

On the Validity of Edgeworth Expansions and Moment Approximations for Three Indirect Inference Estimators

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

December 2, 2013

Abstract

This paper deals with higher order asymptotic properties for three indirect inference estimators. We provide conditions that ensure the validity of locally uniform Edgeworth approximations. When these are of sufficiently high order they also form integrability conditions that validate locally uniform moment approximations. We derive the relevant 2^{nd} order bias and MSE approximations for the three estimators as functions of the respective approximations for the auxiliary estimator. We allow the possibility of stochastic weighting in any of the steps of the estimation procedure. We confirm that in the special case of deterministic weighting and affinity of the binding function, one of them is second order unbiased. The other two estimators do not have this property under the same conditions. Moreover, in this case, the second order approximate MSEs imply the superiority of the first estimator. We generalize to multistep procedures that provide recursive indirect estimators which are locally uniformly unbiased at any given order. Furthermore, in a particular case, we manage to validate locally uniform Edgeworth expansions for one of the estimators without any differentiability requirements for the estimating equations.

KEYWORDS: Locally Uniform Edgeworth Expansions, Locally Uniform Moment Approximations, Bias Approximation, MSE Approximation, Binding Function, Recursive Indirect Estimators.

JEL: C10, C13

*We would like to thank, without implicating them, Gabriele Fiorentini, Enrique Sentana, Eric Renault, Peter Phillips, the seminar participants at the University of Piraeus, University of Cyprus, and Athens University of Economics and Business, the conference participants at SSEE and at EEA-ESEM 2012. Financial support of BRFP II AUEB is gratefully acknowledged.

1 Introduction

Indirect Inference (II) is a set of inferential procedures that employ a possibly misspecified auxiliary statistical model for estimation of the parameter representing the unknown probability measure with which the underlying measure space is equipped. The original motivation for their introduction was largely computational (see e.g. Smith [46]) hence the choice of the auxiliary model is primarily driven by numerical cost considerations. Despite this motivational characteristic, it enriches the theory of (semi-) parametric statistical inference,¹ since it relies on the existence and the invertibility of functions between collections of probability measures defined on the same σ -algebra. These constitute the core notion upon which the estimators are built and are termed binding functions. The resulting estimators are defined by the inversion of parametric representations of those functions.

This paper is concerned with the *establishment* of higher order asymptotic properties of three classes of II estimators (IIEs), introduced by Gallant and Tauchen [19], Gouriéroux, Monfort, and Renault [25], and Smith [46], with a view towards their characterization in terms of their approximate *bias and mean squared error* (MSE). For a detailed description of our motivation please see subsection 1.2 below. In order to facilitate this description the next subsection is concerned with the establishment of our general framework and the definition of the estimators.

1.1 Definition of Estimators

In what follows, when A is a matrix $\|A\|$ denotes a *submultiplicative* matrix norm, such as the Frobenius one (i.e. $\|A\| = \sqrt{\text{tr} A' A}$). We denote with $\mathcal{PD}(k, \mathbb{R})$ the cone of positive definite matrices of dimension $k \times k$. When $x \in \mathbb{R}^k$, $\|x\|_W$ denotes $\sqrt{x' W x}$ with respect to the conformal positive definite matrix W . When W is the identity we again use $\|x\|$. Any subset of a metric space is (by abuse of terminology) considered as a metric space when endowed with the obvious restriction of the underlying metric. Analogously it is considered a measurable space when endowed with the resulting Borel σ -algebra. $\mathcal{O}_\varepsilon(\theta)$ denotes the open ε -ball around θ in a relevant metric space and $\overline{\mathcal{O}_\varepsilon(\theta)}$ its closure. For s^* and s positive integers with $s^* \geq s$, let $a^* = \frac{s^*-1}{2}$ and $a = \frac{s-1}{2}$.

The notions employed in the paper essentially rely on the characteristics of the statistical model at hand. The following assumption sets these up.

¹For the discussion of the computational aspect of finite dimensional indirect inference in a semiparametric framework, see Dridi and Renault [15].

Assumption A.1 For a measurable space (Ω, \mathcal{F}) , the statistical model (SM) is a family of probability distributions on \mathcal{F} parameterized by \mathbf{F} a function that is onto a compact subset $\Theta \subset \mathbb{R}^p$ for some $p \in \mathbb{N}$. SM is considered endowed with the topology of weak convergence and \mathbf{F}^{-1} is continuous in the following manner: for any $\theta \in \Theta$ and $\hat{\theta}_n$ converging to θ then P_n an arbitrary member of $\mathbf{F}^{-1}(\hat{\theta}_n)$ converges to a member of $\mathbf{F}^{-1}(\theta)$.²

We abbreviate with $\theta_0 = \mathbf{F}(P_{\theta_0}) \in \text{Int}(\Theta)$, for P_{θ_0} in SM, where θ_0 the true parameter vector. P_θ denotes an arbitrary member of $\mathbf{F}^{-1}(\theta)$. All three IIEs, considered here, essentially involve two step estimation procedures. In the first step, an *auxiliary criterion* which is associated with part of the structure of an auxiliary, possibly misspecified (semi-) parametric statistical model, is employed. The *auxiliary estimator*, a random element with values in the auxiliary parameter space, is obtained via the minimization of the auxiliary criterion. Under the appropriate conditions, it will converge to the binding function evaluated at θ_0 , under P_{θ_0} .

This motivates the second step. If the inversion of the function at this parameter value is single valued, and an approximation of the binding function is available, an IIE is defined as a measurable selection on the set constructed from the inversion of this approximation at the auxiliary estimator. The auxiliary estimator is denoted in the paper by $\hat{\beta}_n$ whereas $\hat{\theta}_n$ is the collective notation for the II ones. We also employ $b(\theta)$ to denote the binding function. Given $b(\theta)$, differences between IIEs hinge on different *approximations* of the binding function.

The following assumption defines the auxiliary parameter space and criterion without any direct reference to the auxiliary model. It also describes properties that enable the subsequent definitions. Notice that we are concerned with *finite dimensional inference* both in SM and the associated auxiliary model. Obviously the definitions can be extended to infinite dimensional cases.

Assumption A.2 For B a compact subset of \mathbb{R}^q , $Q_n : \Omega \times B \rightarrow \mathbb{R}$ is jointly measurable. Moreover Q_n is continuous on B for P_θ -almost every $\omega \in \Omega$.

We suppress the dependence of the random elements involved on Ω , for notational simplicity.

Definition D.1 The auxiliary estimator is defined as

$$\hat{\beta}_n \in \arg \min_{\beta \in B} Q_n(\beta).$$

²This assumption enables the consideration of semi-parametric models for which the expectations involved in the following assumptions and results are independent of the infinite dimensional parameter.

Q_n could be a likelihood function, a GMM or more generally, a distance type criterion like the ones appearing in the following definitions. It can summarize part of the structure of the auxiliary model. This definition allows the use of any of the IIEs defined below as auxiliary estimators in multi step procedures, hence enables the discussion in section 4.

Assumption A.3 *The binding function $b : \Theta \rightarrow B$ is injective and continuous on Θ .*

The binding function is usually retrieved by the following. Q_n converges uniformly w.r.t. β to a limit auxiliary criterion $Q(\theta, \beta)$ under P_θ for any θ . For any θ , Q is uniquely minimized by $b(\theta)$. Hence $\widehat{\beta}_n$ converges in P_θ probability to $b(\theta)$. In this framework the continuity of b follows from assumptions A.1 and A.2. Injectivity is an *indirect identification* condition.

Given $\widehat{\beta}_n$, the IIEs are defined as minimum distance ones. In our setup the relevant distances are represented by norms with respect to positive definite matrices. As in the context of GMM estimation, we allow these to be stochastic, and/or depend on initial estimators, say θ_n^* . We refer to this general framework as *stochastic weighting*.

Assumption A.4 $W_n^* : \Omega \times \Theta \rightarrow \mathcal{PD}(q, \mathbb{R})$ and $\widetilde{\theta}_n^* : \Omega \rightarrow B$ are jointly measurable.

The first IIE considered here minimizes a distance function between the $\widehat{\beta}_n$ and $b(\theta)$. It is termed GMR1 and was proposed by Gouriéroux et al. [25] in order to relax the numerical burden associated with the second estimator, defined below.

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

$$\widehat{\theta}_n \in \arg \min_{\theta \in \Theta} \left\| \widehat{\beta}_n - b(\theta) \right\|_{W_n^*(\widetilde{\theta}_n^*)}.$$

The second estimator is termed GMR2 and minimizes the previous distance between $\widehat{\beta}_n$ and $E_\theta \widehat{\beta}_n$. First, due to assumptions A.1 and A.2 the following lemma is trivially true.

Lemma 1.1 *Under assumptions A.1 and A.2 $\left\| E_\theta \widehat{\beta}_n \right\| < \infty$ on Θ .*

Given the above lemma, it is possible to define the GMR2 estimator by

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

$$\widehat{\theta}_n \in \arg \min_{\theta \in \Theta} \left\| \widehat{\beta}_n - E_{\theta} \widehat{\beta}_n \right\|_{W_n^*(\widetilde{\theta}_n^*)}.$$

The third IIE estimator, denoted here by GT, was originally proposed by Gallant and Tauchen [19] in cases where the score of the auxiliary likelihood is well defined. This estimator is also associated with smaller numerical costs, as compared to the GMR2.

Assumption A.5 *Let Q_n be differentiable on B for P_{θ} —almost every $\omega \in \Omega$. We denote with c_n the derivative of Q_n except for the case where $Q_n = \|c_n(\beta)\|_{W_n(\widehat{\beta}_n^*)}$, where $c_n : \Omega \times B \rightarrow \mathbb{R}^l$, $W_n : \Omega \times B \rightarrow \mathcal{PD}(l, \mathbb{R})$, and $\widehat{\beta}_n^* : \Omega \rightarrow B$ are jointly measurable. Moreover c_n is continuous on B for P_{θ_0} —almost every $\omega \in \Omega$, $c_n(\beta)$ is P_{θ} —integrable on $\Theta \times B$ and $E_{\theta}(c_n(\beta))$ is continuous on $\Theta \times B$. Also $W_n^{**} : \Omega \times \Theta \rightarrow \mathcal{PD}(l, \mathbb{R})$ and $\widetilde{\theta}_n^*$ are jointly measurable.*

Notice that, given assumption A.1, the joint continuity of $E_{\theta}(c_n(\beta))$ would follow by dominated convergence and a condition of the form $E_{\theta}(\sup_B \|c_n(\beta)\|) < +\infty$ for any θ . The GT estimator minimizes the norm of the expectation of the auxiliary estimating vector. We denote by $E_{\theta}(c_n(\widehat{\beta}_n))$, the quantity $E_{\theta}(c_n(\beta))|_{\beta=\widehat{\beta}_n}$ for notational simplicity. Due to assumption A.5 we have that $\|E_{\theta}(c_n(\widehat{\beta}_n))\| < \infty$. Consequently, the following minimization procedure is well defined.

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

$$\widehat{\theta}_n \in \arg \min_{\theta \in \Theta} \left\| E_{\theta}(c_n(\widehat{\beta}_n)) \right\|_{W_n^{**}(\widetilde{\theta}_n^*)}.$$

The usual definition of the aforementioned estimator is given only when the auxiliary estimator is the MLE of the auxiliary model. The current one is obviously an extension.

Given the form of SM described in the first part of assumption A.1, the existence issue is based on assumption A.2 for $\widehat{\beta}_n$, A.2, A.3, A.4 for the GMR1, the continuity part of assumption A.1, and assumptions A.2, A.4 for the GMR2, and A.2, A.5 for the GT estimator.

Remark R.1 *Some simple cases of almost sure (or possibly asymptotic) coincidence of the estimators are the following. Suppose, first, that $Q_n(\beta) = \|c_n(\beta)\|$,*

$c_n(\beta) = q_n - g(\beta)$ with the random element $q_n \in g(B)$ with P_θ probability 1 for any θ . Suppose that q_n converges in probability to $m(\theta)$ under P_θ , and $g(\beta)$ and $m(\theta)$ are invertible. Then with P_θ probability 1, for any θ , $\widehat{\beta}_n = g^{-1}(q_n)$, $b(\theta) = g^{-1}(m(\theta))$, and $\text{GMR1} = m^{-1}(q_n)$. Also, when $m(\theta) = E_\theta q_n$ then $\text{GMR1} = \text{GT}$ with P_θ probability 1. If moreover g is linear then the GMR2 is also P_θ almost surely equal to the other two. Second, when $q = l$, $m(\theta) \neq E_\theta q_n$ yet $E_\theta q_n$ is injective, $\widehat{\beta}_n$ belongs to its range with P_θ probability $1 - o(1)$ and g is linear then $\text{GMR2} = \text{GT}$ with P_θ probability $1 - o(1)$. In this case, the existence of GMR2 would imply the existence of GT with P_θ probability $1 - o(1)$.

The computation of all three estimators relies on the analytical tractability of the binding function or the engaged expectations. Due to this fact in applications the analytically intractable elements appearing in the definitions are usually approximated by some numerical integration method. Moreover the optimization in every step is performed also numerically. In such a framework, it is easy to show that the numerically approximated counterpart of the GMR2 estimator is the one associated with the maximal computational burden among the three. Finally notice that our definitions of the GMR1 and the GT estimators correspond (and actually in the GT case generalize) to the definitions of the classical minimum distance and the GMM estimators made by Calzolari et. al. [10]. The definition of the GMR2 estimator is made analogously.

1.2 Higher Order Properties and Motivation

Gourieroux, Renault and Touzi [27] show that the GMR2 estimator has zero second order bias, when *i*) $p = q$ and *ii*) the binding function is affine. Notice that both are automatically satisfied, when the auxiliary model coincides with the SM and the binding function is approximated by a consistent estimator of the auxiliary parameter. In this case the particular IIE is said to perform a *bias correction* of the first step one, a result which was only derived using non primitive conditions. First, we provide the validity of this result in a more general framework³ via the use of more primitive assumptions than the ones used in the aforementioned paper, and secondly, we define estimating procedures based on recursions of the GMR2 estimator that are characterized by higher order unbiasedness.

Furthermore, the question of whether the above is also true for the remaining two IIEs follows naturally. To our knowledge, the only result that could provide such a connection is Proposition 4.1 of Gourieroux and Monfort [24], which is not generally true (it is true under additional assumption see e.g. remark R.1).

³for example we allow that $q > p$.

However, their result creates the impression that all three estimators coincide for large enough n and therefore provides an affirmative answer to this question. By validating the expansions at any order and deriving the second order expansion for them, we show that their Proposition does not apply in these cases. Hence we essentially derive their *higher order non-equivalence*. We also provide comparisons between the MSE approximations of the estimators.

In section 2 we establish the validity of locally uniform Edgeworth approximations for the examined estimators. We employ analogous methodologies of those previously applied in Andrews [2], Andrews and Lieberman [3], Hall and Horowitz [29], Götze and Hipp [22] and [23], Lieberman et al. [32], and Robinson [41].

At this point it is possibly useful to stress that in the construction of the GMR2 estimator, see definition D.3 above, we employ the expected value of the auxiliary estimator, $\widehat{\beta}_n$. Consequently, to validate locally uniform Edgeworth expansions of this estimator, employing the above mentioned methodologies, we require the s^* derivative of this expectation to be uniformly bounded, over $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ (see assumption A.10). However, in a particular case (see lemma 2.5.ii) we manage to validate locally uniform Edgeworth expansions for the GMR2 estimator *without any differentiability requirements* for the equations that the estimator satisfies asymptotically. We instead rely on the existence of asymptotic polynomial approximations for these equations that hold for an appropriate class of sequences and properties of the GMR1 which are sufficient for the establishment that the GMR2 estimator belongs to the aforementioned class. Hence we essentially utilize higher order properties of one IIE that could be easily establishable, even if the estimator *per se* is computationally infeasible, in order to verify analogous properties of another.

In section 3 we provide additional assumptions that validate the first and second moment approximations and derive the second order ones as functions of the analogous moment approximations for the auxiliary estimator. In section 4 we generalize the estimation procedures via multistep extensions of the GMR2 estimator that, under the scope of section 3, have desirable higher order properties. At this point we should remark that since their introduction IIEs are becoming increasingly popular. Several applications have been carried out in the context of stochastic volatility and equity return models (e.g. Gallant et al. [17], Garcia et al. [20], Andersen et al. [1], and Sentana et al. [43]), exchange rate models (e.g. Bansal et al. [7], and Chung and Tauchen [12]), commodity price and storage models (e.g. Michaelides and Ng [36]), dynamic panel data (e.g. Gouriéroux et al. [26]), stochastic differential equation models (e.g. Gallant and Long [18] and Gouriéroux and Monfort [24]), and in ARMA models (e.g. Chumacero [11], Ghysels et al. [21], Demos and Kyriakopoulou [14], and Phillips [40]). Consequently, the results presented in these two sections could be useful to applied work, as well.

In section 5 we present two examples and engage into Monte Carlo experiments. We conclude in section 6. As the paper is concentrated on the asymptotic properties of IIEs we make high level assumptions for the auxiliary one, i.e. in section 2 we assume that the auxiliary estimator has a valid Edgeworth expansion. This could be justified by either adopting the weak dependence set up of Götze and Hipp [23] or the strong dependence one in Lieberman, Rousseau and Zucker [32]. Consequently, in this way we are as general as possible, as well. However, in section 7 we discuss low level assumptions that suffice for the high level ones to hold. We collect all proofs and cite some general results in the appendix.

2 Validity of Edgeworth Approximations

In this section we expand the assumption framework presented above, in order to facilitate the validation of Edgeworth approximations.⁴ We initially assume analogous expansions for $\widehat{\beta}_n$, and require differentiability of the random elements appearing in the definitions of $\widehat{\theta}_n$ in a neighborhood of θ_0 . Given these, we prove that the IIEs satisfy first order conditions with probability uniformly bounded by $1 - o(n^{-a^*})$, around θ_0 . Then, a justified application of the mean value theorem implies asymptotic tightness for $\sqrt{n}(\widehat{\theta}_n - \theta)$ with probability uniformly bounded by $1 - o(n^{-a^*})$, around θ_0 . Third, we obtain a local polynomial approximation of $\sqrt{n}(\widehat{\theta}_n - \theta)$ using the first order conditions and bound the relevant remainder by an $o(n^{-a^*})$ real sequence with probability uniformly bounded by $1 - o(n^{-a^*})$, around θ_0 . Then, if valid, $\sqrt{n}(\widehat{\theta}_n - \theta)$ and the polynomial approximation have the same Edgeworth expansion uniformly over the required neighborhood of θ_0 . Finally, the validity of the aforementioned approximation is established by employing an argument analogous to Skovgaard [44], i.e. a Theorem of Invariance of validity of Edgeworth approximations with respect to sequences of smooth transformations (see also Bhattacharya and Ghosh [8] in an iid framework, where the transformation examined is independent of n). This methodology is justified by lemmas AL.4 and AL.5 in the second part of the appendix (for analytical proofs of these results see Arvanitis and Demos [5]).

⁴The Edgeworth distributions discussed in the sequel, are not *necessarily* the formal ones. They comply with the general definition of Magdalinos [34] (see equations 3.7-8).

Assumptions Specific to the Validity of the Edgeworth Approximations

We denote with D^r , the r -derivative operator and with $D^r(h(x_0))(x^r)$ the r^{th} -linear function defined by the evaluation of $D^r h$ at x_0 evaluated at $\underbrace{(x, \dots, x)}_{r \text{ times}}$. Let

M denote a universal positive constant, independent of n and θ , not necessarily taking the same value across and inside assumptions proofs and results. $\text{pr}_{i,j}(x)$ denotes the transformation of an r^{th} dimensional vector, say $x = (x_1, x_2, \dots, x_r)'$, to a vector containing only the elements of x from the i^{th} to the j^{th} coordinate, i.e. $\text{pr}_{i,j}(x) = (x_i, x_{i+1}, \dots, x_j)'$, where naturally $1 \leq i \leq j \leq r$. Finally whenever the assertion "locally" appears in the sequel it implies "for all $\theta \in \bar{\mathcal{O}}_\varepsilon(\theta_0)$ " unless otherwise specified.⁵

Assumption A.6 $\widehat{\beta}_n$ is uniformly consistent for $b(\theta)$ with rate $o(n^{-a^*})$, i.e.

$$\sup_{\theta \in \Theta} P_\theta \left(\left\| \widehat{\beta}_n - b(\theta) \right\| > \varepsilon \right) = o(n^{-a^*}), \forall \varepsilon > 0.$$

Moreover $\widetilde{\theta}_n^*$, see assumption A.4, is uniformly consistent for θ with rate $o(n^{-a^*})$.

This assumption along with the boundeness of B enables the uniform convergence of $E_\theta \widehat{\beta}_n$ to $b(\theta)$, hence the establishment of the analogous property for the GMR2 estimator.⁶ Low level assumptions and ways of verifying this assumption can be found in section 7.

The following concerns the asymptotic behavior of the weighting matrices involved in the II estimation.

