**ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS**

**DEPARTMENT OF ECONOMICS**

**WORKING PAPER SERIES** **03-2011**

# Stochastic Expansions and Moment Approximations for Three Indirect Estimator

Stelios Arvanitis and Antonis Demos

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

# Stochastic Expansions and Moment Approximations for Three Indirect Estimators*

Stelios Arvanitis and Antonis Demos
Athens University of Economics and Business

November 20, 2011

**Abstract**

This paper deals with properties of three indirect estimators that are known to be *(first order) asymptotically equivalent.* Specifically, we examine a) the issue of validity of Edgeworth expansions of arbitrary order. b) Given a), we are concerned with valid moment approximations and employ them to characterize the second order *bias and* MSE structure of the estimators. Our motivation resides on the fact that one of the three is reported by the relevant literature to be second order unbiased. We provide the analytical justification and generalize the conditions under which this holds and prove the *higher order non-equivalence* between the three estimators. We generalize by introducing recursive indirect estimators, that emerge from multistep optimization procedures. We provide conditions ensuring that these are higher order unbiased and retain the MSE.

KEYWORDS: Edgeworth Expansions, Higher Order Bias Approximation, Higher Order MSE Approximation, Binding Function, Recursive Indirect Estimators.

JEL: C10, C13

*We would like to thank, without implicating them, Gabriele Fiorentini, Enrique Sentana, Eric Renault, and the seminar participants at the University of Piraeus, University of Cyprus, and Athens University of Economics and Business. The comments of two anonymous referees and the associate editor have greatly improved the paper. Financial support of BRFP II AUEB is gratefully acknowledged.

# 1 Introduction

Indirect estimation is a set of inferential procedures, that employ a "misspecified" auxiliary model for estimation of the parameters corresponding to the unknown probability measure with which the underlying measure space is equipped. The motivation is largely computational, 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 $\sigma$-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 their parametric representations.

Apart from the initial articles introducing the indirect estimators under consideration (Gallant and Tauchen [15], Gourieroux, Monfort, and Renault [21], and Smith [39]), applications of these estimators have become increasingly popular. They have been applied to stochastic volatility and equity return models (e.g. Gallant et al. [13], Garcia et al. [16], Andersen et al. [1], and Sentana et al. [37]), exchange rate models (e.g. Bansal et al. [4], and Chung and Tauchen [9]), commodity price and storage models (e.g. Michaelides and Ng [31]), dynamic panel data (e.g. Gourieroux et al. [22]), stochastic differential equation models (e.g. Gallant and Long [14] and Gourieroux and Monfort [20]), and in ARMA models (e.g. Chumacero [8], Ghysels et al. [17], Demos and Kyriakopoulou [11], and Phillips [34]).

This paper is concerned with the *derivation* of higher order asymptotic properties of indirect estimators with a view towards their characterization in terms of their approximate *bias and mean squared error (*MSE*).*

## 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. $\|A\| = \sqrt{trA'A}$). $\mathcal{O}_\varepsilon(\theta)$ denotes the open $\varepsilon$-ball around $\theta$ in a relevant metric space and $\overline{\mathcal{O}}_\varepsilon(\theta)$ its closure. We denote with $\mathcal{PD}(k,\mathbb{R})$ the vector space of positive definite matrices of dimension $k \times k$ endowed with the topology of the Frobenius norm. Consider the following real function from $\mathbb{R}^k \times \mathcal{PD}(k,\mathbb{R})$ for $k \in \mathbb{N}$

$$\|x\|_A \to \left(x' Ax\right)^{1/2}.$$

[1] For the discussion of the computational aspect in a semiparametric framework, see Dridi and Renault [12].

For a given matrix the previous function defines a norm on $\mathbb{R}^k$. For $s^*, s \in \mathbb{N}^*$ 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.

**Assumption A.1** *For a measurable space* $(\Omega, \mathcal{F})$*, the statistical model (SM) is a family of probability distributions on* $\mathcal{F}$*, that is absolutely continuous to a dominating measure* $\mu$ *and equipped with the topology of weak convergence, with respect to which it is compact. There exists a homeomorphism* $\mathrm{par}(\cdot)$ *onto* $\Theta \subset \mathbb{R}^p$ *for some* $p \in \mathbb{N}$*. The likelihood function is* $\mu$*-almost everywhere* $(s^* + 2)$ *times differentiable, when restricted to an open neighborhood of* $\theta_0 = \mathrm{par}(P_0) \in \mathrm{Int}(\Theta)$*, for* $P_0$ *in SM.*

All three indirect estimators, considered here, essentially involve two step estimation procedures. In the first step, an estimating equation, that can be associated with part of the structure of an auxiliary, possibly misspecified statistical model, is employed in order for the statistical information to be summarized into an estimator with values in the auxiliary parameter space termed as the *auxiliary estimator*. Under the appropriate conditions it will converge to the binding function evaluated at the true parameter.

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 indirect estimator 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 $\beta_n$ whereas $\theta_n$ is the collective notation for the indirect ones. We also employ $b(\theta)$ to denote the binding function. Given $b(\theta)$, differences between indirect estimators hinge on different approximations of the binding function.

The auxiliary estimator is defined as a minimizer of a criterion formed as the norm of a measurable function with values on a finite dimensional Euclidean space. The following assumption enables the subsequent definition.

**Assumption A.2** *For* $B$ *a compact subset of* $\mathbb{R}^q$*,* $Q_n : \Omega \times B \to \mathbb{R}$ *is jointly measurable. Moreover* $Q_n$ *is continuous on* $B$ *for* $P_{\theta_0}-$*almost every* $\omega \in \Omega$*.*

We suppress the dependence of the random elements involved on $\Omega$, for notational simplicity.

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

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

$Q_n$ could be a likelihood function, a GMM or more generally, a distance type criterion like the ones appearing in the following definitions (see also section 4).

