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

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

September 23, 2014

Abstract

In this paper we define a set of Indirect Inference estimators based on moment approximations of the auxiliary ones. Their introduction is motivated by reasons of analytical and computational facilitation. Their definition provides an indirect inference framework for some "classical" bias correction procedures. We derive higher order asymptotic properties of these estimators. We demonstrate that under our assumption framework and in the special case of deterministic weighting and affinity of the binding function these are second order unbiased. Moreover their second order approximate Mean Square Errors do not depend on the cardinality of the Monte Carlo or Bootstrap samples that our definition may involve. Consequently, the second order Mean Square Error of the auxiliary estimator is not altered. We extend this to a class of multistep Indirect Inference estimators that have zero higher order bias without increasing the approximate Mean Squared Error, up to the same order. Our theoretical results are also validated by three Monte Carlo experiments.

KEYWORDS: Indirect Estimator, Recursive Indirect Estimator, Binding Function, Edgeworth Expansion, Moment Approximation, Higher Order Bias Approximation, Higher Order Mean Square Error Approximation, Approximate Bias Correction, Monte Carlo, Bootstrap, GARCH model, Stationary Gaussian ARFIMA model.

JEL: C10, C13

*We would like to thank, without implicating them, the seminar participants at the University of Ioannina. Financial support of BRFP II AUEB is gratefully acknowledged. Corresponding author S. Arvanitis, Dept. of Economics, Athens University of Economics and Business, Patission 76, Athens 10434, Greece. Tel: +302108203313, mail: stelios@aueb.gr

1 Introduction

Indirect Inference estimators (IIEs) are usually emerging from two-step optimization procedures. 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 mapping connecting the underlying statistical models, termed as *the binding function*. Minimization of the inversion criterion essentially inverts the binding function, which, usually, is not analytically known and is *approximated numerically*. This approximation may involve some kind of Monte Carlo integration of an $\arg\min$ functional with respect to which the auxiliary estimator is defined. Moreover the $\arg\min$ is also usually approximated by some numerical procedure. Hence the derivation of the analogous IIE involves *nested* numerical optimizations that impose a large computational cost (see the $\mathrm{GARCH}\,(1,1)$ example below). The same IIE under a more involved assumption framework can also have desirable high order asymptotic properties.[1] This framework creates a trade off between numerical cost considerations and asymptotic properties that could cast these estimators unattractive. In fact, under some appropriate conditions, the IIE which is most numerically involved possess the desirable property of reducing the bias of the auxiliary one (see e.g. Gourieroux et al. [36], Arvanitis and Demos [8]). This property has been exploited in dynamic panel setup (see e.g. Gourieroux, Phillips and Yu [35]) as well as in time series (see e.g. Demos and Kyriakopoulou [20], Gourieroux and Monfort [33], Phillips [56]).

Part of the scope of the present paper is the introduction of a class of (potentially multistep) IIEs, where the binding functions depend on approximations of the first moment of the auxiliary estimator that avoid the aforementioned numerical cost. This is due to the fact that even when these approximations involve numerical integration, the integrands are analytically tractable,[2] hence the resulting IIEs avoid nested numerical optimizations. This comes at the *fixed* cost of the analytical derivation of the moment approximation. Under a relevant assumption framework, higher order asymptotic properties of these estimators are potentially similar to the ones in the previous paragraph. Hence this class of estimators can reduce the trade off.

Furthermore, the analysis of higher order asymptotic properties of the aforementioned class of IIEs, along with already established results, provide us with an interesting unification of distinct procedures of approximate bias correction. For example, in cases where the auxiliary estimator (say β_n) is consistent, the analysis that follows

[1] See e.g. Gourieroux and Monfort [33], and Gourieroux, Renault and N. Touzi [36].

[2] i.e. they are analytically tractable functions of the Monte Carlo sample. Hence their Monte Carlo integration does not involve any *extra* numerical procedure.

defines the considered IIEs as solutions of the procedure

$$\min_{\theta} \|\beta_n - \theta - K_n(\theta)\|$$

where $\theta + K_n(\theta)$ is a (potentially stochastic) approximation of $E_\theta \beta_n$. Notice that the first order approximation of the solution to the previous problem $\theta_n^* \doteqdot \beta_n - K_n(\beta_n)$ satisfies this definition as an extreme case.[3] θ_n^* is essentially the paradigm of a vast literature of approximate bias correction (see e.g. Bao and Ullah [10], Bao and Ullah [11], Cordeiro and McCullagh [15], Cordeiro and Klein [16], Cox and Hinkley [18], Fernandez-Val and Vella [23], Gourieroux et al. [36], Iglesias and Phillips [40], Linton [46], MacKinnon and Smith [49], and Rilstone, Srivastava and Ullah [57]). The definition also allows for *intermediate cases* in which some of the elements of $K_n(\theta)$ are evaluated in θ_n. In fact, by unifying these procedures in an IIE framework, we are able to provide sufficient conditions for their validity.

Given the statistical model and β_n, the definition of the considered class of IIEs presupposes the existence of a valid asymptotic approximation for $E_\theta \beta_n$, for any θ. We assume that the auxiliary estimator has a standard $\sqrt{n}$ rate of convergence independent of θ, the mean approximation is polynomial in $\frac{1}{\sqrt{n}}$ and the approximation error is $o\left(n^{-\frac{s}{2}}\right)$ for some positive integer s and for any θ. However, IIEs could also be defined and exist even if the restriction on the rate of convergence is weakened, at the potential cost of invalidating the subsequent methodology, employed for the derivation of higher order asymptotic properties.

The asymptotic approximation of $E_\theta \beta_n$ can be established in a variety of ways. In the first instance, one assumes that β_n, with sufficiently high probability, can be represented as a ratio of quadratic forms in normal (see Magnus [51]) or non-normal (see Ullah and Srivastava [70]) random variables and thereby the procedures of these papers can be employed along with an asymptotic approximation for the subsequent integral. Examples are provided by Phillips [56] (see Theorem 4) in the context of the $\mathrm{AR}(1)$ model with β_n the OLSE, and by Bao and Ullah [11] in the context of the maximum likelihood estimator in spatial models. A second way is to employ expansions of β_n, as in e.g. Bao and Ullah [10], MacKinnon and Smith [49], Newey and Smith [53], or Rilstone et al. [57], and then by employing Nagar [52] type arguments (see Rothenberg [59]) approximate $E_\theta \beta_n$ by the expectation of the expansion.

Alternatively, the mean approximations can be derived as Edgeworth means. This can be validated if for any θ, $\sqrt{n}(\beta_n - b(\theta))$ admits an Edgeworth expansion of sufficiently high order due to lemma AL.2 (b denotes the limit binding function). *Given* the rate of convergence this approach is certainly more general than the first case since the question now concerns the validation of sufficient conditions under which the

[3] Ignoring potential difficulties with the parameter space we have that $\theta_n^* = \arg\min_\theta \|\beta_n - \theta - K_n^*\|$ where $K_n^* \doteqdot K_n(\beta_n)$ which is obviously by construction independent of θ.

Edgeworth expansion holds and does not restrict so much the form of β_n. This is the methodology employed here. Theorem 7.1, in appendix B, provides a set of sufficient conditions and it is a pointwise reformulation of Theorem 3.2 of Arvanitis and Demos [9]. Hence the needed result rests upon the verification of the four theorem's conditions (see the discussion in appendix B for more on this). Notice that due to lemma AL.2 the means of the estimators are valid. However, this is not the case for the first two ways (see e.g. Srinivasan [68] and Sargan [60] on this).

The approximation of $E_\theta \beta_n$ by Edgeworth means can yield analytically intractable functions of θ due to the fact that (some of) the corresponding coefficients may depend on nuisance parameters, analytically intractable moments etc. In order to tackle such cases, we require the existence of random elements that in turn approximate the intractable parts of the mean approximation. In our examples, when needed, the existence of these random elements is "natural", as they have, first, the form of a standardized sums of simulable random elements or, second, some asymptotically smooth approximation of such forms.

In any case we construct a (possibly) stochastic binding function that can be employed for the definition of the introduced IIEs. Their existence is facilitated by standard continuity arguments of this binding function w.r.t. (with respect to) θ. In the first case considered above this is satisfied by mere inspection of the moment approximations. In the second case under analytical tractability the theorem in appendix B provides sufficient conditions for these continuity arguments. Under analytical intractability we further suppose that the stochastic parts of the approximations are also almost surely continuous.

Given the existence of the proposed IIEs, we provide their consistency employing continuity conditions along with conditions that restrict the rates of the stochastic parts of the binding function so that they do not become asymptotically non tight. We then derive higher order properties of them, with a view towards approximate bias and MSE functions. At this point our remaining methodology *necessitates* the validity of an Edgeworth expansion of β_n. Given this, some technical conditions concerning the asymptotic properties of the random elements appearing in the weighting matrices, the stochastic approximations of the intractable parts of the binding function and their derivatives, we establish their valid Edgeworth expansions. Finally, given these expansions, we provide their approximate bias and MSE functions. The aforementioned technical conditions can be verified, in general, by the employment of appropriate Edgeworth expansions for those random elements when needed.[4] This is the case for our examples, as well.

In section 2 we review several classes of IIEs that have already been defined. Notice that indirect inference algorithms were initially employed by Smith [64], were formally

[4] Again this procedure can be also easily employed for the IIE of Phillips [56] in the context of the $\mathrm{AR}(1)$ model, with β_n the corresponding OLSE.

introduced by Gourieroux et al. [34], complemented by Gallant and Tauchen [27] and extended by Calzolari, Fiorentini and Sentana [13], and Garcia, Renault and Veredas [28]. Properties similar to those studied here were studied in Gourieroux et al. [36] and were validated and extended in Arvanitis and Demos [8]. In section 3 we define the proposed estimators and derive their asymptotic properties in the following one. In section 5 we extend the procedures to multi step ones. To facilitate the exposition of our results we employ throughout all sections two examples, concerning the $\mathrm{MA}(1)$ and $\mathrm{GARCH}(1,1)$ processes. 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 fractional Gaussian processes, and Monte Carlo experiments. Conclusions and questions for further research are gathered in section 7. In the appendix A we collect our proofs and in appendix B we present some useful tools concerning our derivations.

2 General Framework

In this section, we describe a quite general assumption framework suitable for indirect inference and establish 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,[5] such as the Frobenius one (i.e. $\|W\| = \sqrt{trW'W}$). The relevant metric space of r-dimensional square real matrices is denoted by $M(\mathbb{R}, r)$. $\|x\|_W$ denotes the norm $\sqrt{x'Wx}$ with respect to the conformal positive definite matrix W.

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

2. The limit binding function (lbf) $b : \Theta \to B$, for B a compact subset of $\mathbb{R}^q$, $b(\Theta) \subseteq \mathrm{Int}\, B$.[6] Moreover it is continuous, injective and, for a natural number s^* specified in the sequel, it is $s^* + 1$ times continuously differentiable when

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

[6] Notice here that the restriction on the image of the lbf is essentially only technical. It allows for the avoidance of Egdeworth expansions when the parameter is on the boundary and it can be easily satisfied due to the fact that in most cases given the statistical model the auxiliary parameter space, i.e. B can be chosen arbitrarily.

restricted to $\operatorname{Int}\Theta$.[7] Also, there exists a function $\varsigma_n : \Omega \times B \to \mathbb{R}$ that is jointly measurable and $\varsigma_n(\omega, \beta)$ is (lower semi-) continuous on $b(\Theta)$ for $P_\theta-$almost all ω, for any θ.

3. Let $W_n : \Omega \times \Theta \to M(\mathbb{R}, q)$ be jointly measurable and P_θ-almost surely positive definite, for every $\theta \in \Theta$. Also, let θ_n^+ denote a random element on Ω with values in Θ. When $p = q$ suppose without loss of generality that W_n is the identity.

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 σ-field of the product sets, and that any $P^* \in \operatorname{par}(\Theta)$ is the unique extension on $\mathcal{F}$, of a sequence of probability measures $(P_n^*)_{n=1}^{\infty}$-with P_n^* defined on $\bigotimes_{i=1}^n \mathcal{F}_i$- that is Kolmogorov consistent. Given the Kolmogorov consistency, the existence of P^* is guaranteed when Ω_n is a Hausdorff topological space, $\mathcal{F}_n$ is the relevant Borel algebra, and P_n^* is tight for any n (see corollary 15.28 of Aliprantis and Border [2]). 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.

Notice that assumption A.1.1 considers a family of probability measures that are partially described by a finite dimensional parameter. Hence it allows for finite dimensional inference in semi-parametric models. For assumption A.1.2 continuity of the lbf would follow from $P_\theta\left(\sup_{b(\Theta)} |\varsigma_n(\omega, \beta) - \varsigma(\theta, \beta)| > \varepsilon\right) \to 0, \forall \varepsilon > 0$ for some $\varsigma : \Theta \times b(\Theta) \to \mathbb{R}$ that is jointly continuous and $\varsigma(\theta, b(\theta)) < \varsigma(\theta, \beta)\ \forall \beta \in b(\Theta) - \{b(\theta)\}$ and $\forall \theta \in \Theta$. The differentiability assumptions could follow from analogous assumptions for $\varsigma(\theta, \beta)$ and the implicit function theorem. Finally A.1.3 implies the possibility of *stochastic weighting*. In this framework θ_n^+ is essentially an initial estimator and it can be any of the IIE considered below. In the following we suppress the dependence of the aforementioned binding functions on Ω where unnecessary.

Definitions of Already Known IIEs We can now recall the definitions of the auxiliary and the already established IIEs, entitled here as $\mathrm{GMR1}$ and $\mathrm{GMR2}$. These were initially formalized by Gourieroux et al. [34].

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

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

The $\mathrm{GMR1}$ estimator is defined by

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

[7]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 $\theta \in \operatorname{Int}\Theta$ and appropriate ε.

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

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

By assumption A.1 and the discussion that followed the above estimators are well defined. In most cases b and b_n are analytically intractable hence are usually numerically approximated. Moreover the optimization procedure is also approximated numerically. Thereby, their computation can be associated with *nested numerical optimizations* introducing large computational costs especially in the case of $\mathrm{GMR2}$.

To facilitate the exposition of our results we present two simple examples.

$\mathrm{MA}(1)$ **Example** Consider the invertible $\mathrm{MA}(1)$ process

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

In the language of assumption A.1 par is the set of $\mathrm{MA}(1)$ processes on Θ which is a non-empty compact subset of $(-1, 1)$. Let $\varsigma_n(\omega, \beta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta y_{i-1})^2$, and thereby $\varsigma(\theta, \beta) = (1 + \beta^2)(1 + \theta^2) - 2\beta\theta$, and $B = [-c, c]$, $c \leq \frac{1}{2}$. ς_n is essentially the quasi likelihood function associated to the stationary $\mathrm{AR}(1)$ model and β_n is the QMLE for β. The lbf is $\frac{\theta}{1+\theta^2}$ and thereby the $\mathrm{GMR1}$ estimator is equal to $\frac{1-\sqrt{1-4\beta_n^2}}{2\beta_n}$ when $\beta_n \in (-c, c)$, equal to -1 when $\beta_n \leq -c$ and equal to 1 when $\beta_n \geq c$ (see Gourieroux et al. [34], and Demos and Kyriakopoulou [20]). To evaluate the $\mathrm{GMR2}$ estimator the $E_\theta \beta_n$ is needed. However, it is analytically intractable and it is usually approximated by Monte Carlo integration (see e.g. Gourieroux et al. [34] or Arvanitis and Demos [8]).