Assumption A.7 For $W_n^j(\theta)$ be either $W_n^*(\theta)$ or $W_n^{**}(\theta)$ and

$$k = \begin{cases} q & \text{if } W_n^j(\theta) = W_n^*(\theta) \\ l & \text{if } W_n^j(\theta) = W_n^{**}(\theta) \end{cases},$$

suppose that there exists a sequence of random elements $x_n : \Omega \rightarrow \mathbb{R}^m$, such that $W_n^j(\theta) = \frac{1}{n} \sum W^j(x_i(\omega), \theta)$ for measurable $W^j : \mathbb{R}^m \times \Theta \rightarrow \mathcal{PD}(k, \mathbb{R})$, integrable with respect to P_{θ^*} , such that

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

⁵Notice that due to the fact that the spaces Θ and B are separable and closed, suprema of real random elements over these spaces are typically measurable (see van der Vaart and Wellner [49], example 1.7.5 p. 47 due to the theorem of measurable projections, completeness of the underlying probability space, the compactness of Θ and the continuity of b).

⁶see the first paragraph of the proof of lemma 2.3.

$E_{\theta^*} W^j(\theta)$ is Lipschitz w.r.t. θ , for any θ^* and for the Lipschitz coefficient (say) $\kappa^j(\theta^*)$ we have that $\sup_{\theta^* \in \Theta} \kappa^j(\theta^*) < +\infty$.

b) For $W_n^j(\theta) = W_n^*(\theta)$ if $p \neq q$, and $W_n^j(\theta) = W_n^{**}(\theta)$ if $p \neq l$, $W^j(x, \theta)$ is s^* -differentiable on $\overline{\mathcal{O}}_{\varepsilon_0}(\theta_0)$ for $\varepsilon_0 > \varepsilon$ and

$$\sup_{\theta^* \in \overline{\mathcal{O}}_{\varepsilon}(\theta_0)} P_{\theta^*} \left(\sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon}(\theta_0)} \|D^{s^*+1} W_n^j(\theta)\| > M \right) = o(n^{-a^*}).$$

The first part of a) can be justified by conditions on the asymptotic behavior of $\sup_{\theta^*} E_{\theta^*} (\|W_n^j(\theta) - E_{\theta^*} W^j(\theta)\|^q)$. The second part can be justified by

$$\sup_{\theta^* \in \Theta} E_{\theta^*} \sup_{\theta \in \Theta} \|DW^j(x_i(\omega), \theta)\| < +\infty.$$

Part b) can be justified analogously.

Obviously when $W^j(x, \theta)$ is independent of x and θ the above assumption is trivially satisfied. Let $f(x, \theta)$ denote the vector that contains stacked all the *distinct* components of $W^*(x, \theta)$ and $W^{**}(x, \theta)$ 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*. Furthermore when $p = q$ the elements corresponding to $W^*(x, \theta)$ and its derivatives are also discarded. Analogously when $p = l$ the elements corresponding to $W^{**}(x, \theta)$ and its derivatives are discarded too. Obviously when $f(x_0, \theta) - E_{\theta} f(x_0, \theta)$ equals zero or $p = q = l$, f becomes irrelevant to what follows. Let

$$m_n(\theta) = \widehat{\beta}_n - b(\theta)$$

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

$$m_n(\theta) = \begin{pmatrix} \widehat{\beta}_n - b(\theta) \\ \frac{1}{n} \sum f(x_i) - E_{\theta} \frac{1}{n} \sum f(x_i) \end{pmatrix}$$

when $f(x_0, \theta)$ is independent of θ yet $f(x_0) - E_{\theta} f(x_0)$ is not zero and the involved dimensions do not coincide, and

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

in any other case. $\Psi_{n,s^*}(\theta)$ denotes an Edgeworth measure of order s^* (see for example equations (3.7) and (3.8) of Magdalinos [34]) defined on the Borel algebra of \mathbb{R}^w for $w = \dim(m_n)$. $\pi_{i-1}(z, \theta)$ denotes the polynomial (w.r.t. to z) appearing in the density of $\Psi_{n,s^*}(\theta)$ with coefficient $\frac{1}{n^{\frac{i-1}{2}}}$, for $i = 1, \dots, s^*$ (notice that $\pi_0 = 1$).

The random elements appearing in the paper assume their values in finite dimensional Euclidean spaces. For notational simplicity and without any risk of error, in what follows \mathcal{B}_C generically denotes the class of convex Borel subsets of any of those spaces.

Definition D.5 *We say that a sequence of random elements $(S_n(\theta))_{n \in \mathbb{N}}$ with values in some finite dimensional Euclidean space has an Edgeworth expansion of order s^* if there exists an Edgeworth measure $\Psi_{n,s^*}(\theta)$ of order s^* such that*

$$\sup_{A \in \mathcal{B}_C} |P(S_n(\theta) \in A) - \Psi_{n,s^*}(\theta)(A)| = o(n^{-a^*}).$$

If there exists some $\varepsilon > 0$ and a θ_0 such that the previous holds also uniformly w.r.t. $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ then we say that the expansion is locally uniform (around θ_0).

In some of the cases examined in this paper the uniformity w.r.t. \mathcal{B}_C can be enlarged to include the whole Borel algebra. For a discussion please see section 7. Furthermore, in some cases we describe explicitly the neighborhoods of θ_0 for which we establish the validity of the definition. Given the definition, the following is an extremely important-to what follows-high order assumption concerning the asymptotic behavior of the auxiliary estimator.

Assumption A.8 *$\sqrt{n}m_n(\theta)$ has a locally uniform Edgeworth expansion of order s^* . Furthermore $\pi_i(z, \theta)$ is equicontinuous on $\overline{\mathcal{O}}_\varepsilon(\theta_0) \forall z \in \mathbb{R}^q$, for $i = 1, \dots, a^*$, and if $V(\theta)$ denotes the variance matrix in the density of $\Psi_{n,s}(\theta)$ then it is continuous on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ and positive definite.*

We now briefly discuss the conditions that could validate the previous assumption. We first reformulate Theorem 3.2 of Arvanitis and Demos [5] as follows.

Theorem 2.1 *Suppose that:*

-A. *The random element $M_n(\theta)$ satisfies*

$$0_{p \times 1} = \sum_{i=0}^{s-1} \frac{1}{n^{i/2}} \sum_{j=0}^{i+1} C_{i,j,n}(\theta) \left(M_n(\theta)^j, S_n(\theta)^{i+1-j} \right) + R_n(\theta)$$

with probability $1 - o\left(n^{-\frac{s-1}{2}}\right)$ independent of θ where $C_{i,j,n} : \overline{\mathcal{O}}_\varepsilon(\theta_0) \times \mathbb{R}^{q^{i+1}} \rightarrow \mathbb{R}^p$ is $(i+1)$ -linear $\forall \theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$, $C_{0,0,n}(\theta), C_{0,1,n}(\theta)$ are independent of n and have rank $p \forall \theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$, $C_{i,j,n}$ are equicontinuous on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$, $\forall x^{i+1}$,
 -B. *$S_n(\theta)$ admits a locally uniform Edgeworth expansion with equicontinuous on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ density polynomials $\pi_i(z, \theta)$, and the variance matrix is continuous*

on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ and positive definite,

-C. $\sup_{\theta \in \Theta} P\left(\|M_n(\theta)\| > C \ln^{1/2} n\right) = o\left(n^{-\frac{s-1}{2}}\right)$ for some $C > 0$ independent of θ ,

-D. $\sup_{\theta \in \Theta} P(\|R_n(\theta)\| > \gamma_n) = o\left(n^{-\frac{s-1}{2}}\right)$ for some real sequence $\gamma_n = o\left(n^{-\frac{s-1}{2}}\right)$ independent of θ .

Then $M_n(\theta)$ admits an order s locally uniform Edgeworth expansion with density comprised of polynomials which are equicontinuous on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ and the variance matrix is continuous on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ and positive definite.

When $\widehat{\beta}_n$ is an M-estimator the verification of the assumption A.8 proceeds via the verification of the conditions of the above theorem. When, on the other hand, $\widehat{\beta}_n$ 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. For further discussion as well as counterexamples please see section 7.

Existence of Edgeworth Expansions for the GMR-type Estimators

In this subsection we establish locally uniform Edgeworth expansions for the GMR1 and GMR2 estimators. Without any direct reference to Q_n we utilize additional assumptions concerning the behavior of the coefficients in the asymptotic polynomial approximations of the estimators by the elements of the random vector $m_n(\theta)$ (see assumption A.8).

The GMR1 Case Here the local continuous differentiability of $b(\theta)$ is sufficient.

Assumption A.9 $b(\theta)$ is $s^* + 1$ continuously differentiable and $\text{rank } Db(\theta) = p$, for all θ in $\overline{\mathcal{O}}_{\varepsilon_0}(\theta_0)$ and $\varepsilon_0 > \varepsilon$.

In a special case, i.e. when $Q_n = \|c_n(\beta)\|_{W_n(\widehat{\beta}_n^*)}$ and $q = l$, the previous assumption can be justified when $c(\theta, \beta)$ (the uniform limit of $c_n(\beta)$ defined in assumption A.12 below) is $s^* + 1$ times differentiable on $\mathcal{O}_\eta(\varphi_0)$ where $\varphi_0 = (\theta'_0, b'(\theta_0))'$, for η large enough and $\text{rank } \frac{\partial c(\theta, b(\theta))}{\partial \beta'} = q$ for all θ in $\overline{\mathcal{O}}_\varepsilon(\theta_0)$, via the Implicit Function Theorem.

The next lemma provides the results.

Lemma 2.2 i) Under the assumptions A.1, A.2, A.3, A.4, A.6 and A.7.a the GMR1 is uniformly consistent for θ with rate $o(n^{-a^*})$.

ii) If additionally assumptions A.7.b, A.8 and A.9 hold then, $\sqrt{n}(\text{GMR1} - \theta)$ has a locally Edgeworth expansion of order s^* w.r.t. $\overline{\mathcal{O}}_\varepsilon(\theta_0)$, for $\varepsilon < \varepsilon_0$, where ε_0 as in assumption A.9.

The GMR2 Case The analogous to assumption A.9 is the following.

Assumption A.10 $\sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon_0}(\theta_0)} \left\| D^{s^*} E_{\theta} \widehat{\beta}_n \right\| < M.$

Assumption A.9 along with assumption A.10 imply that for $r = 1, \dots, s^*$, $\sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon}(\theta_0)} \left\| D^r \left(E_{\theta} \widehat{\beta}_n - b(\theta) \right) \right\| < M$, which in turn means that $D^{r-1} \left(E_{\theta} \widehat{\beta}_n - b(\theta) \right)$ are uniformly Lipschitz on $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$, and therefore uniformly equicontinuous on the same ball. This implies the commutativity of the limit, with respect to n and the derivative operator, uniformly over $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$. This along with the second part of assumption A.9 and continuity imply that $\text{rank } DE_{\theta} \widehat{\beta}_n = p$, for all θ in $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$ for n large enough. If the statistical model has an appropriately differentiable likelihood function then this assumption could be verified by existence of moments of suprema of the likelihood derivatives along with restrictions on the asymptotic properties of these derivatives.

The next lemma provides the results.

Lemma 2.3 i) Under the assumptions A.1, A.2, A.3, A.4, A.6 and A.7.a the GMR2 is uniformly consistent for θ with rate $o(n^{-a^*})$.

ii) If additionally assumptions A.7.b, A.8, A.9 and A.10 hold then $\sqrt{n}(\text{GMR2} - \theta)$ has a locally Edgeworth expansion of order $s^* - 1$.

A Case of Validation of the Expansion Without Differentiability for $E_{\theta} \widehat{\beta}_n$. The fact that the Edgeworth approximation of the auxiliary estimator is *locally uniform* enables the possibility that the second part of the previous lemma holds with the strength of assumption A.10 either diminished, or eliminated in particular cases. We investigate such possibilities. We denote with $k_{i_{\beta}}(z, \theta) = z \pi_{i-1}(z, \theta)$ and with $\mathcal{I}_V(k_{i_{\beta}}(z, \theta)) = \int_{\mathbb{R}^q} k_{i_{\beta}}(z, \theta) \varphi_{V(\theta)}(z) dz$ where $\pi_{i-1}(z, \theta)$ and $V(\theta)$ are as in assumption A.8.

Assumption A.11 $\mathcal{I}_V(k_{i_{\beta}}(z, \theta))$ is s^* (equi-) continuously differentiable for $i = 1, \dots, s^* - 1$ over $\mathcal{O}_{\varepsilon}(\theta_0)$.

Due to dominated convergence, the form of the Edgeworth densities and the properties of the normal distribution this assumption would hold if the coefficients of the polynomials (w.r.t. z) $\pi_{i-1}(z, \theta)$ and the quadratic form in the density approximation of the auxiliary estimator are s^* (equi-) continuously differentiable functions of θ (see section 7).

The following lemma enables the derivation of an analogous result to the second part of lemma 2.3 in the special case where $p = q$.

Lemma 2.4 *If assumptions A.8, A.9 and A.11 hold for $s^* > s$ then for any sequence θ_n^+ for which*

$$\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\sqrt{n} \|\theta_n^+ - \theta\| > M \ln^{1/2} n \right) = o(n^{-a^*})$$

we have that for any $\varepsilon_ < \varepsilon$*

$$\sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon_*}(\theta_0)} P_\theta \left(\left\| \sqrt{n} \left(E_{\theta_n^+} \widehat{\beta}_n - E_\theta \widehat{\beta}_n \right) - A_n(\theta) \right\| > \gamma_n \right) = o(n^{-a^*}),$$

where

$$A_n(\theta) = \sum_{i=1}^s \frac{1}{n^{\frac{i-1}{2}} i!} D^i \left(b(\theta) + \sum_{j=1}^{s-i} \frac{\mathcal{I}_V(k_{j\beta}(z, \theta))}{n^{\frac{j}{2}}} \right) \left(\sqrt{n} (\theta_n^+ - \theta)^i \right),$$

$\gamma_n = o(n^{-a})$ independent of θ and $E_{\theta_n^+} \widehat{\beta}_n = E_{\theta} \widehat{\beta}_n \Big|_{\theta=\theta_n^+}$, i.e. the composition of $E_{\theta} \widehat{\beta}_n$ with θ_n^+ , employing the convention that when $s-i=0$ then $\sum_{j=1}^{s-i}$ is empty.

Notice that under the provisions of the following lemma 2.5, the Edgeworth expansion can be established either by restricting assumption A.10 to hold only for the second order derivative so as to establish that $E_{\text{GMR2}} \widehat{\beta}_n = \widehat{\beta}_n$ with P_θ -probability $1 - o(n^{-a})$ that is independent of θ , or does not need to hold at all if the latter is established in some other way. In this case the validity of the Edgeworth approximation goes through without the need to assume differentiability for $E_\theta \widehat{\beta}_n$. This becomes possible since due to the previous lemma the quotients that would define the derivative need only converge for an appropriate class of sequences. Hence lemma 2.5.ii below, enables the validation of an Edgeworth expansion of the GMR2 estimator without the application of Taylor expansions for the first order conditions.

Lemma 2.5 *Suppose that $p = q$ and assumptions A.1, A.2, A.3, A.4, A.6, A.7, A.8, A.9 and A.11 hold for $s^* > s$. i) If $\sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon_0}(\theta_0)} \|D^2 E_\theta \widehat{\beta}_n\| < M$ then $\sqrt{n}(\text{GMR2} - \theta)$ has a locally uniform Edgeworth expansion of order s w.r.t. $\overline{\mathcal{O}}_{\varepsilon_*}(\theta_0)$ for any $\varepsilon_* < \varepsilon$. ii) if $\widehat{\beta}_n = b(\text{GMR1})$ with probability $1 - o(n^{-a})$ uniformly on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ and $\widehat{\beta}_n = E_{\text{GMR2}} \widehat{\beta}_n$ with probability $1 - o(n^{-a})$ uniformly on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ then $\sqrt{n}(\text{GMR2} - \theta)$ has a locally uniform Edgeworth expansion of order s w.r.t. $\overline{\mathcal{O}}_{\varepsilon_*}(\theta_0)$ for any $\varepsilon_* < \varepsilon$.*

The second case of the previous lemma implies that the results on the higher order asymptotics for the GMR1 estimator can be *analytically* useful for the determination of the higher order properties of the GMR2 estimator. Finally notice that the Edgeworth distributions validated in all cases of lemmas 2.3 and 2.5 need not coincide. However due to the triangle inequality the uniform with respect to on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$, convex variational distance (see for example Andrews [2]) between any pair of them must be $o(n^{-a})$. Let us now turn our attention to the GT estimator.

Existence of Edgeworth Expansion for the GT Estimator

We first consider two particular cases which link the asymptotic behaviors of the three estimators.

Lemma 2.6 A. *Suppose that $p = q = l$, $E_{GT} \left(c_n \left(\widehat{\beta}_n \right) \right) = \mathbf{0}_l$ with probability $1 - o(n^{-a^*})$ independent of θ and $E_\theta(c_n(\beta)) = \mathbf{0}_l$ iff $\beta = b(\theta)$. i) If the assumptions of lemma 2.2.i hold then the GT is uniformly consistent for θ with rate $o(n^{-a^*})$. ii) If the assumptions of lemma 2.2.ii hold then $\sqrt{n}(GT - \theta)$ has a locally uniform Edgeworth expansion of order s^* which coincides with the one of lemma 2.2.*

B. *Suppose that $q = l$, $c_n(\beta) = q_n - \beta$ for q_n an appropriate q -dimensional random element and $W_n^* = W_n^{**}$ (P_θ almost everywhere for all θ). i) If the assumptions of lemma 2.3.i hold then the GT is uniformly consistent for θ with rate $o(n^{-a^*})$. ii) If the assumptions of lemma 2.3.ii or the ones of lemma 2.5.i or ii hold then $\sqrt{n}(GT - \theta)$ has a locally uniform Edgeworth expansion of order $s^* - 1$ w.r.t. $\overline{\mathcal{O}}_{\varepsilon_*}(\theta_0)$, for $\varepsilon_* < \varepsilon$, which coincides with the expansions of lemmas 2.3 or 2.5.i or ii, respectively.*

In a more general case, due to the definition of the particular estimator, we utilize the following two assumptions concerning the asymptotic behavior of c_n .

Assumption A.12 *Let $Q_n = \|c_n(\beta)\|_{W_n(\widehat{\beta}_n^*)}$ and*

$$\|c_n(\beta) - c_n(\beta')\| \leq \kappa_n \|\beta - \beta'\|, \text{ for all } \beta, \beta' \quad (1)$$

$\sup_{\theta \in \Theta} E_\theta \kappa_n = O(1)$ and

$$\sup_{\theta \in \Theta} P_\theta \left(\sup_{\beta \in B} \|c_n(\beta) - c(\theta, \beta)\| > \varepsilon \right) = o(n^{-a^*}), \forall \varepsilon > 0, \quad (2)$$

where $c(\theta, \beta)$ is continuous on B and equals zero iff $\beta = b(\theta)$ for any θ . Furthermore

$$\sup_{\theta^* \in \Theta} \limsup_n E_{\theta^*} \|c_n(\beta)\|^2 < +\infty, \text{ for all } \beta. \quad (3)$$

Assumption A.13 For $\varphi = (\theta', \beta')'$, φ_0 as before and η large enough for $\overline{\mathcal{O}}_\eta(\varphi_0) \supset \overline{\mathcal{O}}_{\varepsilon_0}(\theta_0) \times \overline{\mathcal{O}}_{\varepsilon'}(b(\theta_0))$, $\text{rank}\left(\lim_{n \rightarrow \infty} \frac{\partial E_\theta c_n(b(\theta))}{\partial \theta'}\right) = p$, $\text{rank}\left(\lim_{n \rightarrow \infty} \frac{\partial E_\theta c_n(b(\theta))}{\partial \beta'}\right) = q$ on $\overline{\mathcal{O}}_{\varepsilon_0}(\theta_0)$, and $\sup_{\varphi \in \overline{\mathcal{O}}_\eta(\varphi_0)} \|D^{s^*+1} E_\theta c_n(\beta)\| < M$.

We are now ready to provide the analogous to lemmata 2.2 and 2.3 for the GT estimator.

Lemma 2.7 i) Under the assumptions 2.2, A.2, A.3, A.4, A.6, A.7.a and A.12 the GT is uniformly consistent for θ with rate $o(n^{-a})$.
ii) If additionally $c(\theta, \beta) = E_\theta c_n(\beta)$ and assumptions A.7.b, A.8 and A.13 hold then $\sqrt{n}(\text{GT} - \theta)$ has a locally uniform Edgeworth expansion of order s^* .

Notice that locally uniform (around φ_0) expansions of $\sqrt{n}(c_n(\beta) - c(\theta, \beta))$ of appropriate order could enable the result of the second part of the previous lemma without the employment of assumption A.13 in circumstances analogous to the ones described in lemma 2.5. We do not pursue this kind of reasoning in order to economize on space.⁷

3 Valid First and Second Moment Expansions

Having established the validity of Edgeworth expansions we are concerned with the approximation of their first and second moment sequences with a view towards bias-MSE comparisons. The methodology employed to derive the results is similar to the so-called *delta method* of approximations of moments of estimator sequences (see e.g. Linton [33] and McCullagh [35], Phillips [39], and Sargan [42]). To evaluate the approximate biases and MSEs we employ the following, general, lemma which establishes that, if the Edgeworth expansions involved are of an appropriately large order, the needed moment approximations are provided by the analogous moments of the Edgeworth distributions.⁸

⁷Similarly to the GMR2 case, the Edgeworth distributions validated in lemmas 2.6 and 2.7 need not coincide. Again their distance must be of order $o(n^{-a^*})$.

