomial functions of the analogous terms of the original approximation. Hence new $\zeta_{i_n}(\theta)$ can be defined by compositions of the old ones with these functions. Hence the needed assumptions will still hold. For a new compact parameter space $\Theta' \subset \operatorname{Int}\Theta$ apply the $\text{GMR2}(a_*)$ estimator to the latter. Then the resulting indirect estimator is derived from a three-step procedure, in the last step of which the binding function is obviously $\left(\theta', 0'_{q-p}\right)'$. An extension of the three step procedure of the previous remark to an arbitrary number of steps, where the i^{th}-step auxiliary estimator is the $\text{GMR2}(a_*)$ of the previous step embedded to $\mathbb{R}^q$, can provide an unbiased indirect estimator of arbitrary order. This extension is the object of study of the present section. Obviously, the embedding of the auxiliary estimator in any step after the first to $\mathbb{R}^q$ is irrelevant and therefore will be dropped.

We define recursive $\text{GMR2}(a_*)$ estimators as follows.

Definition D.3 Let assumptions A.2, A.3 and A.5 hold for some $a = a_*$, $a^* \geq a + \frac{1}{2}$. For $\kappa \in \mathbb{N}^*$, the recursive $\kappa - \text{GMR2}(a)$ estimator (denoted by $\theta_n^{(\kappa)}(a)$) is defined by the following steps:

1. $\Theta^{(0)} = \Theta$ and $\theta_n^{(0)} = \text{GMR2}(a)$,
2. For $1 \leq j \leq \kappa$, $\Theta^{(j)}$ a compact subset of $\operatorname{Int}\Theta^{(j-1)}$

$$\theta_n^{(j)}(a) = \arg\min_{\theta \in \Theta^{(j)}} \left\| \theta_n^{(j-1)}(a) - \theta - \sum_{i=j+1}^{2a+1} \frac{1}{n^{i/2}} \zeta_{i_n}^{(j)}(\theta) \right\|$$

where $\zeta_{i_n}^{(j)}(\theta)$ is the polynomial transformation of $\zeta_{i_n}^{(j-1)}(\theta)$ implied by the form of $\xi_i(\theta)$ for $\theta_n^{(j)}(a)$ as a function of $\xi_i(\theta)$ for $\theta_n^{(j-1)}(a)$.

We are now able to prove the following lemma.[8]

Lemma 5.1 For any $\theta \in \operatorname{Int}\Theta^{(\kappa)}$ and $\kappa \leq 2a+1$, $\left\| E_\theta\left(\sqrt{n}\left(\theta_n^{(\kappa)}(a) - \theta\right)\right) \right\| = o\left(n^{-\frac{\kappa}{2}}\right)$ hence it is unbiased of order $\kappa + 1$ and has the same approximate MSE with the $\theta_n^{(\kappa-1)}(a)$, up to the same order.

Let us now turn our attention to a further example along with some Monte Carlo experiments.

[8] Notice that both definition D.3 and lemma 5.1 would be valid, if $\theta_n^{(0)}$ was any consistent estimator.

6 Further Examples and Monte Carlo

In this section we employ Monte Carlo experiments for our two examples in order to assess the relevance of our results for finite n. We also present a third example involving $\text{GMR2}\,(a_*)$ in the context of linear, strongly dependent, Gaussian processes, and engage to an analogous Monte Carlo experiment.

Monte Carlo for MA(1)

For the MA (1) process, we draw a random sample of $n = 50$, 100, 150, 250, 500, 750, 1000, 1500 and 3000 observations from a non-central Student's-t distribution with non-centrality parameter equal to 1 and 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 three estimators, i.e. the $\theta_n\,(0)$ or also called GMR1, the commonly biased corrected $\theta_n^*\,(1)$ or also $\text{GMR2}^*\,(1)$, and the $\theta_n\,(1)$ or also $\text{GMR2}\,(1)$. Out of these estimators only the $\text{GMR2}^*\,(1)$ and $\text{GMR2}\,(1)$ ones are 2^{nd} order unbiased. We set the number of replications to 100000. To conserve space, we present the results only for $\theta = -0.4$. The results for $\theta = 0.4$ are qualitatively the same. In a few cases the commonly bias corrected estimator, $\theta_n^*\,(1)$, turned to be outside the interval $[-0.9999, 0.9999]$ (see remark R.8). In fact even when $n = 250$ we noticed 9, out of 100000, where $\theta_n^*\,(1) \geq 1$. In these cases we set $\theta_n^*\,(1) = 0.9999$.

In figure 1 we present the absolute biases of the three estimators. These are multiplied by n, so that the second order bias of the GMR1 equals to 1.252, for all n. Furthermore, the depicted bias of the $\text{GMR2}\,(1)$ is less than the one of the $\text{GMR2}^*\,(1)$, something which more pronounced for $n \leq 500$. This is can be attributed to the fact that in a few cases $\theta_n^*\,(1)$ had to be restricted to the value of 0.9999 (see previous paragraph). The same explanation goes for the behavior of the approximate MSEs, presented in figure 2. All estimators reach their common asymptotic variance, which is 1.796, at least for $n > 250$. However, for $n < 250$, the most variable estimator is the $\text{GMR2}^*\,(1)$.

Monte Carlo for GARCH$(1,1)$

For the $\text{GARCH}\,(1,1)$, we draw a random sample of $n = 150$, 250, 400, 550, 750, 900, 1000, 1500, 2000, 3000, 5000 and 10000 observations, plus

250 for initialization, from a standard normal distribution. We perform 3000 replications. For each random sample, we generate the GARCH$(1,1)$ process y_t with $\theta_1 = 1.0$, $\theta_2 = 0.2$ and $\theta_3 = 0.7$, and we find the MLE of $\theta_i' s$, which is our auxiliary vector estimator β_n. As the auxiliary and true model coincide, the binding function is the identity one. Consequently, the GMR1 and the auxiliary estimators, β_n, coincide. We further consider the feasibly bias corrected estimator, suggested in Linton [37] and Iglesias and Linton [31], which is our $\theta_n^*\left(\frac{1}{2}\right) = \beta_n + \frac{1}{n}\left(\lambda_0\left(\beta_n\right) + \lambda_2\left(\beta_n\right)\right)$, also named GMR2$^*\left(\frac{1}{2}\right)$. The third estimator we employ is the $\theta_n\left(\frac{1}{2}\right)$, also named GMR2$\,(1)$. To evaluate the analytically unknown λ_0 and λ_2, needed for the valuation of $\theta_n^*\left(\frac{1}{2}\right)$ and $\theta_n\left(\frac{1}{2}\right)$, we employ 150 samples of 400 random numbers coming from a standard normal distribution (see Linton [31] for details).