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

Given $\beta_n$ the indirect estimators 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 $\theta_n^*$. We term this general framework as *stochastic weighting.*

**Assumption A.4** *$W_n^* : \Omega \times \Theta \to \mathbb{R}^q$ and $\theta_n^* : \Omega \to B$ are jointly measurable.*

The first indirect estimator considered here minimizes a distance function between the $\beta_n$ and $b(\theta)$. It is termed GMR1 and it was proposed by Gourieroux, Monfort and Renault [21] in order to relax the numerical burden associated with the second estimator, defined below.

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

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

The second is termed GMR2 and it minimizes the previous distance between $\beta_n$ and $E_\theta \beta_n$. First, due to assumptions A.1 and A.3 the following lemma is trivially true.

**Lemma 1.1** *Under assumptions A.1 and A.3, $\|E_\theta \beta_n\| < \infty$ on $\Theta$.*

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

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

$$\theta_n = \arg\min_{\theta \in \Theta} \|\beta_n - E_\theta \beta_n\|_{W_n^*(\theta_n^*)}$$

In order to define the third estimator, denoted by GT and proposed by Gallant and Tauchen [15], we need the following assumption.

**Assumption A.5** *Let $Q_n$ be differentiable on $B$ for $P_\theta-$almost every $\omega \in \Omega$. We denote with $c_n$ the derivative of $Q_n$ except for the case where $Q_n = \|c_n(\beta)\|_{W_n(\beta_n^*)}$, where $c_n : \Omega \times B \to \mathbb{R}^l$, $W_n : \Omega \times B \to \mathcal{PD}(l, \mathbb{R})$, and $\beta_n^* : \Omega \to B$ are jointly measurable. Moreover $c_n$ is continuous on $B$ for $P_{\theta_0}-$almost every $\omega \in \Omega$, $c_n(\beta)$ is $P_\theta-$integrable on $\Theta \times B$ and $E_\theta(c_n(\beta))$ is continuous on $\Theta \times B$. Also $W_n^{**} : \Omega \times \Theta \to \mathbb{R}^l$ is jointly measurable.*

The GT estimator minimizes the norm of the *expectation* of the auxiliary estimating vector. We denote by $E_\theta\left(c_n\left(\beta_n\right)\right)$, the quantity $E_\theta\left(c_n\left(\beta\right)\right)|_{\beta=\beta_n}$ for notational simplicity. Due to assumption A.5 we have that $\left\|E_\theta\left(c_n\left(\beta_n\right)\right)\right\| < \infty$. Consequently, the following minimization procedure can be defined.

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

$$\theta_n = \arg\min_{\theta\in\Theta} \left\|E_\theta\left(c_n\left(\beta_n\right)\right)\right\|_{W_n^{**}(\theta_n^*)}$$

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. The existence issue for any of the aforementioned estimators is resolved by assumption A.2 for $\beta_n$, A.2, A.3, A.4 for the GMR1, A.1, A.2, A.4 for the GMR2, and A.2, A.5 for the GT estimator (see also footnote 4).[2]

The computation of all three estimators relies on the analytical form of the binding function or the engaged expectations, which are usually intractable. Due to this fact, in applications, approximations of these estimators are defined, in which the unknown elements are approximated numerically. It is easily seen that the simulation counterpart of the GMR2 estimator is the one associated with the maximal numerical burden among the three.

## 1.2 Higher Order Properties and Motivation

Gourieroux, Renault and Touzi [23] show that the GMR2 estimator has zero second order bias, when *i)* $p = q$ and *ii)* the binding function is affine. Notice that *ii)* is automatically satisfied, when the auxiliary model coincides with the SM and the binding function is approximated by a consistent estimator of the auxiliary parameters. In this case the particular indirect estimator is said to perform a *bias correction* of the first step one, a result which was only derived in a formal (i.e. non topological) manner. First, we provide this establishment in more general circumstances, and secondly, we define estimating procedures based on recursions of the GMR2 estimator, that are characterized by higher order unbiaseness.

Furthermore, the question of whether the above is also true for the remaining two indirect estimators follows naturally. To our knowledge, the only

[2] When $p = q = l$ and $Q_n\left(\beta\right) = \left\|c_n\left(\beta\right)\right\|$, $c_n\left(\beta\right) = h_n - E_\beta h_n = h_n - g\left(\beta\right)$ with $h_n : \Omega \to \mathbb{R}^p$, integrable on $\Theta$ and $B$, $g\left(\beta\right)$ and $m\left(\theta\right) = E_\theta h_n$ invertible, it is easy to see that a) the GMR1 estimator is a GMM estimator and b) $g$ is linear GMR1 = GMR2. Notice that a) would be valid even if $\beta_n = r \circ g^{-1} \circ h_n$ for $r$ a bijection. Hence the GMR1 can be a GMM estimator even in cases that the auxiliary is an appropriate transformation of a GMM estimator.

result that could provide such a connection is Proposition 4.1 of Gourieroux and Monfort [20], which is **not** generally true, except for cases where the arg min operator commutes with integration. However, their result creates the impression that all three estimators coincide and therefore provide an affirmative answer to this question. By validating the expansions at any order and deriving the second order expansion for them, we show that the previous result does not apply in these cases. Hence we essentially derive their *higher order non-equivalence*. We are also able to 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 then provide additional assumptions that validate the first and second moment approximations and derive the second order ones. 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. In section 5 we present two examples and engage into Monte Carlo experiments. We conclude in section 6. We gather all proofs in an 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.[3] We initially assume analogous expansions for $\beta_n$, and require differentiability of the random elements appearing in the definitions of $\theta_n$ in a neighborhood of $\theta_0$. Given these, we prove that the indirect estimators satisfy first order conditions with uniform probability $1-o\left(n^{-a^*}\right)$. Then, a justified use of the mean value theorem implies $o\left(n^{-a^*}\right)$ asymptotic tightness of $\sqrt{n}$ transformation of the estimators. Third, a local polynomial approximation of the $\sqrt{n}$ transformation is obtained by a Taylor expansion of the first order conditions and then it is proven that the relevant remainder is bounded by an $o\left(n^{-a^*}\right)$ real sequence with probability $1-o\left(n^{-a^*}\right)$. Then, if valid, the $\sqrt{n}$ transformation and the approximation have the same Edgeworth expansion. Finally, the validity of the aforementioned approximation is established employing an argument analogous to Skovgaard [38], i.e. a theorem of invariance of validity of Edgeworth approximations with respect to sequences of smooth transformations (see also Bhattacharya and Ghosh [5] in an iid framework, when the transformation examined is independent of $n$).