⁸Notice that separate methodologies concerning moment approximations (as for example the one in Koenker et al. [30]) are not general enough to cover our framework due to the following reasons. First, these methodologies concern only moment approximations, i.e. does not utilize the Edgeworth expansions, secondly, our auxiliary criterion is more general since our auxiliary model need not be a linear one, and thirdly these methodologies cannot provide analogous results for the GMR2 estimator.

Lemma 3.1 Suppose that K is a m -linear real function on \mathbb{R}^w , the support of an \mathbb{R}^w valued random element (say) ζ_n is bounded by $\mathcal{O}_{\sqrt{n}\rho}(0)$ for some $\rho > 0$, and ζ_n admits an Edgeworth expansion of order $s^* = 2a + m + 1$, then

$$\left| \int_{\mathbb{R}^q} K(z^m) \left(dP_n - \left(1 + \sum_{i=1}^s \frac{\pi_i(z)}{n^{\frac{i}{2}}} \right) \varphi_V(z) dz \right) \right| = o(n^{-a}),$$

where P_n , and $\left(1 + \sum_{i=1}^s \frac{\pi_i(z)}{n^{\frac{i}{2}}} \right) \varphi_V(z)$ denote the distribution of ζ_n and the density of the analogous Edgeworth measure of order $s = 2a + 1$ respectively. Moreover if P_n depends on θ , and $\pi_i(z)$ are continuous on $\overline{\mathcal{O}_\varepsilon}(\theta_0)$ for any z , V is continuous on $\overline{\mathcal{O}_\varepsilon}(\theta_0)$ and the expansion is uniformly valid on $\overline{\mathcal{O}_\varepsilon}(\theta_0)$, the approximation holds uniformly on $\overline{\mathcal{O}_\varepsilon}(\theta_0)$.

Notice that in the case that the support of ζ_n is not bounded the previous result would hold for $s^* = 2a + m + 2$. This follows easily from the proof of the previous lemma by letting $\rho = \ln^\epsilon n$ ($\epsilon \geq 1/2$) and by the fact that the Edgeworth approximation is *uniform w.r.t.* \mathcal{B}_C , and the fact that by construction \mathcal{B}_C contains the class of closed balls centered at zero.

Due to the fact that in the previous results we do not derive the Edgeworth approximations but only prove their valid existence, this lemma is not very practical for the derivation of the approximations of moments of the IIEs. This difficulty can be circumvented as follows. Notice first that the proof of the validity of the Edgeworth approximations in any of the examined cases includes the verification of the condition EXPAND of lemma AL.5 in the second part of the appendix. This implies the verification of condition 2.1–A. This is achieved by properly truncating the polynomial approximations of the *Lagrange inversion* of the first order conditions, that the estimators satisfy asymptotically (see also the proof of Theorem 3.2 in Arvanitis and Demos [5]). Consequently, the resulting functions are integrated with respect to the Edgeworth distribution appearing in assumption A.8 (due to theorem 2.1) delivering, thus, the required moments.

We proceed with the computation of these functions and the analogous integration w.r.t. the Edgeworth distribution appearing in assumption A.8 for the derivation of the first and second moment approximations for the estimators considered above when $s = 2$. These are derived as functions of the analogous approximations for $\widehat{\beta}_n$. This in turn is *sufficient* for the characterization of the 2^{nd} order bias and MSE structure of the IIEs at hand.

In the following we suppress the dependence on θ and z where possible for notational convenience. For the rest of this section we denote by $b = b(\theta)$, $b_{j,j'}$ is the j^{th} element of b , $W^* = E_\theta W^*(\theta)$, $W_{j,j'}^*$ is the (j, j') element of W^* , and analogously for W^{**} and $\mathcal{C} = \frac{\partial b'}{\partial \theta} W^* \frac{\partial b}{\partial \theta'}$. 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.8. $k_{i_\beta}(z, \theta) = \pi_{i-1}(z, \theta) \text{pr}_{1,q}(z)$, for any $i = 1, \dots, s^*$, $k_{1_{\theta^*}}(z, \theta) = \text{pr}_{q+1,p+q}(z)$ if $\theta_n^* - \theta$ appears in the vector $m_n(\theta)$, otherwise it is 0_q . $k_{i_{w^*}}(z, \theta)$ 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 $(W_n^*(\theta) - E_\theta W^*(\theta))_{j,j'}$ in $m_n(\theta)$, if the latter appears there, otherwise it is zero. Analogously, $k_{i_{w^{**}}}(z, \theta)$ is the symmetric $l \times l$ 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.

3.1 Valid 2nd order Bias approximation for the Indirect estimators

We are ready to provide the results for the second order bias approximation of the IIEs. Notice that due to their form, the results in Newey and Smith [38] imply that the bias will depend on the relation between p, q, l , the non linearities of the relevant estimating vectors and the stochastic weighting.

GMR1 Estimator We obtain the following lemma.

Lemma 3.2 *Let $\hat{\theta}_n$ denote the GMR1 estimator. If assumptions A.1, A.2, A.3, A.4, A.6 and A.7, A.8, A.9 and A.10 hold with $s^* \geq 3$ then uniformly over $\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$*

$$\left\| E_\theta \sqrt{n} (\hat{\theta}_n - \theta) - \frac{\xi_1(\theta)}{\sqrt{n}} \right\| = o(n^{-\frac{1}{2}}),$$

where

$$\begin{aligned} \xi_1(\theta) = & \mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^* \mathcal{I}_{\varphi_{V^*}}(k_{2_\beta}) \\ & - \frac{1}{2} \mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^* \mathcal{I}_{\varphi_{V^*}} \left(\left[\left(\mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^* k_{1_\beta} \right)' \frac{\partial b_j}{\partial \theta \partial \theta'} \mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^* k_{1_\beta} \right]_{j=1, \dots, q} \right) \\ & + \mathcal{C}^{-1} \mathcal{I}_{\varphi_{V^*}} \left(\left(\begin{array}{c} \left[\left(\mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^* k_{1_\beta} \right)' \frac{\partial^2 b'}{\partial \theta \partial \theta_j} \right]_{j=1, \dots, p} W^*(\theta) \\ + \frac{\partial b'}{\partial \theta} k_{1_{w^*}} + \frac{\partial b'}{\partial \theta} \left[\frac{\partial}{\partial \theta'} W_{j,j'}^* k_{1_{\theta^*}} \right]_{j,j'=1, \dots, q} \end{array} \right) \times \left(Id_q - \frac{\partial b}{\partial \theta'} \mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^* \right) k_{1_\beta} \right). \end{aligned}$$

The following corollary is trivial and it essentially assumes that the random element in assumption A.8 is only $\sqrt{n}(\hat{\beta}_n - b(\theta))$.

Corollary 1 *When W^* is independent of x and θ and $b(\theta)$ is affine then*

$$\xi_1(\theta) = \mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^* \mathcal{I}_V(k_{2_\beta}).$$

Hence ξ_1 is zero iff $\mathcal{I}_V(k_{2\beta})$ belongs to the kernel of $\mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^*$, which obviously is a $q - p$ dimensional subspace of \mathbb{R}^q for any θ . Suppose for instance that $b(\theta) = (\theta', 0'_{q-p})'$ in a large enough neighborhood of θ_0 .⁹ Then partitioning W^* as $\begin{pmatrix} W_1^* & W_3^* \\ W_3^{*'} & W_2^* \end{pmatrix}$ where W_1^* is $p \times p$, W_2^* is $q - p \times q - p$, and W_3^* is $p \times q - p$ we have that $\mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^* \mathcal{I}_V(k_{2\beta}) = (W_1^*)^{-1} \begin{pmatrix} W_1^* & W_3^* \end{pmatrix} \mathcal{I}_V(k_{2\beta})$ and thereby $\xi_2(\theta)$ locally equals 0_p iff $W_1^* [\mathcal{I}_V(k_{2\beta})_i]_{i=1, \dots, p} + W_3^* [\mathcal{I}_V(k_{2\beta})_i]_{i=p+1, \dots, q-p}$ is locally equal to 0_p . In the case that W_3^* is zero this is equivalent to $[\mathcal{I}_V(k_{2\beta})_i]_{i=1, \dots, p}$ locally equals 0_p . This discussion raises the issue of the optimal choice of W^* so that $\|\xi_2(\theta)\|$ is minimal but this is out of the scope of the present paper.

We continue now with the GMR2 case.

GMR2 Estimator The approximation contains the term $-\mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^* \mathcal{I}_V(k_{2\beta})$ something that is not present in the other two, a fact that is attributed to the existence of $E_\theta \widehat{\beta}_n$ in the definition of the particular estimator.

Lemma 3.3 *Let $\widehat{\theta}_n$ denote the GMR2 estimator. If assumptions A.1, A.2, A.3, A.4, A.6 and A.7, A.8, A.9, A.10 and A.11 hold for $s^* \geq 4$ then uniformly over $\mathcal{O}_{\varepsilon_*}(\theta_0)$ for any $\varepsilon_* < \varepsilon$*

$$\left\| E_\theta \sqrt{n} (\widehat{\theta}_n - \theta) - \frac{\xi_2(\theta)}{\sqrt{n}} \right\| = o(n^{-\frac{1}{2}}),$$

where

$$\xi_2(\theta) = \xi_1(\theta) - \mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^* \mathcal{I}_V(k_{2\beta}).$$

The following corollary is trivial and establishes general conditions under which the GMR2 estimator is second order unbiased.

Corollary 2 *When W^* is independent of x and θ and $b(\theta)$ is affine then $\xi_2(\theta) = 0_p$.*

This result is already known for the case where $p = q$, $\widehat{\beta}_n$ is a consistent estimator of θ , whence the GMR2 obviously performs a second order bias correction.¹⁰ Hence the previous *generalizes* the results in Gourioux and Monfort [24] and Gourioux et al. [27]. Finally it is easy to see that lemma 2.5 for $s^* \geq 4$ implies that the results in lemma 3.3 and corollary 2 also hold when the

⁹For a discussion on when the binding function admits this "canonical" form see section 4.

¹⁰If in addition $E_\theta \widehat{\beta}_n$ is linear, then the estimator is exactly unbiased (see Gourioux et al. [27]).

approximations appearing there are computed by taking into account that $p = q$, thereby becoming independent of the weighting matrix.

We conclude the presentation of the expansions with the last of the three estimators.

GT Estimator We present first a straightforward case implied by the previous results. Denoting with $\mathcal{D} = \frac{\partial b'}{\partial \theta} \frac{\partial c(\theta, b)}{\partial \beta'} W^{**}(\theta) \frac{\partial c(\theta, b)}{\partial \beta'} \frac{\partial b}{\partial \theta'}$, $\mathcal{E} = \frac{\partial b'}{\partial \theta} \frac{\partial c(\theta, b)}{\partial \beta'} W^{**}(\theta)$, $\mathcal{H}_j = \frac{\partial b'}{\partial \theta} \frac{\partial^2 c_j(\theta, b)}{\partial \beta \partial \beta'} \frac{\partial b}{\partial \theta'} - \left[\frac{\partial c_j(\theta, b)}{\partial \beta'} \frac{\partial^2 b}{\partial \theta' \partial \theta_r} \right]_{r=1, \dots, p}$, $\mathcal{J} = k_{1_{w^{**}}} + \left[\frac{\partial}{\partial \theta'} W^{**}(\theta)_{j,j'} k_{1_{\theta^*}} \right]_{j,j'=1, \dots, l}$, $\mathcal{J}^* = \left(\frac{\partial c(\theta, b)}{\partial \beta'} \frac{\partial b}{\partial \theta'} \mathcal{D}^{-1} \mathcal{E} - \text{Id}_l \right) \frac{\partial c(\theta, b)}{\partial \beta'}$ and $q_{1_\beta} = \mathcal{D}^{-1} \mathcal{E} \frac{\partial c(\theta, b)}{\partial \beta'} k_{1_\beta}$ we obtain the following lemma.

Lemma 3.4 Under A.12 suppose that $E_\theta c_n(\beta) = c(\theta, \beta)$. Furthermore let A.1, A.2, A.3, A.4, A.6, A.7, A.8, A.13 hold for $s^* \geq 3$, then uniformly on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$

$$\left\| E_\theta \sqrt{n} (\hat{\theta}_n - \theta) - \frac{\xi_3(\theta)}{\sqrt{n}} \right\| = o(n^{-\frac{1}{2}}),$$

where

$$\begin{aligned} \xi_3(\theta) = & \mathcal{D}^{-1} \mathcal{E} \frac{\partial c(\theta, b)}{\partial \beta'} \mathcal{I}_V(k_{2_\beta}) + \frac{1}{2} \mathcal{D}^{-1} \mathcal{E} \left[\mathcal{I}_V \left(k'_{1_\beta} \frac{\partial^2 c_j(\theta, b)}{\partial \beta \partial \beta'} k_{1_\beta} \right) \right]_{j=1, \dots, l} \\ & - \mathcal{D}^{-1} \mathcal{E} \left[\mathcal{I}_V \left(q'_{1_\beta} \frac{\partial b'}{\partial \theta} \frac{\partial^2 c_j(\theta, b)}{\partial \beta \partial \beta'} k_{1_\beta} \right) \right]_{j=1, \dots, l} + \frac{1}{2} \mathcal{D}^{-1} \mathcal{E} \left[\mathcal{I}_V \left(q'_{1_\beta} \mathcal{H}_j q_{1_\beta} \right) \right]_{j=1, \dots, l} \\ & + \mathcal{D}^{-1} \mathcal{I}_V \left(\left[\mathcal{H}_j q_{1_\beta} - \frac{\partial b'}{\partial \theta} \frac{\partial^2 c_j(\theta, \beta)}{\partial \beta \partial \beta'} |_b k_{1_\beta} \right]_{j=1, \dots, l} W^{**}(\theta) \mathcal{J}^* k_{1_\beta} \right) \\ & - \mathcal{D}^{-1} \frac{\partial b'}{\partial \theta} \frac{\partial c'(\theta, b)}{\partial \beta} \mathcal{I}_V(\mathcal{J} \mathcal{J}^* k_{1_\beta}). \end{aligned}$$

The following corollary proves that under the conditions in corollary 2 the GT estimator is not 2^{nd} order unbiased as opposed to the GMR2 one.

Corollary 3 When W^* is independent of x and θ and $b(\theta)$ is affine then

$$\begin{aligned} \xi_3(\theta) = & \mathcal{D}^{-1} \mathcal{E} \frac{\partial c(\theta, b)}{\partial \beta'} \mathcal{I}_V(k_{2_\beta}) + \frac{1}{2} \mathcal{D}^{-1} \mathcal{E} \left[\mathcal{I}_V \left(k'_{1_\beta} \frac{\partial^2 c_j(\theta, b)}{\partial \beta \partial \beta'} k_{1_\beta} \right) \right]_{j=1, \dots, l} \\ & + \frac{1}{2} \mathcal{D}^{-1} \mathcal{E} \left[\mathcal{I}_V \left(q'_{1_\beta} \frac{\partial b'}{\partial \theta} \frac{\partial^2 c_j(\theta, b)}{\partial \beta \partial \beta'} \left(\frac{\partial b}{\partial \theta'} q_{1_\beta} - 2k_{1_\beta} \right) \right) \right]_{j=1, \dots, l} \\ & + \mathcal{D}^{-1} \mathcal{I}_V \left(\left[\left(q'_{1_\beta} - k'_{1_\beta} \right) \frac{\partial b'}{\partial \theta} \frac{\partial^2 c_j(\theta, b)}{\partial \beta \partial \beta'} \right]_{j=1, \dots, l} W^{**}(\theta) \mathcal{J}^* k_{1_\beta} \right). \end{aligned}$$

Moreover, even under the scope of stochastic weighting, when $p = q = l$ and $b(\theta)$ is affine, then $\xi_3(\theta) = \left(\frac{\partial b}{\partial \theta'}\right)^{-1} \mathcal{I}_V(k_{2\beta})$.

This essentially provides a counterexample concerning the hypothesis of second order asymptotic equivalence between the three estimators. An analogous result applies even when $E_{\theta}c_n(\beta) = c(\theta, \beta)$ does not hold. There are cases where the GT estimator has the same approximate bias with the GMR1, as well as other cases where it has the same approximate bias with the GMR2. This is the topic of the next lemma.

Lemma 3.5 *i). Under the assumptions in lemma 2.6.A and for $s^* \geq 3$ we have that $\xi_1(\theta) = \xi_3(\theta)$ uniformly over $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$.
ii). Under the assumptions in lemma 2.6.B and for $s^* \geq 4$ we have that $\xi_2(\theta) = \xi_3(\theta)$ uniformly over $\overline{\mathcal{O}}_{\varepsilon_*}(\theta_0)$, for $\varepsilon_* < \varepsilon$.*

Notice that lemma 3.5.ii traces a special case under which the GT estimator can be second order unbiased. Finally this case can not be generally derived from corollary 3 since the corollary assumes that $E_{\theta}c_n(b(\theta)) = c(\theta, b(\theta)) = 0$ which is not, in general, true for lemma 3.5.ii as under its assumptions we have that $E_{\theta}c_n(b(\theta)) = E_{\theta}\widehat{\beta}_n - b(\theta) \neq 0$, unless $\widehat{\beta}_n$ is unbiased (see remark R.1, as well). Let us now consider the MSEs.

3.2 MSE 2nd order Approximations for the Indirect Estimators

Given the results of the previous subsection, establishing the conditions under which only the GMR2 estimator is second order unbiased, the question arising concerns the comparison between the analogous MSE approximations between the three.¹¹ Employing these results we obtain the following lemmas.

Lemma 3.6 *Let $\widehat{\theta}_n$ denote either the GMR1, or the GMR2 estimator. If $W^*(x, \theta)$ is independent of x and θ , b is affine and assumptions A.1, A.2, A.3, A.4, A.6 and A.7, A.8, A.9 hold for $s^* \geq 5$, then uniformly on $\overline{\mathcal{O}}_{\varepsilon_*}(\theta_0)$, for any $\varepsilon_* < \varepsilon$,*

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

where

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

¹¹It 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.

This along with corollary 2 establishes the second order superiority of the GMR2 estimator in the particular case. It is also easy to see that lemma 2.5 for $s^* \geq 5$, implies again that the result in lemma 3.6 also holds when the approximations appearing are computed by taking into account that $p = q$, thereby becoming independent of the weighting matrix. For the GT we obtain the following result.

Lemma 3.7 *Let $\hat{\theta}_n$ denote the GT estimator. If $W^{**}(x, \theta)$ is independent of x and θ , b is affine, $E_{\theta}c_n(\beta) = c(\theta, \beta)$ and assumptions A.1, A.2, A.3, A.4, A.6, A.7, A.8, and A.13 hold for $s^* \geq 4$ then, uniformly on $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$*

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

where

$$\begin{aligned} H_1(\theta) &= \mathcal{D}^{-1} \mathcal{E} \frac{\partial c(\theta, b)}{\partial \beta'} V(\theta) \frac{\partial c'(\theta, b)}{\partial \beta} \mathcal{E}' \mathcal{D}^{-1} \\ H_2(\theta) &= \mathcal{D}^{-1} \mathcal{E} \frac{\partial c(\theta, b)}{\partial \beta'} \mathcal{I}_V \left(k_{2\beta} k_{1\beta}' \right) \frac{\partial c'(\theta, b)}{\partial \beta} \mathcal{E}' \mathcal{D}^{-1}. \end{aligned}$$

Notice that even in the case of stochastic weighting, when $W^* = \frac{\partial c(\theta, b)}{\partial \beta'} W^{**}$ then $H_1(\theta)$ coincide for all three estimators. This is in accordance with the conditions implying their first order equivalence (see for example chapter 4 of Gourieroux and Monfort [24]). Moreover under the assumptions of both lemmas and if $p = q = l$ then, $H_2(\theta)$ coincide for all three estimators *establishing the superiority of the GMR2 estimator* in the light of corollaries 1, 2, 3 and lemmas 3.6, 3.7 as it is second order unbiased while having the same second order MSE with the other two. Finally it is easy to see that analogous results could also be obtained in the light of assumptions validating lemma 2.7.A-B, if additionally $W^{**}(x, \theta)$ is independent of x and θ , and b is affine. In any of these cases we would have that the second order MSE s coincide for the three estimators. In case **A** the GMR2 would be generally superior and in the case **B** the GMR2 would be equivalent to the GT and these would be superior to the GMR1.

4 Recursive GMR2

The previous section highlights the fact that the second order bias properties of the GMR2 estimator depend among others on the local behavior of the binding function. Due to its injectivity as prescribed by assumption A.3, it is easy to see that B can always be chosen so that $b(\theta)$ is in the *canonical* form $(\theta', 0'_{q-p})'$.¹²

¹²Assumption A.9, and Theorem 10.2 of Spivak [47] (p. 44) imply a local version of this result.