As in the previous example, in a few cases the feasibly corrected estimator, GMR2$^*\,(1)$, turned out to be outside the admissible region, i.e. $\theta_{n,2}^*\left(\frac{1}{2}\right)$ could be non-positive or $\theta_{n,2}^*\left(\frac{1}{2}\right) + \theta_{n,3}^*\left(\frac{1}{2}\right)$ could be greater than 1. In these cases we adjust the estimators accordingly, i.e. we set $\theta_{n,2}^*\left(\frac{1}{2}\right)$ to a small positive number etc.

In figure 3 we present the norm of the biases of the three estimators multiplied by n, i.e. $n\left\|E\left(\widehat{\theta}-\theta\right)\right\|$ where $\widehat{\theta}$ any of the three estimators. For all n the $\theta_n\left(\frac{1}{2}\right)$ is less biased than $\theta_n^*\left(\frac{1}{2}\right)$, with the exemption of $n = 2500$, and both of them have almost half the bias of GMR1 (the MLE). Furthermore, the approximate MSEs of the estimator are presented in figure 4. It seems that 1500 observations are enough for the estimators to reach their common asymptotic variance.

Let us now turn our attention to another popular process in economic applications, this of the fractional Gaussian noise.

Example: Gaussian ARFIMA Process

In this section we demonstrate how the suggested estimators can be applied to a simple, but popular, model of the ARFIMA class. Let us consider the fractional Gaussian process, i.e. the ARFIMA$\,(0,d,0)$, given by:

$$(1-L)^d y_t = u_t, \qquad t = ..., -1, 0, 1, ..., \quad 0 < d < \frac{1}{2}, \quad u_t \overset{iid}{\backsim} N(0,1).$$

The estimation of d is presented in Sowell [48] (see also Doornik and Ooms [15]), for the time domain. The asymptotic distribution of the estimator is introduced in Dahlhaus [13], whereas the validity of the Edgeworth expansion of this estimator is established in Lieberman, Rousseau and Zucker [36] with analytic formulae presented in Lieberman and Phillips [35].

Let d_n denote the MLE of d, i.e. d_n maximizes

$$l = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\det(G) - \frac{1}{2}\mathbf{y}'G^{-1}\mathbf{y}$$

on a compact subset of $\left(0, \frac{1}{2}\right)$ that includes d in its interior, where

$$G = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-2) & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \ddots & \ddots & \gamma(n-2) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \gamma(n-2) & \ddots & \ddots & \ddots & \gamma(1) \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(1) & \gamma(0) \end{pmatrix}$$

and

$$\gamma(0) = \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}, \qquad \gamma(h) = \gamma(0)\frac{\Gamma(h+d)\Gamma(1-d)}{\Gamma(1+h-d)\Gamma(d)}.$$

It is known that the first order asymptotic distribution of $\sqrt{n}(d_n - d)$ is $N\left(0, \frac{6}{\pi^2}\right)$ (see Dalhaus [13] and Yajima [53]) and has similar asymptotic properties to the Whittle (Fox and Taqqu [19], and Giraitis and Surgailis [24]) but superior to the semiparametric one (Geweke and Porter-Hudak [22], and Robinson [44]), given correct model specification. However, there is evidence that the bias of d_n can be severe (see Cheung and Diebold [9], Smith, Sowell and Zin [47], Hauser [30], Lieberman [34], and Doornik and Ooms [15]).

From the formulae in Lieberman and Phillips [35] we have that

$$E\left[\sqrt{n}(d_n - d)\right] = \frac{1}{\sqrt{n}k_n}\left(C_{1,n} + 3\frac{C_{3,n}}{k_n}\right) + o\left(n^{-\frac{1}{2}}\right)$$

where

$$k_n = \frac{1}{2n} tr\left(G^{-1}\dot{G}\right)^2, \quad C_{1,n} = -\frac{tr\left(\left(G^{-1}\dot{G}\right)^3 - G^{-1}\dot{G}G^{-1}\ddot{G}\right)}{tr\left(G^{-1}\dot{G}\right)^2}, \quad and$$

$$C_{3,n} = \frac{1}{12n} tr\left(2\left(G^{-1}\dot{G}\right)^3 - 3G^{-1}\dot{G}G^{-1}\ddot{G}\right)$$

Due to normality and the relevant differentiability of the Gamma function, we have the Edgeworth expansions of d_n are valid for any order of s^* (see Lieberman et al. [36]) and our assumption framework is satisfied. Further, as the auxiliary and true models are the same, the binding function is the inclusion and consequently, $d_n = \text{GMR1}$. It follows that

$$E\left(\text{GMR1} - d\right) = -\frac{tr\left(G^{-1}\dot{G}G^{-1}\ddot{G}\right)}{\left[tr\left(G^{-1}\dot{G}\right)^2\right]^2} + o\left(n^{-1}\right)$$

where $\dot{G}$ and $\ddot{G}$ the first and second derivative of G.

Consequently $d_n\left(\frac{1}{2}\right) = \text{GMR2}\left(\frac{1}{2}\right)$ is defined as

$$d_n\left(\frac{1}{2}\right) = \text{GMR2}\left(\frac{1}{2}\right) = \arg\min_d \left(\text{GMR1} - d + \frac{tr\left(G^{-1}\dot{G}G^{-1}\ddot{G}\right)}{\left[tr\left(G^{-1}\dot{G}\right)^2\right]^2}\right).$$

and by corollary 1 it is 2^{nd} order unbiased and has the same MSE, up to $O\left(n^{-\frac{1}{2}}\right)$, with d_n, (see lemma 4.10).