[3] The Edgeworth distributions discussed in the sequel, are not *necessarily* the formal ones. They comply with the general definition of Magdalinos [29] (see equations 3.7-8).

This methodology has been applied in Andrews [2], Andrews and Lieberman [3], Hall and Horowitz [25], Götze and Hipp [18] and [19], Lieberman et al. [27], and Robinson [35].

### Assumptions Specific to the Validity of the Edgeworth Approximations

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

Let $M$ denote a universal positive constant, independent of $n$ and $\theta$, not necessarily taking the same value across and inside assumptions proofs and results. $\mathrm{pr}_{i,j}\left(x\right)$ denotes the transformation of an $r^{th}$ dimensional vector, say $x=\left(x_1,x_2,...,x_r\right)'$, to a vector containing only the elements of $x$ from the $i^{th}$ to the $j^{th}$ coordinate, i.e. $\mathrm{pr}_{i,j}\left(x\right)=\left(x_i,x_{i+1},...,x_j\right)'$, where naturally $1\leq i\leq j\leq r$. Finally whenever the assertion "local" appears in the sequel it implies "local to $\theta_0$" unless otherwise specified.[4]

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

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

*Moreover $\theta_n^*$ is uniformly consistent for $\theta$ with rate $o\left(n^{-a^*}\right)$.*

**Remark R.1** *This assumption along with the boundeness of $B$ enables the uniform convergence of $E_\theta\beta_n$ to $b\left(\theta\right)$, hence the establishment of the analogous property for the* GMR2 *estimator. Given assumption A.2, assumption A.6 can be justified by*

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

*where $Q\left(\theta,\beta\right)$ is continuous on $B$ and it is uniquely minimized on $b\left(\theta\right)$ for any $\theta$. When $Q_n=\left\|c_n\left(\beta\right)\right\|_{W_n\left(\beta_n^*\right)}$ as in assumption A.5, the latter would be satisfied if*

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

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

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

[4] Notice that due to the fact that the spaces $\Theta$ and $B$ are separable and closed, suprema of real random elements over these spaces are typically measurable (see van der Vaart and Wellner [42], example 1.7.5 p. 47 due to the theorem of measurable projections, completeness of the underlying probability space, the compactness of $\Theta$ and the continuity of $b$).

*where $c(\theta,\beta)$ is continuous on $B$ and equals zero* iff $\beta = b(\theta)$, *where $c(\theta,\beta)$ is continuous on $B$ and equals zero* iff $\beta = b(\theta)$ *for any $\theta$, $W_n$ satisfies a similar assumption to A.7 below and $\beta_n^*$ satisfies the current assumption. In this case it is obvious that $\beta_n^*$ can be defined similarly to $\beta_n$. Finally when $\kappa_n$ is in the form of an arithmetic mean the asymptotic boundness with uniform probability $1 - o\left(n^{-a^*}\right)$ 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 forming $\kappa_n$ (consider for example the Yokoyama-Doukhan inequality in the case of weak dependence-see Andrews [2],7.3.3).*

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

**Assumption A.7** *For $j = *, **$, suppose that there exists a sequence of random elements $x_n : \Omega \to \mathbb{R}^m$, such that $W_n^j(\theta) = \frac{1}{n}\sum W^j(x_i(\omega),\theta)$ for measurable $W^* : \mathbb{R}^m \times \Theta \to \mathcal{PD}(q,\mathbb{R})$, $W^{**} : \mathbb{R}^m \times \Theta \to \mathcal{PD}(l,\mathbb{R})$ integrable with respect to $P_{\theta^*}$, such that a)*

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

*$E_{\theta^*}W^j(\theta)$ is Lipschitz w.r.t. $\theta$, for any $\theta^*$ and for the analogous Lipschitz coefficient (say) $\kappa^j(\theta^*)$ we have that $\sup_{\theta^*\in\Theta}\kappa^j(\theta^*) < +\infty$. b) Moreover $W^j(x,\theta)$ is $s^*+1$-differentiable on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ and*

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

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

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

Part b) can be justified analogously.

Obviously when $W^j(x,\theta)$ is independent of $x$ and $\theta$ the above is trivially satisfied. Now, 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 $s^*+1$.

**Assumption A.8** $\sqrt{n}m_n(\theta)$ *has an Edgeworth expansion of order* $s^*$ *uniformly on* $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ *where*

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

*or*

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

*when* $W^*(x,\theta)$ *and* $W^{**}(x,\theta)$ *are independent of* $x$ *and* $\theta$.