$\mathrm{GARCH}(1,1)$ **Example** Consider the second order stationary $\mathrm{GARCH}(1,1)$ model (Bollerslev [12])

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

Θ is a non-empty compact subset of $\mathbb{R}^3$ with elements that satisfy the above inequalities and par is the set of $\mathrm{GARCH}(1,1)$ processes on Θ. Let $\varsigma_n(\omega, \beta)$ be the conditional likelihood function and $\beta_n = (\beta_{1,n}, \beta_{2,n}, \beta_{3,n})'$ represent the MLE of $\theta = (\theta_1, \theta_2, \theta_3)'$ where B is a compact superset of Θ that obeys assumption A.1.2. $\varsigma(\theta, \beta)$ is the expectation of the likelihood evaluated at β, w.r.t. the process defined by θ. It is well known that β_n is consistent and asymptotically normal (see e.g. Lee and Hansen [42] or Lumsdaine [47]) and thereby the lbf is the identity (more precisely the inclusion).

However, it is also known that in small samples it can be severely biased (see e.g. Linton [46] or Lumsdaine [48]). Now treating β_n as the auxiliary estimator we have that due to the form of the lbf, $\mathrm{GMR1} = \beta_n$ at least with probability converging to 1. Again for the evaluation of the $\mathrm{GMR2}$ estimator the needed $E_\theta \beta_n$ is analytically intractable and is approximated via Monte Carlo integration. Now the Monte Carlo integration involves the *extra* numerical procedure of the approximate maximization of the likelihood function for each Monte Carlo sample. Further the derivation of the $\mathrm{GMR2}$ estimator necessarily involves an analogous optimization (maximization of the likelihood). Consequently, the $\mathrm{GMR2}$ estimator optimization nests the optimization for each Monte Carlo sample. It is in such cases that the evaluation of the $\mathrm{GMR2}$ estimator involves nested optimizations. Notice that for the $\mathrm{MA}(1)$ example nested optimizations are avoided as there is an explicit solution to the auxiliary optimization.

Let us now turn our attention to the assumptions concerning the proposed IIEs.

Assumptions Specific to the Proposed IIEs The following assumptions enable the definition and the derivation of properties of a new class of IIEs. 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 $\sum_{i=1}^{s^*} \frac{\pi_{i-1}(z,\theta)}{n^{\frac{i-1}{2}}} \varphi_{V_\theta}(z)$, with $\pi_0(z,\theta) = 1$ 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 [50] eq. 3.7-8, p. 348). The distinction among the three orders is employed so that the setup is not only as general as possible, but also facilitates the derivation of recursive estimators defined in section 5. 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 $\underbrace{(x, \ldots, x)}_{r \quad \text{times}}$.[8] The first assumption essentially builds the binding function upon which the definitions of IIEs that follow will be based.

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

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

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

[8] On multilinear functions see e.g. Northcot [55].

The above assumption holds when $\sqrt{n}\left(\beta_n - b\left(\theta\right)\right)$ admits an Edgeworth expansion of order $s^* > s$ due to 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 $\theta \in \Theta$. For a discussion on the assumptions needed for a valid Edgeworth expansion to exist please see appendix B. Due to assumption A.1.2, $b\left(\theta\right)$ lies in the interior of B hence, $\xi_1 = 0_q$. Alternatively, in some instances, expressions of the $\xi_i's$ can be found in Bao and Ullah [10] Rilstone et al. [57] and Ullah [69]. Finally the $o\left(n^{-\frac{s}{2}}\right)$ rate of the remainder and the analogous discrepancy with s_* enables the possibility that equation (3) is not valid beyond s and it is included due to the fact that the properties of the estimators (to be defined) up to order s would be identical whether the $\xi_i\left(\theta\right)$ terms are included or not for $i > s$. For all practical purposes one can consider that $s_* = s$.

In cases where (some of) the $\xi_i's$ are analytically unknown, i.e. 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. Consequently, the next assumption, enables the *stochastic approximations* of the $\xi_i\left(\theta\right)' s$ in equation (3).

Assumption A.3 For some $\delta > 0$ small enough:

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 jointly measurable, Q_θ-almost everywhere continuous on Θ, where $Q_\theta = P_\theta \times P'_\theta$. **If** $\xi_i = 0_q$ **then** $\zeta_{i_n} = 0_q$.

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 $s^* + 1$ 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 $i = 2, \ldots, s_*$, $j = 1, \ldots, s^* + 1$.

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 (not necessarily all) elements of ζ_{i_n} that depend on θ are evaluated at a stochastic point, e.g. at an initial estimator θ_n^+.

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 (of analogous order) continuous differentiability of the moments appearing in the Edgeworth polynomials. For a discussion on low order assumptions that would validate this see Arvanitis and Demos [8] (Discussion on Assumption A.11). 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

$$E_\theta\left\|\sup_{\theta'} D^{s^*+1}\zeta_i(y_1,\theta') - E_\theta \sup_{\theta'} D^{s^*+1}\zeta_i(y_1,\theta')\right\|^p < +\infty$$

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. This can be achieved by employing results such as the Yokoyama moment inequality (see Andrews [4], proof of lemma 3).

Either the GMR1 and GMR2 estimators or the ones to be defined depend on stochastic norms based on weighting matrices. In principle *continuous updating* versions of those estimators could also be defined. We avoid such definitions since in such cases results as corollary 1 below would generally cease to be valid. Our 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 $W : \mathbb{R}^m \times \Theta \to M(\mathbb{R}, q)$ integrable with respect to P_θ for any $\theta \in \operatorname{Int}\Theta$, such that

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

$W(\theta')$ 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 $\theta \in \operatorname{Int}\Theta$.

$b)$ If $q > p$, W_n is s^*-continuously differentiable P_θ-almost surely on $\operatorname{Int}\Theta$ and for any $\theta' \in \operatorname{Int}\Theta$

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

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

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

Part b) can be justified analogously. Obviously when $W(x,\theta)$ 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 $q_n(\theta)$ be the vector containing the elements of $\zeta_{i_n}(\omega,\omega',\theta)-\xi_i(\theta)$ for all i. Let $f(x,\theta)$ denote the vector that contains stacked all the *distinct* components of $W(x,\theta)$ as well as their derivatives up to the order s^*-1. If $f(x_0,\theta)-E_\theta f(x_0,\theta)$ contains zero elements then these are *discarded*. Let

$$m_n^*(\theta)=\beta_n-b(\theta)$$

when $f(x_0,\theta)-E_\theta f(x_0,\theta)$ is zero or $p=q$,

$$m_n^*(\theta)=\begin{pmatrix}\beta_n-b(\theta)\\ \frac{1}{n}\sum f(x_i)-E_\theta\frac{1}{n}\sum f(x_i)\end{pmatrix}$$

when $f(x_0,\theta)$ is independent of θ, and

$$m_n^*(\theta)=\begin{pmatrix}\beta_n-b(\theta)\\ \theta_n^+-\theta\\ \frac{1}{n}\sum f(x_i,\theta)-E_\theta\frac{1}{n}\sum f(x_i,\theta)\end{pmatrix}$$

in any other case.

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. $Q_\theta\left(\|q_n(\theta)\|>o(1)\right)=o\left(n^{-a^*}\right)$ for some $o(1)$ real sequence.

The first part can be established by the verification of conditions such as those appearing in theorem 7.1. Due to Lemma 2 of Magdalinos [50] the second part can also be established if $\sqrt{n}q_n(\theta)$ under Q_θ admit Edgeworth expansions of order s^*. Notice that in this case the expansions involved **need not hold jointly**.

The following lemmas are useful and immediate (proofs are presented in appendix A).

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 $\sqrt{n}\left(\mathrm{GMR1}-\theta\right)$ under P_θ admits an Edgeworth expansion of order s^*.

Given the definition of the vector f and since in any of the following examples $p = q$ we can assume that W_n is the identity *without loss of generality (see assumption A.1.3)*. Hence in those cases assumption A.4 follows trivially and m_n^* is $\beta_n - b(\theta)$. In the first example $\zeta_{i_n} = \xi_i = \mathcal{I}_{V_\theta}(k_i(z,\theta))$ and thereby assumption A.3.2 and the second part of 3 follow trivially given the relevant smoothness. Analogously assumption A.5.2 follows also trivially. Hence, the rest of the assumption framework lies on the validity of Edgeworth expansions for the distribution of $\sqrt{n}(\beta_n - b(\theta))$ as well as on the smoothness properties of the coefficients of the Edgeworth densities as functions of θ. For these a general procedure is described in the Appendix B. Validity would follow in the context of theorem 7.1 along with the validation of the conditions of Gotze and Hipp [32] (Assumptions 1-4) for the derivatives of ς_n. For details see Demos and Kyriakopoulou [20]. Smoothness follows from the smoothness of the moments appearing as coefficients in the Edgeworth densities due to the smoothness of the process itself, as well as due to dominated convergence. The results are established without any further parameter restrictions but by employing restrictions on the distribution of u_0 (see below).

The second example is essentially more complex due to the fact that the ζ_{i_n} are stochastic approximations of the relevant ξ_i. Assumptions A.2 and A.5.1 follow from the validation and the derivation of the formal Edgeworth expansions for the distribution of $\sqrt{n}(\beta_n - b(\theta))$. This is established in Corradi and Iglesias [17], Linton [46] and Iglesias and Linton [38] again by a procedure similar to the one described in the context of theorem 7.1 along with the validation of the conditions of Gotze and Hipp [32] (Assumptions 1-4) for the derivatives of ς_n. For further details see the following discussion concerning each of the employed examples. Again the results are established without any further restrictions on the parameters while given the assumption of conditional normality (see above) no further restrictions are needed for the behavior of z_0.

$\mathrm{MA}(1)$ **Example Cont.** From the results of Arvanitis and Demos [8] and Demos and Kyriakopoulou [20] we have that if $E|u_0|^{14} < \infty$ and if $D(0,\sigma^2)$ has a smooth density, then the β_n estimator admits a 5^{th} order valid Edgeworth expansion, uniformly over Θ. Further, by lemma AL.2 in 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 [20] for their expressions). 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 $\mathrm{GMR1}$ estimator, by lemma 2.2.

$\mathrm{GARCH}(1,1)$ **Example Cont.** From Corradi and Iglesias [17] we have that, under conditional normality, the Edgeworth expansion of the MLE, β_n, is valid for any order s^* when Θ and B are defined as previously. This naturally implies also assumption A.5.1. Employing the formulae for the Edgeworth expansion, given in Linton [46] and Iglesias and Linton [38], we have that assumption A.2 is verified by $E(\beta_n - \theta) =$

$-\frac{1}{n}(\lambda_0+\lambda_2)+o(n^{-1})$, 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 are not analytically tractable. 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 $\left\{E_\theta\left(u_t^{-2}u_{t;i}u_{t;j}\right)\right\}_{i,j=1,\ldots,3}$. These expectations are approximated by Monte Carlo integration in the obvious way. 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, i.e. $u_t(\theta)\geq\frac{\theta_1}{1-(\theta_2+\theta_3)}$ $\forall t$. Now let $l_{2_{t,s}}(\theta)=\mathrm{vec}\left\{\left(u_{t,s}^{-2}u_{t,s;i}u_{t,s;j}\right)\right\}_{i,j=1,\ldots,3}$ where $s=1,\ldots,S$ denotes the relevant path of a resampling procedure. In the same way as in lemma A.1 of Corradi and Iglesias [17], it is possible to prove that the vector $S_n(\theta)=\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(l_{2_t}(\theta)-El_{2_t}(\theta)\right)\right)_{s=1,\ldots,S}$ has a valid Edgeworth expansion of order s^* under Q_θ without any further parameter restrictions and due to smoothness something analogous holds for $\frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(l_{2_{t,s}}(\theta)-El_{2_{t,s}}(\theta)\right)\right)$ for finite S. Smoothness of λ_0, λ_2 imply the applicability of theorem 7.1 and thereby establishes that $\sqrt{n}q_n(\theta)$ admits an analogous Edgeworth expansion of order s^*. Hence A.5.2 follows.

We are now ready to define our estimators.

3 Definition of the $\mathrm{GMR2}(a_*)$ Estimators

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

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

$\theta_n(a_*)$ exists due to assumptions A.1 and A.3.1. Its feasibility depends on i) the analytical knowledge of b, and ii) the 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. When β_n is $\theta-$consistent then b is the identity or the inclusion. For more general cases please see the discussions at the second and the semifinal paragraphs of section 5.1. Given this, the $\mathrm{GMR2}(a_*)$ estimator can surpass the nested optimization burden associated with the $\mathrm{GMR2}$ estimator. Finally it is easy to deduce, from assumptions A.2 and A.3, that the $\mathrm{GMR1}$ estimator can be identified as our $\mathrm{GMR2}(0)$.

Remark R.1 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. $\theta^*_n(a_*)$ is also denoted by $\mathrm{GMR2}^*(a_*)$. $\theta^*_n(a_*)$ admits another interesting characterization. It is the first term of a sequence defined by $\theta^{(i)}_n = \theta_n(0) - \sum_{i=2}^{s_*} \frac{1}{n^{i/2}} \zeta_{i_n}\left(\theta^{(i-1)}_n\right)$, for $i = 0, 1, 2, \ldots$, and $\theta^{(0)}_n = \theta_n(0)$. It is easy to prove that under assumptions A.3 and A.5 for any $\theta \in \mathrm{Int}\,\Theta$ and $\delta > 0$

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

Hence $\theta^*_n(a_*)$ is in fact the first term of a "Newton-Raphson" sequence that need not converge to $\theta_n(a_*)$ as $i \to \infty$. It is widely employed in the literature of bias correction especially in the case where $a_* = \frac{1}{2}$ (see e.g. Bao and Ullah [10] and [11], Cox and Hinkley [18], Linton [46], MacKinnon and Smith [49], Newey and Smith [53] 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 $\mathrm{GMR2}(a)$ estimators, i.e. cases $\zeta^*_{i_n}$ depend on θ for some i.

$\mathrm{MA}(1)$ **Example Cont.** From Demos and Kyriakopoulou [20] we have that under the assumptions described above

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

for any $\theta \in (-1, 1)$. Notice that the $\frac{1}{n}$ term is zero. Treating the $\mathrm{GMR1}$ ($\theta_n(0)$ in our terminology, see remark R.1) estimator as the auxiliary one we can define the $\mathrm{GMR2}(1)$ as:

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

Notice that in this case the binding function is the identity (or more precisely the inclusion from Θ to $(-1, 1)$), i.e. $b(\theta) = \theta$. Further, notice that, what is commonly known as the bias corrector is given by $\theta_n(0) - \frac{1}{n}\theta_n(0) \frac{1 + 5\theta^2_n(0) + 2\theta^4_n(0) + \theta^6_n(0) - \theta^8_n(0)}{\left(1 - \theta^2_n(0)\right)^3}$ (see e.g. Bao and Ullah [10] or Demos and Kyriakopoulou [20]). We refer to this estimator as $\theta_n^*(1)$ (see remark R.1) and it is defined as

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

$\mathrm{GARCH}(1,1)$ **Example Cont.** Given the previous considerations the $\mathrm{GMR2}\left(\frac{1}{2}\right)$ estimator is defined as

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

The commonly employed bias corrector (see e.g. Bao and Ullah [10] Linton [46]), $\theta_n^*\left(\frac{1}{2}\right)$ (see remark R.1), 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 $\mathrm{GMR2}(a_*)$ estimator lies in the class of estimators considered by MacKinnon and Smith [49] 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 they constitute a subclass of the 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 the sequel.

4 Higher Order Asymptotic Theory

In this section we derive high order asymptotic properties of the estimators. The results on consistency, valid Edgeworth and moment expansions are established in this order.

4.1 Consistency

In this section, it is initially proven that the $\mathrm{GMR2}(a_*)$ under Q_θ is contained in an arbitrary neighborhood of $\theta \in \mathrm{Int}\,\Theta$ with probability $1 - o\left(n^{-a^*}\right)$. Given this, it is shown that the particular estimator has a very convenient characterization as a near minimizer of the $\mathrm{GMR2}$ criterion. Analogous relations are established between $\mathrm{GMR2}\left(a_*\right)$ and $\mathrm{GMR2}\left(a_*'\right)$, for potentially different a_*, a_*' including the case of the $\mathrm{GMR1}$. For notational simplicity we denote $E_\theta W\left(\theta\right)$ with $W\left(\theta\right)$.