Furthermore, let us call d_n^* the approximate bias corrected estimator of Lieberman [34], adapted for this case from Firth [18]. d_n^* is given by:

$$\frac{\partial l}{\partial d} + \frac{18\zeta(3)}{\pi^2} = -\frac{1}{2}tr\left(G^{-1}\dot{G}\right) + \frac{1}{2}\mathbf{y}'G^{-1}\dot{G}G^{-1}\mathbf{y} + \frac{18\zeta(3)}{\pi^2} = 0$$

where $\zeta(.)$ is the Riemann zeta function and $\frac{18\zeta(3)}{\pi^2} \simeq 2.1923$. In fact, if we consider the score ,$\frac{\partial l}{\partial d}$, as our auxiliary estimator, then its approximate bias is given by $\frac{18\zeta(3)}{\pi^2}$. Hence the Lieberman [34] estimator, d_n^*, is nothing else but our $d_n^*\left(\frac{1}{2}\right) = \text{GMR2}^*\left(\frac{1}{2}\right)$.

Monte Carlo In figure 5 we present the $n \times |biases|$, of the three estimators, for $n = 20, 40, 50, 60, 70, 80, 90, 100$ and 120, and $d = 0.4$.[9] All estimators were calculated by a simple grid, of length 0.001, search on the interval $[-0.499, 0.499]$. The interval $[-0.499, 0]$ is included to avoid a pile-up at the origin (see Lieberman and Phillips [35]). 10,000 replications were performed. It is obvious that $d_n\left(\frac{1}{2}\right)$ is less biased than $d_n^*\left(\frac{1}{2}\right)$ which in turn is less so than d_n for all examined sample sizes. However, the approximate MSE of

[9] In fact, Monte Carlo exercises where performed for $d = \{0.1, 0.2, 0.3, 0.4\}$. We present only the results of $d = 0.4$ to conserve space.

$d_n^*\left(\frac{1}{2}\right)$ is smaller as compared to the one of $d_n\left(\frac{1}{2}\right)$, which in turn is smaller than the one of GMR1 (see figure 6). The results are consistent with those for $d_n^*\left(\frac{1}{2}\right)$ and d_n in Lieberman [34]. Further the same results appear for the rest of the values of d, i.e. for $d = \{0.1, 0.2, 0.3\}$. The only exception is when $d = 0.2$, where the bias of $d_n^*\left(\frac{1}{2}\right)$ appears to be smaller for $40 \leq n \leq 100$.

Notice that employing the results in Giraitis and Robinson [23], the same procedure can be applied to the semiparametric Whittle estimator.

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 can be also true in cases where the analytical form of the approximation is unknown and is in turn numerically approximated. It is mostly useful in cases where the lbf is the identity (or the inclusion function in our framework). This is true when the auxiliary is itself a consistent estimator of the parameters under interest.

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, some of their higher order asymptotic properties coincide with those of the GMR2 estimator, yet it *seems* that they can be derived with respect to more general sets of assumptions. For example, the rates of convergence of the derivatives of $E_\theta \beta_n$ needed for GMR2 can be established by locally uniform Edgeworth approximations along with conditions resembling our assumption framework. Obviously our results indicate that this need not be the case for the GMR2 (a_*) estimators.

We confirm that under our assumption framework and in the special case of deterministic weighting and affinity of the binding function, the GMR2 (a_*)

estimator for any $a_* > 0$, is second order unbiased. Furthermore, by generalizing to multistep procedures we are able to provide recursive indirect estimators that are unbiased at any given order when analogous conditions hold. Moreover, the approximate MSE of the unbiased GMR2 (a_*) presented in this paper, even when these are derived via the use of Monte Carlo and/or Bootstrap sampling techniques do not depend on the number of these samples. At the same time their practical implementation does not seem as numerically involved as an analogous procedure defining recursive GMR2 estimators due to reasons stated above.

The definition and the derived properties of the GMR2 (a_*) estimators, provide with a general framework for some well known "approximate bias correction" procedures. Notice that it is already established by Gourieroux, Renault and Touzi [28] that the GMR2 when derived as the solution of $\theta_n - E_\theta \theta_n = \mathbf{0}$ has as a first step of an sequential approximation $\theta_n - E_{\theta_n} \theta_n$ which can, under relevant conditions, be approximated by a bootstrap procedure. Hence they interpret the bootstrap estimator as an one step numerical approximation of the indirect one with equivalent second order asymptotic properties.

In a direct analogy, when the previous framework is considered, some of 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 $E_\theta \theta_n$. Despite the previous remark the analogous first step approximation, i.e. $\theta_n - K(\theta_n, a)$ lies also in the premises of our definition, as an extreme case, hence can be characterized as an IE. The definition allows for *intermediate cases* in which some of the elements of the approximation are evaluated in θ_n. Despite the minimal numerical cost with which $\theta_n - K(\theta_n, a)$ is associated, it can be outperformed by the other GMR2 (a_*) estimators for finite n due to its behavior in the boundary of the parameter space. We leave for future work the question of the definition of GMR2 (a_*) estimators as well as the derivation of higher order asymptotic properties of any IE discussed when θ and/or $b(\theta)$ lies in the boundary of the relevant parameter space.[10]

[10] This could constitute a case under which $\theta_n - K(\theta_n, a)$ properly defined, can be characterized by inferior asymptotic properties compared to the other GMR2 (a_*) estimators.

References

[1] Aliprantis, C.D. & 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. & O. 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. & A. Demos (2010), "Stochastic Expansions and Moment Approximations for Three Indirect Estimators", IEES Working Paper No 1004, AUEB.

[5] Bao, Y. & A. Ullah (2007), "The second-order bias and mean squared error of estimators in time-series models" *Journal of Econometrics* 140, 650–669.

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

[7] Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics* 31, 307-327.

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

[9] Cheung, Y.W. & F.X. Diebold (1994), "On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean", *Journal of Econometrics* 62, 301-16.

[10] Cordeiro G.M. & R. Klein (1994) "Bias correction in ARMA models", *Statistics and Probability Letters* 19, 169-176.

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

[12] Cox, D.R. & D.V. Hinkley (1974) "*Theoretical Statistics*" Chapman and Hall.

[13] Dahlhaus, R. (1989), "Efficient parameter estimation for self-similar processes" The Annals of Statistics 17, 1749-1766.

[14] Demos, A. & Kyriakopoulou D. (2011), "Edgeworth and Moment Approximations: The Case of MM and QML Estimators for the MA(1) Models", Communications in Statistics-Theory and Methods (forthcoming), IEES Working Paper No 1003, AUEB.