**Remark R.2** *Lieberman et al. [3] and Andrews and Lieberman [3] provide conditions that validate this assumption in the context of linear, Gaussian, strongly dependent time series models, when* $b$ *is the identity and* $\beta_n$ *is the* MLE *or the Whittle* MLE. *In the context of weakly dependent time series models, as described in Götze and Hipp [19] one can show that when the random vector considered here is in the form of their* $S_n$, *i.e. as a scaled sum, if their condition (2) still holds when the supremum with respect to* $\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$ *is considered, their (3)-(4) hold with constants independent of* $\theta$ *and* $\inf_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} \|Var_\theta(S_n)\| > 0$, *then the above assumption follows. This can be proven if one first shows that the aforementioned conditions imply that the conditions (2.2)-(2.6) of Götze and Hipp [18] hold with constants independent of* $\theta$, *due to a generalization of lemma 2.1 of Götze and Hipp [19]. Second it can be proven that lemma 3.33 and consequently 3.3, theorem 2.8 and corollary 2.9 of Götze and Hipp [18] hold with constants and asymptotic bounds independent of* $\theta$, *due to the fact that in this case lemma 5 of Sweeting [41] provides a bound independent of* $\theta$. *Notice that the uniform version of condition (4) in Götze and Hipp [19] would follow if their conditions (2.3.i-ii) hold with constants independent of* $\theta$ *and if the infimum of the determinant in their condition (2.3.iii) is positive on a set of positive probability. If the previous hold for* $S_n$, *and* $\sqrt{n}m_n(\theta) = f_n(\theta, S_n) + R_n(\theta)$, *uniformly with probability* $1-o\left(n^{-\frac{s^*}{2}}\right)$, *with* $\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta(\|R_n(\theta)\| > \alpha_n) = o\left(n^{-\frac{s^*}{2}}\right)$, $\alpha_n = o\left(n^{-\frac{s^*}{2}}\right)$ *independent of* $\theta$ *and* $f_n(\theta, S_n)$ *satisfy conditions (3.1.I-II) in Skovgaard [38] uniformly with respect to* $\theta$, *then again the assumption follows. This can be proven by a slight generalization of lemma 4.6 in Skovgaard [38] with his* $\beta_{n,s} = O\left(n^{-\frac{s^*}{2}}\right)$.

### Existence of Edgeworth Expansions for the GMR-type Estimators

In this paragraph 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 in assumption A.8.

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

**Assumption A.9** *$b(\theta)$ is $s^*+2$ continuously differentiable and* $\operatorname{rank} Db(\theta) = p$*, for all $\theta$ in $\overline{\mathcal{O}}_\varepsilon(\theta_0)$.*

In the special case considered in remark R.1 with $q = l$, the previous assumption can be justified when $c(\theta, \beta)$ is $s^* + 3$ times differentiable on $\mathcal{O}_\eta(\varphi_0)$ where $\varphi_0 = (\theta_0', b'(\theta_0))'$, for $\eta$ large enough and rank $\frac{\partial c(\theta, b(\theta))}{\partial \beta'} = q$ for all $\theta$ in $\overline{\mathcal{O}}_\varepsilon(\theta_0)$, via the use of the implicit function theorem.

The next lemma provides the results.

**Lemma 2.1** *i) Under the assumptions A.1, A.2, A.3, A.4, A.6 and A.7a) the* GMR1 *is uniformly consistent for $\theta$ with rate $o(n^{-a})$. ii) If additionally assumptions A.7b), A.8 and A.9 hold then, $\sqrt{n}$* (GMR1 $-\theta$) *has an Edgeworth expansion of order $s^*$ uniformly on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$.*

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

**Assumption A.10** $\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} \left\| D^{s^*+2} E_\theta \beta_n \right\| < M$.

**Remark R.3** *Assumption A.9 along with Assumption A.10 imply that for $r = 1, \ldots, s^* + 2$, $\sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} \| D^r (E_\theta \beta_n - b(\theta)) \| < M$, which in turn means that $D^{r-1}(E_\theta \beta_n - b(\theta))$ 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* $\operatorname{rank} DE_\theta \beta_n = p$*, for all $\theta$ in $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ for $n$ large enough. Due to Assumption A.1, this assumption is verified via conditions which enable the use of dominated convergence, hence the commutativity of the integral and the derivative, and then conditions of the form*

$$\limsup_n \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} E_\theta \left\| \sqrt{n} \left( \beta_n - b(\theta) \right) \right\|^2 < +\infty$$

$$\limsup_n \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} E_\theta \left\| \sqrt{n} l_n(\theta) \right\|^2 < +\infty$$

*where $\bar{l}_n(\theta)$ depends on derivatives of the average likelihood function. For example when $r = 1$ we have that $\bar{l}_n(\theta) = \bar{s}_n(\theta)$, $\bar{s}_n(\theta)$ being the average score, and $r = 2$, $\bar{l}_n(\theta) = \bar{s}_n(\theta)\bar{s}'_n(\theta) + \overline{H}_n(\theta)$, with $\overline{H}_n(\theta)$ denoting the average Hessian. In the light of the results in section 3 the first condition holds when $s^* \geq 3$ in assumption A.8 and a locally uniform Edgeworth expansion of order greater than or equal to 3 holds for the random vector containing the elements of $\sqrt{n}\bar{l}_n(\theta)$.*

Again, 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.7a) the* GMR2 *is uniformly consistent for $\theta$ with rate $o(n^{-a})$. ii) If additionally assumptions A.7b), A.8, A.9 and A.10 hold then $\sqrt{n}(\text{GMR2} - \theta)$ has an Edgeworth expansion of order $s^*$ uniformly on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$.*

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 $k_{i_\beta}(z,\theta) = z\pi_{i-1}(z,\theta)$ where $\pi_{i-1}(z,\theta)$ is the polynomial in the density of the Edgeworth distribution in assumption A.8 with coefficient $\frac{1}{n^{\frac{i-1}{2}}}$, for $i = 1, \ldots, s^*$, and with $\mathcal{I}_V\left(k_{i_\beta}(z,\theta)\right) = \int_{\mathbb{R}^q} k_{i_\beta}(z,\theta)\varphi_{V(\theta)}(z)\,dz$ where $V(\theta)$ is the asymptotic variance of $\sqrt{n}m_n(\theta)$ under $P_\theta$. We make the following assumption on the behavior of $\mathcal{I}_V\left(k_{i_\beta}(z,\theta)\right)$ for $i = 1, \ldots, s$.