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

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

and therefore

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

From lemma 4.1 we obtain the following results concerning interesting characterizations of the estimator under examination. First, we consider the characterization of the new estimators in respect to the $\mathrm{GMR2}$ one.

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

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

with $Q_\theta\left(\eta_n > \varepsilon\right) = o\left(n^{-a^*}\right)$, for any $\varepsilon > 0$ and η_n is Q_θ-almost surely non negative, and where $E_{\theta_n\left(a_*\right)}\beta_n = E_\theta\beta_n$ evaluated at $\theta_n\left(a_*\right)$, i.e. $E_{\theta_n\left(a_*\right)}\beta_n = E_\theta\beta_n|_{\theta=\theta_n\left(a_*\right)}$, and $E_{\mathrm{GMR2}}\beta_n = E_\theta\beta_n|_{\theta=\mathrm{GMR2}}$.

The condition $\sup_{\theta\in\Theta}\left\|E_\theta\beta_n - b\left(\theta\right)\right\| = o\left(1\right)$ would follow from the uniform (pseudo) consistency of the auxiliary estimator given the compactness of B. Given assumption A.1.2, if the lbf is a bijection $\sup_{\theta\in\Theta}P_\theta\left(\sup_{\beta\in B}\left\|\varsigma_n\left(\beta\right) - \varsigma(\theta,\beta)\right\| > \delta\right) = o\left(1\right)$ $\forall\delta > 0$ would be sufficient for this. The examined estimator is essentially an η_n-$\mathrm{GMR2}$ estimator, i.e. an approximate minimizer of the $\mathrm{GMR2}$ criterion defined by $\left\|\beta_n - E_{\eta_n-\mathrm{GMR2}}\beta_n\right\|_{W_n\left(\theta_n^+\right)} \leq \inf_{\theta\in\Theta}\left\|\beta_n - E_\theta\beta_n\right\|_{W_n\left(\theta_n^+\right)} + \eta_n$. Unfortunately, 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 on the analogous rate of uniform convergence of $b_n\left(\theta\right)$ to $b\left(\theta\right)$. Obviously, the $\mathrm{GMR1}$ estimator is an approximate $\mathrm{GMR2}$.

The previous reasoning can also establish analogous relations between $\mathrm{GMR2}\left(a_*\right)$ and $\mathrm{GMR2}\left(a_*'\right)$ estimators, which are defined by different ζ_{i_n}, and a_* and a_*' are not necessarily the same. This would allow us to provide another interesting interpretation of the $\mathrm{GMR2}^*\left(a_*\right)$ estimator, apart from the one provided in remark R.1.

Lemma 4.3 Suppose that assumptions A.1, A.2, A.3.1, A.4 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

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

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

Therefore any $\mathrm{GMR2}\left(a'_*\right)$ is an approximate $\mathrm{GMR2}\left(a_*\right)$ estimator in this sense. Obviously this relation holds between a given $\mathrm{GMR2}\left(a_*\right)$ and the $\theta_n^*\left(a_*\right)$ or its variants discussed in R.1. The following lemma deals with a more special case where $a_* > a'_*$ and, as opposed to the previous lemma, the ζ_{i_n} defining the two estimators coincide up to $i = 2a'_*$.

Lemma 4.4 Suppose that assumptions A.1, A.2, A.3, A.4 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

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

with $Q_\theta\left(\eta_n > \gamma_n\right) = o\left(n^{-a^*}\right)$, 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.

Obviously, the $\mathrm{GMR1}$ estimator is an approximate $\mathrm{GMR2}\left(a_*\right)$, for any a_*.

4.2 Validity of Edgeworth Approximation

In this subsection, we are concerned with the higher order approximation of the distribution of $\mathrm{GMR2}\left(a_*\right)$ for $a_* > 0$.[9] We essentially rely on the previous results, the local differentiability of the criterion, from which it emerges, and lemma AL.1 presented in appendix B.

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

[9] For the case $a_* = 0$, i.e the $\mathrm{GMR1}$ case, see Arvanitis and Demos [8],

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

Lemma 4.6 Under assumptions A.1, A.4, A.2, A.3 and A.5 for any $\theta \in \operatorname{Int}\Theta$ $i)$ the distribution of $\sqrt{n}\left(\theta_n\left(a_*\right)-\theta\right)$ under Q_θ has an Edgeworth expansion of order s^*, and $ii)$ the distribution of $\sqrt{n}\left(\theta_n\left(a_*\right)-\theta\right)$ as described in the first part of the proof of lemma AL.4 in appendix B is a valid order s^* Edgeworth expansion.

The valid Edgeworth expansions are generally not unique. However the previous lemma and the triangle inequality, force the distance between the evaluations on the same Borel set of any two such Edgeworth measures to be $o\left(n^{-a^*}\right)$, uniformly over the Borel sets on $\mathbb{R}^p$.

4.3 Valid Moment Approximations

Lemma 4.6 in the light of lemmas AL.2 and AL.4 under the correct relation between a^* and a, provides us with an approximation of the sequence of moments (of any order) of the defined estimator. The next lemma clarifies this relation.

Lemma 4.7 Under assumptions A.1, A.2, A.3, A.4, A.5 and if $a^* \geq a+\frac{m}{2}$, then for any $\theta \in \operatorname{Int}\Theta$ and K an m-linear real function

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

where g_n as in the proof of lemma 4.6 and $\left(1+\sum_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i\left(z,\theta\right)\right)$ is the density of the Edgeworth approximation of order s of $\sqrt{n}m_n^*\left(\theta\right)$.

The function g_n is essentially the *Lagrange inversion*-truncated up to the s^{th} order- of the polynomial approximations of the first order conditions that the estimator asymptotically satisfies. We essentially derive the g_n function and integrate its composition with K w.r.t. the Edgeworth distribution in assumption A.5 in order to derive the moment approximations at hand. The generality of A.5 implies that these are expressed as functions of the analogous approximations for the $\sqrt{n}m_n^*\left(\theta\right)$ random vector. This is in turn sufficient for the approximate bias-MSE characterizations that we pursue.

In the following we explicitly provide this type of approximation for the mean and the mean squared error for the $\mathrm{GMR2}\left(a_*\right)$ for any a_* when $a=\frac{1}{2}$. We suppress the dependence on θ and z where possible for notational convenience. We denote by b_j the j^{th} element of b, $W_{j,j'}$ the $\left(j,j'\right)$ element of W, and $\mathcal{C}=\frac{\partial b'}{\partial\theta}W\frac{\partial b}{\partial\theta'}$. Let $\mathrm{pr}_{i,j}\left(x\right)$ denotes the transformation of an r^{th} dimensional vector, say $x=\left(x_1,x_2,\ldots,x_r\right)'$, to a vector containing only the elements of x from the i^{th} to the j^{th} coordinate, i.e. $\mathrm{pr}_{i,j}(x) = (x_i, x_{i+1}, ..., x_j)'$, where naturally $1 \leq i \leq j \leq r$. z denotes a variable with values in the Euclidean space of dimension equal to the dimension of the random vector $m_n^*(\theta)$ in assumption A.5, $k_{i_\beta}(z,\theta) = \pi_{i-1}(z,\theta)\,\mathrm{pr}_{1,q}(z)$, for any $i = 1, \ldots, s$, $k_{1_{\theta+}}(z,\theta) = \mathrm{pr}_{q+1,p+q}(z)$ if $\theta_n^+ - \theta$ appears in the vector $m_n^*(\theta)$, otherwise it is 0_q. Analogously, $k_{i_w}(z,\theta)$ is the symmetric $q \times q$ matrix, defined as follows: for $j' \geq j$, $(k_{i_w}(z,\theta))_{j,j'} = z_q$ where q is the position of $(W_n(\theta) - E_\theta W(\theta))_{j,j'}$ if the latter appears in $m_n^*(\theta)$ otherwise it is zero.

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 [53] 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.[10]

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

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

where for $a_* = 0$ $\varsigma_0(\theta)$ equals

$$\begin{aligned}
&\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W(\theta)\,\mathcal{I}_{\varphi_{V^*}}\left(k_{2_\beta}\right)\\
&-\frac{1}{2}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W(\theta)\,\mathcal{I}_{\varphi_V}\left(\left[\left(\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W(\theta)\,k_{1_\beta}\right)'\frac{\partial b_j}{\partial\theta\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W(\theta)\,k_{1_\beta}\right]_{j=1,\ldots,q}\right)\\
&+\mathcal{C}^{-1}\mathcal{I}_{\varphi_V}\left(\left\{\begin{array}{c}\left[\left(\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W(\theta)\,k_{1_\beta}\right)'\frac{\partial^2 b'}{\partial\theta\partial\theta_j}\right]_{j=1,\ldots,p}W(\theta)\\ +\frac{\partial b'}{\partial\theta}\left\{k_{1_w} + \left[\frac{\partial}{\partial\theta'}W(\theta)_{j,j'}\,k_{1_{\theta+}}\right]_{j,j'=1,\ldots,q}\right\}\end{array}\right\}\left\{Id_q - \frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W(\theta)\right\}k_{1_\beta}\right),
\end{aligned}$$

whereas for $a_* > 0$

$$\varsigma_{a_*}(\theta) = \varsigma_0(\theta) - \mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W(\theta)\,\mathcal{I}_{\varphi_{V^*}}\left(k_{2_\beta}\right).$$

[10] 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 employed for comparisons between the estimators in the spirit of Magdalinos [50] (pp. 347-48).

Notice that, as $\mathrm{GMR1} = \mathrm{GMR2}(0)$ (at least with probability converging fast enough to 1), the approximate bias of the $\mathrm{GMR2}(0)$ estimator is the same as the bias of the $\mathrm{GMR1}$ one presented in Arvanitis and Demos [8]. Furthermore, under the assumptions employed in both papers and when the asymptotic weighting matrices and the *assumed* Edgeworth measures coincide, we have that the second order bias of the $\mathrm{GMR2}(a_*)$ estimator, for $a_* > 0$, is the same to the one of the $\mathrm{GMR2}$ estimator (see Lemma 3.2-3 and Corollary 2 in Arvanitis and Demos [8]). This is a direct consequence of the fact that the assumption framework employed there implies that the analogous truncated Lagrange inversion for the $\mathrm{GMR2}$ coincides with the g_n function employed here (see lemma 4.7). This also implies the concurrence between the second order Edgeworth approximations, hence the second order equivalence between the $\mathrm{GMR2}$ and the $\mathrm{GMR2}(a_*)$ estimator, for $a_* > 0$. We can now trivially obtain the following corollary which establishes conditions implying that $\mathrm{GMR2}(a_*)$ estimator, for $a_* > 0$ is second order unbiased.

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

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

The framework of non stochastic weighting and of affinity of the lbf is the most general known for this kind of results to hold. For a discussion on conditions establishing the linearity of b see section 5.

$\mathrm{MA}(1)$**,** $\mathrm{GARCH}(1,1)$ **Examples Cont.** From the above corollary, and for the $\mathrm{MA}(1)$ example, it is obvious $E_\theta \theta_n(1) = E_\theta \theta_n^*(1) = \theta + o\left(n^{-\frac{3}{2}}\right)$. Furthermore, for the $\mathrm{GARCH}(1,1)$ case, we have that $E_\theta \theta_n\left(\frac{1}{2}\right) = E_\theta \theta_n^*\left(\frac{1}{2}\right) = \theta + o\left(n^{-1}\right)$.

4.5 MSE 2nd order Approximations of the $\mathrm{GMR2}(a_*)$

Given the results of the previous subsection an arising question concerns the comparison between the second order MSE approximations of the $\mathrm{GMR2}(a_*)$ for different a_*.[11] We obtain the following lemmas.

Lemma 4.9 If $W(x,\theta)$ is independent of x and θ, b is affine, assumptions A.1, A.2, A.3, A.4 and A.5 hold and $a^* \geq \frac{3}{2}$ then, for any $\theta \in \operatorname{Int}\Theta$

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

[11] It is a matter of trivial calculation to show that the implied second order approximation of the variance coincides with the analogous approximation of the MSE.

where

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

for all a_*.

Notice that it is a matter of a trivial calculation to show that the implied second order approximation of the variance coincides with the analogous approximation of the MSE. Hence this along with corollary 4.8, when $a_* > 0$, the $\mathrm{GMR2}(a_*)$ estimators have zero second order asymptotic bias while possessing the same second order asymptotic variance with the $\mathrm{GMR1}$ estimator. In this respect under the premises of the aforementioned results, the $\mathrm{GMR2}(a_*)$ estimators are considered second order equivalent to each other for positive a_*'s, as well as superior to the $\mathrm{GMR1}$ estimator w.r.t. second order bias-MSE comparisons. The discussion in the previous section implies that the same approximations hold also for the $\mathrm{GMR2}$ estimator when the asymptotic weighting matrices and the *assumed* Edgeworth measures coincide (see Arvanitis and Demos [8]-Lemma 3.6).

$\mathrm{MA}(1)$ **Example Cont.** Recall that $\theta_n(1)$ is the $\mathrm{GMR2}(1)$ with auxiliary estimator the $\mathrm{GMR1} = \theta_n(0)$ one. Now from Demos and Kyriakopoulou [20] we have that $\mathcal{I}_V\left(k_{2_{\mathrm{GMR1}}}k'_{1_{\mathrm{GMR1}}}\right) = 0$, hence by the above lemma

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

which is the asymptotic variance of $\theta_n(0)$ for any $\theta \in \mathrm{Int}\,\Theta$ (see Fuller [26]).

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

$$\begin{aligned} E\left\|\sqrt{n}\left(\theta_n\left(\frac{1}{2}\right)-\theta\right)\right\|^2 &= E\left\|\sqrt{n}\left(\theta_n^*\left(\frac{1}{2}\right)-\theta\right)\right\|^2 \\ &= E\left\|\sqrt{n}\left(\beta_n-\theta\right)\right\|^2 + o\left(n^{-\frac{1}{2}}\right) \text{ for any } \theta \in \mathrm{Int}\,\Theta. \end{aligned}$$

Due to the structure of the mean approximations the approximate MSE of the unbiased $\mathrm{GMR2}(a_*)$ presented in this paper, even when these are derived via of Monte Carlo and/or Bootstrap sampling techniques, does not depend on the cardinality of these samples, due to assumption A.5.2 This is not obviously the case with other simulation based bias correctors whence the corresponding MSEs may be inflated by a factor depending on this number (see MacKinnon and Smith [49]). In these cases, and under suitable 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 $\mathrm{GMR2}\,(a_*)$ estimators.

5 Recursive Estimators

The previous results imply that the second order asymptotic properties of the $\mathrm{GMR2}\,(a_*)$ depend, among other factors, on the local behavior of the lbf. Due to its injectivity as prescribed by assumption A.1.2, it is easy to see that B can always be chosen so that $b\,(\theta)$ is of the form $\left(\theta', 0'_{q-p}\right)'$.[12] This along with non stochastic weighting and corollary 1 imply that there always exists an auxiliary parametrization such that the $\mathrm{GMR2}\,(a_*)$ estimators for $a_* > 0$ are second order unbiased. Usually, this reparameterization 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 $\mathrm{GMR1}$, let $\beta_n = \left(\mathrm{GMR1}', 0'_{q-p}\right)'$. Given the validity of our assumption framework, lemma 4.8 implies the validation of assumption A.2 for β_n. For a compact $\Theta' \subset \mathrm{Int}\,\Theta$ apply the $\mathrm{GMR2}\,(a_*)$ estimator on β_n Then the resulting indirect estimator is derived from a three-step procedure, in the last step of which the binding function is obviously $\left(\theta', 0'_{q-p}\right)'$. Obviously, the embedding of the auxiliary estimator in any step after the first to $\mathbb{R}^q$ is irrelevant and therefore will be dropped. An extension of this three step procedure to an arbitrary number of steps, where the current step auxiliary estimator is the indirect estimator of the previous step, can provide an unbiased indirect estimator of arbitrary order. This extension is the object of this section.