[15] Doornik, J.A. & M. Ooms (2003) "Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models", *Computational Statistics and Data Analysis* 42, 333-348.

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

[17] Fernandez-Val, I. & F. Vella (2011) "Bias corrections for two-step fixed effects panel data estimators", *Journal of Econometrics* 163, 144-162.

[18] Firth, D. (1993) "Bias reduction of maximum likelihood estimates", *Biometrika* 80, 27-38.

[19] Fox, R. & M.S. Taqqu (1986), "Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series", *Annals of Statistics* 14, 517-532.

[20] Fuller, W.A. (1976) "*Introduction to Statistical Time Series*" J. Wiley & Sons, New York.

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

[22] Geweke, J. & S. Porter-Hudak (1983), "The estimation and application of long-memory time series models" Journal of Time Series Analysis 4, 221-238.

[23] Giraitis, L. & P.M. Robinson (2003), "Edgeworth Expansions for Semiparametric Whittle Estimation of Long Memory", *The Annals of Statistics*, Vol. 31, 1325-1375.

[24] Giraitis, L. & D. Surgailis (1990), "A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle's estimate" *Probability Theory and Related Fields* 86, 87-104.

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

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

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

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

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

[30] Hauser, M.A. (1999), "Maximum likelihood estimators for ARMA and ARFIMA models: A Monte Carlo study", *Journal of Statistical Plannning and Infererence* 80, 229-55.

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

[32] Iglesias, E.M. & 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.

[33] Lee, S.W. & B.E. Hansen (1994) "Asymptotic Theory for the Garch (1,1) Quasi-Maximum Likelihood Estimator", *Econometric Theory* 10, 29-52.

[34] Lieberman, O. (2001) "Penalised Maximum Likelihood Estimation for Fractional Gaussian Processes", *Biometrika* 88, 888-894.

[35] Lieberman, O. & P.C.B. Phillips (2004) "Expansions for the Distribution of the Maximum Likelihood Estimator of the Fractional Difference Parameter", *Econometric Theory* 20, 464-484.

[36] Lieberman O., J. Rousseau & D.M. Zucker (2003), "Valid Asymptotic Expansions for the Maximum Likelihood Estimator of the Parameter of a Stationary, Gaussian, Strongly Dependent Process", *The Annals of Statistics* 31, 586-612.

[37] Linton, O. (1997) "An Asymptotic Expansion in the Garch(1,1) Model", *Econometric Theory* 13, 558-581.

[38] Lumsdaine, R.L. (1996) "Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models", *Econometrica* 64, 575-596.

[39] Lumsdaine, R.L. (1995) "Finite-Sample Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models: A Monte Carlo Investigation", *Journal of Business & Economic Statistics* 13, 1-10.

[40] MacKinnon, J.G. & A.A. Smith Jr. (1998) "Approximate bias correction in econometrics" *Journal of Econometrics* 85, 205-230.

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

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

[43] Rilstone, P., V.K. Srivastava & A. Ullah (1996) "The second-order bias and mean squared error of nonlinear estimators", *Journal of Econometrics* 75, 369-395.

[44] Robinson, P.M. (1995), "Log-periodogram regression of time series with long range dependence", The Annals of Statistics 23, 1048-1072.

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

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

[47] Smith, A.A., F.B. Sowell & S.E. Zin (1997). Fractional integration with drift: Estimation in small samples. Emnpirical Econi. 22, 103-16.

[48] Sowell, F. (1992) "Maximum likelihood estimation of stationary univariate fractionally integrated time series models" *Journal of Econometrics* 53, 165-188.

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

[50] Tanizaki, H. (2000) "Bias correction of OLSE in the regression model with lagged dependent variables" *Computational Statistics and Data Analysis* 34, 495-511.

[51] Tang, C.Y. & S.X. Chen (2009) "Parameter estimation and bias correction for diffusion processes" *Journal of Econometrics* 149, 65-81.

[52] Wang, X., P.C.B. Phillips & J. Yu (2011) "Bias in estimating multivariate and univariate diffusions", *Journal of Econometrics* 161, 228-245.

[53] Yajima, Y. (1985), "On estimation of long-memory time series models", *Australian Journal of Statistics* 27, 303-320.

Appendices

A Proofs of Lemmas and Corollaries

Proof of Lemma 2.1. Due to assumption A.5.1 θ_n^+ lies in $\overline{\mathcal{O}}_\varepsilon(\theta)$ with P_θ-probability $1-o\left(n^{-a^*}\right)$. Then a Taylor expansion of order s^*+1 of $W_n\left(\theta_n^+\right)$ around θ along with assumptions A.4, A.5.1, implies that for any $\delta>0$, exist $\delta_i>0$, $i=0,\ldots,d$ such that

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

■

Proof of Lemma 2.2. First notice that due to A.5 the estimator lies in $\overline{\mathcal{O}}_\varepsilon(\theta)$ with Q_θ-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 along with A.1.2 and A.5.1 implies that $Q_\theta\left(\sqrt{n}\left\|\theta_n(0)-\theta\right\|>C\sqrt{\ln n}\right)=o\left(n^{-a^*}\right)$, for some $C>0$. A Taylor expansion of order d of the first order conditions implies that with P_θ-probability $1-o\left(n^{-a^*}\right)$ $\sqrt{n}\left(\theta_n(0)-\theta\right)=L\sqrt{n}m_n^*(\theta)+\frac{1}{\sqrt{n}}\rho_n\left(\sqrt{n}m_n^*(\theta)\right)+R_n$ where L is an $p\times\dim\left(m_n(\theta)\right)$ matrix of rank p due to A.1, A.4, ρ_n is a polynomial function with absolutely bounded coefficients due to A.1, and $Q_\theta\left(\left\|R_n\right\|>\gamma_n\right)=o\left(n^{-a^*}\right)$ for some $\gamma_n=o\left(n^{-a^*}\right)$ that might depend on θ, due to $Q_\theta\left(\sqrt{n}\left\|\theta_n(0)-\theta\right\|>C\sqrt{\ln n}\right)=o\left(n^{-a^*}\right)$. Hence from lemma AL.1 the result would follow if $L\sqrt{n}m_n^*(\theta)+\frac{1}{\sqrt{n}}\rho_n\left(\sqrt{n}m_n^*(\theta)\right)$ has a valid Edgeworth expansion of the respective order. This is established by lemma 3 of Magdalinos [41] and assumption A.5.1. ■