**Assumption A.11** *$\mathcal{I}_V\left(k_{i_\beta}(z,\theta)\right)$ is $s^* + 2$ continuously differentiable for $i = 1, \ldots, s^*$ and $V(\theta)$ is continuous on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$.*

Employing assumption A.8, dominated convergence and the properties of the density of the normal distribution it is easy to see that the continuous differentiability of the relevant order assumption on the $\mathcal{I}_V\left(k_{i_\beta}(z,\theta)\right)$ would follow from the same assumption on $\pi_{i-1}(z,\theta)$ with respect to $\theta$ in the particular neighborhood of $\theta_0$ for any $z$.

This assumption essentially enables the following polynomial approximations of $E_{\theta_n^*}\beta_n$ for special sequences $\theta_n^*$ converging to $\theta$.

**Lemma 2.3** *If assumptions A.8, A.9 and A.11 hold for $s^* > s$ then for any sequence $\theta_n^*$ for which*

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

*we have that*

$$\sup_{\theta\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)} P_\theta\left(\left\|\sqrt{n}\left(E_{\theta_n^*}\beta_n - E_\theta\beta_n\right) - A_n(\theta)\right\| > \gamma_n\right) = o\left(n^{-a^*}\right)$$

*where*

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

$\gamma_n = o\left(n^{-a}\right)$ *using the convention that when* $s - i = 0$*, then* $\sum_{j=1}^{s-i}$ *is empty.*

This lemma enables the derivation of an analogous result to the second part of lemma 2.2 in the special case where $p = q$.

**Lemma 2.4** *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* $\sup_{\theta\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)} \left\|D^2 E_\theta \beta_n\right\| < M$ *then* $\sqrt{n}\,(\mathrm{GMR2} - \theta)$ *has an Edgeworth expansion of order* $s$ *uniformly on* $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$*. ii) if* $\beta_n = b(\mathrm{GMR1})$ *with probability* $1 - o\left(n^{-a}\right)$ *uniformly on* $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$ *and* $\beta_n = E_{\mathrm{GMR2}}\beta_n$ *with probability* $1 - o\left(n^{-a}\right)$ *uniformly on* $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$ *then* $\sqrt{n}\,(\mathrm{GMR2} - \theta)$ *has an Edgeworth expansion of order* $s$ *uniformly on* $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$*.*

Notice that the Edgeworth distributions validated in all cases of lemmas 2.2 and 2.4 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 $o\left(n^{-a}\right)$.

### Existence of Edgeworth Expansion for the GT Estimator

We first consider a particular case which links the asymptotic behaviors of the GMR1 and the GT estimators.

**Lemma 2.5** *Suppose that* $p = q = l$*,* $E_{\mathrm{GT}}\left(c_n\left(\beta_n\right)\right) = \mathbf{0}_l$ *with probability* $1 - o\left(n^{-a^*}\right)$ *independent of* $\theta$ *and* $E_\theta\left(c_n(\beta)\right) = \mathbf{0}_l$ iff $\beta = b(\theta)$*. i) Under the assumptions A.1, A.2, A.3, A.4, A.6 and A.7a) the* GT *is uniformly consistent for* $\theta$ *with rate* $o\left(n^{-a}\right)$*. ii) If additionally assumptions A.7b), A.8 and A.9 hold then* $\sqrt{n}\,(\mathrm{GT} - \theta)$ *has an Edgeworth expansion of order* $s^*$ *uniformly on* $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$*. Moreover this expansion coincides with the one of lemma 2.1.*

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 $c_n$ and $\kappa_n$ be as in remark R.1 and such that* $\sup_\theta E_\theta \kappa_n < +\infty$ *and*

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

**Assumption A.13** *For $\varphi = (\theta', \beta')'$, $\varphi_0$ as before and $\eta$ large enough for $\overline{\mathcal{O}}_\eta(\varphi_0) \supseteq \overline{\mathcal{O}}_\varepsilon(\theta_0) \times \overline{\mathcal{O}}_{\varepsilon'}(b(\theta_0))$,* $\operatorname{rank}\left(\lim_{n\to\infty} \frac{\partial E_\theta c_n(b(\theta))}{\partial \theta'}\right) = p$ *on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$,* $\sup_{\varphi \in \overline{\mathcal{O}}_\eta(\varphi_0)} \left\| D^{s^*+2} E_\theta c_n(\beta) \right\| < M$.

**Lemma 2.6** *i) Under the assumptions A.1, A.2, A.3, A.4, A.6, A.7a) and A.12 the* GT *is uniformly consistent for $\theta$ with rate $o(n^{-a})$. ii) If additionally assumptions A.7b), A.8 and A.13 hold then $\sqrt{n}(\mathrm{GT} - \theta)$ has an Edgeworth expansion of order $s^*$ uniformly on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$.*

We finally note that locally uniform (around $\varphi_0$) 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 use of assumption A.13 in circumstances analogous to the ones described in lemma 2.4. We do not pursue this kind of reasoning in order to economize on space.[5]

# 3 Validity of 1st and 2nd 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 the approximation of their bias-MSE comparisons. The expansions employed to derive the moment results, concern the so-called *delta method* of approximations of moments of estimator sequences (see e.g. Linton [28] and McCullagh [30], Phillips [33], and Sargan [36]). We provide a general lemma, which establishes that if the Edgeworth expansions involved are of an appropriately large order then, the needed moment approximations are provided by the analogous moments of the Edgeworth distributions.[6]

[5] Similarly to the GMR2 case, the Edgeworth distributions validated in lemmas 2.5 and 2.6 need not coincide. Again their distance must be of order $o\left(n^{-a^*}\right)$.