In most cases this reparametrization of the auxiliary model is analytically intractable. However there exists at least one II estimation procedure that can be employed in order to approximate it. Given the GMR1, let $\beta'_n = (\text{GMR1}', 0'_{q-p})$ and apply the GMR2 estimator to the latter. Then the resulting IIE is derived from a three-step procedure, in the last step of which the binding function is obviously in *canonical* form. An extension of the three step procedure of the previous remark to an arbitrary number of steps, where the k^{th} -step auxiliary estimator is the the GMR2 of the previous step embedded to \mathbb{R}^q , can provide an unbiased IIE of arbitrary order when k is large enough. This extension is the object of study of the present section. Obviously, the embedding of the auxiliary estimator in any step after the first to \mathbb{R}^q is irrelevant and therefore will be dropped.

We define recursive II estimation procedures as follows. Let $\widehat{\theta}_n^{(0)}$ denote any consistent estimator of θ .

Definition D.6 Let $k \in \mathbb{N}$, the recursive k – GMR2 estimator (denoted by $\theta_n^{(k)}$) is defined in the following steps:

1. $\widehat{\theta}_n^{(1)} = \arg \min_{\theta} \left\| \widehat{\theta}_n^{(0)} - E_{\theta} \widehat{\theta}_n^{(0)} \right\|,$
2. for $k > 1$ $\widehat{\theta}_n^{(k)} = \arg \min_{\theta} \left\| \widehat{\theta}_n^{(k-1)} - E_{\theta} \widehat{\theta}_n^{(k-1)} \right\|.$

Using the results of the previous section, we are now able to prove the following lemma.¹³

Lemma 4.1 Suppose that assumptions A.6, A.8, A.11 hold for $\theta_n^{(0)}$ for $s^* \geq 2k + 4$. Moreover suppose that $E_{\theta} \sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon}(\theta_0)} \|n\overline{s}_n\|^2 < +\infty$ and $E_{\theta} \sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon}(\theta_0)} \|n\overline{H}_n\| < +\infty$ for all $\theta \in \overline{\mathcal{O}}_{\varepsilon}(\theta_0)$ and $\sqrt{n}\overline{s}_n(\theta)$ admits a locally uniform Edgeworth expansion of order 6, where \overline{s}_n and \overline{H}_n denote the average score and Hessian. Then the k – GMR2 estimator is of order $s = 2k + 1$ unbiased and has the same MSE with the $(k - 1)$ – GMR2, up to $2k$ order, uniformly over $\overline{\mathcal{O}}_{\varepsilon_*}(\theta_0)$ for any $\varepsilon_* < \varepsilon$.

The previous lemma implies the existence of IIEs that are higher order unbiased when obtained in this recursive structure. The investigation of algorithms

¹³The proof of lemma 4.1 indicates that the conditions on the behavior of derivatives of the likelihood function are solely employed for the establishment that at any step of the recursion $\widehat{\theta}_n^{(i)}$ satisfies $\widehat{\theta}_n^{(i-1)} = E_{\theta_n^{(i)}} \widehat{\theta}_n^{(i-1)}$ with probability $1 - o(n^{-a^*})$ independent of θ . If the latter is adopted as an assumption a more general form of this lemma is easily deduced.

concerning feasible numerical approximations of those estimators with similar properties is delegated to future research. For now, consider again the case where $k = 1$. Then $1 - \text{GMR2}$ is actually 3^{rd} order unbiased at θ_0 hence the previous results are essentially expanded under the conditions of the lemma. Furthermore, the $1 - \text{GMR2}$ has the same second order MSE as the GMR2 (and/or GMR1) one.

Further, notice that the conclusions of corollary 2 and lemma 4.1 are valid provided that the expectation involved in the definitions of the GMR2 and $k\text{-GMR2}$ estimator can be evaluated explicitly. However, as we have already mentioned, these are approximated by a Monte Carlo method, based on, say, H replications. Of course, unless H is infinite the aforementioned conclusions hold only approximately. Nevertheless, although the expectations' approximations are outside the scope of the present paper, the results in Gouriéroux and Monfort [24] and Gouriéroux et al. [27] can be employed to evaluate the minimum number of H so that the bias of the second step estimators have a lower absolute approximate bias than the first ones (see the examples section for more on this). *In any case, our results can be viewed as a benchmark for the finite H cases, as well as directions for the design of IIEs that possess those properties while avoiding the shortcomings of the ones defined by "crude" approximations of the intractable expectations via resampling* (see last paragraph of conclusions section, as well). Let us now turn our attention to two examples.

5 Examples and Monte Carlo Experiments

In this section we verify our conditions to ensure the validity of locally uniform Edgeworth expansions concerning the estimators presented here. We do this by considering two classes of models, i.e. invertible $\text{MA}(1)$ and $\text{GARCH}(1, 1)$. Given the fact that in both examples $p = q = l$ we can assume that the weighting matrices are identities *without loss of generality*. Furthermore, we investigate the finite n effect on our asymptotic results of the bias and MSE. Hence, our Monte Carlo exercise is by no means exhaustive.

5.1 The $\text{MA}(1)$ Case

Consider the set of invertible $\text{MA}(1)$ processes defined by

$$y_j = \varepsilon_j + \theta \varepsilon_{j-1},$$

where the (ε_j) are iid, with $E\varepsilon_0 = 0$, $E\varepsilon_0^2 = 1$, $E(\varepsilon_0^{2s+4}) < +\infty$ the distribution of ε_0 admits a positive continuous density and $\theta \in \text{Int } \Theta$ for $\Theta = [\underline{\eta}_{\text{MA}}, \bar{\eta}_{\text{MA}}]$ where $\underline{\eta}_{\text{MA}} > -1, \bar{\eta}_{\text{MA}} < 1$.

Let

$$b(\theta) = \frac{\theta}{1 + \theta^2}$$

and for some compact $B \supseteq b(\Theta)$ define

$$\widehat{\beta}_n \in \arg \min_{\beta \in B} \frac{1}{2} \|\rho_n - \beta\|^2$$

where $\rho_n = \frac{\sum_{j=1}^n y_j y_{j-1}}{\sum_{j=1}^n y_j^2}$. Notice that β_n is also the QMLE of the coefficient of an AR(1) auxiliary model (see Gourieroux et al. [25], and Demos and Kyriakopoulou [14]) and the binding function is $b(\theta) = \frac{\theta}{1+\theta^2}$. Furthermore define

$$\text{GMR1} \in \arg \min_{\theta \in \Theta} \frac{1}{2} \left\| \widehat{\beta}_n - b(\theta) \right\|^2$$

and $c_n(\beta) = \rho_n - \beta$. To evaluate the GMR2 estimator we have to evaluate the $E_{\theta} \widehat{\beta}_n$ (see definition D.3).

Applying now GMR2 on the GMR1 we get the $1 - \text{GMR2}$ (or equivalently $\theta_n^{(1)}$ employing the notation of section 4). The existence of any of the GMR2 type estimators can be easily deduced via the continuity part of assumption A.1 when $\varepsilon_0 \sim N(0, 1)$. Given the form of $c_n(\beta)$ the existence of the GT estimator also follows with probability that converges to 1 sufficiently fast.

Proposition 4 *If $E\varepsilon_i^{14} < \infty$ and if $D(0, 1)$ has a positive continuous density, then the $\widehat{\beta}_n$, and GMR1 admit 4^{th} order valid Edgeworth expansions, uniformly over Θ . Furthermore if $\varepsilon_0 \sim N(0, 1)$ the GT, GMR2 and the $1 - \text{GMR2}$ admit 4^{th} order valid Edgeworth expansions, uniformly over any compact subset of Θ .*

Given the results of the above proposition and those in section 3, the 2^{nd} order first and second moment approximations are valid, for all estimators (see lemma 3.1). Furthermore, since the binding function is not linear, neither the GMR1 nor the GMR2 and the GT are 2^{nd} order unbiased (see results in section 3.1). However, the $1 - \text{GMR2}$ is 3^{rd} order unbiased by lemma 4.1. Also lemmas 3.6 and 3.7 imply the second order superiority of $1 - \text{GMR2}$ w.r.t. the GMR1, GMR2 and the GT uniformly over Θ .¹⁴

¹⁴Notice that the expansions of the $\widehat{\beta}_n$ and GMR1 estimators are available from the work of Demos and Kyriakopoulou [14]. These formulae imply that both estimators, GMR1 and GMR2, are biased, unless $\theta = 0$.

Monte Carlo Experiment To evaluate the GMR2 estimator we need $E_\theta \widehat{\beta}_n$ (see definition D.3) which is analytically intractable. We approximate this expectation numerically by

$$\frac{1}{H} \sum_{i=1}^H \frac{\sum_{t=2}^n y_t^{(i)} y_{t-1}^{(i)}}{\sum_{t=2}^n \left(y_{t-1}^{(i)}\right)^2},$$

where $H = 1600$. The same applies for E_θ (GMR1) needed for the $1 - \text{GMR2}$ estimator. In this respect we obtain approximate GMR2 and $1 - \text{GMR2}$ estimators. The choice of H rests on two facts with opposite effects. First, higher H delivers a better approximation to the expectations. On the opposite, the higher H is, the higher is the time cost of obtaining the GMR2 and $1 - \text{GMR2}$ estimators, especially in a Monte Carlo exercise. Furthermore, equation (4.29) in Gouriéroux and Monfort [24] (equation (28) in Gouriéroux et al. [27]) provides the minimum value of H so that the absolute value of 2^{nd} order bias of $1 - \text{GMR2}$ is smaller than the one of GMR1. Employing this equation and taking the asymptotic variance formula of the GMR1 from Demos and Kyriakopoulou [14] we get that, for $\theta = 0.4$, if $H \geq 4$ the absolute 2^{nd} order bias of $1 - \text{GMR2}$ is smaller than the one of GMR1. However, notice that as the estimator is univariate and consequently the evaluation of $1 - \text{GMR2}$ is not too expensive, in terms of computing time, we have chosen a quite high value of H , i.e. $H = 1600$.

To assess the impact of these numerical approximations on our results as well as the performance of the resulting estimators for finite n , we engage to the following Monte Carlo experiment. We draw a sample of $n \in \{30, 50, 100, 250, 500, 750, 1000\}$ observations from a standard normal. For each random sample, we generate the MA(1) process y_t for $\theta \in \{-0.5, 0.4\}$. We evaluate $\widehat{\beta}_n$ and if the estimate is in the $[-0.499999, 0.499999]$ interval we retain the sample, otherwise we throw it away and draw another one. For each retained sample we evaluate the three estimators, i.e. the GMR1, GMR2, and $1 - \text{GMR2}$.

In figure 1 the absolute biases, multiplied by n , of all three estimators are presented, for $\theta = 0.4$. According to the results of Demos and Kyriakopoulou [14], for this value of θ , we should have that $n |E_\theta(\text{GMR1}) - \theta| - 1.252 = o(1)$, $n |E_\theta(\text{GMR2}) - \theta| - 2.094 = o(1)$ and the $1 - \text{GMR2}$ estimator should be 2^{nd} order unbiased. It is obvious that, for $\theta = 0.4$, the theoretical results are validated for $n \geq 500$. In terms of MSEs, in figure 2, it seems again that for $n \geq 500$ the MSE of all three estimators are close to their theoretical value, which is 1.796. Consequently, the $1 - \text{GMR2}$ estimator appears second order superior to the GMR1 and GMR2 ones. Similar results emerge for $\theta = -0.5$, not presented here for reasons of economy of space.

5.2 The GARCH(1,1) Case

Consider the set of stationary ergodic and covariance stationary GARCH (1, 1) processes (Bollerslev [9]) defined by the recursion

$$\begin{aligned} y_j &= \varepsilon_j h_j^{1/2}, \\ h_j &= \theta_1 (1 - \theta_2 - \theta_3) + (\theta_2 \varepsilon_{j-1}^2 + \theta_3) h_{j-1}, \end{aligned}$$

where the (ε_j) are iid, $E\varepsilon_0 = 0$, $E\varepsilon_0^2 = 1$, $E\varepsilon_0^{28} < +\infty$ the distribution of ε_0 admits a positive continuous density and $\theta = (\theta_1, \theta_2, \theta_3)' \in \Theta = [\underline{\eta}_\omega, \bar{\eta}_\omega] \times [\underline{\eta}_\alpha, \bar{\eta}_\alpha] \times [\underline{\eta}_\beta, \bar{\eta}_\beta]$ where $\underline{\eta}_\omega, \underline{\eta}_\alpha, \underline{\eta}_\beta > 0$ and for any $\theta \in \Theta$, $E(\theta_2 \varepsilon_0^2 + \theta_3)^{14} < 1$.
Let

$$b(\theta) = \left(\theta_1, \frac{\theta_2 (1 - (\theta_2 + \theta_3) \theta_3)}{1 - 2\theta_2 \theta_3 - \theta_3^2}, \theta_2 + \theta_3 \right)',$$

and for some compact $B \supseteq b(\Theta)$ and $c_n(\beta) = \left((\bar{y}^2, \hat{\rho}_1, \frac{\hat{\rho}_2}{\hat{\rho}_1}) - \beta \right)'$ define

$$\widehat{\beta}_n \in \arg \min_{\beta \in B} \frac{1}{2} \|c_n(\beta)\|^2,$$

where $\bar{y}^2 = \frac{1}{n} \sum_{j=1}^n y_j^2$, $\hat{\rho}_i = \frac{\frac{1}{n} \sum_{t=1}^n (y_t^2 y_{t-i}^2) - (\bar{y}^2)^2}{\frac{1}{n} \sum_{t=1}^n (y_t^4) - (\bar{y}^2)^2}$. Furthermore define

$$\text{GMR1} \in \arg \min_{\theta \in \Theta} \frac{1}{2} \left\| \widehat{\beta}_n - b(\theta) \right\|^2.$$

Notice that $\widehat{\beta}_n$ is the estimator of the constant, the first order autocorrelation and the autoregressive parameter of the ARMA (1, 1) reparametrization of the GARCH (1, 1) (commonly called the Pantula reparametrization, see Bollerslev [9]), i.e. the auxiliary model is:

$$y_j^2 = \theta_1 (1 - \theta_2 - \theta_3) + (\theta_2 + \theta_3) y_{j-1}^2 + v_j - \theta_3 v_{j-1},$$

where $v_j = y_j^2 - h_j$, a martingale difference sequence. Further, the GMR1 estimator is the one introduced and its asymptotic properties analyzed in Kristensen and Linton [31].

Now employing the GMR2 estimator, treating the GMR1 as an auxiliary one, we get the 1 - GMR2 estimator. The issue of existence of any of the GMR2 type and the GT estimators follows exactly as in the previous example when $\varepsilon_0 \sim N(0, 1)$.

Proposition 5 *If the distribution of ε_0 admits a positive and continuous density then $\widehat{\beta}_n$ and GMR1 admit 4th order valid Edgeworth expansions, uniformly over Θ . Furthermore if the distribution of ε_0 is standard normal, then GMR2, 1 – GMR2 and GT admit 4th order valid Edgeworth expansions, uniformly over any compact subset of Θ .*

Given the results of the previous proposition and those in section 3 the 2nd order first and second moment approximations are valid, for all estimators. Furthermore, since the binding function is not linear, neither the GMR1 nor the GMR2 nor the GT are 2nd order unbiased (see results in section 3.1). However, the 1 – GMR2 is 3rd order unbiased and has the same second order MSE with the other three (see lemma 4.1), due to lemmas 3.6, and 3.7 indicating the second order superiority of 1 – GMR2 w.r.t. to GMR1, GMR2 and GT uniformly over Θ^* .

Monte Carlo Experiment As we have already mentioned, we have to evaluate $E_\theta \widehat{\beta}_n$ and E_θ (GMR1) which are analytically intractable. Consequently, we approximate these expectations numerically, i.e. we approximate $E_\theta \widehat{\beta}_n$ by

$$\frac{1}{H} \sum_{i=1}^H \left(\overline{y}^{(i)}, \widehat{\rho}_1^{(i)}, \frac{\widehat{\rho}_2^{(i)}}{\widehat{\rho}_1^{(i)}} \right)',$$

$$\text{where } H = 60, \widehat{\rho}_i^{(i)} = \frac{\sum_{t=i+1}^n \left((y_t^{(i)})^2 - \overline{y}^{(i)} \right) \left((y_{t-i}^{(i)})^2 - \overline{y}^{(i)} \right)}{\sum_{t=i+1}^n \left((y_t^{(i)})^2 - \overline{y}^{(i)} \right)^2}, \overline{y}^{(i)} = \frac{1}{n} \sum_{t=1}^n (y_t^{(i)})^2,$$

and similarly for the E_θ (GMR1). The choice of $H = 60$ is much lower than the one employed in the previous MA (1) example. However, notice that in this case 1 – GMR2 is a three dimensional vector and consequently its evaluation is very costly, in terms of computing resources, within the context of the Monte Carlo experiment. Furthermore, the calculations of the asymptotic variance of the GMR1 estimator, presented in Kristensen and Linton [31], and its partial derivatives are extremely difficult to derive. Nevertheless, these quantities can be approximated by simulations. In fact, we approximated them by setting $n = 100000$, $(\theta_1, \theta_2, \theta_3) = (1.0, 0.05, 0.7)$ and performing 5000 Monte Carlo replications. Consequently, employing these approximations and equation (4.29) in Gouriéroux and Monfort [24] (or equation (28) in Gouriéroux et al. [27]) we estimate that if $H \geq 10$ the absolute 2nd order bias of 1 – GMR2 is smaller than the one of GMR1. Hence, taking into consideration the computational cost associated with the evaluation of the 1 – GMR2 estimator, we decided that a value of H as high as six times the Gouriéroux and Monfort [24] lower limit would offer a satisfactory approximation of E_θ (GMR1).

Again, we perform the following Monte Carlo experiment to assess the impact of these approximations on our results as well as the finite sample behavior of the resulting estimators. However, notice that the analytic results of sections 3 and 4 cannot be applied since there are no Edgeworth expansions' analytic results for the first step estimator $\widehat{\beta}_n$. Of course this is not a problem for the simulations. We generate, with $n \in \{250, 500, 1000, 5000, 10000, 20000, 50000, 100000\}$, the GARCH(1, 1) model, plus 250 observations to initialize the process. We choose $(\theta_1, \theta_2, \theta_3) = (1.0, 0.05, 0.7)$ and perform 5000 Monte Carlo replications. Notice that under the following conditions the $\widehat{\beta}_n$ estimator is always defined:

$$\widehat{\rho}_1, \widehat{\rho}_2 > 0, \quad \frac{\widehat{\rho}_2}{\widehat{\rho}_1} < 1, \quad \text{and} \quad \widehat{\rho}_1 - \frac{\widehat{\rho}_2}{\widehat{\rho}_1} < 0.$$

In case that any of these conditions is not satisfied the random drawing is thrown away and we draw a new one.

In figure 3 the norms of the biases of the three estimators, multiplied by n , are presented. It is obvious that for this model the asymptotic results are validated for large n . Notice that for any n the bias of the 1 – GMR2 estimator is almost zero. Furthermore, the norms of, the MSEs of the three estimators are almost equal, for large sample values (see figure 4). Hence we can say that first, the results of this paper are validated, for the chosen GARCH(1, 1) process, and second, that the approximations to the expectations are satisfactory. As in the previous example the 1 – GMR2 estimator appears second order superior to the other three estimators.

6 Conclusions

Our results can be summarized as follows: we provide conditions that ensure the validity of locally uniform Edgeworth approximations for the three IIEs, of arbitrary order. Notice that the conditions validate *a fortiori* the first order theory providing a rigorous framework for the derivation of the GMR2 properties. Then, we provide integrability conditions that validate moment approximations of the aforementioned estimators. We derive the relevant 2^{nd} order bias and MSE approximations for the three IIEs under quite general conditions. These enable differences in the dimensions of the auxiliary estimating equations and/or the parameter spaces employed, and consequently the possibility of stochastic weighting in any of the steps of the estimation procedure. We confirm that under our assumption framework and in the special case of deterministic weighting and affinity of the binding function, the GMR2 estimator is second order unbiased. This result can be easily generalized when the auxiliary model is properly reparametrized. The GMR1 and GT estimators do not have this property under

the same conditions. Moreover the second order approximations of the MSE in this case imply the superiority of the GMR2 estimators. Hence we establish conditions that ensure the *existence* of IIEs that are second order unbiased and have the same second order MSE with second order biased estimators of the same class.

Furthermore, by generalizing to multistep procedures we are able to provide recursive IIEs that are locally uniformly unbiased at any given order when analogous conditions hold. However, the practical implementation of these seems numerically involved. Nevertheless, the construction of algorithms to implement these estimators, possibly employing results in Andrews [2], could be the object of further research.