Proof of Lemma 4.1. Notice that due to the triangle inequality and submultiplicativity

$$\begin{aligned} & Q_\theta\left(\sup_{\theta'\in\Theta}\left|\left\|\beta_n-b_n\left(\theta',\zeta_n\left(\theta',a_*\right)\right)\right\|_{W_n\left(\theta_n^+\right)}-\left\|b(\theta)-b\left(\theta'\right)\right\|_{W(\theta)}\right|>\varepsilon\right) \\ \leq\ & Q_\theta\left(\sup_{\theta'\in\Theta}\left\|\beta_n-b_n\left(\theta',\zeta_n\left(\theta',a_*\right)\right)\right\|\left\|W_n\left(\theta_n^+\right)-W(\theta)\right\|>\frac{\varepsilon}{2}\right) \\ & +Q_\theta\left(\left\|\beta_n-b(\theta)\right\|_{W(\theta)}>\frac{\varepsilon}{4}\right)+Q_\theta\left(\sup_{\theta'\in\Theta}\left\|\sum_{i=2}^{s_*}\frac{1}{n^{i/2}}\zeta_{i_n}\left(\theta'\right)\right\|_{W(\theta)}>\frac{\varepsilon}{4}\right) \end{aligned}$$

and that due A.1, A.3.1, A.5.1, and lemma 2.1 all the probabilities in the second part of this display are $o\left(n^{-a^*}\right)$ for any $\theta\in\operatorname{Int}\Theta$. The result follows by assumption A.1. ■

Proof of Lemma 4.2. First notice that for any $\theta \in \operatorname{Int}\Theta$

$$\begin{aligned} & Q_\theta\left(\sup_{\theta' \in \Theta} \left\|b_n(\theta') - b_n(\theta', \zeta_n(\theta', a))\right\| > \varepsilon\right) \\ \leq \; & P_\theta\left(\sup_{\theta' \in \Theta} \left\|b_n(\theta') - b(\theta')\right\| > \frac{\varepsilon}{2}\right) + \sum_{i=1}^{s_*} Q_\theta\left(\sup_{\theta' \in \Theta} \left\|\zeta_{i_n}(\theta)\right\| > o\left(n^{(i-1)/2-\delta}\right)\right) \end{aligned}$$

which due to A.1, A.3.1-2, A.5 and the hypothesis that $\sup_{\theta \in \Theta} \|b_n(\theta) - b(\theta)\| = o(1)$ are $o\left(n^{-a^*}\right)$ for any $\theta \in \operatorname{Int}\Theta$. From the definition of the two estimators we obtain that

$$\begin{aligned} & \left\|\beta_n - E_{\theta_n(a)}\beta_n\right\|_{W_n\left(\theta_n^+\right)} - \left\|\beta_n - E_{\mathrm{GMR2}}\beta_n\right\|_{W_n\left(\theta_n^+\right)} \\ \leq \; & 2\sup_{\theta' \in \Theta}\left|\left\|\beta_n - E_{\theta'}\beta_n\right\|_{W_n\left(\theta_n^+\right)} - \left\|\beta_n - b(\theta') + \sum_{i=2}^{s_*}\frac{1}{n^{i/2}}\zeta_{i_n}(\theta')\right\|_{W_n\left(\theta_n^+\right)}\right| = \eta_n \end{aligned}$$

due to the fact that for any $\theta \in \operatorname{Int}\Theta$

$$\begin{aligned} & Q_\theta(\eta_n > \varepsilon) \\ \leq \; & 2Q_\theta\left(\sup_{\theta' \in \Theta}\left\|b_n(\theta') - b_n(\theta', \zeta_n(\theta', a_*))\right\| > \varepsilon_*\right) + Q_\theta\left(\left\|W_n\left(\theta_n^+\right) - W(\theta)\right\| > K\right) \end{aligned}$$

for $K > 0$ and $\varepsilon_* = \frac{\varepsilon}{2}\min\left(\frac{1}{\sqrt{\|W^*(\theta)\|}}, \frac{1}{\sqrt{K}}\right)$ and the result follows from the previous lemma and 2.1. ■

Proof of Lemma 4.3. Likewise to the previous proof set

$$\eta_n = 2\sup_{\theta \in \Theta}\left|\left\|\beta_n - b(\theta') - \sum_{i=2}^{s'_*}\frac{1}{n^{i/2}}\zeta'_{i_n}(\theta')\right\|_{W_n\left(\theta_n^+\right)} - \left\|\beta_n - b(\theta') + \sum_{i=2}^{s_*}\frac{1}{n^{i/2}}\zeta_{i_n}(\theta')\right\|_{W_n\left(\theta_n^+\right)}\right|$$

Then for any $\theta \in \operatorname{Int}\Theta$

$$\begin{aligned} & Q_\theta(\eta_n > \gamma_n) \\ \leq \; & Q_\theta\left(\sup_{\theta' \in \Theta}\left\|\sum_{i=2}^{s'_*}\frac{1}{n^{i/2}}\zeta'_{i_n}(\theta')\right\| + \sup_{\theta' \in \Theta}\left\|\sum_{i=2}^{s'_*}\frac{1}{n^{i/2}}\zeta'_{i_n}(\theta')\right\| > c_*\frac{\gamma_n}{2}\right) \\ & + Q_\theta\left(\left\|W_n\left(\theta_n^+\right) - W(\theta)\right\| > K\right) \end{aligned}$$

for $K > 0$ and $c_* = \frac{1}{2}\min\left(\frac{1}{\sqrt{\|W^*(\theta)\|}}, \frac{1}{\sqrt{K}}\right)$ and the result follows from A.3.1 and 2.1. ■