[6] Notice that separate methodologies concerning moment approximations (as for example the one in Koenker et al. [26]) are not general enough to cover our framework due to the following reasons. First, the latter 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) $\zeta_n$ is bounded by $\mathcal{O}_{\sqrt{n}\rho}(0)$ for some $\rho > 0$, and $\zeta_n$ admits an Edgeworth expansion of order $s^* = 2a + m + 1$ then*

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

*where $P_n$, and $\left(1 + \sum\nolimits_{i=1}^{s} \frac{\pi_i(z)}{n^{\frac{i}{2}}}\right)\varphi_V(z)$ denote the distribution of $\zeta_n$ and the density of the analogous Edgeworth measure of order $s = 2a+1$ respectively. Moreover if $P_n$ depends on $\theta$, 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 $\overline{\mathcal{O}}_\varepsilon(\theta_0)$.*

Due to the fact that we do not derive the Edgeworth approximations described in the previous results, this lemma is not very practical for the derivation of the approximations of moments of the indirect estimators. This difficulty can be circumvented if additionally to the above we consider the following. Suppose that we can invert the Taylor expansion of the first order condition that with high probability satisfies each one of the estimators considered. *By similar arguments as in lemma 4.7 of Skovgaard [38] the distance between the integral of any multilinear function, with respect to the valid Edgeworth distribution of the estimator under consideration, and the integral of the same function evaluated at the inverted Taylor polynomial, with respect to the Edgeworth distribution described in A.8 is of the required order.*[7] We proceed on the derivation of the first and second moment approximations for the estimators considered above when $s = 2$.

In the following we suppress the dependence on $\theta$ and $z$ where possible for notational convenience. For the rest of this section we denote by $b = b(\theta)$, $b_{,j}$ 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 $W^{**}$. Moreover, $\mathcal{C} = \frac{\partial b'}{\partial \theta} W^* \frac{\partial b}{\partial \theta'}$, $k_{i_\beta}(z,\theta) = \mathrm{pr}_{1,q}(z)\,\pi_{i-1}(z,\theta)$, $k_{i_{\theta^*}}(z,\theta) = \mathrm{pr}_{q+1,p+q}(z)\,\pi_{i-1}(z,\theta)$, $k_{i_{w^*}}(z,\theta)$ is the matrix containing the elements $\mathrm{pr}_{p+q+1,q^2}(z)\,\pi_{i-1}(z,\theta)$ and $k_{i_{w^*}}(z,\theta)$ is the matrix containing the elements of $\mathrm{pr}_{q^2+1,2q^2}(z)\,\pi_{i-1}(z,\theta)$ at the appropriate orders.

## 3.1 Valid $2^{nd}$ order Bias approximation for the Indirect estimators

We are ready to provide the results for the second order bias approximation of the indirect estimators. Notice that due to their form, the results in

[7] Lemma 3.1 is analytically derived since we are not acquainted with any reference to similar results in the relevant literature. The extension of Lemma 4.7 of Skovgaard is only sketched due to the fact that its proof is similar to the one of that lemma.

Newey and Smith [32] 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 $\theta_n$ 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 for any $\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)$*

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

*where*

$$\xi_1(\theta) = \mathcal{C}^{-1}\frac{\partial b'}{\partial \theta}W^*\mathcal{I}_V\left(k_{2_\beta}\right) - \mathcal{C}^{-1}\mathcal{I}_V\left(\left[k'_{1_{\theta^*}}\frac{\partial^2 b_j}{\partial\theta\partial\theta'}\right]_{j=1,\ldots,l} W^*\frac{\partial b}{\partial \theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W^*k_{1_\beta}\right)$$
$$-\frac{1}{2}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W^*\mathcal{I}_V\left(\left[\frac{\partial b'}{\partial\theta}k'_{1_\beta}W^*\mathcal{C}^{-1}\frac{\partial b_j}{\partial\theta\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W^*k_{1_\beta}\right]_{j=1,\ldots,q}\right)$$
$$-\mathcal{C}^{-1}\mathcal{I}_V\left(\left(\frac{\partial b'}{\partial\theta}\frac{1}{\sqrt{n}}k_{1_{w^*}} + \left[\frac{\partial}{\partial\theta'}W^*_{j,j'}k_{1_{\theta^*}}\right]_{j,j'=1,\ldots,l}\right)\frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W^*k_{1_\beta}\right)$$

*If moreover assumption A.11 the previous holds uniformly over $\overline{\mathcal{O}}_\varepsilon(\theta_0)$.*

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

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

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

Hence $\xi_1$ is zero iff $\mathcal{I}_V\left(k_{2_\beta}\right)$ is.

GMR2 **Estimator** We continue with the GMR2 case. The approximation contains the term $-\Gamma\frac{\partial b'}{\partial\theta}W^*\mathcal{I}_V\left(k_{2_\beta}\right)$ something that is not present in the other two, a fact that is attributed to the existence of $E_\theta\beta_n$ in the definition of the particular estimator.

**Lemma 3.3** *Let $\theta_n$ 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 $s^* \geq 3$ then uniformly over $\overline{\mathcal{O}}_\varepsilon(\theta_0)$*

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

*where*

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

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 $\theta$ and $b(\theta)$ is affine then $\xi_2(\theta) = \mathbf{0}_p$.*

This result is already known for the case where $p = q$, $\beta_n$ is a consistent estimator of $\theta$, whence the GMR2 obviously performs a second order bias correction.[8] Hence the previous *generalizes* the results in Gourieroux and Monfort [20] and Gourieroux et al. [23].