Possible further extensions are the following: First, the derivation of the analogous approximations when the true parameter value and/or its image w.r.t. the binding function lie at the boundary of the parameter spaces (see Calzolari et al. [10]). This could also imply the first order asymptotic non-equivalence between the three IIEs. Second, an application of the Edgeworth approximations could lay in the derivation of higher order properties of II testing procedures. Third, as stated before the IIEs studied are usually analytically intractable, and in applications they are approximated via the approximations of the binding function and/or the employed expectations by simulators. Our results can be *also* considered as *guidelines* for the construction of appropriate simulators that could lead to estimators that share the same properties with the GMR2 estimator. One example could be the definition of IIEs via the actual *use* of the Edgeworth approximations for the auxiliary one. For example, an IIE could be defined by substituting $E_{\theta} \widehat{\beta}_n$ with $\mathcal{I}_{\varphi_V(\theta)}(k_{2\beta})$ in definition D.3 or some *numerical approximation* of it (see Arvanitis and Demos [6]). We leave all these questions for future work.

7 Discussion on Assumptions

In this section, by employing primitive assumptions, we discuss ways of verifying the high level assumptions made mainly for the auxiliary estimator in section 2.

Discussion on Assumption A.6 Given assumption A.2, assumption A.6 can be justified as in lemma AL.1 in the second part of the appendix. When $Q_n = \|c_n(\beta)\|_{W_n(\widehat{\beta}_n^*)}$ as in assumption A.5, the requirements of this lemma would be satisfied if

$$\|c_n(\beta) - c_n(\beta')\| \leq \kappa_n \|\beta - \beta'\|, \text{ for all } \beta, \beta'$$

with $\sup_{\theta^* \in \Theta} P_{\theta^*}(\kappa_n > M) = o(n^{-a^*})$ and

$$\sup_{\theta^* \in \Theta} P_{\theta^*}(\|c_n(\beta) - c(\theta^*, \beta)\| > \varepsilon) = o(n^{-a^*}), \forall \varepsilon > 0,$$

where $c(\theta, \beta)$ is jointly continuous on $\Theta \times B$ and equals zero iff $\beta = b(\theta)$, W_n satisfies a similar assumption to A.7 below and $\widehat{\beta}_n^*$ satisfies the current assumption due to lemmas AL.2 and AL.3 in the second part of the appendix. In this case it is obvious that $\widehat{\beta}_n^*$ can be defined similarly to $\widehat{\beta}_n$. Finally when κ_n is in the form of an arithmetic mean the asymptotic boundness with uniform probability $1 - o(n^{-a^*})$ can be obtained by conditions on the asymptotic behavior of $\sup_{\theta^* \in \Theta} E_{\theta^*} \kappa_n^q$ where q is a function of a^* implied by the form of dependence between the elements of κ_n (consider for example the Yokoyama-Doukhan inequality in the case of weak dependence-see Andrews [2], 7.3.3). Alternatively, under the Assumptions A1-A3 in Newey [37] for $Q_n(\beta)$ assumption A.6 holds. $\widehat{\theta}_n^*$ could be any of the three IIEs.

Discussion on Assumption A.8 We now briefly discuss conditions that could validate assumption A.8 as well as provide some references for counterexamples. When $\widehat{\beta}_n$ is an M-estimator the verification of the assumption A.8 proceeds via the verification of conditions 2.1–A, 2.1–B, 2.1–C and 2.1–D (for a descriptive justification of the procedure see the initial paragraph of this section and for a general application see also lemma AL.5 in the appendix). 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 first order conditions (f.o.c.) that the estimator asymptotically satisfies with high probability.

In the case that $\widehat{\beta}_n$ 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. In this case the establishment of the condition 2.1–B relies on the properties of those random elements. For example in the context of weakly dependent time series models, as described in Götze and Hipp [23], Assumptions A.1-3, Theorem 2.1 and Lemma 2.6 of Arvanitis and Demos [5] provide conditions that imply 2.1–B. For a more general set of sufficient conditions see also Durbin [16]. These conditions along with corollary 3.3 of Skovgaard [45] actually imply that the Edgeworth expansion holds uniformly over the whole Borel algebra. The approach of Durbin [16] was utilized for the establishment of 2.1–B in Lieberman et al. [32] and Andrews and Lieberman [3] in the context of linear, Gaussian, strongly dependent time series models, when b is the identity and $\widehat{\beta}_n$ is the MLE or the Whittle MLE. When $\widehat{\beta}_n$ is itself an IIE then the verification of

2.1– B renders relevant the results in the subsequent section of the paper (see also the examples section).

Condition **2.1– C** can be verified via the employment of **2.1– B** and a mean value expansion of the f.o.c. in many cases (for a different approach see the proof of lemma 2.4.ii in the case of the GMR2 estimator). Finally condition **2.1– D** can be verified due to the form of the remainder in the approximation of **2.1– A**, which is usually a remainder emerging from the employment of the Mean Value Theorem, and it is bounded in norm by $C \|S_n\|^d$ for d large enough and $C > 0$ and **2.1– B**.

Assumption **A.8** might fail if $\sqrt{n}(\widehat{\beta}_n - b(\theta))$ is not asymptotically normal. This in turn can happen either because S_n is not asymptotically normal, in which case **2.1– B** cannot hold (a plethora of examples can be obviously constructed in the context of non stationary time series models), or in cases where $b(\theta)$ lies in the boundary of B and thereby **2.1– A** cannot hold (notice that $b(\theta)$ can be a boundary point even if θ is an interior point). Finally notice that this assumption might fail to hold (essentially due to failure of **2.1– B**) even in cases where S_n is asymptotically normal and $b(\theta)$ is 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) of Corradi and Iglesias [13] fail to hold for the QMLE in the context of a semi-parametric GARCH(1, 1) model. Assumption **A.8** would follow for the entire $\sqrt{nm_n}(\theta)$ by analogous arguments. For the equicontinuity part of **2.1. B** see the paragraph entitled as "Discussion on Assumption **A.11**" below and consider the case where the differentiability order is zero.

Discussion on Assumptions **A.9 and **A.10**** For an appropriately differentiable likelihood function, a condition such as $\sup_{\theta \in \Theta} \|D^2 E_{\theta} \widehat{\beta}_n\| < M$ could be justified according to the following. Conditions such as $E_{\theta} \sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon}(\theta_0)} \|n \overline{s}_n\|^2 < +\infty$ and $E_{\theta} \sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon}(\theta_0)} \|n \overline{H}_n\| < +\infty$ for all θ would enable the commutativity of differentiation with integration by dominated convergence and the fact that B is bounded. Then if $\sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon}(\theta_0)} E_{\theta} \|\widehat{\beta}_n - b(\theta)\|^2 = O(\frac{1}{n})$ something that in the light of the results in section 3 holds when $s^* \geq 3$ in assumption **A.8**, it would suffice that

$$\sup_{\theta \in \Theta} E \|n \overline{s}_n(\theta) \overline{s}'_n(\theta) + E \overline{H}_n(\theta)\|^2 = O(1),$$

$$\sup_{\theta \in \Theta'} E \|\overline{H}_n(\theta) - E \overline{H}_n(\theta)\|^2 = O(1),$$

where \overline{s}_n and \overline{H}_n denote the average score and Hessian. It is easy to see that these would hold if assumptions A.1, A.2-M, A.2-WD in Arvanitis and Demos

[4] hold for the random vector consisting of the log-likelihood derivatives. For an implementation of this in the scope of a GARCH (1, 1) example see Lemma AL.9 in Arvanitis and Demos [4].

Discussion on Assumption A.11. Tracing the proof of Theorem 3.2 in Arvanitis and Demos [5], we obtain that for the quadratic form in the density approximation of the auxiliary estimator it would suffice that $C_{i,j,n}(\theta)$ and the analogous polynomials in the Edgeworth densities of $S_n(\theta)$ are s^* (equi-) continuously differentiable functions of θ . When, for example, S_n is a random vector consisting of the random elements appearing in the derivatives of the likelihood function or the moment conditions then it is of the form of a normalized sum that satisfies Assumption A.2 of Arvanitis and Demos [5]. In this case, as it can be seen from the proof of Theorem 2.1 in the same paper, remark (2.12) and paragraph 4 of Gotze and Hipp [22], it would suffice that the cumulant spectral densities up to order $s^* - 1$ and their derivatives w.r.t. to the angular frequencies be s^* times continuously differentiable over $\mathcal{O}_\varepsilon(\theta_0)$.

References

- [1] Andersen, T.G., L. Benzoni and J. Lund (2002), "An empirical investigation of continuous-time equity return models", *The Journal of Finance*, LVII, pp. 1239-1284.
- [2] Andrews, D.W.K. (2002), "Equivalence of the Higher-order Asymptotic Efficiency of k-step and Extremum Statistics", *Econometric Theory*, 18, pp. 1040–1085.
- [3] Andrews, D.W.K. and O. Lieberman (2005), "Valid Edgeworth expansions for the Whittle maximum likelihood estimators for stationary long-memory Gaussian time series" , *Econometric Theory*, 21, pp. 710-734.
- [4] Arvanitis, S. and A. Demos (2013) "On the Validity of Edgeworth Expansions and Moment Approximations for Three Indirect Estimators (Extended Revised Appendix)", working paper No. 1330, Dept. of IEES, AUEB (<http://wpa.deos.aueb.gr/docs/bias-ii-revis-jspi-ta.pdf>).
- [5] Arvanitis, S. and A. Demos (2012) "Valid Locally Uniform Edgeworth Expansions Under Weak Dependence and Sequences of Smooth Transformations", *Journal of Time Series Econometrics* (forthcoming) working paper No. 1229, Dept. of IEES, AUEB (<http://wpa.deos.aueb.gr/docs/Unif-Edg-fin-wp.pdf>).

- [6] Arvanitis, S. and A. Demos (2011) "A New Class of Indirect Estimators and Bias Correction" (working paper No. 05-11, Dept Economics, AUEB (<http://www.econ.aueb.gr/uploadfiles/052011AIDP.pdf>))
- [7] Bansal, R.P., A. R. Gallant, R. Hussey and G. Tauchen (1995), "Nonparametric estimation of structural models for high-frequency currency market data" *Journal of Econometrics*, 66, 251-287.
- [8] Bhattacharya, R.N. and J.K. Ghosh (1978), "On the validity of the formal Edgeworth expansion", *The Annals of Statistics*, Vol. 6, No. 2, pp. 434-451.
- [9] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* 31, 307-328.
- [10] Calzolari, C., Fiorentini G. and E. Sentana (2004), Constrained Indirect Inference Estimation, *Review of Economic Studies* **71**, pp. 945-973.
- [11] Chumacero, R.A. (2001), "Estimating ARMA Models Efficiently", *Studies in Nonlinear Dynamics and Econometrics*, 5, 103-114.
- [12] Chung, C. and G. Tauchen (2001), "Testing target-zone models using efficient method of moments" *Journal of Business and Economic Statistics*, 19, 255-269.
- [13] Corradi, V. and E.M. Iglesias (2008), "Bootstrap refinements for *QML* estimators of the *GARCH* (1,1) parameters", *Journal of Econometrics*, 144, 500-510.
- [14] Demos, A. and D. Kyriakopoulou (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.
- [15] Dridi, R., and E. Renault (2000), "Semi-parametric Indirect Inference", mimeo, LSE and CREST-INSEE and Universite de Paris IX Dauphine.
- [16] Durbin, J. (1980), "Approximations for Densities of Sufficient Estimators", *Biometrika* 67-2, pp. 311-333.
- [17] Gallant, R.A., D. Hsieh and G. Tauchen (1997), "Estimation of stochastic volatility models with diagnostics", *Journal of Econometrics*, 81, 159-192.
- [18] Gallant, R.A. and J.R. Long (1997) "Estimating stochastic differential equations efficiently by Minimum Chi-Squared", *Biometrika*, 84, 125-141.

- [19] Gallant, R.A. and G.E. Tauchen (1996), "Which Moments to Match," *Econometric Theory*, Vol. 12, pp. 657-681.
- [20] Garcia, R., E. Renault and D. Veredas (2011) "Estimation of stable distributions by indirect inference", *Journal of Econometrics*, 161, 325-337.
- [21] Ghysels, I., L. Khalaf and C. Vodounou (2003), "Simulation based inference in moving average models" *Annales d' Economie et de Statistique*, 69, 85-99.
- [22] Götze, F. and C. Hipp (1983), "Asymptotic expansions for sums of weakly dependent random vectors", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 64, pp. 211-239.
- [23] Götze, F. and C. Hipp (1994), "Asymptotic distribution of statistics in time series", *The Annals of Statistics*, 22, pp. 2062-2088.
- [24] Gouriéroux, C., and A. Monfort (1996), "*Simulation-based Econometric Methods*", CORE Lectures, Ox. Un. Press.
- [25] Gouriéroux, C., A. Monfort, and E. Renault (1993), "Indirect Inference", *Journal of Applied Econometrics*, Volume 8 Issue 1, pp. 85-118.
- [26] Gouriéroux, C., P.C.B. Phillips and J. Yu (2010), "Indirect inference for dynamic panel models", *Journal of Econometrics*, 157, 68-77.
- [27] Gouriéroux, C., E. Renault and N. Touzi, (2000), "Calibration by simulation for small sample bias correction", in "*Simulation-Based Inference in Econometrics: Methods and Applications*", edited by R. Mariano, T. Schuermann and M.J. Weeks, Cambridge University Press, Cambridge.
- [28] Gradshteyn, I. S., and I.M. Ryzhik (1994), "Table of Integrals, Series, and Products" A. Jeffrey Editor, Academic Press.
- [29] Hall, P. and J.L. Horowitz (1996), "Bootstrap critical values for tests based on generalized-method-of-moment estimators", *Econometrica* 64, 891-916.
- [30] Koenker, R., J.A.F. Machado, C.L. Skeels, and A.H. Welsh (1994) "Momentary lapses: Moment expansions and the robustness of minimum distance estimation" , *Econometric Theory* 10, 172-197.
- [31] Kristensen, D. and O. Linton (2006) "A closed-form estimator for the GARCH(1,1) model", *Econometric Theory* 22, 323-337.

- [32] Lieberman, O., J. Rousseau and D.M. Zucker (2003), "Valid asymptotic expansions for the maximum likelihood estimator of the parameter of a stationary, Gaussian, strongly dependent process", *Annals of Statistics* 31, 586–612.
- [33] Linton, O. (1997), "An asymptotic expansion in the GARCH(1,1) model", *Econometric Theory* 13, 558-581.
- [34] Magdalinos, M.A. (1992), "Stochastic Expansions and Asymptotic Approximations", *Econometric Theory*, Vol. 8, No. 3, pp. 343-367.
- [35] McCullagh, P. (1987), "*Tensor Methods in Statistics*" Chapman and Hall, London, New York.
- [36] Michaelides, A. and S. Ng (2000), "Estimating the rational expectations model of speculative storage: A Monte Carlo comparison of three simulation estimators" *Journal of Econometrics*, 96, 231-266.
- [37] Newey, W.K. (1991) "Uniform Convergence in Probability and Stochastic Equicontinuity", *Econometrica* 59, 1161-1167.
- [38] Newey, W.K., and R.J. Smith (2001), "*Asymptotic Bias and Equivalence of GMM and GEL Estimators*", mimeo, Department of Economics, MIT.
- [39] Phillips, P.C.B. (1980), "Finite sample theory and the distribution of alternative estimators of the marginal propensity to consume", *Review of Economic Studies* 47, 183-224.
- [40] Phillips, P.C.B. (2012), "Folklore Theorems, Implicit Maps and Indirect Inference" *Econometrica* 80, 425-454.
- [41] Robinson, P.M. (1988) "The stochastic difference between econometric statistics", *Econometrica* 56, 531-548.
- [42] Sargan, J.D. (1976) "Econometric estimators and the Edgeworth approximation", *Econometrica* 44, 421-448.
- [43] Sentana, E., G. Calzolari and G. Fiorentini (2008), "Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks" *Journal of Econometrics*, 146, 10-25.
- [44] Skovgaard I.M. (1981), "Transformation of an Edgeworth expansion by a sequence of smooth functions", *Scandinavian Journal of Statistics* 8, 207-217.

- [45] Skovgaard I.M. (1986), "On Multivariate Edgeworth Expansions", *International Statistical Review / Revue Internationale de Statistique*, Vol. 54, No. 2, pp. 169-186.
- [46] Smith, A.A. (1993), "Estimating nonlinear time-series models using Simulated Vector Autoregressions", *Journal of Applied Econometrics* 8, 63-84.
- [47] Spivak, M. (1999), "*A Comprehensive Introduction to DIFFERENTIAL GEOMETRY*", Vol. 1, Publish or Perish, Inc, Texas.
- [48] Sweeting, T.J. (1977), "Speeds of convergence for the multidimensional central limit Theorem", *The Annals of Statistics* 5, 28-41.
- [49] van der Vaart, A.W. and J.A. Wellner (2000), "*Weak convergence and empirical processes*" Springer.

Appendix

Proofs

Proof of Lemma 2.2. i) Due to the triangle inequality and assumption A.6 we have that for $\varepsilon > 0$

$$\begin{aligned} & \sup_{\theta^* \in \Theta} P_{\theta^*} \left(\sup_{\theta \in \Theta} \left| \left\| \widehat{\beta}_n - b(\theta) \right\| - \left\| b(\theta^*) - b(\theta) \right\| \right| > \varepsilon \right) \\ & \leq \sup_{\theta^* \in \Theta} P_{\theta^*} \left(\left\| \widehat{\beta}_n - b(\theta^*) \right\| > \varepsilon \right) = o(n^{-a^*}) \end{aligned}$$

Hence for $q_n(\theta) = \widehat{\beta}_n - b(\theta)$, $q(\theta^*, \theta) = b(\theta^*) - b(\theta)$ and by assumption A.7.a) lemma AL.3 applies. Hence for $\gamma(\theta) = \theta$ due to assumption A.3 lemma AL.1 also applies implying the result.

ii) Given i), we have that $\widehat{\theta}_n \in \mathcal{O}_\varepsilon(\theta)$ with P_θ -probability $1 - o(n^{-a^*})$ that is locally independent of θ for any $\varepsilon > 0$. For some ε small enough, such that $\mathcal{O}_\varepsilon(\theta) \subset \mathcal{O}_{\varepsilon_0}(\theta_0)$ (which exists due to the fact that $\varepsilon_0 > \varepsilon$) due to assumption A.9, we have that condition FOC of the appendix lemmas AL.4 and AL.5 is satisfied by the GMR1 estimator with $Q_n \doteq \frac{\partial b'}{\partial \theta}$. Furthermore assumption A.9 implies conditions HUB ($\gamma(\theta) = \theta$ hence set $\delta = \varepsilon_0$) and RANK of the same lemma. Condition TIGHT follows from A.8, as under this assumption there is $C^* > 0$ locally independent of θ such that

$$\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\sqrt{n} \left\| \widehat{\beta}_n - b(\theta) \right\| > C^* \ln^{1/2} n \right) = o(n^{-a^*}) \quad (4)$$