Proof of Lemma 4.4. Argue as in the previous proof and notice that for any $\theta \in \operatorname{Int}\Theta$

$$Q_\theta\left(\sum_{i=s'_*+1}^{s_*} \frac{1}{n^{i/2}} \sup_{\theta\in\Theta} \left\|\zeta_{i+1_n}(\theta)\right\| > c_* \frac{\gamma_n}{2}\right) = o\left(n^{-a^*}\right)$$

for $\gamma_n \leq \frac{2M}{c_*} \sum_{i=s'_*+1}^{s_*} \frac{1}{n^{i/2}}$. ■

Proof of Lemma 4.5. Due to lemma 4.1 we have that for any $\theta \in \operatorname{Int}\Theta$, $\theta_n(a_*)$ and $\theta_n(0)$ are in $\overline{\mathcal{O}}_\varepsilon(\theta)$ with Q_θ-probability $1-o\left(n^{-a^*}\right)$. Applying the mean value theorem on the gradient of $J_n(\theta) = \left\|\beta_n - b(\theta) - \sum_{i=2}^{s_*} \frac{1}{n^{i/2}} \zeta_{i_n}(\theta)\right\|^2_{W_n^*(\theta_n^+)}$, we get

$$\sqrt{n}\left(\theta_n(0) - \theta_n(a_*)\right) = \left(D_n^2 J\left(\theta_n^{++}\right)\right)^{-1} \sqrt{n} D J_n\left(\theta_n(0)\right)$$

with θ_n^{++} a random element lying in the line segment between $\theta_n(a)$ and $\theta_n(0)$ with P_θ-probability $1-o\left(n^{-a^*}\right)$. It suffices to prove that $Q_\theta\left(\sqrt{n}\left\|\theta_n(0) - \theta_n(a)\right\| > \gamma'_n\right) = o\left(n^{-a^*}\right)$, for some $\gamma'_n = o\left(n^{-\delta}\right)$ whence the choice of η''_n is possible. Due to the norm submultiplicativity we have that

$$\begin{aligned} & Q_\theta\left(\sqrt{n}\left\|\theta_n(0) - \theta_n(a)\right\| > \gamma'_n\right) \\ \leq\ & Q_\theta\left(\left\|\left(D^2 J_n\left(\theta_n^{++}\right)\right)^{-1}\right\| \left\|\sqrt{n} D J_n\left(\theta_n(0)\right)\right\| > \gamma'_n\right) + o\left(n^{-a^*}\right) \end{aligned}$$

Now, from the definition of GMR1, the triangle inequality, norm submultiplicativity, A.1, A.3, A.5, 2.1 and the subsequent 2.1 and 4.1, and by choosing appropriately $C, M > 0$ we obtain

$$\begin{aligned} & Q_\theta\left(\left\|\sqrt{n} D J_n\left(\theta_n(0)\right)\right\| > \rho_n\right) \\ \leq\ & Q_\theta\left(o\left(n^{-\delta}\right) > \frac{\rho_n}{M}\right) + 2Q_\theta\left(\theta_n(0) \in \overline{\mathcal{O}}_\varepsilon(\theta)\right) + Q_\theta\left(\left\|W_n\left(\theta_n^+\right)\right\| > M_W\right) \\ & + \sum_{i=1}^{s_*} Q_\theta\left(\left\|\zeta_{i_n}(\theta)\right\| > o\left(n^{(i-1)/2-\delta}\right)\right) + \sum_{i=1}^{s_*} Q_\theta\left(\left\|D\zeta_{i_n}(\theta)\right\| > o\left(n^{(i-1)/2-\delta}\right)\right) \end{aligned}$$

which is $o\left(n^{-a^*}\right)$ for $\rho_n \leq o\left(n^{-\delta}\right)$ (which might depend on θ). In an analogous manner we can prove that there exists a positive constant C, such that $Q_\theta\left(\left\|\left(D^2 J_n\left(\theta_n^{++}\right)\right)^{-1}\right\| > C\right) = o\left(n^{-a^*}\right)$ and therefore we obtain the needed result if we choose $\gamma'_n \leq C^* \rho_n$. ■

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. First notice that from lemma 4.6 and lemma 2 of Magdalinos [41] we have that for any $\theta \in \operatorname{Int}\Theta$, $Q_\theta\left(\sqrt{n}\left\|\theta_n(a_*)-\theta\right\| > C\ln^{1/2} n\right) = o\left(n^{-a^*}\right)$. A Taylor expansion of order s^* of the first order conditions implies that with Q_θ-probability $1-o\left(n^{-a^*}\right)$ $\sqrt{n}\left(\theta_n(a_*)-\theta\right) = \kappa_n\left(\sqrt{n}m_n^*(\theta)\right)+R_n$ where κ_n is a polynomial function that for which we have that

$$Q_\theta\left(\kappa_n\left(\sqrt{n}m_n^*(\theta)\right)\in A\right) = Q_\theta\left(g_n\left(\sqrt{n}m_n^*(\theta)\right)\in A\right)+o\left(n^{-a^*}\right)$$

and g_n as in equation 4, lemma AL.4, due to assumptions A.3, A.5 while $Q_\theta\left(\|R_n\|>\gamma_n\right) = o\left(n^{-a^*}\right)$ for some $\gamma_n = o\left(n^{-a^*}\right)$ due to the previous. Hence the result follows from the first part of lemma AL.4. ■

Proof of Lemma 4.8. The result follows from lemmas 4.7, AL.2, AL.4, and the fact that Θ is compact. ■

Proof of Lemma 4.9. Lemma 4.8 applies for any $\theta \in \operatorname{Int}\Theta$, 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 when $a_*=0$

$$\begin{aligned}
&\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta} - \frac{1}{\sqrt{n}}\mathcal{C}^{-1}\left[k_{1_\theta}'\frac{\partial^2 b_j}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,l}W\frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\\
&-\frac{1}{2\sqrt{n}}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\left[\frac{\partial b'}{\partial\theta}k_{1_\beta}'W\mathcal{C}^{-1}\frac{\partial b_j}{\partial\theta\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\right]_{j=1,\ldots,q}\\
&-\frac{1}{\sqrt{n}}\mathcal{C}^{-1}\left(\frac{\partial b'}{\partial\theta}k_{1_w}+\left[\frac{\partial}{\partial\theta'}W_{j,j'}k_{1_{\theta+}}\right]_{j,j'=1,\ldots,l}\right)\frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}
\end{aligned}$$

and when $a_*>0$

$$\begin{aligned}
&\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\left(k_{1_\beta}-\frac{\xi_2}{\sqrt{n}}\right) - \frac{1}{\sqrt{n}}\mathcal{C}^{-1}\left[k_{1_\theta}'\frac{\partial^2 b_j}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,l}W\frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\\
&-\frac{1}{2\sqrt{n}}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\left[\frac{\partial b'}{\partial\theta}k_{1_\beta}'W\mathcal{C}^{-1}\frac{\partial b_j}{\partial\theta\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\right]_{j=1,\ldots,q}\\
&-\frac{1}{\sqrt{n}}\mathcal{C}^{-1}\left(\frac{\partial b'}{\partial\theta}k_{1_w}+\left[\frac{\partial}{\partial\theta'}W_{j,j'}k_{1_{\theta+}}\right]_{j,j'=1,\ldots,l}\right)\frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}
\end{aligned}$$