GT **Estimator** We conclude the presentation of the expansions with the last of the three estimators. We present first a straightforward case implied by the previous results. Denoting with $\mathcal{D} = \frac{\partial b'}{\partial \theta} \frac{\partial c(\theta,b)}{\partial \beta'} W^{**}(\theta) \frac{\partial c(\theta,b)}{\partial \beta'} \frac{\partial b}{\partial \theta'}$, $\mathcal{E} = \frac{\partial b'}{\partial \theta} \frac{\partial c(\theta,b)}{\partial \beta'} W^{**}(\theta)$, $\mathcal{H}_j = \frac{\partial b'}{\partial \theta} \frac{\partial^2 c_j(\theta,b)}{\partial \beta \partial \beta'} \frac{\partial b}{\partial \theta'} - \left[\frac{\partial c_j(\theta,b)}{\partial \beta'} \frac{\partial^2 b}{\partial \theta' \partial \theta_r}\right]_{r=1,\ldots,p}$, $\mathcal{J} = k_{1_{w^{**}}} + \left[\frac{\partial}{\partial \theta'} W^{**}(\theta)_{j,j'} k_{1_{\theta^*}}\right]_{j,j'=1,\ldots,l}$, $\mathcal{J}^* = \left(\frac{\partial c(\theta,b)}{\partial \beta'} \frac{\partial b}{\partial \theta'} \mathcal{D}^{-1} \mathcal{E} - \mathrm{Id}_l\right) \frac{\partial c(\theta,b)}{\partial \beta'}$ 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** *Using 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 $\overline{\mathcal{O}}_\varepsilon(\theta_0)$*

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

[8] If in addition $E_\theta \beta_n$ is linear, then the estimator is totally unbiased (see Gourieroux et al. [23]).

*where*

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

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 $\theta$ and $b(\theta)$ is affine then*

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

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

An analogous result applies even when $E_\theta c_n(\beta)=c(\theta,\beta)$ does not hold. Under the assumptions in lemma 2.5 we have that the approximate biases of GT and GMR1 are the same.

**Lemma 3.5** *Let $\theta_n$ be either* GT *or the* GMR1 *estimator. Under the assumptions in lemma 2.5 and for $s^*\geq 3$ we have that*

$$
\xi_2(\theta)=\xi_3(\theta)=\left(\frac{\partial b}{\partial\theta'}\right)^{-1}\mathcal{I}_V\left[k_{2_\beta}-\frac{1}{2}\left\{k'_{1_\beta}\left(\frac{\partial b'}{\partial\theta}\right)^{-1}\frac{\partial^2 b_j}{\partial\theta\partial\theta'}\left(\frac{\partial b}{\partial\theta'}\right)^{-1}k_{1_\beta}\right\}_{j=1,\ldots,l}\right]
$$

*Furthermore, under assumption A.11 the previous holds uniformly over $\overline{\mathcal{O}}_\varepsilon(\theta_0)$.*

This essentially provides a counterexample concerning the hypothesis of second order asymptotic equivalence between the three estimators.

## 3.2 MSE $2^{nd}$ 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. Using these results we obtain the following lemmas.

**Lemma 3.6** *Let $\theta_n$ denote either the* GMR1*, or the* GMR2 *estimator. If $W^*(x,\theta)$ is independent of $x$ and $\theta$, $b$ is affine and assumptions A.1, A.2, A.3, A.4, A.6 and A.7, A.8, A.9 and A.11 hold for $s^* \geq 4$ then,*

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

*where*

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

This along with corollary 2 establishes the second order superiority of the GMR2 estimator in the particular case. For the GT we obtain the following.

**Lemma 3.7** *Let $\theta_n$ denote the* GT *estimator. If $W^{**}(x,\theta)$ is independent of $x$ and $\theta$, $b$ 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 $\overline{\mathcal{O}}_\varepsilon(\theta_0)$*

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

*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 [20]). 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.*

# 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 to $\theta_0$ behavior of the binding function. Due to assumption A.9, and theorem 10.2 of Spivak [40] (p. 44) as $p \leq q$ $B$ **can always be chosen** so that the binding function $b$ is of the form $\begin{pmatrix} \theta \\ \mathbf{0}_{q-p} \end{pmatrix}$ at least in a small enough neighborhood of $\theta_0$. This along with non stochastic weighting and corollary 2 imply that there always exists an auxiliary parametrization such that the GMR2 estimator is second order unbiased. Usually, the reparametrization of the auxiliary model is analytically intractable.

However there exists at least one indirect estimation procedure that can be employed in order to approximate this "canonical" parameterization. Given the GMR1, let $\beta_n' = (\text{GMR1}', 0_{q-p})$and apply the GMR2 estimator to the latter. Then the resulting indirect estimator is derived from a three-step procedure, in the last step of which the binding function is obviously $\left(\theta_1, \theta_2, \ldots, \theta_p, \underbrace{0, \ldots, 0}_{q-p}\right)'$. An extension of the three step procedure of the previous remark to an arbitrary number of steps, where the $i^{th}$-step auxiliary estimator is the the GMR2 of the previous step embedded to $\mathbb{R}^q$, can provide an unbiased indirect estimator of arbitrary order when $i$ 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 indirect estimation procedures as follows. Let $\theta_n^{(0}$ denote any estimator of $\theta$.

**Definition D.5** *Let $\zeta \in \mathbb{N}$, the recursive $\zeta -$ GMR2 estimator (denoted by $\theta_n^{(\zeta}$) is defined in the following steps:*

1. $\theta_n^{(1} = \arg\min_\theta \left\| \theta_n^{(0} - E_\theta \theta_n^{(0} \right\|$,

2. *for* $\zeta > 1$ $\theta_n^{(\zeta} = \arg\min_\theta \left\| \theta_n^{(\zeta-1} - E_\theta \theta_n^{(\zeta-1} \right\|$.

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

**Lemma 4.1** *Suppose that assumptions A.6, A.8, A.11 hold for $\theta_n^{(0}$ for $s^* \geq 2\zeta+3$. Moreover suppose that* $\limsup_n \sup_{\theta \in \overline{\mathcal{O}}_\varepsilon(\theta_0)} E_\theta \left\| \sqrt{n} \bar{l}_n(\theta) \right\|^2 < \infty$ *where*

$\bar{l}_n(\theta)$ *contains the elements of* $\overline{s}_n(\theta)$ *and* $\overline{s}_n(\theta)\,\overline{s}'_n(\theta)+\overline{H}_n(\theta)$. *Then the* $\zeta-$ GMR2 *estimator is of order* $s=2\zeta+1$ *unbiased and has the same* MSE *with the* $(\zeta-1)-$ GMR2, *up to* $2\zeta$ *order, uniformly on* $\overline{\mathcal{O}}_{\varepsilon}(\theta_0)$.