(see Lemma AL.2 of Arvanitis and Demos [5]). Hence lemma AL.4 applies ensuring that

$$\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\sqrt{n} \|\text{GMR1} - \theta\| > C \ln^{1/2} n \right)$$

for some $C > 0$ locally independent of θ . Hence condition UTIGHT of lemma AL.5 holds. Moreover assumption A.8 implies condition UEDGE of the same lemma for $M_n(\theta) = \sqrt{n}m_n(\theta)$. Due to assumption A.9 for any $\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$ and any θ_* sufficiently close to θ , $\frac{\partial b'}{\partial \theta}(\theta_*)$ admits a Taylor expansion of order $s^* - 1$ around θ of the form

$$\begin{aligned} \frac{\partial b'}{\partial \theta}(\theta_*) &= \sum_{i=0}^{s^*-1} \frac{1}{i!} D^i \frac{\partial b'}{\partial \theta}(\theta) \left((\theta_* - \theta)^i \right) \\ &+ \frac{1}{(s^* - 1)!} \left(D^{s^*-1} \frac{\partial b'}{\partial \theta}(\theta^+) - D^{s^*-1} \frac{\partial b'}{\partial \theta}(\theta) \right) \left((\theta_* - \theta)^{s^*-1} \right) \end{aligned}$$

where θ^+ lies between θ_* and θ . This implies that for any $\widehat{\theta}_n = \text{GMR1}$ due to condition UTIGHT we have that with P_θ -probability $1 - o(n^{-a^*})$ locally independent of θ

$$\frac{\partial b'}{\partial \theta}(\widehat{\theta}_n) = \sum_{i=0}^{s^*-1} \frac{1}{i!} \frac{1}{n^{i/2}} D^i \frac{\partial b'}{\partial \theta}(\theta) \left(\left(\sqrt{n}(\widehat{\theta}_n - \theta) \right)^i \right) + R_n^*(\widehat{\theta}_n, \theta)$$

$$\text{where } R_n^*(\widehat{\theta}_n, \theta) = \frac{1}{(s^*-1)!} \frac{1}{n^{(s^*-1)/2}} \left(D^{s^*-1} \frac{\partial b'}{\partial \theta}(\theta_n^+) - D^{s^*-1} \frac{\partial b'}{\partial \theta}(\theta) \right) \left(\left(\sqrt{n}(\widehat{\theta}_n - \theta) \right)^{s^*-1} \right),$$

and θ_n^+ lies between $\widehat{\theta}_n$ and θ . Now by assumption A.9 $\frac{\partial b'}{\partial \theta}(\theta)$ has full rank for any $\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$ and by submultiplicativity, the relation of θ_n^+ to $\widehat{\theta}_n$ and condition UTIGHT

$$\begin{aligned} & \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| \frac{1}{(s^*-1)!} \frac{1}{n^{(s^*-1)/2}} \left(D^{s^*-1} \frac{\partial b'}{\partial \theta}(\theta_n^+) - D^{s^*-1} \frac{\partial b'}{\partial \theta}(\theta) \right) \times \left(\left(\sqrt{n}(\widehat{\theta}_n - \theta) \right)^{s^*-1} \right) \right\| > \gamma_n^* \right) \\ & \leq \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\frac{1}{(s^*-1)!} \frac{1}{n^{(s^*-1)/2}} \sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon_0}(\theta_0)} \left\| D^{s^*} \frac{\partial b'}{\partial \theta}(\theta) \right\| \|\theta_n^+ - \theta\| \left\| \sqrt{n}(\widehat{\theta}_n - \theta) \right\|^{s^*-1} > \gamma_n^* \right) \\ & \leq \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\frac{M}{(s^*-1)!} \frac{C^{s^*}}{n^{s^*/2}} \ln^{s^*/2} n > \gamma_n^* \right) + o(n^{-a^*}) \end{aligned}$$

which is of order $o(n^{-a^*})$ for $\gamma_n^* = \frac{M}{(s^*-1)!} \frac{C^{s^*}}{n^{s^*/2}} \ln^{s^*/2} n = o(n^{-a^*})$ and locally independent of θ . Analogously, due to assumption A.9 for any $\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$ and any θ_* sufficiently close to θ , $b(\theta_*)$ admits a Taylor expansion of order $s^* - 1$ around θ of the form

$$\begin{aligned} q_n &= \widehat{\beta}_n - b(\theta_*) = \widehat{\beta}_n - b(\theta_*) - \sum_{i=1}^{s^*} \frac{1}{i!} D^i b(\theta) \left((\theta_* - \theta)^i \right) \\ &\quad - \frac{1}{s^*!} \left(D^{s^*} b(\theta^+) - D^{s^*} b(\theta) \right) \left((\theta_* - \theta)^{s^*} \right) \end{aligned}$$

where θ^+ lies between θ_* and θ . This implies that for $\widehat{\theta}_n$ we have that with P_θ -probability $1 - o(n^{-a^*})$

$$\begin{aligned} & \sqrt{n} \left(\widehat{\beta}_n - b(\widehat{\theta}_n) \right) \\ &= \sqrt{n} \left(\widehat{\beta}_n - b(\theta) \right) \\ &\quad + \sum_{i=0}^{s^*-1} \frac{1}{(i+1)!} \frac{1}{n^{i/2}} D^{i+1} b(\theta) \left(\left(\sqrt{n}(\widehat{\theta}_n - \theta) \right)^{i+1} \right) + R_n^\#(\widehat{\theta}_n, \theta) \end{aligned}$$

where $R_n^\#(\widehat{\theta}_n, \theta) = \frac{1}{s^*!} \frac{1}{n^{(s^*-1)/2}} (D^{s^*} b(\theta_n^+) - D^{s^*} b(\theta)) \left(\left(\sqrt{n} (\widehat{\theta}_n - \theta) \right)^{s^*} \right)$, and θ_n^+ lies between $\widehat{\theta}_n$ and θ . Now by assumption A.9 $\frac{\partial b'}{\partial \theta}(\theta)$ has full rank for any $\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$ and so does the identity matrix in front of $\sqrt{n}(\widehat{\beta}_n - b(\theta))$, and thereby due to submultiplicativity, the relation of θ_n^+ to $\widehat{\theta}_n$ and condition UTIGHT

$$\begin{aligned} & \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| \frac{1}{s^*!} \frac{1}{n^{(s^*-1)/2}} (D^{s^*} b(\theta_n^+) - D^{s^*} b(\theta)) \times \left(\left(\sqrt{n} (\widehat{\theta}_n - \theta) \right)^{s^*} \right) \right\| > \gamma_n^\# \right) \\ & \leq \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\frac{1}{s^*!} \frac{1}{n^{(s^*-1)/2}} \sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon_0}(\theta_0)} \|D^{s^*+1} b(\theta)\| \|\theta_n^+ - \theta\| \times \left\| \sqrt{n} (\widehat{\theta}_n - \theta) \right\|^{s^*} > \gamma_n^\# \right) \\ & \leq \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\frac{M C^{s^*+1}}{s^*!} \frac{1}{n^{s^*/2}} \ln^{(s^*+1)/2} n > \gamma_n^\# \right) + o(n^{-a^*}) \end{aligned}$$

which is of order $o(n^{-a^*})$ for $\gamma_n^\# = \frac{M C^{s^*+1}}{s^*!} \frac{1}{n^{s^*/2}} \ln^{(s^*+1)/2} n = o(n^{-a^*})$ and locally independent of θ . Finally due to lemma AL.6 which applies by assumptions A.8 and A.7 condition EXPAND holds and the result follows by the same lemma. ■

Proof of Lemma 2.3. The proof is in the same spirit as the proof of lemma 2.2. For $\varepsilon > 0$, let $E(\varepsilon, \theta) = \left\{ \omega \in \Omega : \left\| \widehat{\beta}_n - b(\theta) \right\| > \frac{\varepsilon}{2} \right\} \in \mathcal{F}$, then

$$\sup_{\theta \in \Theta} \left\| E_\theta \widehat{\beta}_n - b(\theta) \right\| \leq \sup_{\theta \in \Theta} E_\theta \left\| \widehat{\beta}_n - b(\theta) \right\| 1_{E(\varepsilon, \theta)} + \frac{\varepsilon}{2}.$$

As B is bounded, due to assumption A.2 and by assumption A.6 there exists an n^* such that $\sup_{\theta \in \Theta} P_\theta \left(\left\| \widehat{\beta}_n - b(\theta) \right\| > \frac{\varepsilon}{3} \right) \leq \frac{\varepsilon}{2M}$ where M denotes the diameter of B . Hence

$$\sup_{\theta \in \Theta} \left\| E_\theta \widehat{\beta}_n - b(\theta) \right\| \leq \varepsilon \text{ for all } n \geq n^*$$

and since ε is arbitrary

$$\sup_{\theta \in \Theta} \left\| E_\theta \widehat{\beta}_n - b(\theta) \right\| = o(1) \quad (5)$$

Due to the triangle inequality and assumption A.6 we have that for $\varepsilon > 0$

$$\begin{aligned} & \sup_{\theta^* \in \Theta} P_{\theta^*} \left(\sup_{\theta \in \Theta} \left| \left\| \widehat{\beta}_n - E_\theta \widehat{\beta}_n \right\| - \|b(\theta^*) - b(\theta)\| \right| > \varepsilon \right) \\ & \leq \sup_{\theta^* \in \Theta} P_{\theta^*} \left(\left\| \widehat{\beta}_n - b(\theta^*) \right\| + \sup_{\theta \in \Theta} \left\| E_\theta \widehat{\beta}_n - b(\theta) \right\| > \varepsilon \right) = o(n^{-a^*}) \end{aligned}$$

For $q_n(\theta) = \widehat{\beta}_n - E_\theta \widehat{\beta}_n$, $q(\theta^*, \theta) = b(\theta^*) - b(\theta)$ and by assumption A.7.a) lemma AL.3 applies. Hence for $\gamma(\theta) = \theta$ due to assumption A.3 lemma AL.1 also applies implying the result.

ii) Given i), we have that $\widehat{\theta}_n \in \mathcal{O}_\epsilon(\theta)$ with P_θ -probability $1 - o(n^{-a^*})$ that is locally independent of θ for any $\epsilon > 0$. For some ϵ small enough, such that $\mathcal{O}_\epsilon(\theta) \subset \mathcal{O}_{\varepsilon_0}(\theta_0)$ (which exists due to the fact that $\varepsilon_0 > \varepsilon$) due to assumption A.10, we have that condition FOC of the appendix lemmas AL.4 and AL.5 is satisfied by the GMR2 estimator with $Q_n \doteq \frac{\partial E_\theta(\widehat{\beta}_n)'}{\partial \theta}$. Furthermore assumption A.10 and A.9 imply conditions HUB ($\gamma(\theta) = \theta$ hence set $\delta = \varepsilon_0$) and RANK of the same lemma due to the fact that since $D^2 E_\theta \widehat{\beta}_n$ is uniformly bounded on $\mathcal{O}_{\varepsilon_0}(\theta_0)$, $D E_\theta \widehat{\beta}_n$ converges uniformly to $D b(\theta)$ due to lemma AL.7 and therefore the rank condition is implied by A.10 for large enough n . Condition TIGHT follows from A.8 and Lemma AL.2 of Arvanitis and Demos [5] and the fact that as $a^* > a \geq 0$ we have that $a^* > 0$ and there exists a $C_2 > 0$ locally independent of θ and for $E^* = \left\{ \omega \in \Omega : \left\| \widehat{\beta}_n - b(\theta) \right\| > C_2 \frac{\ln^{1/2} n}{n^{1/2}} \right\} \in \mathcal{F}$

$$\begin{aligned}
& \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} \left\| E_\theta \widehat{\beta}_n - b(\theta) \right\| \\
& \leq \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} E_\theta \left[\left\| \widehat{\beta}_n - b(\theta) \right\| 1_{E^*} \right] + \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} E_\theta \left[\left\| \widehat{\beta}_n - b(\theta) \right\| 1_{\overline{E^*}} \right] \\
& \leq M \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| \widehat{\beta}_n - b(\theta) \right\| > C_2 \frac{\ln^{1/2} n}{n^{1/2}} \right) \\
& \quad + C_2 \frac{\ln^{1/2} n}{n^{1/2}} \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| \widehat{\beta}_n - b(\theta) \right\| \leq C_2 \frac{\ln^{1/2} n}{n^{1/2}} \right) \\
& = o(n^{-a^*}) + C_2 \frac{\ln^{1/2} n}{n^{1/2}} (1 - o(n^{-a^*})) = O\left(\frac{\ln^{1/2} n}{n^{1/2}}\right),
\end{aligned}$$

where the last line comes from equation 4, above, and therefore

$$\begin{aligned}
& \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| \widehat{\beta}_n - E_\theta \widehat{\beta}_n \right\| > C_1 \frac{\ln^{1/2} n}{n^{1/2}} \right) \\
& \leq \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| \widehat{\beta}_n - b(\theta) \right\| + \left\| E_\theta \widehat{\beta}_n - b(\theta) \right\| > C_1 \frac{\ln^{1/2} n}{n^{1/2}} \right) \\
& \leq \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| \widehat{\beta}_n - b(\theta) \right\| + O\left(\frac{\ln^{1/2} n}{n^{1/2}}\right) > C_1 \frac{\ln^{1/2} n}{n^{1/2}} \right) = o(n^{-a^*})
\end{aligned}$$

Hence lemma AL.4 applies ensuring that

$$\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\sqrt{n} \|\text{GMR2} - \theta\| > C \ln^{1/2} n \right) = o(n^{-a^*})$$

for some $C > 0$ locally independent of θ , hence condition UTIGHT of lemma AL.5 holds. Moreover assumption A.8 along with the fact that the support of $\widehat{\beta}_n - b(\theta)$ is uniformly bounded by $\overline{\mathcal{O}}_{3\eta}(0)$ for any η greater or equal than the diameter of B , and the fact that $\sqrt{n} \left(\widehat{\beta}_n - E_\theta \widehat{\beta}_n \right)$ admits a locally uniform Edgeworth expansion of order $s^* - 1$ (see Lemma 4.1 of Arvanitis and Demos [5]) imply condition UEDGE of lemma AL.5 for $M_n(\theta) = \sqrt{n} \left(\frac{m_n(\theta)}{\widehat{\beta}_n - E_\theta \widehat{\beta}_n} \right)$ with order $s^* - 1$. Due to assumption A.10 for any $\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$ and any θ_* sufficiently close to θ , $\frac{\partial E_\theta(\widehat{\beta}_n)'}{\partial \theta}(\theta_*)$ admits a Taylor expansion of order $s^* - 1$ around θ and the rest follows as in the proof of lemma 2.2 above with the difference that here $\frac{\partial E_{\theta_*}(\widehat{\beta}_n)'}{\partial \theta}$ is considered instead of $\frac{\partial b'}{\partial \theta}(\theta_*)$ and assumption A.10 is employed instead of A.9 (see Arvanitis and Demos [4] for details). ■

Proof of Lemma 2.4. By assumption A.8, lemma 3.1, below, adding subtracting

$$\begin{aligned} & \sqrt{n} \left(b(\theta) + \sum_{i=1}^s \frac{\mathcal{I}_V(k_{i\beta}(z, \theta))}{n^{\frac{i}{2}}} \right) \text{ and } \sqrt{n} \left(b(\theta_n^+) + \sum_{i=1}^s \frac{\mathcal{I}_V(k_{i\beta}(z, \theta_n^+))}{n^{\frac{i}{2}}} \right), \text{ we get} \\ & \sqrt{n} \left(E_{\theta_n^*} \widehat{\beta}_n - E_\theta \widehat{\beta}_n \right) - A_n(\theta) = \\ & \sqrt{n} \left(E_{\theta_n^*} \widehat{\beta}_n - b(\theta_n^+) - \sum_{i=1}^s \frac{\mathcal{I}_V(k_{i\beta}(z, \theta_n^+))}{n^{\frac{i}{2}}} \right) - \sqrt{n} \left(E_\theta \widehat{\beta}_n - b(\theta) - \sum_{i=1}^s \frac{\mathcal{I}_V(k_{i\beta}(z, \theta))}{n^{\frac{i}{2}}} \right) + \\ & \sqrt{n} \left(b(\theta_n^+) - b(\theta) - \sum_{i=1}^s \frac{1}{i!} D^i b(\theta) \left((\theta_n^+ - \theta)^i \right) \right) + B_n \text{ where} \\ & B_n = \sum_{i=1}^s \frac{\mathcal{I}_V(k_{i\beta}(z, \theta_n^+))}{n^{\frac{i-1}{2}}} - \sum_{i=1}^s \frac{\mathcal{I}_V(k_{i\beta}(z, \theta))}{n^{\frac{i-1}{2}}} - \sum_{i=1}^s \frac{1}{i!} \sum_{j=1}^{s-i} D^i \frac{\mathcal{I}_V(k_{j\beta}(z, \theta))}{n^{\frac{j-1}{2}}} \left((\theta_n^+ - \theta)^j \right). \end{aligned}$$

Employing the mean value Theorem for $\mathcal{I}_V(k_{i\beta}(z, \theta_n^+))$, and for θ_n^{++} such that $\|\theta_n^{++} - \theta\| < \|\theta_n^+ - \theta\|$, and collecting terms we get:

$$B_n = \sum_{i=1}^s \frac{1}{n^{\frac{i-1}{2}}} \frac{1}{(s-i+1)!} D^{s-i+1} \mathcal{I}_V(k_{i\beta}(z, \theta)) \left((\theta_n^{++} - \theta)^{s-i+1} \right).$$

Taking into account that $\theta_n^+ \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$ with probability $1 - o(n^{-a^*})$ and em-

plying the triangular inequality we have that, for $s < s^*$,

$$\begin{aligned}
& \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| \sqrt{n} \left(E_{\theta_n^+} \widehat{\beta}_n - E_\theta \widehat{\beta}_n \right) - A_n(\theta) \right\| > \gamma_n \right) \\
& \leq \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} \left\| \sqrt{n} \left(E_\theta \widehat{\beta}_n - b(\theta) - \sum_{i=1}^s \frac{\mathcal{I}_V(k_{i\beta}(z, \theta))}{n^{\frac{i}{2}}} \right) \right\| > \frac{\gamma_n}{6} \right) \\
& \quad + \sum_{i=1}^s \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\frac{1}{n^{\frac{i-1}{2}}} \|B_n\| > \frac{\gamma_n}{3s} \right) \\
& \quad + \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| \sqrt{n} \left(b(\theta_n^+) - b(\theta) - \sum_{i=1}^s \frac{1}{i!} D^i b(\theta) \left((\theta_n^+ - \theta)^i \right) \right) \right\| > \frac{\gamma_n}{3} \right) + o(n^{-a^*}).
\end{aligned}$$

Now we have that

$$a_n = \sqrt{n} \left\| E_\theta \widehat{\beta}_n - b(\theta) - \sum_{i=1}^s \frac{\mathcal{I}_V(k_{i\beta}(z, \theta))}{n^{\frac{i}{2}}} \right\| = o(n^{-a})$$

independent of θ , due to lemma 3.1.

Further, due to the continuity of $D^{s-i+1} \mathcal{I}_V(k_{i\beta}(z, \theta))$, assumption A.11, and the assumption of the asymptotic behavior of θ_n^+ we get

$$\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\frac{1}{n^{\frac{i-1}{2}}} \|B_n\| > \frac{\gamma_n}{3s} \right) \leq o(n^{-a^*})$$

provided that $\gamma_n \geq \gamma_n^{(1)} = \frac{\ln \frac{s-i+1}{2} n}{n^{\frac{s}{2}}} \frac{3s \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} \|D^{s-i+1} \mathcal{I}_V(k_{i\beta}(z, \theta))\|}{(s-i+1)!}$.

Furthermore using the same reasoning as above

$$\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| \sqrt{n} \left(b(\theta_n^+) - b(\theta) - \sum_{i=1}^s \frac{1}{i!} D^i b(\theta) \left((\theta_n^+ - \theta)^i \right) \right) \right\| > \frac{\gamma_n}{3} \right) \leq o(n^{-a^*})$$

when $\gamma_n \geq \gamma_n^{(2)} = \frac{3 \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} \|D^{s+1} b(\theta)\|}{(s+1)!} \frac{\ln \frac{s+1}{2} n}{n^{\frac{s}{2}}}$. Hence for

$\gamma_n = \max \left(\gamma_n^{(2)}, 6a_n, \gamma_n^{(1)}, i = 1, \dots, s \right)$ the result follows for large enough n . ■

Proof of Lemma 2.5. i) Notice that the uniform consistency follow for GMR1 and GMR2 as in the first parts of lemmas 2.2, 2.3. Assumption A.9 along with i) imply that for $r = 1, 2$, $\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} \left\| D^r \left(E_\theta \widehat{\beta}_n - b(\theta) \right) \right\| < M$, which in turn means that $D^{r-1} \left(E_\theta \widehat{\beta}_n - b(\theta) \right)$ are uniformly Lipschitz on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$, and therefore uniformly equicontinuous on the same ball. This implies the commutativity of the limit, with respect to n and the derivative operator, uniformly over $\overline{\mathcal{O}}_\varepsilon(\theta_0)$.