In order to define recursive $\mathrm{GMR2}\,(a_*)$ estimators we need an assumption that would make possible the stochastic approximation of the Edgeworth mean of the IIE employed as an auxiliary one at each step of the procedure.

Assumption A.6 For any $\theta \in \Theta$ and a probability space $(\Omega', \mathcal{F}', P'_\theta)$ and each $j = 1, \ldots, \kappa \in \mathbb{N}^*$ there exist $\zeta_{j_n} : \Omega \times \Omega' \times \Theta^{(j)} \to \mathbb{R}^q$, for $\Theta^{(j)}$ a compact subset of $\mathrm{Int}\,\Theta^{(j-1)}$, that is jointly measurable, Q_θ-almost everywhere continuous on $\Theta^{(j)}$ where $Q_\theta = P_\theta \times P'_\theta$ that satisfy A.3 and A.5.2 in which ξ_j now denotes the j^{th} element of the moment approximation of $\theta_n^{(j-1)}$ that is defined below. The order of derivatives for ζ_{j_n} appearing in A.3 need not be greater than 2.

[12] Given the full rank condition of the jacobian of b at the true θ, theorem 10.2 of Spivak [67] (p. 44) implies a local version of this result.

This assumption enables the following definition.

Definition D.3 Let assumptions A.2, A.5 .1 and A.6 hold for $a = a_* = \frac{\kappa-1}{2}$, and $a^* > \frac{\kappa}{2}$. For $\Theta^{(1)} = \Theta$ and $\theta_n^{(1)}$ any consistent estimator, the recursive $(\kappa - 1) - \mathrm{GMR2}\left(\frac{\kappa-1}{2}\right)$ estimator (denoted by $\theta_n^{(\kappa)}$) is defined by

$$\text{for } 1 < j \leq \kappa,\ \theta_n^{(j)} = \arg\min_{\theta \in \Theta^{(j)}} \left\| \theta_n^{(j-1)} - \theta - \frac{1}{n^{j/2}} \zeta_{j_n}(\theta) \right\|.$$

Remark R.2 (Discussion on Assumption D.3 and Definition D.3) The restrictions on the sequence of $\Theta^{(j)}$ are needed so that assumption A.1 is satisfied (see also footnote 6). If $\theta_n^{(1)}$ is an indirect estimator, assumption A.6 is easily satisfied when assumptions A.2, A.3 and A.5 are satisfied for the auxiliary estimator β_n upon which $\theta_n^{(1)}$ is defined. This is due to the fact that the coefficients appearing in the Edgeworth means, in any step of the recursion defined above, are smooth (multilinear) transformations of the coefficients appearing in the previous step. This follows from that these expansions are derived via Lagrange inversions of approximations of the first order conditions that the estimators asymptotically satisfy with sufficiently high probability in each step. This also means that the principle of analogy immediately implies "natural" approximations (ζ_{j_n}) of the intractable parts of the Edgeworth means at any given step of the recursion given the availability of analogous approximations for the moment expansion of β_n. Hence the difficulty in employing the procedure above rests in the analytical derivation of those Lagrange inversions as functions of the coefficients in the moment expansion of β_n. The following lemma essentially implies that as j converges to κ given the provision of assumption A.3.1 the terms appearing in those inversions become fewer. Finally, the above definition also clarifies the distinction between s^* and s. This is due to the fact that due to Lemma 4.7 the validity of an order s^* Edgeworth expansion for the auxiliary estimator β_n can in principle facilitate the derivation of expansions of order s of $\theta_n^{(j)}$, for any $j \leq \kappa$ and any $s \leq \kappa \leq s^* - 1$ in order for the results on the bias properties of the following lemma to hold, or $s \leq \kappa \leq s^* - 2$ in order for the totality of the results of the following lemma to hold. In this respect κ and thereby a is determined by s^*, i.e. the largest order of a valid Edgeworth expansion available for β_n. Hence κ is the largest order for which an analytically known mean approximation of β_n is available and it less than or equal to the largest order of a valid Edgeworth expansion available for β_n minus 2.

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

Notice first that this lemma partially incorporates the results of corollary 1 and lemma 4.9 since those are specific instances for $\kappa = 2$ and $b = \mathrm{id}$. The lemma says that $\theta_n^{(\kappa)}(a)$ has null asymptotic bias of order κ yet has the same κ^{th} order asymptotic MSE with the $\theta_n^{(\kappa-1)}(a)$, and thereby the same κ^{th} order asymptotic variance again due to the form of the asymptotic bias and the fact that both estimators are asymptotically unbiased up to the $\kappa - 1$ order. This however does not imply that the κ^{th} order asymptotic MSE of $\theta_n^{(\kappa)}(a)$ coincides with the one of $\theta_n^{(1)}$. In any case given the validity of a moment approximation of large enough order for $\theta_n^{(1)}$ this recursive procedure would provide with an asymptotically unbiased estimator of the same order, with (at least) second order asymptotic variance that equals the one of the estimator upon which bias correction is performed.

Second, Arvanitis and Demos [8] define another recursive IIE based on the $\mathrm{GMR2}$ estimator, with similar properties in terms of bias and MSE to the ones of $\theta_n^{(\kappa)}$. However, for $\kappa \geq 2$ their estimator involves, at least, $\kappa - 1$ nested optimizations, casting doubt on the applicability of their estimator for $\kappa \geq 3$. On the other hand $\theta_n^{(\kappa)}$ involves κ sequential optimizations, avoiding all together the nested-optimization problem.

Third, the definition of recursive estimators and lemma 5.1 essentially imply that the dependence of the moment approximation in equation (3) on the (possibly analytically intractable) lbf *is not as restrictive* as it appears to be. In principle given the validity of an Edgeworth approximation (of order greater than three) for β_n depending on possibly analytically intractable, yet simulable, elements, the $1 - \mathrm{GMR2}\left(\frac{1}{2}\right)$ could be defined w.r.t. to a version of $\mathrm{GMR1}$ based on a feasible (possibly numerical) approximation of b. Under our assumption framework this would be 2^{nd} order unbiased while retaining the second order MSE of the $\mathrm{GMR1}$.

Finally, the above justify our *methodological choice* of obtaining moment approximations via the validation of Edgeworth expansions. Even though this is not the only available approach (see for example the approach in Kristensen and Salanie [41]), our methodology enables the result of lemma 5.1 due to the fact that theorem 7.1 does not require the definition of the estimator at hand to rely on a criterion that is of the form of an *arithmetic mean*.[13] Let us now turn our attention to a further example along with some Monte Carlo experiments.

6 Further Example and Monte Carlo Experiments

In this section we employ Monte Carlo experiments for our two examples in order to assess the relevance of our results for finite n and under the influence of numerical optimization errors. We also present a third example involving the $\mathrm{GMR2}(a_*)$ in the

[13]Furthermore the Edgeworth approximations per se could also be useful in other steps of the inferential procedure.

context of a stationary $\mathrm{ARFIMA}\,(0,d,0)$ process and engage to an analogous Monte Carlo study.

Monte Carlo Experiment for the MA(1) Example

For the $\mathrm{MA}\,(1)$ process, we draw random samples of sizes 50, 100, 150, 250, 500, 750, 1000, 1500 and 3000 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 $\mathrm{MA}(1)$ 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.[14] For each retained sample we evaluate four estimators, i.e. the $\theta_n\,(0)$, which is the $\mathrm{GMR1}$, the commonly biased corrected $\theta_n^*\,(1)$ (also called $\mathrm{GMR2}^*\,(1)$), the $\theta_n\,(1)$ (also called $\mathrm{GMR2}\,(1)$), and the $\mathrm{GMR2}-H$, i.e. the $\mathrm{GMR2}$ estimator employing H Monte Carlo experiments to approximate $E_\theta\beta_n$, where $H = \{1, 10, 50, 100, 500, 1000, 1500\}$. We set the number of replications to 100000.

Out of these estimators only the $\mathrm{GMR2}^*\,(1)$ and $\mathrm{GMR2}\,(1)$ ones are 2^{nd} order $\mathrm{unbiased}$. 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.1). 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 $\mathrm{figure}\ 1a$ we present the absolute biases of the $\mathrm{GMR1}$, $\mathrm{GMR2}^*\,(1)$ and $\mathrm{GMR2}\,(1)$ estimators multiplied by n to focus, in such a way, on the second order bias. The analogous second order term corresponding to -0.4 of the benchmark $\mathrm{GMR1}$ equals 1.252. Furthermore, the depicted bias of the $\mathrm{GMR2}\,(1)$ is less than the one of the $\mathrm{GMR2}^*\,(1)$, 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 $\mathrm{figure}\ 2a$. 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 $\mathrm{GMR2}^*\,(1)$. To verify that the skewness and kurtosis do not affect the 2^{nd} order biases and MSEs of the three estimators, we repeated the whole exercise drawing from a standard normal distribution. The results are essentially the same, for more than 250 observations, and are presented in $\mathrm{figure}\ 1b$ and $\mathrm{figure}\ 2b$.

Returning now to the non-central Student's-t distribution errors' case and for $\theta = -0.4$ we have that β_n is outside the $[-0.499999, 0.499999]$ interval in 7377,

[14] Even though this estimator is not defined for sample space regions of possibly positive probability, under our assumption framework this seazes to happen with probability tending to one fast enough and the higher asymptotic order theory described above is valid.

2195, 725, and 67 cases for 50, 100, 150, and 250 number of observations, respectively. For 500 and more number of observations we have no such cases. Consequently, including these cases in the Monte Carlo experiment by setting the $\mathrm{GMR1} = -0.9999$ if $\beta_n < -0.499999$ and $\mathrm{GMR1} = 0.9999$ if $\beta_n > 0.499999$ the estimators' second order biases and MSEs are not affected (see figures $1c$ and $2c$). However, if the true parameter is closer to the boundary a much bigger number of observations than 3000 is needed for our asymptotic results to get through, e.g. if $\theta = -0.9$ we need at least 45000 observations. This is potentially due to the fact that the exact distribution of $\sqrt{n}\,(\mathrm{GMR1} + 0.9)$ for moderate n attributes a large probability to the boundary of the centered and rescaled parameter space (i.e. $\sqrt{n}\,(\Theta + 0.9)$)[15], a property that cannot be captured by the limiting normal distribution or even by the higher order Edgeworth approximations. It has analogous consequences for the quality of the approximation of the bias function of the $\mathrm{GMR1}$ by the Edgeworth mean and thereby on the bias properties of the resulting estimator for the *particular* n.[16] This could be potentially remedied by the consideration of approximations based on appropriate projections of the Edgeworth distribution on $\sqrt{n}\,(\Theta + 0.9)$. It is also possible that the limiting distributions of the so called "local to non-invertibility" asymptotics (see Section 6 in Demos and Kyriakopoulou [20] and Arvanitis [7]) could provide with better approximations of the exact distribution for the aforementioned n. In any case we do not further pursue this matter but choose to leave it for further research.

The evaluation of the $\mathrm{GMR2} - H$ estimators gives us the opportunity to investigate two important questions. The obvious one, i.e. the effect of approximating the $E_\theta \beta_n$ by H Monte Carlo experiments, as well as the time cost associated with the evaluation of these estimators as compared to the one associated with the $\mathrm{GMR2}\,(1)$. We investigate these questions in the sequel. However, a question arises on the choice of the distribution employed for the H Monte Carlo draws. Three possibilities come immediately in mind; the true non-central Student's-t distribution we employ to generate the $y_t' s$ in equation (1), the standard normal distribution, or the empirical distribution constructed by bootstrapping the residuals implied by the $\mathrm{GMR1}$ estimator. We explored all three possibilities and the results are similar, at least for $\mathrm{H} \geq 50$. For $\mathrm{H} = 1$ and $\mathrm{H} = 10$ the biases and MSEs of the Student's-t and normal estimators are further away from their theoretical values as compared to the bootstrap ones. Consequently, in the sequel, and to conserve space, we present only the $\mathrm{GMR2} - H$ estimators emerging from bootstrap results.

In $\mathrm{figure}\,3$ we present the absolute biases, multiplied by n, of the $\mathrm{GMR2} - H$, for $H = 1$, 10 and 1500, estimators, and, for comparison the $\mathrm{GMR1}$. Notice first the erratic behavior of the biases of the $\mathrm{GMR2} - 1$ and $\mathrm{GMR2} - 10$ estimators, as compared to those of the $\mathrm{GMR1}$ and $\mathrm{GMR2} - 1500$ ones, something that could be

[15] Something that is vividly manifested in the Monte Carlo experiment.
[16] This does not contradict the previous results as those are asymptotic.

attributed to the small H. For $\theta = -0.4$, the, multiplied by n, second order bias of the $\mathrm{GMR2}$ estimator equals to 2.094 (see Arvanitis and Demos [8]) and we could expect that $\mathrm{GMR2}-H$ estimator approximates this number for large H. However, even for the $\mathrm{GMR2}-1500$ estimator, more the $n = 1000$ observations are needed to reach this limit. The erratic behavior of the $\mathrm{GMR2}-H$ biases, for small H, can be explained by the MSEs of these estimators, presented in figure 4 in comparison to the analogous MSE of the $\mathrm{GMR1}$ one. From Gourieroux et al. [34] it is known that the asymptotic variance of the $\mathrm{GMR2}-H$ estimator equals that of $\mathrm{GMR1}$ inflated by a factor $1/H$, something apparent in figure 4, for $H = 1$ and $H = 10$. On the other hand the $\mathrm{GMR2}-1500$ MSE approaches very fast the common asymptotic variance of all estimators. In short, it seems that $H > 10$ and $n > 250$ is needed for the $\mathrm{GMR2}-H$ to be close to its theoretical 2^{nd} order bias and MSE.

Finally, in Table 1 we present the average CPU times per iteration of the numerical minimization part of the $\mathrm{GMR2}\,(1)$ and the $\mathrm{GMR2}-H$ estimators, for $H = 1$, 10 and 1500, as representations for the computational costs involved in the analogous numerical derivations. As expected, the numerical optimization concerning the derivation of the $\mathrm{GMR2}\,(1)$ is less costly than that of the $\mathrm{GMR2}-H$.[17] This relative cost is increasing with n and H, something that is straightforwardly attributed to the fact that the $\mathrm{GMR2}-H$ estimator needs H Monte Carlo drawings of length n. Consequently, the larger n is the more time the routine needs to produce each sample. It is worth mentioning that for the $\mathrm{MA}\,(1)$ model the evaluation of the $\mathrm{GMR2}-H$ estimators does not require nested optimizations, as the first step estimator, β_n, is analytically known. We would expect that in more complex cases (the ones that involve numerical procedures also for the derivation of β_n) the relative numerical costs would be even more profoundly in favor of the estimators presented here.

Monte Carlo Experiment for GARCH$(1,1)$ Example

For the $\mathrm{GARCH}\,(1,1)$, we draw random samples of size 150, 250, 400, 550, 750, 900, 1000, 1500, 2000, 3000, 5000 and 10000, plus 250 for initialization, from a standard normal distribution. We perform 3000 replications. For each random sample, we generate the $\mathrm{GARCH}(1,1)$ process y_t with $\theta_1 = 0.1$, $\theta_2 = 0.2$ and $\theta_3 = 0.7$, and we find the MLE of $\theta_i^{\prime}s$, which is our auxiliary vector estimator β_n. As the auxiliary and true model coincide, the binding function is the identity. Consequently, the $\mathrm{GMR1}$ and the auxiliary estimators, β_n, coincide at least asymptotically. We further consider the feasibly bias corrected estimator, suggested in Linton [46] and Iglesias and Linton [38], 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 $\mathrm{GMR2}^*\left(\frac{1}{2}\right)$. The third estimator we employ is the $\theta_n\left(\frac{1}{2}\right)$, also named $\mathrm{GMR2}\left(\frac{1}{2}\right)$. To evaluate the analytically

[17] All experiments have been performed with an Intel i7 processor computer. For all optimizations we employed the E04JBF routine of the NAG Foundation Library Release 1 and for the boostrap in the evaluation of the $\mathrm{GMR2}-H$ we employ the G05DYF routine of the same library.