Integrating with respect to $\left(1+\frac{\pi_1(z,\theta)}{\sqrt{n}}\right)\varphi_{V(\theta)}(z)$, using the fact that $k_{i_\beta}(z,\theta) = \mathrm{pr}_{1,q}(z)\,\pi_{i-1}(z,\theta)$ for $i=1,2$ and that due to A.1, A.2, A.5 and AL.2, $\left\|\mathcal{I}_{\varphi_V}\left(k_{2_\beta}\right)-\xi_2\right\| = o(1)$ we obtain the result. ■

Proof of Lemma 4.10. Argue as in the previous proof and note that now $K_{i,j}(x) = x_i x_j$, $i,j=1,\ldots,p$. ■

Proof of Lemma 5.1. First notice that in any step of the procedure the binding function is the inclusion of the current parameter space to the interior of the one in the previous step. Next definition D.3 implies the validity of assumptions A.2, A.3 and A.5 and thereby of lemmas 4.1, 4.6, 4.7 and 4.8, for the given a^* and for any $\theta \in \operatorname{Int}\Theta^{(i)}$ for any $i = 1, \ldots, \kappa$. The proof for the moment approximations for the case $i = 1$ follows easily as a special case of corollary 1 and lemma 4.10. Using induction if these hold for some i, then notice that since ζ_j would be zero for any j less than or equal to i the analogous expansion to the one used in lemmas 4.9 and 4.10 corresponding to the inversion of the the Taylor approximations of order $i+1$ of the first order conditions for $\sqrt{n}\left(\theta_n^{(i+1)} - \theta\right)$ is

$$k_{1_{\theta(i}} - \frac{\xi_{i+1}^{(i)}}{n^{\frac{i+1}{2}}} - \frac{Z_{1,i+1}^{(i)}\left(k_{1_{\theta(i}}\right)}{n^{\frac{i+1}{2}}}$$

where $\xi_{i+1}^{(i)}$ and $Z_{1,i+1}^{(i)}$ are defined as in assumptions A.2, A.3 for $\theta_n^{(i)}$. The result follows by integrating this and its exterior product with respect to the Edgeworth distribution of $\sqrt{n}\left(\theta_n^{(i)} - \theta\right)$ and by holding the relevant terms, by noting that due to that due to A.1, A.2, A.5 and AL.2, $\left\|\mathcal{I}_{\varphi_V}\left(k_{i+1_{\theta(i}}\right) - \xi_{i+1}\right\| = o(1)$. Notica that for $\kappa > 2a+1$ we have that $\theta_n^{(\kappa)} = \theta_n^{(\kappa-1)}$. with probability $1 - o\left(n^{-a^*}\right)$. ■

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.

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

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

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

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

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

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

as $A-a_n$ is convex. Now,

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

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

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

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,

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

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

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

and

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

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

$$\sup_{A \in \mathcal{B}_C} \left| R_n - \int_A q_n(z) dz \right| \geq \left| \sup_{A \in \mathcal{B}_C} |R_n| - \sup_{A \in \mathcal{B}_C} \left| \int_A \varphi_V(z) q_n(z) dz \right| \right| = o\left(n^{-a}\right)$$

and $\sup_{A \in \mathcal{B}_C} \left| P(\zeta_n + x_n \in A) - \int_A \varphi_V(z) \left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}} \pi_i^*(z)\right) dz \right| = o(n^{-a})$ 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 AL.2 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 $\sup_{A\in\mathcal{B}_C} n^a |P_n(A) - Q_n(A)| = O(n^{-\eta})$ for $\eta > 0$, we have that

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

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 [41]. 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 [29] formula 8.357. ■

The following lemmas enable the approximation of the Edgeworth moments by transformations.

Lemma AL.3 Suppose that ζ_n admits a valid Edgeworth expansion of order s^*. Then for any $i < j : 1, \ldots q$, $\mathrm{pr}_{i,j}(\zeta_n)$ admits an analogous expansion of the same order.

Proof. Let $\mathcal{B}$ denotes the class of Borel sets on $\mathbb{R}^q$. Then for $A \in \mathcal{B}$ we have that since $\mathrm{pr}_{i,j}^{-1}(A)$ has the ε-neighborhood property with respect to the relevant normal distribution, hence

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

where $v = \mathrm{pr}_{i,j}(x)$, $x = (v, v^*)$, $I = (0, \mathrm{Id}_{j-i+1\times j-i+1}, 0)$, and $\pi_i^*(v) = \int_{\mathbb{R}^{q-p}} \pi_i(v, v^*)\varphi_V(v, v^*)\,dv^*$. ■

Lemma AL.4 Suppose that ζ_n admits a valid Edgeworth expansion of order s^*. Let also $g_n : \mathbb{R}^q \to \mathbb{R}^p$ $(p \leq q)$ with

$$g_n(x) = Lx + \sum\nolimits_{i=2}^{s_*}\left(n^{-\left(\frac{i-1}{2}\right)}\xi_i + o\left(n^{-\left(\frac{i-1}{2}\right)}\right)\right) + \sum\nolimits_{j=2}^{s^*}\frac{1}{n^{\frac{j-1}{2}}}\left(B_{j_n} + o\left(n^{-\delta}\right)\right)x^j \tag{4}$$

for large enough n, with $\operatorname{rank} L = p$. If for any $A \in \mathcal{B}$, $P(x_n \in A) = P(g_n(\zeta_n) \in A) + o\left(n^{-a^*}\right)$, then x_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

$$P(x_n \in A) = \int_A \left(1 + \sum\nolimits_{i=1}^{s} \frac{1}{n^{\frac{i}{2}}}\pi_i^*(x)\right)\varphi_{LVL'}(x)\,dx + o\left(n^{-a^*}\right)$$

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

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

where $\mathcal{B}$ denotes the class of Borel sets on $\mathbb{R}^p$ ξ_i and B_{j_n} are $O(1)$ and $x^m = \underbrace{(x, x, \ldots x)}_{m}$.