This along with the second part of assumption A.9, i.e. $\text{rank } Db(\theta) = p$ for all θ in $\overline{\mathcal{O}}_{\varepsilon_0}(\theta_0)$, and continuity imply that $\text{rank } DE_{\theta}\widehat{\beta}_n = p$, for all θ in $\mathcal{O}_{\varepsilon_0}(\theta_0)$ for n large enough. As now $p = q$, by the definition of GMR2 we get that $\widehat{\beta}_n = E_{\text{GMR2}}\widehat{\beta}_n$ with probability $1 - o(n^{-a^*})$ uniformly on $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$. This implies condition FOC of lemma AL.5. Furthermore by the second part of lemma 2.2 we have that the GMR1 is locally uniformly tight with probability $1 - o(n^{-a^*})$ locally independent of θ . By the mean value Theorem of $b(\text{GMR1})$ around GMR2 we have that with P_{θ} -probability $1 - o(n^{-a^*})$ locally independent of θ

$$\text{GMR1} - \text{GMR2} = \left(\frac{\partial b'(\theta_n^+)}{\partial \theta} \right)^{-1} (b(\text{GMR1}) - b(\text{GMR2})),$$

where θ_n^+ is such that $\|\theta_n^+ - \text{GMR2}\| < \|\text{GMR1} - \text{GMR2}\|$. As now $p = q$, by the definition of GMR1 we get that $b(\text{GMR1}) = \widehat{\beta}_n$ with probability $1 - o(n^{-a^*})$ uniformly on $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$. Hence with probability $1 - o(n^{-a^*})$ uniformly on $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$

$$\|\text{GMR1} - \text{GMR2}\| \leq M \left\| E_{\text{GMR2}}\widehat{\beta}_n - b(\text{GMR2}) \right\| = O\left(\frac{1}{n}\right)$$

and the last equality is true (as $\widehat{\beta}_n$ has a uniform Edgeworth expansion on $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$, assumption A.8, and apply lemma 3.1). Taking into account that the GMR1 is locally uniformly tight we get that, for some $C > 0$, locally independent of θ

$$\sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon}(\theta_0)} P_{\theta} \left(\sqrt{n} \|\text{GMR2} - \theta\| > C \ln^{1/2} n \right) = o(n^{-a^*}).$$

This implies condition UTIGHT of lemma AL.5. It also, along with lemmas 2.4 and 3.1, implies that for any $\varepsilon_* < \varepsilon$

$$\sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon_*}(\theta_0)} P_{\theta} \left(\left\| \sqrt{n} \left(\widehat{\beta}_n - E_{\text{GMR2}}\widehat{\beta}_n \right) - \Gamma_n(\theta) \right\| > \gamma_n \right) = o(n^{-a})$$

where $\gamma_n = o(n^{-a})$ independent of θ and

$$\begin{aligned} \Gamma_n(\theta) &= \sqrt{n} \left(\widehat{\beta}_n - E_{\theta}\widehat{\beta}_n \right) - \sum_{i=1}^{s-1} \frac{1}{n^{\frac{i}{2}}} \mathcal{I}_V(k_{i\beta}(z, \theta)) \\ &\quad - \sum_{i=1}^s \frac{1}{n^{\frac{i-1}{2}} i!} D^i \left(b(\theta) + \sum_{j=1}^{s-i} \frac{\mathcal{I}_V(k_{j\beta}(z, \theta))}{n^{\frac{j}{2}}} \right) \left(\sqrt{n} (\text{GMR2} - \theta)^i \right) \end{aligned}$$

which validates condition EXPAND lemma AL.5 of for $Q_n = W_n^j = \text{Id}_{p \times p}$. Moreover assumption A.8 along with the fact that the support of $\widehat{\beta}_n - b(\theta)$

is uniformly bounded by $\overline{\mathcal{O}}_{3\eta}(0)$ for any η greater or equal than the diameter of B , and Lemma 4.1 of Arvanitis and Demos [5] imply condition UEDGE of the same Lemma for $M_n(\theta) = \sqrt{n} \left(\widehat{\beta}_n(\theta) - E_\theta \widehat{\beta}_n \right)$ with order $s^* - 1$. Hence the conditions of lemma AL.5 are satisfied and the result follows. ii) follows the same way as i) except now $\left\| \widehat{\beta}_n - E_{\text{GMR2}} \widehat{\beta}_n \right\|$ is zero with probability $1 - o(n^{-a^*})$ independent of θ . ■

Proof of Lemma 2.6. **A.** It is easy to see that this special assumption implies that GMR1 = GT with probability $1 - o(n^{-a^*})$ independent of θ . The rest are trivial consequences of lemma 2.2. **B.** Similarly this special assumption implies that GMR2 = GT (P_θ almost surely for all θ). The rest are trivial consequences of lemmas 2.3 or 2.5 i) or ii). ■

Proof of Lemma 2.7. i) By assumption A.12.1-3., uniform integrability and via the triangle inequality we get that

$$\sup_{\theta^* \in \Theta} P_{\theta^*} \left(\sup_{\theta \in \Theta} \left\| E_{\theta} c_n \left(\widehat{\beta}_n \right) - c(\theta, b(\theta^*)) \right\| > \varepsilon \right) = o(n^{-a^*})$$

Hence for $q_n(\theta) = E_{\theta} c_n \left(\widehat{\beta}_n \right)$, $q(\theta^*, \theta) = c(\theta, b(\theta^*))$ and by assumptions A.7.a) lemma AL.3 applies. Hence for $\gamma(\theta) = \theta$ due to assumptions A.3, A.12 lemma AL.1 also applies proving the result.

ii) Given i), we have that $\widehat{\theta}_n \in \mathcal{O}_\epsilon(\theta)$ with P_θ -probability $1 - o(n^{-a^*})$ that is locally independent of θ for any $\epsilon > 0$. For some ϵ small enough, such that $\mathcal{O}_\epsilon(\theta) \subset \mathcal{O}_{\varepsilon_0}(\theta_0)$ (which exists due to the fact that $\varepsilon_0 > \varepsilon$) due to assumption A.13, we have that condition FOC of lemma AL.4 (in the Appendix) is satisfied by the GT estimator for $Q_n = \frac{\partial E_\theta(c_n(\widehat{\beta}_n))}{\partial \theta}$. Furthermore assumption A.13 implies conditions HUB ($\gamma(\theta) = \theta$ hence set $\delta = \varepsilon_0$) and RANK of the same lemma. Condition TIGHT follows from A.8 Lemma AL.2 of Arvanitis and Demos [5] and as $E_{\theta} c_n(b(\theta)) = 0$ the fact that

$$\begin{aligned} & \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| E_{\theta} c_n \left(\widehat{\beta}_n \right) \right\| > C_1 \frac{\ln^{1/2} n}{n^{1/2}} \right) \\ & \leq \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| \widehat{\beta}_n - b(\theta) \right\| > \frac{C_1}{\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} E_\theta(\kappa_n)} \frac{\ln^{1/2} n}{n^{1/2}} \right) \end{aligned}$$

we get that there exist $C_1 > 0$ large enough locally independent of θ for which the last term in the display is of order $o(n^{-a^*})$. Hence lemma AL.4 applies ensuring that

$$\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\sqrt{n} \|\text{GT} - \theta\| > C \ln^{1/2} n \right) = o(n^{-a^*})$$

for some $C > 0$ independent of θ . Hence condition UTIGHT of lemma AL.5 holds. Moreover assumption A.8 implies condition UEDGE of the same lemma

for $M_n(\theta) = \sqrt{n}m_n(\theta)$. Due to assumption A.13 for any $\varphi = \begin{pmatrix} \theta \\ b(\theta) \end{pmatrix}$ for any $\theta \in \overline{\mathcal{O}}_{\varepsilon_0}(\theta_0)$ and as $\widehat{\theta}_n = \widehat{\theta}_n$ is sufficiently close to φ , $\frac{\partial E_{\theta_n} c'_n(\widehat{\beta}_n)}{\partial \theta}$ admits with P_θ -probability $1 - o(n^{-a^*})$ that is independent of θ a Taylor expansion of order $s^* - 1$ around φ , with remainder

$$R_n^*(\widehat{\theta}_n, \theta) = \frac{1}{(s^* - 1)!} \left(D^{s^*-1} \left(\frac{\partial E_{\theta_n^+} c_n(\beta_n^+)' }{\partial \theta} \right) - D^{s^*-1} \left(\frac{\partial E_\theta c_n(b(\theta))' }{\partial \theta} \right) \right) ((\varphi_n - \varphi)^{s^*-1}),$$

and θ_n^+, β_n^+ lie between $\widehat{\theta}_n$ and θ and $\widehat{\beta}_n$ and $b(\theta)$ respectively (see Arvanitis and Demos [4] for details). Due to assumptions A.13, A.8, Lemma AL.2 of Arvanitis and Demos [5] and by submultiplicativity, the relation of θ_n^+ to $\widehat{\theta}_n$ and condition UTIGHT

$$\begin{aligned} & \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| \frac{1}{n^{(s^*-1)/2}} R_n^*(\widehat{\theta}_n, \theta) \right\| > \gamma_n^* \right) \\ & \leq \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\frac{2^{s^*-1} M \max^{s^*}(C, C^+)}{(s^* - 1)!} \ln^{s^*/2} n > \gamma_n^* \right) + o(n^{-a^*}) \end{aligned}$$

which is of order $o(n^{-a^*})$ for $\gamma_n^* = \frac{2^{s^*-1} M \max^{s^*}(C, C^+)}{(s^* - 1)!} \ln^{s^*/2} n = o(n^{-a^*})$ and independent of θ . Furthermore, due to the same assumption and the fact that $c(\theta, b(\theta)) = \mathbf{0}$ we have that with P_θ -probability $1 - o(n^{-a^*})$ locally independent of θ the $s^* - 1$ Taylor expansion of $\sqrt{n}E_{\theta_n} c_n(\widehat{\beta}_n)$ around φ has remainder

$$R_n^\#(\widehat{\theta}_n, \theta) = \frac{1}{s^*!} \frac{1}{n^{(s^*-1)/2}} (D^{s^*} E_{\theta_n^+} c_n(\beta_n^+) - D^{s^*} E_\theta c_n(b(\theta))) \left((\sqrt{n}(\varphi_n - \varphi))^{s^*} \right),$$

and θ_n^+ lies between $\widehat{\theta}_n$ and θ . Hence analogously to the previous

$$\begin{aligned} & \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\left\| R_n^\#(\widehat{\theta}_n, \theta) \right\| > \gamma_n^\# \right) \\ & \leq \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\frac{2^{s^*} M \max^{s^*+1}(C, C^+)}{s^*!} \ln^{(s^*+1)/2} n > \gamma_n^\# \right) + o(n^{-a^*}) \end{aligned}$$

which is of order $o(n^{-a^*})$ for $\gamma_n^\# = \frac{2^{s^*} M \max^{s^*+1}(C, C^+)}{s^*!} \ln^{(s^*+1)/2} n = o(n^{-a^*})$ and independent of θ (see Arvanitis and Demos [4] for details). Then due to assumption A.13 and the fact that $E_\theta c_n(\beta) = c(\theta, \beta)$, $\frac{\partial E_\theta c'_n(b(\theta))}{\partial \theta}$, $\frac{\partial E_\theta c'_n(b(\theta))}{\partial \beta}$ are of full rank for any $\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$. Finally due to lemma AL.6 which applies by assumptions A.8 and A.7 condition EXPAND holds and the result follows by the same lemma. ■

Proof of Lemma 3.1. Let Q_n denote the measure with density $\left(1 + \sum_{i=1}^s \frac{\pi_i(z)}{n^{\frac{i}{2}}}\right) \varphi_V(z)$. Since $2a+m+1 > 2a+1$, we have that $\sup_{A \in \mathcal{B}_C} |P_n(A) - Q_n(A)| = O(n^{-a-\eta})$, where $\eta > 0$. Hence

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

Due to the hypothesis for the support of P_n

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

Hence

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

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

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

and $n^{a+\frac{m}{2}} \rho^m q^m P(\|\zeta_n\| > c(\ln n)^\epsilon) = o(1)$ if $\epsilon \geq \frac{1}{2}$ and $c \geq \sqrt{2a+m+1}$ by Lemma 2 of Magdalinos [34]. Finally $n^a \int_{\mathbb{R}^q \setminus \mathcal{O}_{c(\ln n)^\epsilon}(0)} |K(x^m)| |dQ_n| = o(1)$ due to Gradshteyn and Ryzhik [28] formula 8.357. For the uniform case first notice that

$$\sup_{\theta \in \mathcal{O}_\epsilon(\theta_0)} P_\theta \left(\|\zeta_n\| > M \ln^{1/2} n \right) = o(n^{-a^*})$$

This is due to the fact that the set $\{x \in \mathbb{R}^q : \|x\| \leq M \ln^{1/2} n\}$ has boundary of Lebesgue measure zero and

$$\begin{aligned} & \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} \int_{\|x\| > M \ln^{1/2} n} \left(1 + \sum_{i=1}^{s^*} \frac{1}{n^{\frac{i}{2}}} |\pi_i(x, \theta)| \right) \varphi_{V(\theta)}(x) dx \\ & \leq \int_{\|z\| > \frac{M}{\lambda_{\max}(\theta^*)} \ln^{1/2} n} \left(1 + \sum_{i=1}^{s^*} \frac{1}{n^{\frac{i}{2}}} |\pi_i(V^{1/2}(\theta_i^*)z, \theta_i^*)| \right) \varphi(z) dz \end{aligned}$$

where $\lambda_{\max}(\theta)$ denotes the maximum absolute eigenvalue of $V^{1/2}(\theta)$ and $\theta_i^* \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$ exist for all $i = 1, \dots, s^*$ due to the continuity and are independent of z due to the positivity and the fact that π_i are polynomials in x , and θ^* exists due to continuity of V and the compactness of $\overline{\mathcal{O}}_\varepsilon(\theta_0)$. For $M \geq s^* \lambda_{\max}(\theta^*)$ the result follows from Lemma 2 of Magdalinos [34]. The rest follows in the same spirit of the first part. ■

Proof of Lemma 3.2. Our assumptions and lemmas 2.2, 3.1 ensure the validity of the mean approximation. Then from Theorem 3.1 of Arvanitis and Demos [5] we have that the relevant moment approximation can be obtained if we employ the appropriate approximations of $\sqrt{n}(\widehat{\beta}_n - b(\widehat{\theta}_n))$, $W_n^*(\theta_n^*)$, and $\frac{\partial b'(\widehat{\theta}_n)}{\partial \theta}$ in terms of $\sqrt{n}(\widehat{\theta}_n - \theta)$ and $\sqrt{n}(\theta_n^* - \theta)$. Therefore an appropriate approximation for $\sqrt{n}(\widehat{\theta}_n - \theta)$ is obtained by inverting the product of these terms. Finally, integrating with respect to $\left(1 + \frac{\pi_1(z, \theta)}{\sqrt{n}}\right) \varphi_{V^*(\theta)}(z)$, noting that $k_{1_\beta}(z, \theta) = z$, $k_{2_\beta}(z, \theta) = z\pi_1(z, \theta)$ we obtain result (see Arvanitis and Demos [4] for details). ■

Proof of Lemma 3.3. The assumptions and lemmas 2.2, 3.1 ensure the validity of the mean approximation uniformly over $\overline{\mathcal{O}}_{\varepsilon_*}(\theta_0)$. Furthermore from lemma AL.7 we have that $\sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon_*}(\theta_0)} \left\| DE_{\theta} \widehat{\beta}_n - Db(\theta) - \frac{1}{n} D\mathcal{I}_{\varphi_{V^*}}(k_{2_\beta})(\theta) \right\| = o(1)$ (recall that $\mathcal{I}_{\varphi_{V^*}}(k_{1_\beta}) = 0$). Then the proof proceeds as in the proof of lemma 3.2 (see Arvanitis and Demos [4] for details). ■

Proof of Lemma 3.4. First, notice that our assumptions and lemmas 2.2, 3.1 ensure the validity of the mean approximation. Then employing the implicit function Theorem, as $c(\theta, b(\theta)) = 0_l$ and $\frac{\partial c(\theta, b(\theta))}{\partial \theta'} = 0_{l \times p}$, appropriate order Taylor expansions of $\sqrt{n}c(\widehat{\theta}_n, \widehat{\beta}_n)$, $W_n^{**}(\theta_n^*)$ and $\frac{\partial c'(\widehat{\theta}_n, \widehat{\beta}_n)}{\partial \theta}$, and inverting the product of these terms we get an appropriate approximation of $\sqrt{n}(\widehat{\theta}_n - \theta)$. Integrating this approximation w.r.t. $\left(1 + \frac{\pi_1(z, \theta)}{\sqrt{n}}\right) \varphi_{V(\theta)}(z)$ and employing Theorem 3.1 of Arvanitis and Demos [5] we get the result (see Arvanitis and Demos [4] for details). ■

Proof of Corollary 3. When W^{**} is independent of x and θ and $b(\theta)$ is affine then $\mathcal{J}=0$ and $\mathcal{H}_j=\frac{\partial b'}{\partial \theta} \frac{\partial^2 c_j(\theta,b)}{\partial \beta \partial \beta'} \frac{\partial b}{\partial \theta'}$. Substituting these expressions in the expansion of $\sqrt{n}(\hat{\theta}_n - \theta)$ and integrating w.r.t. $\left(1 + \frac{\pi_1(z,\theta)}{\sqrt{n}}\right) \varphi_{V(\theta)}(z)$ we get the result. On the other hand, when $p = q = l$ and $b(\theta)$ is affine then $\mathcal{D}^{-1}\mathcal{E} = \left(\frac{\partial c(\theta,b)}{\partial \beta'} \frac{\partial b}{\partial \theta'}\right)^{-1}$, $\mathcal{J}^* = 0$, $q_{1_\beta} = \left(\frac{\partial b}{\partial \theta'}\right)^{-1} k_{1_\beta}$, $q_{1_\beta} = \left(\frac{\partial b}{\partial \theta'}\right)^{-1} k_{1_\beta}$, and $\mathcal{H}_j=\frac{\partial b'}{\partial \theta} \frac{\partial^2 c_j(\theta,b)}{\partial \beta \partial \beta'} \frac{\partial b}{\partial \theta'}$. The result follows by integrating, again, w.r.t. $\left(1 + \frac{\pi_1(z,\theta)}{\sqrt{n}}\right) \varphi_{V(\theta)}(z)$ (see Arvanitis and Demos [4] for details). ■

Proof of Lemma 3.5. i). The result follows from lemmas 2.6.A and 3.2. Notice that as $p = q = l$, we have that $\mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^* = \left(\frac{\partial b}{\partial \theta'}\right)^{-1}$. ii). The result follows from lemmas 2.6.B and 3.3. ■

Proof of Lemma 3.6. For both estimators we have that due to lemma 3.1, Theorem 3.1 of Arvanitis and Demos [5] along with the approximations employed in lemmas 3.2, 3.3 we have $E_\theta \left(n(\hat{\theta}_n - \theta) (\hat{\theta}_n - \theta)' \right) =$

$\int_{\mathbb{R}^q} \mathcal{C}^{-1} \frac{\partial b'}{\partial \theta} W^* \left(k_{1_\beta}(z, \theta) + \frac{k_{2_\beta}(z, \theta)}{\sqrt{n}} \right) \left(k_{1_\beta}(z, \theta) + \frac{k_{2_\beta}(z, \theta)}{\sqrt{n}} \right)' W^* \frac{\partial b}{\partial \theta'} \mathcal{C}^{-1} \varphi_{V(\theta)}(z) dz$, where $k_{1_\beta}(z, \theta) = z$, $k_{2_\beta}(z, \theta) = z\pi_1(z, \theta)$. Keeping the relevant order terms, the result follows. ■

Proof of Lemma 3.7. Again we have that due to lemma 3.1, Theorem 3.1 of Arvanitis and Demos [5] along with the approximations used in lemma 3 when W^{**} is independent of x and θ and $b(\theta)$ is affine, we get the appropriate expression from the proof of lemma 3. Integrating w.r.t. $\left(1 + \frac{\pi_1(z,\theta)}{\sqrt{n}}\right) \varphi_{V(\theta)}(z)$ we get the result (see Arvanitis and Demos [4] for details). ■

Proof of Lemma 4.1. First notice that in any step of the procedure the binding function is the identity. Next the $o(n^{-a^*})$ uniform consistency of $\hat{\theta}_n^{(0)}$ ensures the analogous for any step of the recursion. Then validity of the Edgeworth expansion for $\sqrt{n}\bar{s}_n(\theta)$ along with lemma 3.1 and the discussion in the paragraph immediately after lemma 3.1 imply that $\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} E \left\| n\bar{s}_n(\theta) \bar{s}_n'(\theta) + E\bar{H}_n(\theta) \right\|^2 = O(1)$ and since by the same lemma $\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} E \left\| \left(\hat{\theta}_n^{(0)} - \theta \right) \right\|^2 = O\left(\frac{1}{n}\right)$ and $E_\theta \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} \|n\bar{s}_n\|^2 < +\infty$ and $E_\theta \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} \|n\bar{H}_n\| < +\infty$ we have that $\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} \left\| D^2 E_\theta \hat{\theta}_n^{(1)} \right\| < M$. Hence lemma 2.5 applies and accordingly $\hat{\theta}_n^{(1)}$ admits a locally uniform Edgeworth expansion of order s^* . Given this the exact same reasoning implies the same result for $\hat{\theta}_n^{(k)}$ for any k . Moreover assumption A.11 follows for the expansions in every step of the procedure due to the previous. The proof for the moment approximations for the case $k = 1$ follows easily.