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 [38] for details).

As in the previous example, in a few cases the $\mathrm{GMR2}^*(1)$ 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 $\mathrm{figure}\ 5a$ we present the norm of the biases of the three estimators multiplied by n, i.e. $n\times\left|\widehat{\mathrm{bias}}\right|$. 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 $\mathrm{GMR1}$ (the MLE). Furthermore, the approximate MSEs of the estimators are presented in $\mathrm{figure}\ 6a$. It seems that 1500 observations are enough for the estimators to reach their common asymptotic variance. The same, more or less, results we get for a second set of parameter values, i.e. for $\theta_1=0.1$, $\theta_2=0.05$ and $\theta_3=0.85$ (see $\mathrm{figure}\ 5b$ and figure $6b$). However, near non-stationarity, i.e. when $\theta_2+\theta_3\lessapprox 1$, we are faced with the same behavior as in the $\mathrm{MA}(1)$ near non-invertibility case.

For example, when $\theta_2+\theta_3=0.2+0.75=0.95$ and for $n=5000$ the bias of the $\theta_n\left(\frac{1}{2}\right)$ is only 10% lower than the one of $\mathrm{GMR1}$. Again the issue seems to be the high probability attributed by the exact distribution of $\sqrt{n}\,(\mathrm{GMR1}-\theta)$ on the boundary of the parameter space for this n. This in this case not only affects the quality of the Edgeworth mean approximation to the true bias function, but also the quality of the stochastic approximation to the Edgeworth mean. The latter complication, that was not present in the $\mathrm{MA}\,(1)$ case, could imply even "larger" approximating errors to the true bias function, further undermining the properties of the resulting $\theta_n\left(\frac{1}{2}\right)$ for the *particular* n. This is in accordance with Lumsdaine [48] from whom we know that the finite-sample distributions of $\theta_{n,2}$ and $\theta_{n,3}$ are skewed, when the parameters are constrained as above. As previously we suspect that this phenomenon could be partially alleviated by the consideration of appropriate projections of the Edgeworth approximations to the centered and rescaled parameter space and the appropriate redefinition of the $\theta_n\left(\frac{1}{2}\right)$. As those projections are currently unavailable, in the near $\mathrm{IGARCH}(1,1)$ case a large value of n is needed to avoid the finite-sample skewness and a large number of random draws are needed to approximate them. These two facts make the evaluation of $\theta_n\left(\frac{1}{2}\right)$ a formidable task, especially in a simulation exercise where a large number of Monte Carlo repetitions is also required. Nevertheless, for large n, say $n\geq 10000$, the bias of the $\mathrm{GMR1}$ (the MLE) estimator is small and consequently, the extra effort for the evaluation of $\theta_n^*\left(\frac{1}{2}\right)$ and/or $\theta_n\left(\frac{1}{2}\right)$, when we suspect that $\theta_2+\theta_3\simeq 1$, is put on doubt.

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

6.1 Example: Stationary Gaussian ARFIMA Process

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

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

In the language of assumption A.1 par is the set of $\mathrm{ARFIMA}\,(0, d, 0)$ processes on Θ which is a compact subset of $\left(0, \frac{1}{2}\right)$. Let ς_n is the likelihood function, and thereby $\varsigma\,(\theta, \beta)$ its expectation. The estimator of d, presented in Sowell [66] (see also Doornik and Ooms [21]), for the time domain is the MLE considered as the auxiliary estimator in our context. Let us denote it with d_n. It maximizes by definition

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

on Θ where G is the variance covariance matrix of $\mathbf{y}$ (see e.g. Lieberman and Phillips [44] for explicit formula of G).

The $\sqrt{n}$-consistency and asymptotic normality of the estimator with asymptotic variance equal to $\frac{6}{\pi^2}$ was established in Dahlhaus [19] (see also Yajima [71]). Thereby the lbf in this example is also the inclusion. Hence, $d_n = \mathrm{GMR1}$ at least with probability tending to one.

d_n has similar asymptotic properties to the Whittle (Fox and Taqqu [25], and Giraitis and Surgailis [31]) but superior to the semiparametric one (Geweke and Porter-Hudak [29], and Robinson [58]). However, there is evidence that the bias of d_n can be severe (see Cheung and Diebold [14], Smith, Sowell and Zin [65], Hauser [37], Lieberman [43], and Doornik and Ooms [21]) and this is our motivation for the consideration of our proposed estimators defined on d_n.

The validity of the Edgeworth expansion for any given order of this estimator is established in Lieberman, Rousseau and Zucker [45] with analytic formulae for the coefficients of the formal expansion presented in Lieberman and Phillips [44] for $s = 2$. Validity is essentially established in a similar manner to theorem 7.1 along with the validation of the Assumptions 2-4 of Durbin [22] for the derivatives of ς_n. Hence, from the formulae in Lieberman and Phillips [44] 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

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

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

where $\dot{G}$ and $\ddot{G}$ the first and second derivative of G. Thereby assumption A.2 follows for $s = s_* = 2$.

Due to $p = q = 1$ we can again assume that W_n is the identity *without loss of generality*. Hence assumption A.4 follows trivially and m_n^* is $d_n - d$. Also $\zeta_{i_n} = \xi_i = \mathcal{I}_{V_\theta}\left(k_i\left(z,\theta\right)\right)$ and thereby assumption A.3.2 and the second part of 3 follow trivially given the relevant smoothness. Hence A.5 follows also trivially from the previous discussions. Smoothness for the coefficients of the Edgeworth densities as functions of d due to the relevant smoothness of the Gamma function imply finally the validity of A.3.1 and 3. Hence our assumption framework is satisfied.

Consequently $d_n\left(\frac{1}{2}\right)$, also denoted $\mathrm{GMR2}\left(\frac{1}{2}\right)$, is defined as

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

Due to the previous corollary 1 and lemma 4.9 apply and therefore $d_n\left(\frac{1}{2}\right)$ is 2^{nd} order unbiased and has the same MSE, up to $O\left(n^{-\frac{1}{2}}\right)$, with d_n.

Furthermore, let us call d_n^* the approximate bias corrected estimator of Lieberman [43], adapted for this case from Firth [24]. d_n^* is given by:

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

where $\boldsymbol{\zeta}\left(.\right)$ 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 [43] estimator, d_n^*, is our $d_n^*\left(\frac{1}{2}\right)$, and it is also denoted by $\mathrm{GMR2}^*\left(\frac{1}{2}\right)$.

Monte Carlo Experiment

In figure $7a$ we present the $n \times \left|\widehat{\text{bias}}\right|$ for the three estimators, for sample sizes 20, 40, 50, 60, 70, 80, 90, 100 and 120, and $d = 0.4$.[18] All estimators were obtained by a simple grid search, of length 10^{-3}, 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 [44]). 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 $d_n^*\left(\frac{1}{2}\right)$ is smaller as compared to the one of $d_n\left(\frac{1}{2}\right)$, which in turn is smaller than the one of $\mathrm{GMR1}$ (see figure $8a$). The results are consistent with those for $d_n^*\left(\frac{1}{2}\right)$ and d_n in Lieberman [43]. 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 than the one of $d_n\left(\frac{1}{2}\right)$ for $40 \leq n \leq 100$.

Comparing the Local Whittle estimator, in Shimotsu and Phillips [62], approximate bias and MSE with those of $d_n\left(\frac{1}{2}\right)$ we see that the estimated approximate bias and MSE of $d_n\left(\frac{1}{2}\right)$ is smaller, i.e. the bias of the their estimator is 2.45 and its MSE is 3.10 as compared to 0.31 and 0.47 of $d_n\left(\frac{1}{2}\right)$, respectively (see Table 1 in Shimotsu and Phillips [62]). Further, from Table 1 in Shimotsu and Phillips [61] we see that, for $d = 0.3$, the exact Whittle estimator approximate bias and MSE are 1.00 and 3.05, respectively. These values are higher than the equivalent ones of $d_n\left(\frac{1}{2}\right)$ (see figure $7b$ and figure $8b$), which are, respectively, 0.22 and 0.60. Finally, from Table 1 in Nielsen [54] we see that the approximate bias and MSE of the Local Polynomial Whittle estimator of Andrews and Sun [5], for $d = 0.3$, are 24.06 and 133.12, respectively, whereas the equivalent values for the Extended Local Whittle estimator of Abadir, Distaso and Giraitis [1] are 3.63 and 40.60 and those of the Extended Local Polynomial Whittle estimator of Nielsen [54] are 32.46 and 194.56. All these values are bigger than the equivalent one for the $d_n\left(\frac{1}{2}\right)$ estimator (see figure $7b$ and figure $8b$).

Hence it seems that $d_n\left(\frac{1}{2}\right)$ has smaller bias and MSE as compared to the Whittle-type estimators. However, these type of estimators have other advantages as compared to the time domain MLE ones (see the above mentioned articles). Notice that employing the results in Giraitis and Robinson [30], the same procedure can be applied to the semiparametric Whittle estimator. Nevertheless, space conservation considerations prohibits an extensive comparison between the two types of estimators and/or the bias corrected ones.

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

7 Conclusions

In this paper we define a set of Indirect Inference 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 possess:

First, computational facility, as they are derived from procedures avoiding the nested numerical optimization burden. This is due to the fact that any Monte Carlo integration involved in their definition is with respect to analytically tractable integrands. This is not the case with the simulated analog of the $\mathrm{GMR2}$ estimator since this involves analogous integration with respect to analytically intractable $\arg\min$ functionals. This numerical facility comes at the *fixed cost* of the analytical derivation of the approximation. It is mostly useful in cases where the lbf is (locally) the identity (or the inclusion function in our framework).

Second, the lbf is (locally) the identity when it has locally full rank via a canonical reparameterization of the auxiliary model. The analytical intractability of the reparameterization can be overcome via the *recursive* employment of those estimators. The latter exhibit an even greater computational advantage since they involve sequential optimizations as opposed to the nested ones associated with the recursive $\mathrm{GMR2}$ estimator.

Third, the $\mathrm{GMR1}$ 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 the considered estimators and the $\mathrm{GMR2}$ one, as well as any pair of these estimators.

Fourth, and more generally, some of their higher order asymptotic properties coincide with those of the $\mathrm{GMR2}$ estimator. However it *seems* that these properties can be established via assumption frameworks that contain less restrictive requirements, for the asymptotic behavior of the random elements involved. For example, consider the case where $p = q$, we have that $\beta_n = E_{\mathrm{GMR2}}\beta_n$ and $\beta_n = b_n\left(\mathrm{GMR2}\left(a_*\right), \zeta_n\left(\mathrm{GMR2}\left(a_*\right), a_*\right)\right)$ with P_θ probability bounded by $1 - o\left(n^{-a^*}\right)$ independent of θ. When $a_* > 0$ and $\zeta_{i_n} = \xi_i$ for all i, it can be seen by a direct comparison of the current assumption framework and the one employed in Lemmas 2.5.ii, 3.6 and Corollary 2 of Arvanitis and Demos [8] that the two estimators are second order equivalent. However, this result is obtained for the $\mathrm{GMR2}$ estimator via locally uniform Edgeworth approximations for the auxiliary estimator, something that need not be the case for the $\mathrm{GMR2}\left(a_*\right)$ ones. Notice that first, the two assumption frameworks cannot be "globally" compared due to the fact that the one corresponding to the $\mathrm{GMR2}\left(a_*\right)$ involves also restrictions on the asymptotic properties of several stochastic approximations not present in the $\mathrm{GMR2}$ case. Second, the assumptions are only sufficient.

Finally, we demonstrate that under our assumption framework and in the special case of deterministic weighting and affinity of the binding function, the $\mathrm{GMR2}(a_*)$ estimator for any $a_* > 0$, is second order unbiased. Furthermore, for a given order of approximation, we provide a procedure that yields recursive Indirect Inference estimators that are approximately unbiased of that order. Moreover, the approximate MSE of the unbiased $\mathrm{GMR2}(a_*)$ presented in this paper, even when these are derived via the use of Monte Carlo and/or Bootstrap sampling techniques, does not depend on the cardinality of those samples. At the same time their practical implementation does not appear as numerically involved as an analogous procedure defining recursive $\mathrm{GMR2}$ estimators.

Furthermore, the $\mathrm{GMR2}(a_*)$ estimators provide an IIE framework that incorporate some "classical" bias correction procedures. These can be perceived as extreme cases in our definitions. They are associated with minimal numerical cost however they can be outperformed by other $\mathrm{GMR2}(a_*)$ estimators for finite n due to their behavior on the boundary of the parameter space.

Our methodology can not be applied in cases where θ and/or $b(\theta)$ lies on the boundary of the relevant parameter space. In these cases, the definition of the $\mathrm{GMR2}(a_*)$ estimators as well as the derivation of their asymptotic properties need a different approach. In this respect Andrews [3] may be useful. We leave these issues for future research.

References

[1] Abadir, K.M, W. Distaso and L. Giraitis (2007), "Non-stationaity-extended local Whittle estimation, *Journal of Econometrics* 141, 1353 - 1384.

[2] Aliprantis, C.D. & K.C. Border (2005), "*Infinite Dimensional Analysis*", Third Edition, Springer.

[3] Andrews, D.W.K. (2002), "Estimation When a Parameter is on a Boundary" *Econometrica* 67, 1341–1383.

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

[5] Andrews, D.W.K. and Sun, Y. (2004), "Adaptive local polynomial whittle estimation of long-range dependence" *Econometrica* 72, 569-614.

[6] 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.

[7] Arvanitis, S. (2014), "A simple example of an Indirect Estimator with discontinuous limit theory in the MA(1) model" forthcoming *Journal of Time Series Analysis* DOI: 10.1111/jtsa.12080.

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

[9] Arvanitis S. and A. Demos (2014) "Valid Locally Uniform Edgeworth Expansions under Weak Dependence and sequences of Smooth Transformations", *Journal of Time Series Econometrics* 6, 183-235.

[10] 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.

[11] Bao, Y. & A. Ullah (2007), "Finite sample properties of maximum likelihood estimator in spatial models", *Journal of Econometrics* 137, 396–413.

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

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

[14] 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.

[15] Cordeiro G.M. & P. McCullagh (1991) "Bias correction in generalized linear models", *Journal of the Royal Statistical Society B* 53, 629-643.

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

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

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

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

[20] Demos, A. & Kyriakopoulou D. (2013), "Edgeworth and Moment Approximations: The Case of MM and QML Estimators for the MA(1) Models", *Communications in Statistics-Theory and Methods* 42, 1713-1747.

[21] 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.

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

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

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

[25] 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.

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

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

[28] Garcia, R., E. Renault and D. Veredas (2011), "Estimation of stable distributions by indirect inference", *Journal of Econometrics* 161, 325–337.

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

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

[31] 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.

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

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

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

[35] Gourieroux, C., P.C.B. Phillips and J. Yu (2010), "Indirect inference for dynamic panel models" *Journal of Econometrics* 157, 68-77.

[36] 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.

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

[38] 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.

[39] Iglesias, E.M. & G.D.A. Phillips (2011), "Small Sample Estimation Bias in GARCH Models with any number of Exogenous Variables in the Mean Equation", *Econometric Reviews* 30, 303–336.

[40] Iglesias, E.M. & G.D.A. Phillips (2012), "Estimation, Testing, and Finite Sample Properties of Quasi-Maximum Likelihood Estimators in GARCH-M Models", *Econometric Reviews* 31, 532–557.

[41] Kristensen and Salanie (2010), "Higher order improvements for approximate estimators", CAM Working Papers with number 2010-04, University of Copenhagen.

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

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

[44] 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.

[45] 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.

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

[47] 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.

[48] 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.

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

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

[51] Magnus, J.R. (1986), "The Exact Moments of a Ratio of Quadratic Forms in Normal Variables", Annales d'Economie et de Statistique, 4, 95-109.