Proof. We can assume that $p = q$ without loss of generality, for if $p < q$ then we can consider

$$g_n^*(x) = \begin{pmatrix} L & 0 \\ 0 & \mathrm{Id} \end{pmatrix} x + \sum_{i=2}^{s_*} \left(n^{-\left(\frac{i-1}{2}\right)} \begin{pmatrix} \xi_i + o\left(n^{-\left(\frac{i-1}{2}\right)}\right) \\ 0 \end{pmatrix} \right) + \sum_{j=2}^{s^*} \frac{1}{n^{\frac{j-1}{2}}} \begin{pmatrix} B_{j_n} + o\left(n^{-\delta}\right) \\ 0 \end{pmatrix} x^j$$

and then apply the previous lemma. Now for $A \in \mathcal{B}$ we have that

$$P(x_n \in A) = P(g_n(\zeta_n) \in A) + o\left(n^{-a^*}\right) = P\left(\zeta_n \in g_n^{-1}(A)\right) + o\left(n^{-a^*}\right) = \int_{g_n^{-1}(A)} \left(1 + \sum_{i=1}^{s^*} \frac{1}{n^{\frac{i}{2}}} \pi_i(x)\right) \varphi_V(x)\, dx + o\left(n^{-a^*}\right)$$

Due to the rank condition on L we have that

$$g_n^{-1}(y) = L^{-1}y + \sum_{i=2}^{s_*} \left(n^{-\left(\frac{i-1}{2}\right)} \xi_i^* + o\left(n^{-\left(\frac{i-1}{2}\right)}\right)\right) + \sum_{j=2}^{s^*} \frac{1}{n^{\frac{j-1}{2}}} \left(B_{j_n}^* + o\left(n^{-\delta}\right)\right) y^j$$

for any $y \in H_n(C) = \left\{x \in \mathbb{R}^q : \|x\| < C \ln^{1/2} n\right\}$ for $C > 4a+2$ from lemma 2 of Magdalinos [41] with ξ_i^* and $B_{j_n}^*$ defined analogously. Moreover due to the same lemma

$$\int_{g_n^{-1}(A) \cap H_n^c(C)} \left(1 + \sum_{i=1}^{s^*} \frac{1}{n^{\frac{i}{2}}} \pi_i(x)\right) \varphi_V(x)\, dx = o\left(n^{-a^*}\right)$$

hence

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

Due to the proof of lemma 3.5 of Skovgaard [45] $H_n(C_*) \subseteq g_n(H_n(C))$ for some $C_* > C$, 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 $o\left(n^{-a^*}\right)$.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 [45] 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 estimate the remainders as $o\left(n^{-a^*}\right)$ terms. The second part follows from analogous considerations to the previous and/or the ones in the proof of lemma 4.7 of Skovgaard [45]. ■

Figures

The MA (1) **Case**, $\theta = -0.4$.

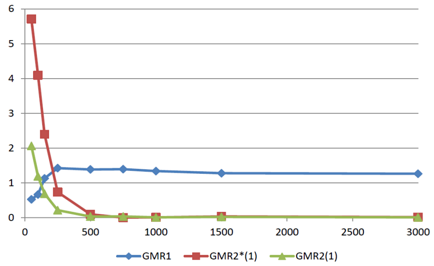

Figure 1: $n \times \left|\widehat{\text{Bias}}\right|$.

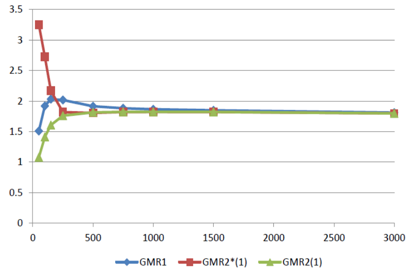

Figure 2: $n \times \widehat{\text{MSE}}$.

The GARCH (1, 1) **Case**, $(\theta_1, \theta_2, \theta_3) = (1.0, 0.2, 0.7)$.

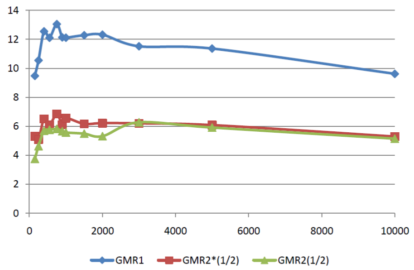

Figure 3: $n \times \left|\widehat{\text{Bias}}\right|$.

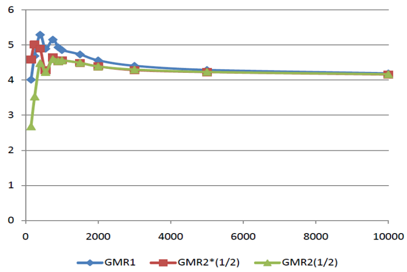

Figure 4: $n \times \widehat{\text{MSE}}$.

The ARFIMA $(0, d, 0)$ **Case**, $d = 0.4$.

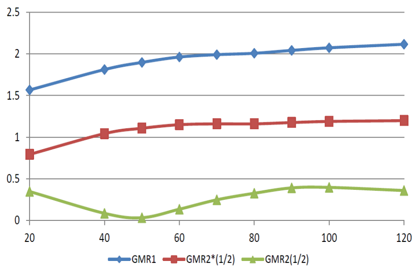

Figure 5: $n \times \left|\widehat{\text{Bias}}\right|$.

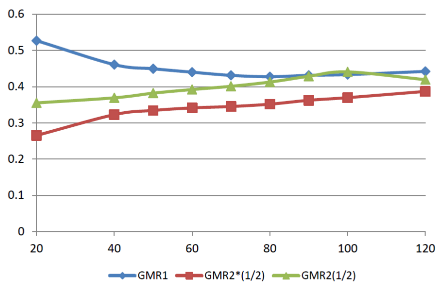

Figure 6: $n \times \widehat{\text{MSE}}$.