Using induction, let us assume that the result holds for some $k = h$, i.e. assume that the appropriate expression for $\sqrt{n} \left(\widehat{\theta}_n^{(h)} - \theta \right)$ is given by:

$$E_\theta \sqrt{n} \left(\widehat{\theta}_n^{(h)} - \theta \right) = \frac{1}{n^{\frac{2h+1}{2}}} \mathcal{I}_V(k_{2h+2}) + \frac{1}{n^{\frac{2h+2}{2}}} \mathcal{I}_V(k_{2h+3}) + o\left(n^{-\frac{2h+2}{2}}\right).$$

uniformly over $\overline{\mathcal{O}}_\varepsilon(\theta_0)$. Hence for $\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$, by lemma 2.4 it follows that $\sqrt{n} \left(E_{\widehat{\theta}_n^{(h+1)}} \widehat{\theta}_n^{(h)} - E_\theta \widehat{\theta}_n^{(h)} \right) - \left(\text{Id}_p + \frac{1}{n^{\frac{2h+2}{2}}} \frac{\partial \mathcal{I}_V(k'_{2h+2})}{\partial \theta} \right) \sqrt{n} \left(\widehat{\theta}_n^{(h+1)} - \theta \right)$ is bounded by a real sequence of order $o\left(n^{-\frac{2h+3}{2}}\right)$ that is independent of θ , with P_θ -probability $1 - o\left(n^{-\frac{2h+3}{2}}\right)$ independent of θ . The $h+1^{\text{st}}$ -step GMR2 estimator satisfies with P_θ -probability $1 - o\left(n^{-\frac{2h+3}{2}}\right)$ independent of θ , $\widehat{\theta}_n^{(h)} = E_{\widehat{\theta}_n^{(h+1)}} \widehat{\theta}_n^{(h)}$. Hence lemma 3.1 and Theorem 3.1 of Arvanitis and Demos [5] imply that the required approximation would be given by the integration of the Edgeworth density in the h^{th} step of the following approximation

$$\sqrt{n} \left(\widehat{\theta}_n^{(h)} - \theta \right) - \left(\frac{1}{n^{\frac{2h+1}{2}}} \mathcal{I}_V(k_{2h+2}) + \frac{1}{n^{\frac{2h+2}{2}}} \mathcal{I}_V(k_{2h+3}) \right).$$

This integration gives an error of $o\left(n^{-\frac{2h+2}{2}}\right)$ as $\int_{\mathbb{R}^p} \left(1 + \sum_{i=1}^{2h+2} \frac{\pi_i(z, \theta)}{n^{i/2}} \right) \varphi_{V(\theta)}(z) dz = 1 + o\left(n^{-\frac{2h+2}{2}}\right)$ due to the validity of the Edgeworth approximation of the distribution of $\sqrt{n} \left(\widehat{\theta}_n^{(h)} - \theta \right)$ and the result follows. For the MSE approximation the result follows analogously, by simply noticing that $\left(\text{Id}_p + \frac{1}{n^{\frac{2h+2}{2}}} \frac{\partial \mathcal{I}_V(k'_{2h+2})}{\partial \theta} \right)^{-1} = \text{Id}_p + o(1)$. ■

Proof of Proposition 4. Let $s^* = 5$, y_0^2 is observed and for $t = 1, \dots, n$,

$$x_t = (y_t^2, y_t y_{t-1}) \text{ and } c(x_t, \beta) = \begin{pmatrix} y_t y_{t-1} - \beta_1 \\ y_{t-1}^2 - \beta_2 \end{pmatrix}.$$

We have that $c_n(\beta) = \begin{pmatrix} \frac{1}{n} \sum_i y_i y_{i-1} - \beta_1 \\ \frac{1}{n} \sum_i y_{i-1}^2 - \beta_2 \end{pmatrix}$ and $W^* = \text{Id}_2$. Notice that

$$b(\theta) = E_\theta \begin{pmatrix} y_t y_{t-1} \\ y_{t-1}^2 \end{pmatrix} = \begin{pmatrix} \theta \\ 1 + \theta^2 \end{pmatrix}, B = b(\Theta) \text{ compact and assumption A.9.}$$

Now as $E(\varepsilon_i^{14}) < \infty$, the continuity of $D(0, 1)$ and the 1-Dependence of the MA(1) model all assumptions of section 2 in Arvanitis and Demos [5] are fulfilled (see also Götze and Hipp [23]). Consequently $S_n = \sqrt{n} \begin{pmatrix} \frac{1}{n} \sum_i y_i y_{i-1} - \theta \\ \frac{1}{n} \sum_i y_{i-1}^2 - 1 - \theta^2 \end{pmatrix}$ has a valid 5^{th} order Edgeworth expansion uniformly over Θ . Furthermore, as

$\sup_{\theta} E_{\theta}(y_t^2)^7 < \infty$ and as $E_{\theta}(y_t^m)$, for any $m = 1, \dots, 14$, are continuous function w.r.t. θ , we have that the polynomials of the Edgeworth density are equicontinuous functions of these moments, on Θ (i.e. Assumption A.4.EEQ in Arvanitis and Demos [5] is satisfied). For the auxiliary estimator, $\widehat{\beta}_n = \frac{\sum_{t=1}^n y_t y_{t-1}}{\sum_{t=1}^n y_{t-1}^2}$, consider the function $f(x, y) = \frac{x}{y}$. All derivatives, up to 5th order, of f evaluated at $x = \theta$ and $y = 1 + \theta^2$ are of the form either $\frac{\theta}{(1+\theta^2)^k}$ or $\frac{1}{(1+\theta^2)^k}$. Hence, due to the continuity of $D^5 f$ on Θ , from Theorem 3.1 in Arvanitis and Demos [5] we get that $\sqrt{n} \left(\widehat{\beta}_n - \frac{\theta}{1+\theta^2} \right)$ has a valid Edgeworth expansion uniformly on Θ with polynomials which are equicontinuous function on Θ , implying also assumption A.6. Notice that assumption A.11 also follows from analogous differentiability of moments of powers of products of elements of X_0 implied by dominated convergence the $E(\varepsilon_i^{14}) < \infty$ and the compactness of Θ , and furthermore from the differentiability of $f(x, y)$ and dominated convergence.

This implies that GMR1, and GMR2 are uniformly consistent with rate $o(n^{-2})$. Assumption A.9 applies for the function $b(\theta) = \frac{\theta}{1+\theta^2}$ and therefore due to lemma 2.2 $\sqrt{n}(\text{GMR1} - \theta)$ has a 5th order Edgeworth expansion, uniformly over Θ . This implies also that the $1 - \text{GMR2}$ is uniformly consistent with rate $o(n^{-2})$. An invocation of dominated convergence along with the same moment condition implies that $\sqrt{n} \bar{l}_n(\theta)$ (where $\bar{l}_n(\theta)$ is the average log-likelihood) has an Edgeworth expansion of third order uniformly over Θ and therefore $DE_{\theta} \widehat{\beta}_n$ is non degenerate for all θ for large enough n . Hence due to lemma 2.5.i) GMR2 has a 5th order Edgeworth expansion, uniformly over Θ which is also true for $1 - \text{GMR2}$ due to lemma 4.1. Notice that as $c_n(\beta) = \rho_n - \beta$, then $E_{\theta}(c_n(\widehat{\beta}_n)) = E_{\theta} \rho_n - \widehat{\beta}_n$, i.e. lemma 2.6.B holds. ■

Proof of Proposition 5. (for analytic proof see Arvanitis and Demos [4]) For any $\theta \in \Theta$ let $X_j(\theta) = (y_j^2 \ y_j^4 \ y_j^2 y_{j-1}^2 \ y_j^2 y_{j-2}^2)'$, and $S_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i(\theta) - EX_0(\theta))$. The results for $\widehat{\beta}_n$ and GMR1 follow from Propositions 1 and 2 in Arvanitis and Demos [5]. The results for GMR2 and $1 - \text{GMR2}$ follow from Proposition 3 in Arvanitis and Demos [5], as well. Finally, notice that by the definition of $c_n(\beta)$ we have that $c_n(\beta) = E_{\theta} \widehat{\beta}_n - \widehat{\beta}_n$, i.e. lemma 2.6.B holds. ■

General Results

The following results essentially justify the methodologies used for the establishment of the validity of the locally uniform Edgeworth expansions in all the previous cases. They are cited without proofs. Their analytical proofs can be found in Arvanitis and Demos [4].

Lemma AL.1 Suppose that:

-UUC

$$\sup_{\theta \in \Theta} P_{\theta} \left(\sup_{\beta \in B} |c_n(\beta) - c(\theta, \beta)| > \varepsilon \right) = o(n^{-a}), \quad \forall \varepsilon > 0$$

-AB $c(\theta, \beta)$ is jointly continuous and $\gamma(\theta) = \arg \min_{\beta \in B} c(\theta, \beta)$, then

$$\sup_{\theta \in \Theta} P_{\theta} \left(\left\| \widehat{\beta}_n - \gamma(\theta) \right\| > \varepsilon \right) = o(n^{-a}), \quad \forall \varepsilon > 0$$

where $\widehat{\beta}_n \in \arg \min_{\beta \in B} c_n(\beta)$.

Lemma AL.2 Let assumptions A.7.a) and A.6 hold. Then for $j = *, **$

$$\sup_{\theta \in \Theta} P_{\theta} \left(\left\| W_n^j(\theta_n^*) - E_{\theta} W^j(\theta) \right\| > \varepsilon \right) = o(n^{-a^*}), \quad \forall \varepsilon > 0$$

Furthermore, there exists $K > 0$ for which

$$\sup_{\theta \in \Theta} P_{\theta} \left(\left\| W_n^j(\theta_n^*) \right\| > K \right) = o(n^{-a^*})$$

Lemma AL.3 Suppose that

$$c_n(\beta) = \sqrt{q'_n(\beta) W_n^j(\theta_n^*) q_n(\beta)}$$

for some appropriate random element q_n where W_n^j, θ_n^* satisfy assumptions A.7.a), A.6 and for q an appropriate jointly continuous function on $\Theta \times B$

$$\sup_{\theta \in \Theta} P_{\theta} \left(\sup_{\beta \in B} \|q_n(\beta) - q(\theta, \beta)\| > \varepsilon \right) = o(n^{-a^*}), \quad \forall \varepsilon > 0$$

Then AL.1.UUC holds for $c(\theta, \beta) = \sqrt{q'(\theta, \beta) E_{\theta} W^j(\theta) q(\theta, \beta)}$ which is jointly continuous.

Lemma AL.4 Suppose $W_n^j, \widetilde{\theta}_n^*$ satisfy assumptions A.7, A.6, $\widehat{\beta}_n$ and $\gamma(\theta)$ are as in lemma AL.1, γ is continuous on $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$ and that:

-FOC $\widehat{\beta}_n$ satisfies

$$\frac{\partial q'_n(\widehat{\beta}_n)}{\partial \beta} W_n^j(\widetilde{\theta}_n^*) q_n(\widehat{\beta}_n) = \mathbf{0}$$

with P_{θ} -probability $1 - o(n^{-a^*})$ that is independent of θ ,

-HUB for some $\delta, M > 0$ independent of θ such that $\gamma(\overline{\mathcal{O}}_{\varepsilon}(\theta_0)) \subset \overline{\mathcal{O}}_{\delta}(\gamma(\theta_0))$

and for all i , $\sup_{\theta \in \overline{\mathcal{O}}_{\varepsilon}(\theta_0)} P_{\theta} \left(\sup_{\beta \in \overline{\mathcal{O}}_{\delta}(\gamma(\theta_0))} \left\| \frac{\partial^2 q'_n(\widehat{\beta}_n)}{\partial \beta \partial \beta_i} \right\| > M \right) = o(n^{-a^*}),$

-RANK for any $\beta \in \overline{\mathcal{O}}_\delta(\gamma(\theta_0))$, $\frac{\partial q_n(\beta)}{\partial \beta'}$ is of full rank with P_θ -probability $1 - o(n^{-a^*})$ that is independent of θ and,

-TIGHT for some $C > 0$ independent of θ , $\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\sqrt{n} \|q_n(\gamma(\theta))\| > C \ln^{1/2} n \right) = o(n^{-a^*})$, then

$$\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\sqrt{n} \left\| \widehat{\beta}_n - \gamma(\theta) \right\| > C^+ \ln^{1/2} n \right) = o(n^{-a^*})$$

for some $C^+ > 0$ independent of θ .

Lemma AL.5 Suppose that:

-FOC $\widehat{\beta}_n$ satisfies

$$Q_n(\widehat{\beta}_n) W_n^j(\widehat{\theta}_n^*) q_n(\widehat{\beta}_n) = \mathbf{0}$$

with P_θ -probability $1 - o(n^{-a^*})$ that is independent of θ ,

-UTIGHT There exists a $C^+ > 0$ independent of θ for which

$$\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta \left(\sqrt{n} \left\| \widehat{\beta}_n - \gamma(\theta) \right\| > C^+ \ln^{1/2} n \right) = o(n^{-a^*})$$

-UEDGE There exists a random element $M_n(\theta)$ with values in an Euclidean space, containing the elements of $\sqrt{n}(\widehat{\theta}_n^* - \theta)$, the distribution of which admits a uniform over $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ Edgeworth expansion $\Psi_{n,s}(\theta)$. The i^{th} polynomial, say, $\pi_i(z, \theta)$ of $\Psi_{n,s}(\theta)$ is equicontinuous on $\overline{\mathcal{O}}_\varepsilon(\theta_0) \forall z \in \mathbb{R}^q$, for $i = 1, \dots, s-2$, and if $\Sigma(\theta)$ denotes the variance matrix in the density of $\Psi_{n,s}(\theta)$ then it is continuous on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ and positive definite.

-EXPAND The following hold with P_θ -probability $1 - o(n^{-a^*})$ that is independent of θ

$$Q_n(\widehat{\beta}_n) = \sum_{i=0}^{s^*-1} \frac{1}{n^{i/2}} \sum_{j=0}^i C_{i,j,n}^*(\theta) \left(M_n(\theta)^j, S_n(\theta)^{i-j} \right) + R_n^*(\widehat{\beta}_n, \theta)$$

$$W_n^j(\widehat{\theta}_n^*) = \sum_{i=0}^{s^*-1} \frac{1}{n^{i/2}} C_{i,n}^{**}(\theta) \left(M_n(\theta)^i \right) + R_n^{**}(\widehat{\theta}_n^*, \theta)$$

$$\sqrt{n} q_n(\widehat{\beta}_n) = \sum_{i=0}^{s^*-1} \frac{1}{n^{i/2}} \sum_{j=0}^{i+1} C_{i,j,n}^\#(\theta) \left(M_n(\theta)^j, S_n(\theta)^{i+1-j} \right) + R_n^\#(\widehat{\beta}_n, \theta)$$

where $S_n(\theta) = \sqrt{n}(\widehat{\beta}_n - \gamma(\theta))$, $C_{i,j,n}^* : \overline{\mathcal{O}}_\varepsilon(\theta_0) \times \mathbb{R}^q \rightarrow \mathbb{R}^p$, $C_{i,n}^{**} : \overline{\mathcal{O}}_\varepsilon(\theta_0) \times \mathbb{R}^q \rightarrow \mathbb{R}^p$ are i -linear, $C_{i,j,n} : \Theta \times \mathbb{R}^{q^{i+1}} \rightarrow \mathbb{R}^p$ is $(i+1)$ -linear $\forall \theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$,

$C_{0,0,n}^*, C_{0,n}^{**}, C_{0,0,n}^\#(\theta), C_{0,1,n}^\#(\theta)$ are independent of n and have full rank $\forall \theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$, $C_{i,n}^*, C_{i,n}^{**}, C_{i,j,n}^\#$ are equicontinuous on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$, and

$$\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta(\|R_n^l\| > \gamma_n^l) = o(n^{-a^*}), \quad l = *, **, \#$$

for real sequence $\gamma_n^l = o(n^{-a^*})$ independent of θ , for $l = *, **, \#$.

Then $\sqrt{n}(\widehat{\beta}_n - \gamma(\theta))$ admits a locally uniform Edgeworth expansion, $\Psi_{n,s}^*(\theta)$, over $\overline{\mathcal{O}}_\varepsilon(\theta_0)$. The i^{th} polynomial, say, $\pi_i^*(z, \theta)$ of the density of $\Psi_{n,s}^*(\theta)$ is equicontinuous on $\overline{\mathcal{O}}_\varepsilon(\theta_0) \forall z \in \mathbb{R}^q$, for $i = 1, \dots, s-2$, and if $\Sigma^*(\theta)$ denotes the variance matrix in the density of $\Psi_{n,s}^*(\theta)$ then it is continuous on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ and positive definite.

Lemma AL.6 Under assumptions A.7 and A.8 condition EXPAND hold for $W_n^j(\tilde{\theta}_n^*)$ where $M_n(\theta) = \sqrt{n}m_n(\theta)$.

Lemma AL.7 For real valued functions f_n, f defined on $\Theta' \supseteq \Theta$, suppose that: $\sup_{\theta \in \Theta} |f_n - f| = o(1)$, and $\sup_{\theta \in \Theta} \|D^2 f_n\|, \sup_{\theta \in \Theta} \|D^2 f\| < M$. Then $\sup_{\theta \in \Theta} \|Df_n - Df\| = o(1)$.

Figures

The MA (1) Model, $\theta = 0.4$.

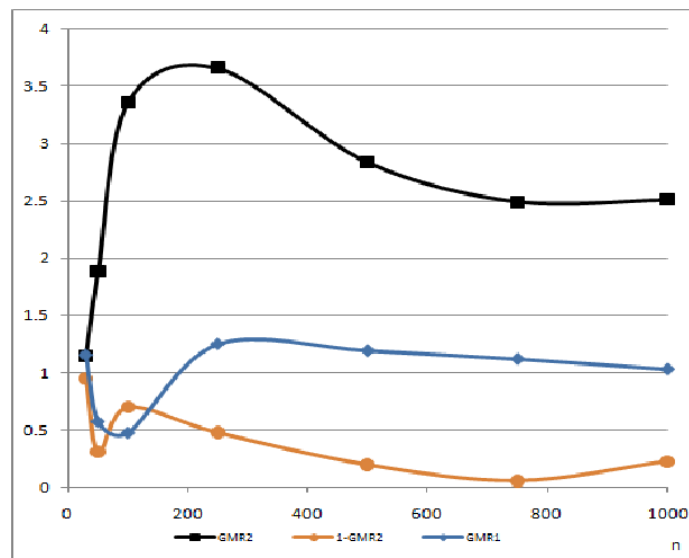


Figure 1: $n \times |\widehat{\text{Bias}}|$.

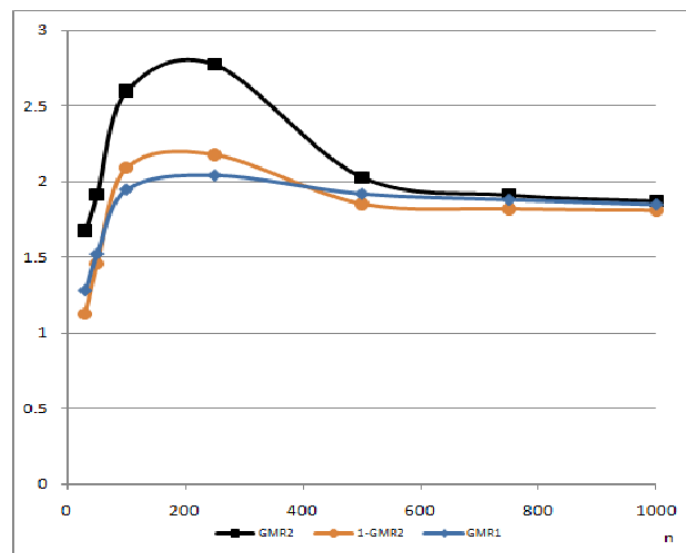


Figure 2: $n \times \widehat{\text{MSE}}$.

The GARCH (1, 1) Case, $(\theta_1, \theta_2, \theta_3) = (1.0, 0.05, 0.7)$.

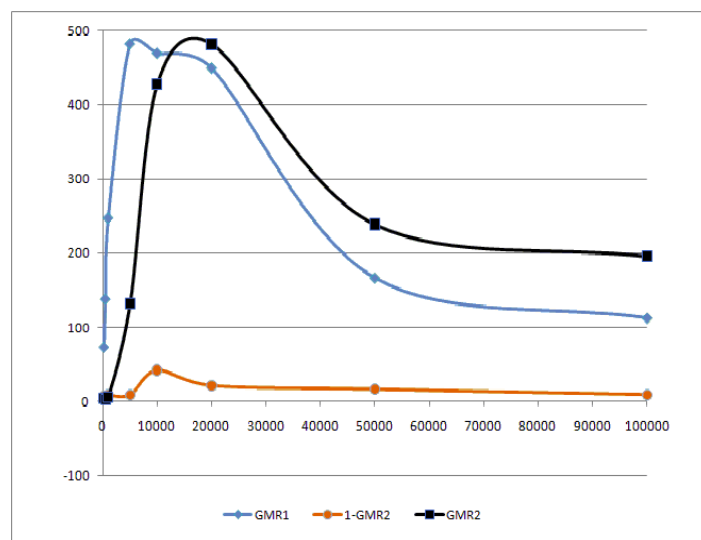


Figure 3: $n \times \|\widehat{\text{Bias}}\|$.

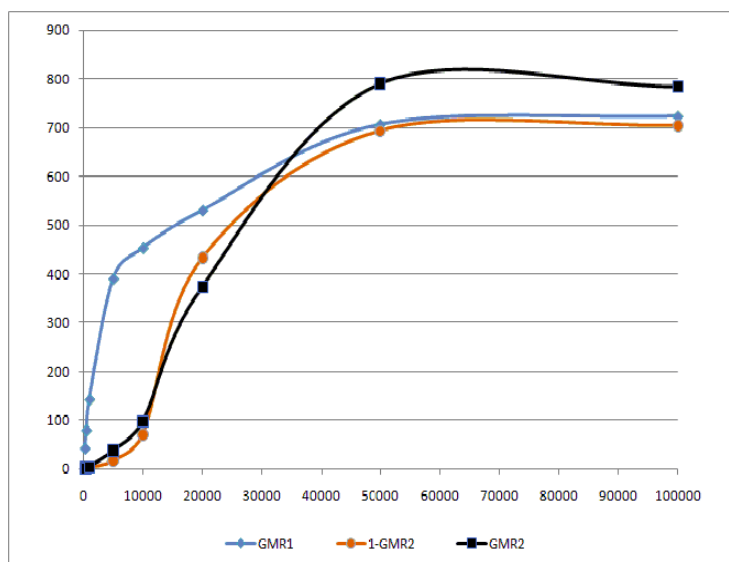


Figure 4: $n \times \|\widehat{\text{MSE}}\|$.