[52] Nagar, A.L. (1959), "The bias and moment Matrix of the general k-class Estimators of the parameters in simultaneous equations", *Econometrica* 27, 573-595.

[53] Newey, W.K. and R.J. Smith (2004), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators", *Econometrica* 72, 219-255.

[54] Nielsen, F.S. (2008), "Local polynomial Whittle estimation covering non-stationary fractional processes" CREATES Research paper 2008-28.

[55] Northcott, D.G. (2009), "*Multilinear Algebra*", Cambr. Univ. Press, N. York.

[56] Phillips P.C.B. (2012), "*Folklore Theorems, Implicit Maps and Indirect Inference*", *Econometrica* 80, 425-454.

[57] 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.

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

[59] Rothenberg, T.J. (1984), "Approximating the distributions of econometric estimators and test statistics", in *Handbook of Econometrics* Vol. 2, ed. by Z. Griliches and M.D. Intriligator, New York, North-Holland.

[60] Sargan, J.D. (1974), "The validity of Nagar's expansion for the moments of econometric estimators", *Econometrica* 42, 169-176.

[61] Shimotsu, K. and P.C.B. Phillips (2005), "Exact Local Whittle estimation of fractional integration", The Annals of Statistics 33, 1890–1933.

[62] Shimotsu, K. and P.C.B. Phillips (2006), "Local Whittle estimation of fractional integration and some of its variants" *Journal of Econometrics* 130, 209–233

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

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

[65] Smith, A.A., F.B. Sowell & S.E. Zin (1997), "Fractional integration with drift: Estimation in small samples", *Empirical Economics* 22, 103-16.

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

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

[68] Srinivasan, T.N. (1970), "Approximations to finite sample moments of estimators whose exact sampling distributions are unknown", *Econometrica* 18, 533-541.

[69] Ullah, A. (2004). "*Finite Sample Econometrics*", Oxford University Press, N. York.

[70] Ullah, A. and V.K. Srivastava (1994), "Moments of the ratio of quadratic forms in non-normal variables with econometric examples", *Journal of Econometrics* 62, 129-141.

[71] 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

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

■

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 [50] 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 $\theta \in \operatorname{Int}\Theta$

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

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

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

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

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

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

Proof of Lemma 4.3. Likewise to the previous proof set

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

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

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

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

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

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

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

Now, from the definition of $\mathrm{GMR1}$, the triangle inequality, norm submultiplicativity, assumptions A.1, A.3 and A.5, lemma 2.1 and the subsequent lemmas 2.2 and 4.1, and by choosing appropriately $C, M>0$ we obtain

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

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 $Q_\theta\left(\left\|\left(D^2J_n\left(\theta_n^{++}\right)\right)^{-1}\right\|>C\right)=o\left(n^{-a^*}\right)$ and therefore we obtain the needed result if we choose $\gamma'_n \leq C^*\rho_n$. ■

Proof of Lemma 4.64.6. For the $i)$ part notice that he result follows directly from AC.1 in appendix B due to lemma 4.5. For part $ii)$, first notice that from lemma 4.6 and Lemma 2 of Magdalinos [50] 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 $1-o\left(n^{-a^*}\right)$ $\sqrt{n}\left(\theta_n\left(a_*\right)-\theta\right) = \kappa_n\left(\sqrt{n}m_n^*\left(\theta\right)\right)+R_n$ where κ_n is a polynomial function for which we have that

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

and g_n as in equation 6, lemma AL.4, due to assumptions A.3, A.5 while $Q_\theta\left(\left\|R_n\right\|>\gamma_n\right)=o\left(n^{-a^*}\right)$ for some $\gamma_n=o\left(n^{-a^*}\right)$ due to the previous. Hence the result follows from the first part of lemma AL.4. ■

Proof of Lemma 4.7. The result follows from lemmas 4.6, AL.2, AL.4, and the fact that Θ is compact. ■

Proof of Lemma 4.8. The proof is based on lemma 4.7 for $m=1$. We essentially compute $K_j\left(g_n\left(z\right)\right)\left(1+\sum_{i=1}^{s}\frac{1}{n^{\frac{i}{2}}}\pi_i\left(z\right)\right)$ for $K_j\left(x\right)=x_j$, $j=1,\ldots,p$ for each of the estimators at hand. Using assumptions A.1, A.3.2-3, A.4, A.5, lemma 4.6 and Lemma 2 of Magdalinos [50] we obtain by a second order Taylor expansion that $\sqrt{n}\left[\beta_n-b_n\left(\theta_n,\zeta_n\left(\theta_n,a_*\right)\right)\right]$, if $a_*>0$, is approximated by

$$\sqrt{n}\left(\beta_n-b\left(\theta\right)\right)-\frac{\partial b}{\partial\theta'}\sqrt{n}\left(\theta_n-\theta\right)-\frac{1}{2\sqrt{n}}\left[\sqrt{n}\left(\theta_n-\theta\right)'\frac{\partial b_j}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n-\theta\right)\right]_{j=1,\ldots,q}$$
$$-\sum_{i=2}^{s_*}\frac{1}{n^{\frac{i-1}{2}}}\zeta_{i_n}\left(\theta\right)-\sum_{i=2}^{s_*}\frac{1}{n^{i/2}}\frac{\partial\zeta_{i_n}\left(\theta\right)}{\partial\theta'}\sqrt{n}\left(\theta_n-\theta\right)+R_{1,n}\left(\theta_n,\theta\right),$$

in the sense that $Q_\theta\left(\sup_{\theta'\in\overline{\mathcal{O}}_\varepsilon\left(\theta\right)}\left\|R_{1,n}\left(\theta_n,\theta'\right)\right\|>o\left(n^{-\frac{1}{2}-\delta}\right)\right)=o\left(n^{-a_*}\right)$, for $\delta>0$. Analogously the term $\frac{\partial b_n'\left(\theta_n,\zeta_n\left(\theta_n,a_*\right)\right)}{\partial\theta}$ is approximated by

$$\frac{\partial b'}{\partial\theta}+\frac{1}{\sqrt{n}}\left[\sqrt{n}\left(\theta_n-\theta\right)'\frac{\partial^2 b'}{\partial\theta\partial\theta_j}\right]_{j=1,\ldots,p}$$
$$+\sum_{i=2}^{s_*}\frac{1}{n^{i/2}}\frac{\partial\left(\zeta_{i_n}'\left(\theta\right)\right)}{\partial\theta}+\sum_{i=2}^{s_*}\frac{1}{n^{\frac{i+1}{2}}}\left[\sqrt{n}\left(\theta_n-\theta\right)'\frac{\partial^2\zeta_{i_n}'\left(\theta\right)}{\partial\theta\partial\theta_j}\right]_{j=1,\ldots,p}+R_{2,n}\left(\theta_n,\theta\right),$$

where again $Q_\theta\left(\sup_{\theta''\in\overline{\mathcal{O}}_\varepsilon\left(\theta\right)}\left\|R_{2,n}\left(\theta_n,\theta''\right)\right\|>o\left(n^{-\frac{1}{2}-\delta'}\right)\right)=o\left(n^{-a_*}\right)$, for $\delta'>0$. Finally, $W_n\left(\theta_n^+\right)$ is approximated by

$$W\left(\theta\right)+\frac{1}{\sqrt{n}}k_{1_w}+\frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\theta'}W\left(\theta\right)_{j,j'}k_{1_{\theta^+}}\right]_{j,j'=1,\ldots,q}+R_{3,n}\left(\theta\right),$$

where $P_\theta\left(\sup_{\theta'''\in\overline{\mathcal{O}}_\varepsilon\left(\theta\right)}\left\|R_{2,n}\left(\theta'''\right)\right\|>o\left(n^{-\frac{1}{2}}\right)\right)=o\left(n^{-a_*}\right)$. Therefore an asymptotic polynomial approximations of the first order conditions which the estimator satisfies with P_θ probability $1-o\left(n^{-a_*}\right)$, for $a_*>0$, is given by

$$
\begin{aligned}
0 \;=\;& \mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\sqrt{n}\left(\beta_n-b\left(\theta\right)\right)-\frac{1}{\sqrt{n}}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\zeta_{2_n}\left(\theta\right) \qquad (4)\\
&+\frac{1}{\sqrt{n}}\mathcal{C}^{-1}\left[\begin{array}{c}\left[\sqrt{n}\left(\theta_n-\theta\right)'\frac{\partial^2 b'}{\partial\theta\partial\theta_j}\right]_{j=1,\ldots,p}W\\ +\frac{\partial b'}{\partial\theta}\left\{k_{1_w}+\left[\frac{\partial}{\partial\theta'}W_{j,j'}k_{1_{\theta+}}\right]_{j,j'=1,\ldots,q}\right\}\end{array}\right]\left\{\sqrt{n}\left(\beta_n-b\left(\theta\right)\right)-\frac{\partial b}{\partial\theta'}\sqrt{n}\left(\theta_n-\theta\right)\right\}\\
&-\frac{1}{2\sqrt{n}}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\left[\sqrt{n}\left(\theta_n-\theta\right)'\frac{\partial b_j}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta_n-\theta\right)\right]_{j=1,\ldots,q}+R_n\left(\theta_n,\theta\right)
\end{aligned}
$$

where $W=W\left(\theta\right)$, $\mathcal{C}=\frac{\partial b'}{\partial\theta}W\frac{\partial b}{\partial\theta'}$ and $Q_\theta\left(\sup_{\theta^*\in\overline{\mathcal{O}}_\varepsilon(\theta)}\left\|R_n\left(\theta_n,\theta^*\right)\right\|>o\left(n^{-\frac{1}{2}}\right)\right)=o\left(n^{-a_*}\right)$. Denote by $g_n\left(z,a\right)$ the $g_n\left(z\right)$ corresponding to $\mathrm{GMR2}\left(a\right)$. Obtaining the Laplace inversion of the expansion in 4 w.r.t. $\sqrt{n}\left(\theta_n-\theta\right)$ and discarding terms we get that, for $a_*>0$,

$$
\begin{aligned}
g_n\left(z,a_*\right) \;=\;& \mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\left(k_{1_\beta}-\frac{\xi_2}{\sqrt{n}}\right)\\
&+\frac{1}{\sqrt{n}}\mathcal{C}^{-1}\left(\begin{array}{c}\left[\left(\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\right)'\frac{\partial^2 b'}{\partial\theta\partial\theta_j}\right]_{j=1,\ldots,p}W\\ +\frac{\partial b'}{\partial\theta}k_{1_{w^*}}+\frac{\partial b'}{\partial\theta}\left[\frac{\partial}{\partial\theta'}W_{j,j'}k_{1_{\theta^*}}\right]_{j,j'=1,\ldots,q}\end{array}\right)\left(Id_q-\frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\right)k_{1_\beta}\\
&-\frac{1}{2\sqrt{n}}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\left[\left(\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\right)'\frac{\partial b_j}{\partial\theta\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}Wk_{1_\beta}\right]_{j=1,\ldots,q}.
\end{aligned}
$$

The result for $a_*=0$ is obtained analogously to the previous case simply by setting ζ_{2_n} equal to zero wherever this term appears in equation 4, and consequently

$$
g_n\left(z,0\right)=g_n\left(z,a_*\right)+\frac{1}{\sqrt{n}}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W\xi_2.
$$

Integrating in each case with respect to $\left(1+\frac{\pi_1(z,\theta)}{\sqrt{n}}\right)\varphi_{V(\theta)}\left(z\right)$,noting that $k_{2_\beta}\left(z,\theta\right)=z\pi_1\left(z,\theta\right)$ we obtain the needed results due to lemma AL.2 and to the fact that by construction $\left\|\mathcal{I}_{\varphi_V}\left(k_{2_\beta}\right)-\xi_2\right\|=o\left(1\right)$. ■

Proof of Lemma 4.9. Argue as in the previous proof and note that now $K_{i,j}\left(x\right)=x_ix_j$, $i,j=1,\ldots,p$. ■

Proof of Lemma 5.1. First notice that the existence of $\theta_n^{(j)}$ at any step of the procedure is ensured by the continuity implied in assumption A.6. Then the assumptions stated in the lemma along with the bound of the rate of divergence for ζ_{j_n} implied again by A.6 establish that for any j, any θ in $\mathrm{Int}\,\Theta^{(j)}$ and any $\varepsilon>0$, $Q_\theta\left(\left\|\theta_n^{(j)}-\theta\right\|>\varepsilon\right)=o\left(n^{-a^*}\right)$ due to lemma 4.1. Then a recursive application of lemmas 4.6 and 4.7 following from the assumptions in the current lemma and A.6 and the relation between $j \leq \kappa$ and s_*, implies that the distribution of $\sqrt{n}\left(\theta_n^{(j)} - \theta\right)$ under Q_θ admit a valid Edgeworth expansion of order s^* and that the ξ_j in assumption A.6 are well defined. 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. The proof for the moment approximations for the case $i = 2$ follows easily as a special case of corollary 1 and lemma 4.9. Using induction if these hold for some i, then notice that since ξ_j would be zero for any j less than a first order Taylor expansion the first order conditions around θ satisfied by $\theta_n^{(i+1)}$ with P_θ probability $1 - o\left(n^{-a^*}\right)$ is of the form

$$
\begin{aligned}
0_p &= \sqrt{n}\left(\theta_n^{(i)} - \theta\right) - \frac{\xi_{i+1}(\theta)}{n^{\frac{i}{2}}} - \left(\frac{\zeta_{i+1_n}(\theta)}{n^{\frac{i}{2}}} - \frac{\xi_{i+1}(\theta)}{n^{\frac{i}{2}}}\right) \\
&\quad + \left(\mathrm{Id}_p + \frac{D\zeta_{i+1_n}(\theta)}{n^{\frac{i+1}{2}}}\right)\sqrt{n}\left(\theta_n^{(i+1)} - \theta\right) + u_n
\end{aligned} \tag{5}
$$

where $Q_\theta\left(\|u_n\| > o\left(n^{-\frac{i}{2}}\right)\right) = o\left(n^{-a^*}\right)$ since

$$
\|u_n\| \leq n^{-\frac{i+1}{2}} \sup_{\theta' \in \overline{O}_\varepsilon(\theta)} \left\|\mathrm{vec}\, D^2\zeta_{i+1_n}^{(i)}(\theta')\right\| \left\|\sqrt{n}\left(\theta_n^{(i+1)} - \theta\right)\right\|
$$

due to A.6. The result follows by as in the previous proof by obtaining the Laplace inversion of 5 w.r.t. $\sqrt{n}\left(\theta_n^{(i+1)} - \theta\right)$ discarding terms of the appropriate order, integrating what remains and its exterior product with respect to the Edgeworth distribution of $\sqrt{n}\left(\theta_n^{(i)} - \theta\right)$ and noting that due to that due to A.1, A.2, A.5 and lemma AL.2 $\left\|\mathcal{I}_{\varphi_V}\left(k_{i+1_{\theta(i}}\right) - \xi_{i+1}\right\| = o(1)$. ■

B General Theorems, 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. Recall that $M_n(\theta) = \sqrt{n}(\beta_n - b(\theta))$. Then

Theorem 7.1 Suppose that:
-POLFOC $M_n(\theta)$ satisfies $0_{p\times 1} = \sum_{i=0}^{s-1} \frac{1}{n^{i/2}} \sum_{j=0}^{i+1} C_{ij_n}(\theta)\left(M_n(\theta)^j, S_n(\theta)^{i+1-j}\right) +$ $R_n(\theta)$ with probability $1 - o\left(n^{-\frac{s-1}{2}}\right)$ where $C_{ij_n} : \Theta \times \mathbb{R}^{q^{i+1}} \to \mathbb{R}^p$ is $(i+1)$-linear $\forall\theta \in \Theta$, $C_{00_n}(\theta), C_{01_n}(\theta)$ are independent of n and have rank p $\forall\theta \in \Theta$, C_{ij_n} are equicontinuous on Θ, $\forall x^{i+1}$,
-LUE $S_n(\theta)$ is a sequence of random elements admitting an Edgeworth expansion of

order s with polynomials (of its density) which are equicontinuous on Θ and a variance matrix which is continuous on Θ and positive definite,
-UAT $P\left(\|M_n(\theta)\| > C\ln^{1/2} n\right) = o\left(n^{-\frac{s-1}{2}}\right)$ for some $C > 0$,
-USR $P\left(\|R_n(\theta)\| > \gamma_n\right) = o\left(n^{-\frac{s-1}{2}}\right)$ for some real sequence $\gamma_n = o\left(n^{-\frac{s-1}{2}}\right)$.
Then $M_n(\theta)$ admits an Edgeworth expansion of order s the polynomials of the density of which are equicontinuous on Θ and the variance matrix is continuous on Θ and positive definite.

The first two imply the existence of a random vector $S_n(\theta)$ for which an analogous Edgeworth expansion exists. The first is usually derived by asymptotic polynomial approximations of the f.o.c. (first order conditions) that the estimator asymptotically satisfies with high probability.

Hence the needed result rests upon the verification of conditions POLFOC, LUE, UAT and USR in the above theorem. In the case that β_n is an MLE or a GMM estimator S_n is a random vector consisting of the random elements appearing in the derivatives of the likelihood function or the moment conditions and therefore is of the form of a normalized sum. Then the establishment of the condition LUE relies on the properties of these random elements. For example in the context of weakly dependent time series models the conditions of Gotze and Hipp [32] (Assumptions 1-4) suffice for LUE. For a more general set of sufficient conditions see also the Assumptions 2-4 of Durbin [22] (Andrews and Lieberman [6] establish these for S_n comprised by the elements of the derivatives of Whittle likelihood function in the context of Gaussian ARFIMA processes). In the case where β_n is itself an indirect estimator this can be established by a re-iteration of this procedure.

Condition UAT can be verified via the use of LUE and a mean value expansion of the f.o.c. Finally condition USR can be verified due to the form of the remainder in the approximation of POLFOC (which is usually a remainder emerging from the application of the mean value Theorem) and is bounded in norm by $C\|S_n\|^d$ for d large enough and $C > 0$ and LUE.

This procedure obviously fails if $\sqrt{n}\left(\beta_n - b(\theta)\right)$ is not asymptotically normal. This in turn can happen because $b(\theta)$ lies in the boundary of B and thereby POLFOC cannot hold (notice that $b(\theta)$ can be a boundary point even if θ is an interior point). Finally notice that this procedure might fail to hold (essentially due to failure of LUE) even in cases where S_n is asymptotically normal and $b(\theta)$ in an interior point. Consider for example the case where the necessary conditions of Theorem 1 (see the final pair of paragraphs of the proof in page 508) in Corradi and Iglesias [17] fail to hold for the QMLE in the context of a semi-parametric $\mathrm{GARCH}(1,1)$ model.

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 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 $a_n = o\left(n^{-\varepsilon}\right)$ and $P\left(\|x_n\| > a_n\right) = o\left(n^{-a}\right)$. Then any η_n, such that $P\left(\zeta_n + x_n = \eta_n\right) = 1 - o\left(n^{-a}\right)$, 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,

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

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

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

which is $o\left(n^{-a}\right)$ since $A - a_n$ is convex. Now, for an appropriate $C > 0$, which exists due to Lemma 2 of Magdalinos [50], and $\mathcal{H}_n\left(C\right) = \left\{x \in \mathbb{R}^q : \|x\| > C \ln^{1/2} n\right\}$.

$$\begin{aligned}
& \int_{A-a_n} \left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}} \pi_i\left(y\right)\right) \varphi_V\left(y\right) dy \\
=\ & \int_{A\cap\mathcal{H}_n(C)} \left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}} \pi_i\left(z - a_n\right)\right) \varphi_V\left(z - a_n\right) dz + o\left(n^{-a}\right)
\end{aligned}$$

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

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

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,

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

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

$$\begin{aligned} &\int_{A\cap\mathcal{H}_n(C)}\left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}}\pi_i(z - a_n)\right)\varphi_V(z - a_n)\,dz \\ = &\int_{A\cap\mathcal{H}_n(C)}\varphi_V(z)\left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}}\pi_i^*(z)\right)dz + \int_A q_n(z)\,dz \end{aligned}$$

and

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

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 and that $\int_{\mathcal{H}_n^c(C)}\varphi_V(z)\left(1 + \sum_{i=1}^{2a+1} n^{-\frac{i}{2}}\pi_i^*(z)\right)dz = o(n^{-a})$, 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. See the proof of Lemma 3.1 of Arvanitis and Demos [8]. ■

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. See Lemma AL.1 of Arvanitis and Demos [9]. ■

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

$$g_n(x) = Lx + \sum_{i=2}^{s_*}\left(n^{-\left(\frac{i-1}{2}\right)}\xi_i + o\left(n^{-\left(\frac{i-1}{2}\right)}\right)\right) + \sum_{j=2}^{s^*}\frac{1}{n^{\frac{j-1}{2}}}\left(B_{j_n} + o\left(n^{-\delta}\right)\right)x^j \tag{6}$$

for large enough n, with $\mathrm{rank}\, L = p$. 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

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

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

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

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

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

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 [50] 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

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

Due to the proof of lemma 3.5 of Skovgaard [63] $H_n(C_*) \subseteq g_n(H_n(C))$ for some $C_* > C$, hence this equals

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

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

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

as in the proof of the first part of lemma 4.6 of Skovgaard [63] 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 [63]. ■

Figures and Table

The $\mathrm{MA}(1)$ Model, $\theta=-0.4$ non-central Student's-t.

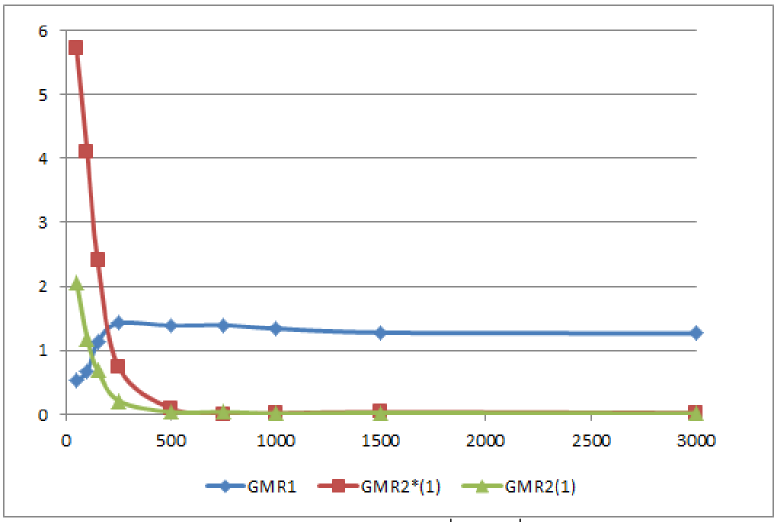

Figure 1a: $n\times\left|\widehat{\mathrm{Bias}}\right|$.

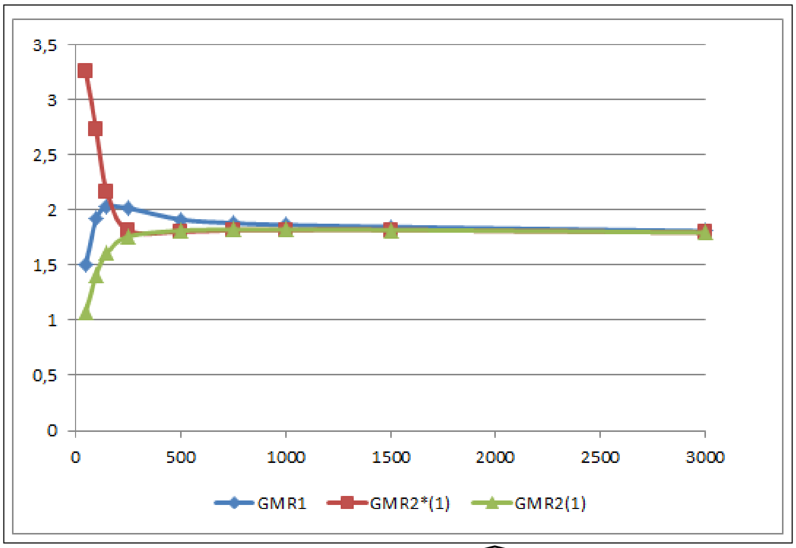

Figure 2a: $n \times \widehat{\mathrm{MSE}}$.

The $\mathrm{MA}\,(1)$ **Model,** $\theta = -0.4$ **normal.**

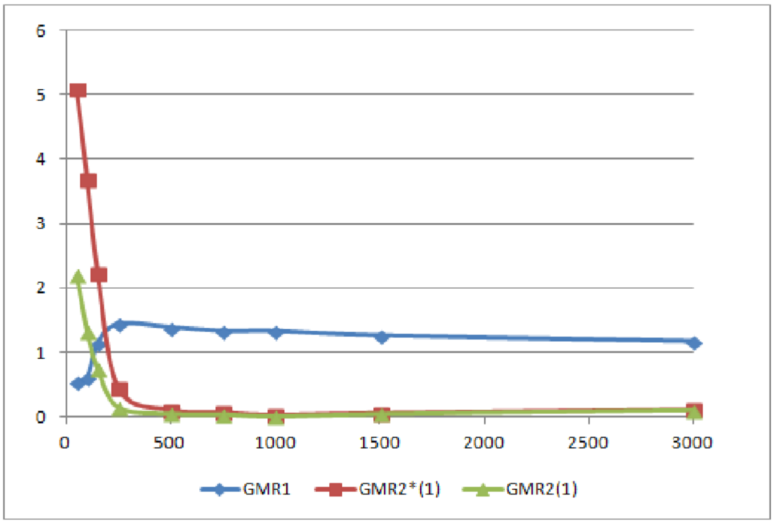

Figure 1b: $n \times \left|\widehat{\mathrm{Bias}}\right|$.

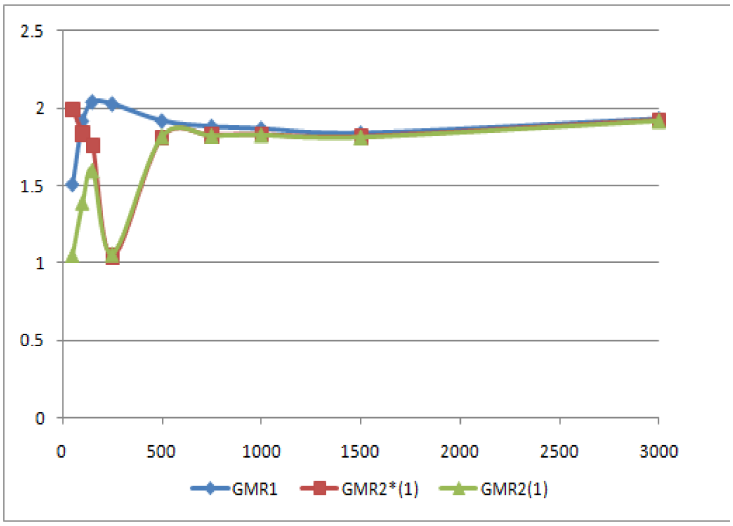

Figure 2b: $n \times \widehat{\mathrm{MSE}}$.

The $\mathrm{MA}\,(1)$ **Model,** $\theta = -0.4$ **non-central Student's-**t**.**

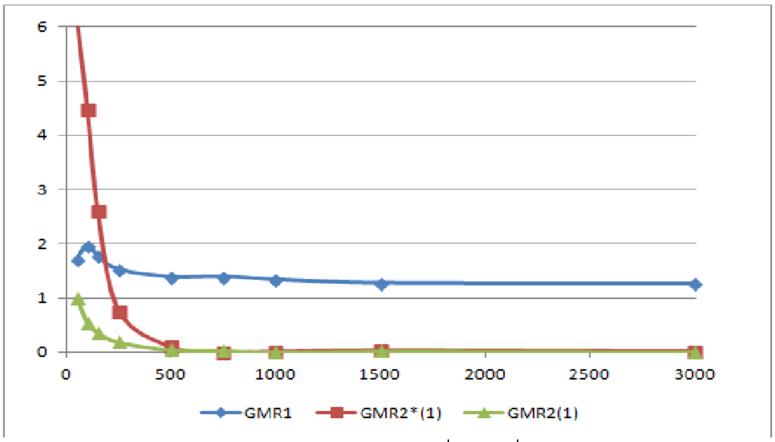

Figure 1c: $n \times \left|\widehat{\mathrm{Bias}}\right|$.

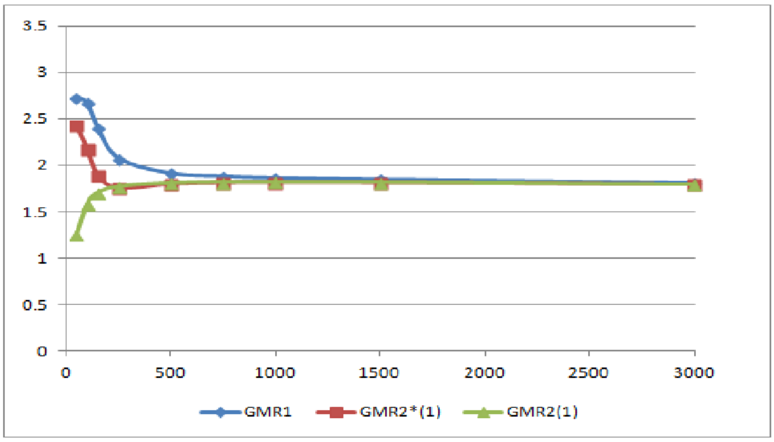

Figure 2c: $n \times \widehat{\mathrm{MSE}}$.

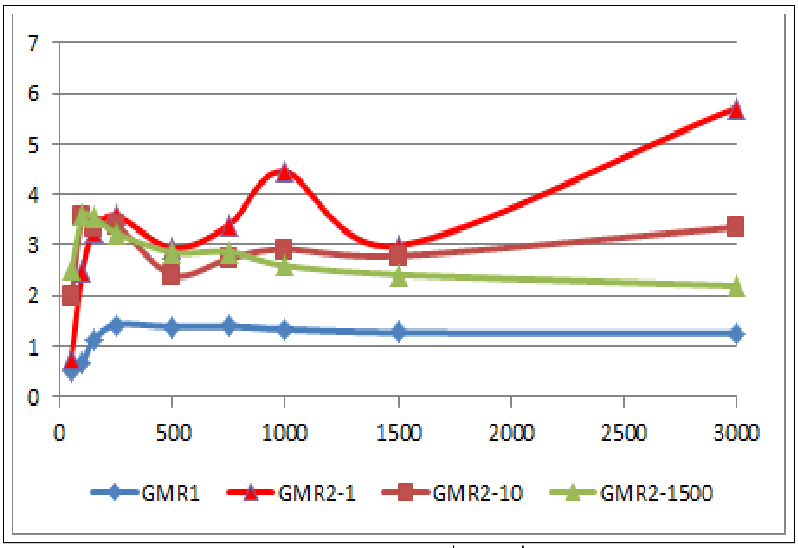

Figure 3: $n \times \left|\widehat{\mathrm{Bias}}\right|$.

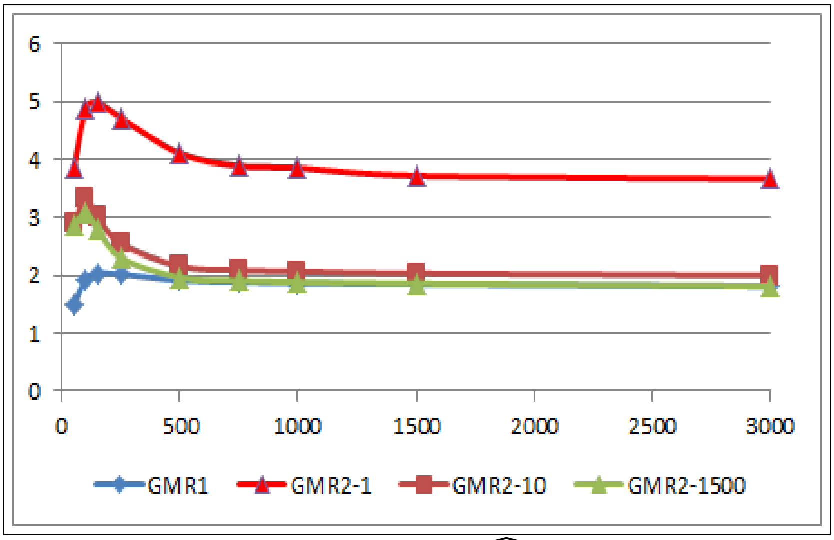

Figure 4: $n \times \widehat{\mathrm{MSE}}$.

The $\mathrm{GARCH}\,(1,1)$ **Model,** $(\theta_1, \theta_2, \theta_3) = (0.1, 0.2, 0.7)$**.**

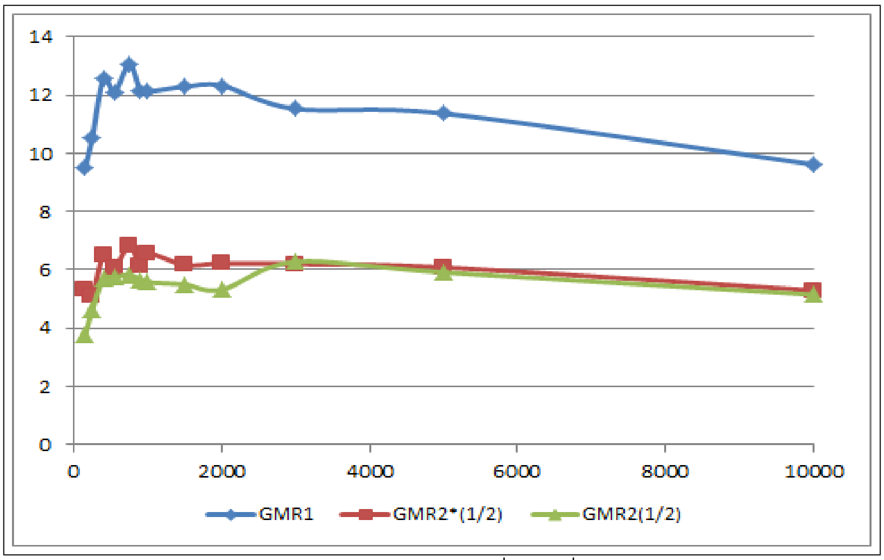

Figure 5a: $n \times \left|\widehat{\mathrm{Bias}}\right|$.

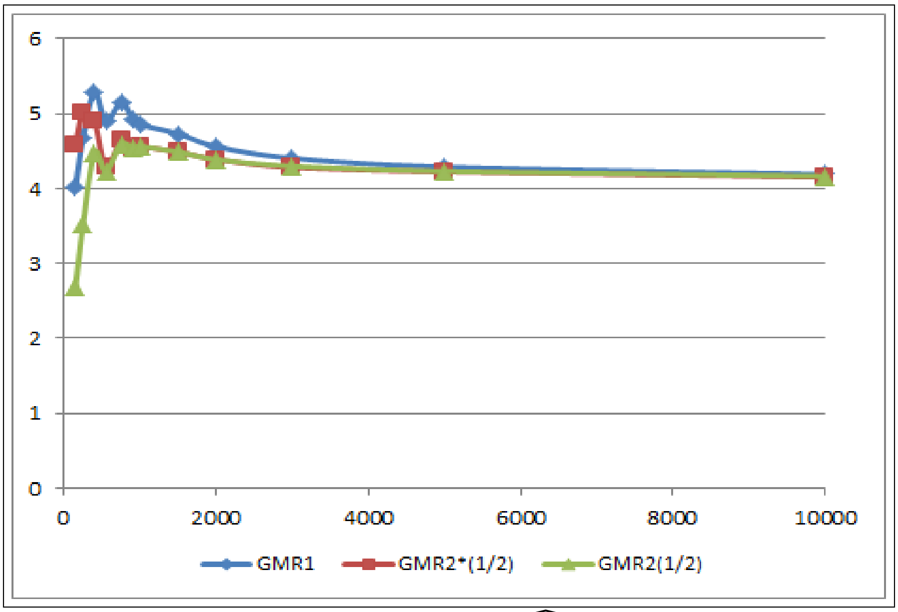

Figure 6a: $n \times \widehat{\mathrm{MSE}}$.

The $\mathrm{GARCH}\,(1,1)$ **Model,** $(\theta_1, \theta_2, \theta_3) = (0.1, 0.05, 0.85)$**.**

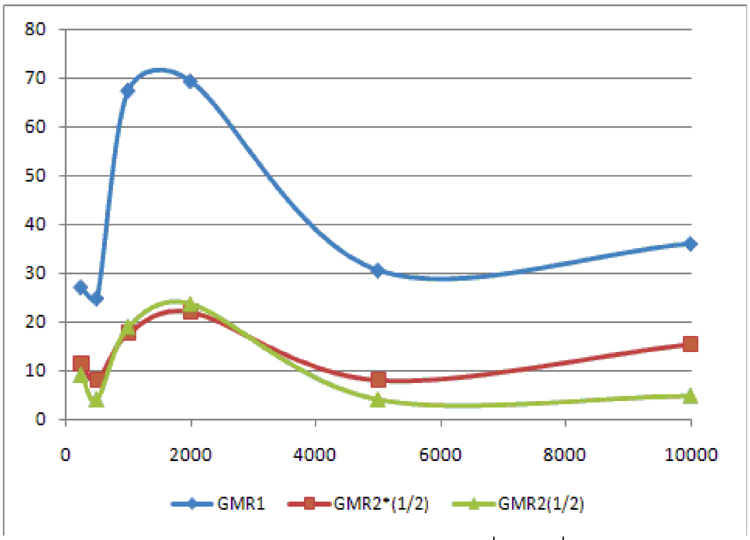

Figure 5b: $n \times \left|\widehat{\mathrm{Bias}}\right|$.

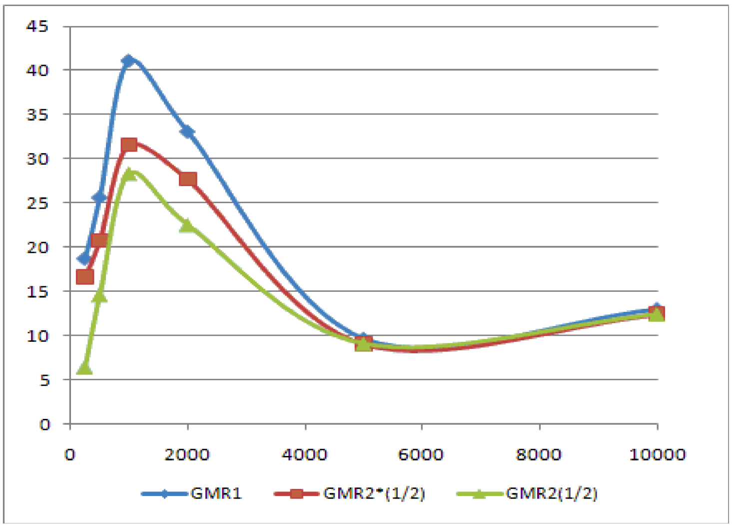

Figure 6b: $n \times \widehat{\mathrm{MSE}}$.

The $\mathrm{ARFIMA}\,(0, d, 0)$ **Model,** $d = 0.4$**.**

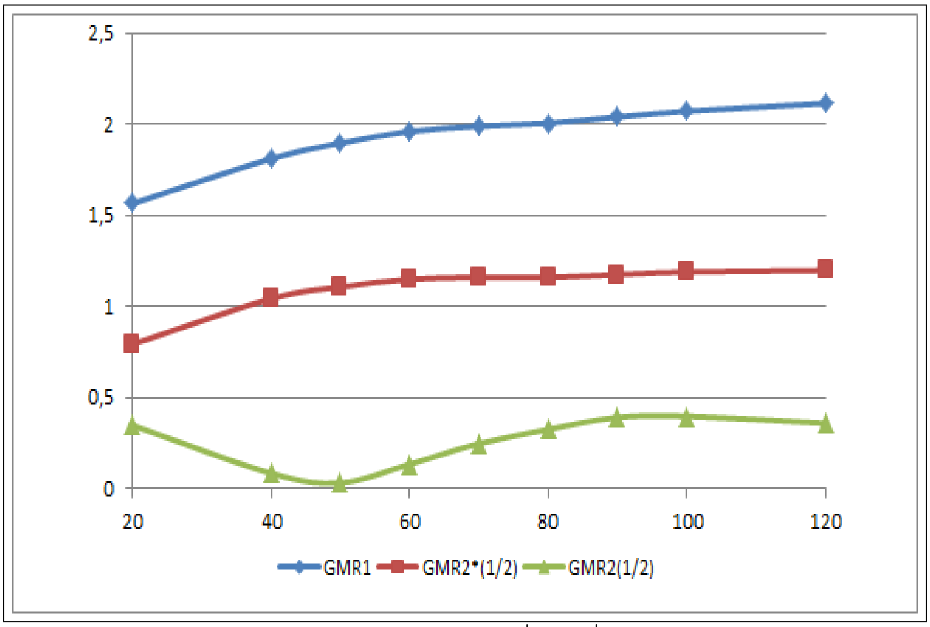

Figure 7a: $n \times \left|\widehat{\mathrm{Bias}}\right|$.

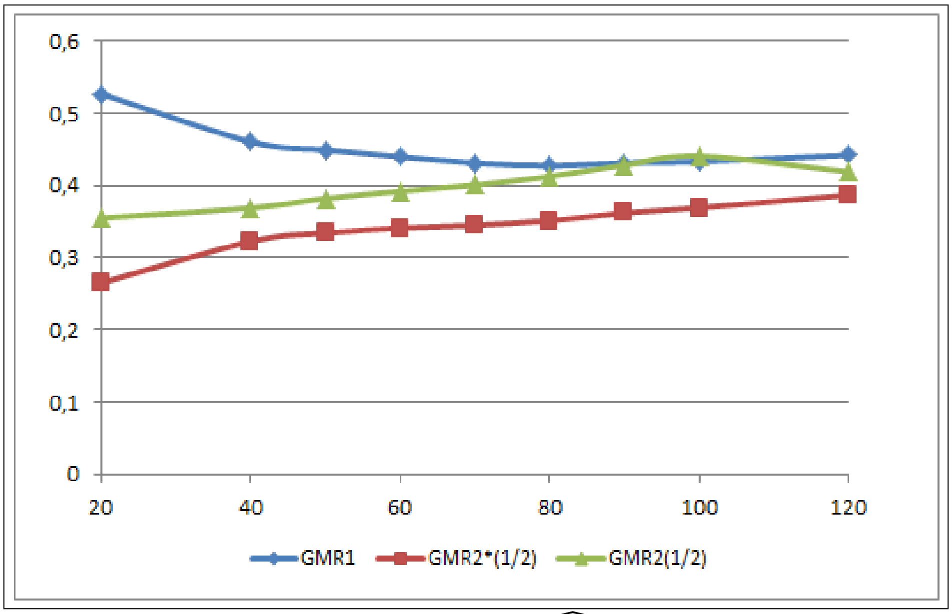

Figure 8a: $n \times \widehat{\mathrm{MSE}}$.

The $\mathrm{ARFIMA}\,(0, d, 0)$ **Model,** $d = 0.3$**.**

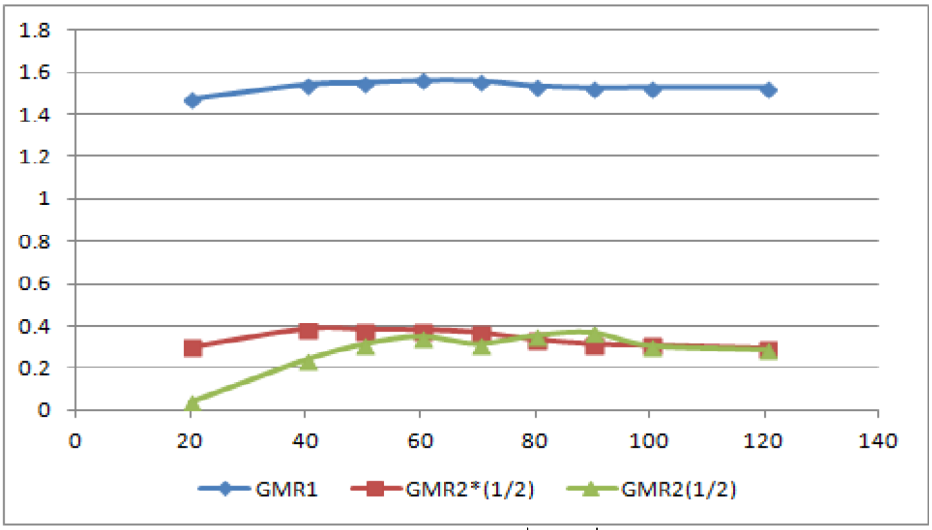

Figure 7b: $n \times \left|\widehat{\mathrm{Bias}}\right|$.

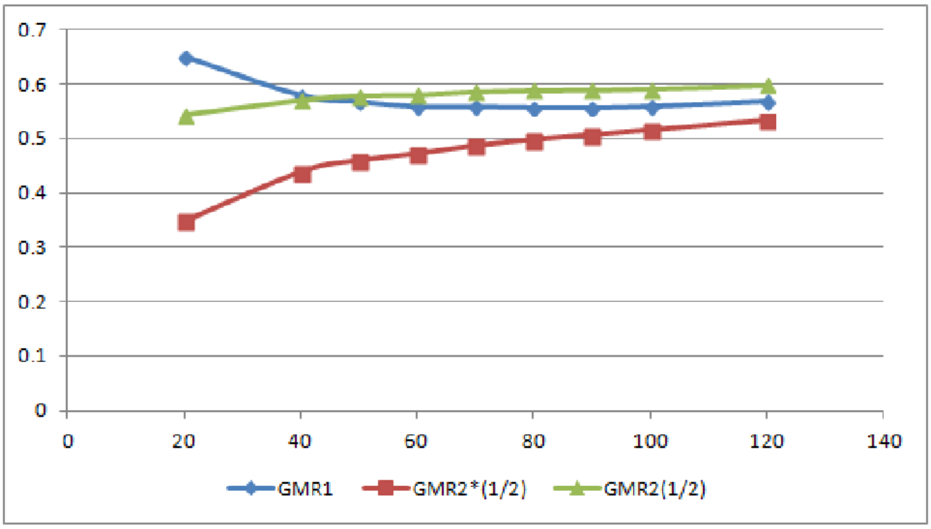

Figure 8b: $n \times \widehat{\mathrm{MSE}}$.

The $\mathrm{MA}\,(1)$ **Model,** $\theta = -0.4$**.**

Table 1: Average CPU times (in seconds)

No. Obs.	$\mathrm{GMR2}\,(1)$	$\mathrm{GMR2}-1$	$\mathrm{GMR2}-10$	$\mathrm{GMR2}-1500$
50	0.0014	0.0022	0.0014	0.2338
100	0.0023	0.0031	0.0033	0.3893
150	0.0041	0.0050	0.0047	0.8121
250	0.0025	0.0096	0.0066	1.2611
500	0.0010	0.0124	0.0133	2.6108
750	0.0010	0.0180	0.0210	3.0978
1000	0.0014	0.0229	0.0282	5.3569
1500	0.0011	0.0520	0.0410	9.1059
3000	0.0014	0.0654	0.0934	17.070

Note: $\mathrm{GMR2}\,(1)$ and $\mathrm{GMR2}-1$ times have been multiplied by 10.