Consider again the case where $\zeta=1$. Then $1-$GMR2 is actually $3^{rd}$ order unbiased at $\theta_0$ hence the previous results are essentially expanded under the conditions of the lemma. Furthermore, the $1-$ GMR2 has the same second order MSE as the $1-$ GMR2 one.

It is worth mentioning that the recursive GMR2 procedure is a generalization of iterated bootstrap. To elaborate on this, consider the GMR1 estimator. Bootstrapping this estimator is equivalent to one $-$ step GMR2 estimation (one $-$ step in the spirit of Andrews [2]) on GMR1 (see Gourieroux et al. [23] section 1.5). Bootstrapping the bootstrapped GMR1 is equivalent to one $-$ step GMR2 on one $-$ step GMR2 on GMR1 etc. Consequently, the iterated bootstrap estimator is a recursive one $-$ step GMR2, on every recursion.[9] Let us now turn our attention to two examples.

# 5 Examples and Monte Carlo Experiments

In this section we present a set of examples concerning the further specification of our assumption framework in statistical models and indirect estimators.

## 5.1 The MA(1) Case

Let the statistical model in assumption A.1 be described by the set of stationary ergodic process defined by the recursion

$$y_t=\varepsilon_t+\theta\varepsilon_{t-1},\ \varepsilon_t\overset{i.i.d.}{\sim}D\left(0,1\right)\tag{1}$$

for some distribution $D$. For $t\in\mathbb{Z}$, $\theta\in\Theta$ a compact subset of $(-1,1)$ with $\theta_0\in\operatorname{int}\Theta$.

Consider now the auxiliary estimator $\beta_n=\frac{\sum_{t=1}^{n}y_ty_{t-1}}{\sum_{t=1}^{n}y_{t-1}^2}$, which can be also interpreted as the OLSE or the conditional QMLE of an $\mathrm{AR}(1)$ auxiliary model, over $B=b(\Theta)$ and $b(\theta)=\frac{\theta}{1+\theta^2}$. Notice that the GMR1 $=\frac{1-\sqrt{1-4\beta_n^2}}{2\beta_n}$ and coincides with the GT estimator. To evaluate the GMR2 estimator we have to evaluate the $E_\theta\beta_n$ (see definition D.3).

[9] It is easy to see that analogous recursive GMR1 and/or GT type estimators would simply coincide with $\theta_n^{(0}$.

Applying now GMR2 on the GMR1 we get the $1-$GMR2 (or equivalently $\theta_n^{(1}$ employing the notation of section 4). Again, its expected value must be evaluated.

**Proposition 4** *If $E\varepsilon_i^{14} < \infty$ and if $D(0,1)$ is a smooth continuous density, then the $\beta_n$,* GMR1*,* GT*,* GMR2 *and* $1-$ GMR2 *admit* $5^{th}$ *order valid Edgeworth expansions, uniformly over* $\Theta$*.*

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 $1-$ GMR2 is $3^{rd}$ order unbiased by lemma 4.1. Also lemmas 3.6 and 3.7 imply the second order superiority of $1-$ GMR2 w.r.t. the GMR1, the GMR2 and the GT uniformly over $\Theta$.[10]

**Monte Carlo Experiment** To evaluate the GMR2 estimator we need $E_\theta\beta_n$ (see definition D.3) which is analytically intractable. We approximate this expectation numerically, i.e.

$$E_\theta\beta_n \simeq \left( \frac{1}{H} \sum_{h=1}^{H} \frac{\sum_{i=2}^{n} y_i^{(h} y_{i-1}^{(h}}{\sum_{i=2}^{n} \left( y_{i-1}^{(h} \right)^2} \right),$$

where $H = 1600$ and $y_t^{(h}$ is given as in equation (1). The same applies for $E_\theta$ GMR1 needed for the $1-$ GMR2 estimator. In this respect we obtain approximate GMR2 and $1-$ GMR2 estimators. 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 $\beta_n$ 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 $1-$GMR2. We perform 50000 Monte Carlo replications. To approximate $E_\theta\beta_n$ and $E_\theta$ GMR1 we chose $H = 1600$.

[10] Notice that the expansions of the $\beta_n$ and GMR1 estimators are available from the work of Demos and Kyriakopoulou [11]. These formulae imply that both estimators, GMR1 and GMR2, are biased, unless $\theta = 0$.

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 [11], for this value of $\theta$, we should have that $n\left|E_{\theta}\left(\text{GMR1}\right) - \theta\right| - 1.252 = o\left(1\right)$, $n\left|E_{\theta}\left(\text{GMR2}\right) - \theta\right| - 2.094 = o\left(1\right)$ 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 MSE $s$, 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

Let, again, the statistical model in assumption A.1 be described by the set of stationary ergodic and covariance stationary processes defined by the recursion

$$\begin{aligned} y_t &= z_t\sqrt{h_t}, \quad h_t = \theta_1\left(1 - \theta_2 - \theta_3\right) + \left(\theta_2 z_{t-1}^2 + \theta_3\right)h_{t-1} \\ \theta_1, \theta_2, \theta_3 &> 0, \quad \theta_2 + \theta_3 < 1, \quad z_t \overset{i.i.d.}{\sim} N\left(0, 1\right), \end{aligned} \tag{2}$$

and $\theta = \left(\theta_1, \theta_2, \theta_3\right)' \in \Theta$ a compact subset of $\mathbb{R}^{++} \times \left[0, 1\right]^2$ such that $\sup_{\theta \in \Theta} Eh_0^{14}\left(\theta\right) < +\infty$. Employing the Pantula reparameterization, the conditional variance equation can be written as an ARMA$(1, 1)$ process in $y_t^2$

$$y_t^2 = \theta_1\left(1 - \theta_2 - \theta_3\right) + \left(\theta_2 + \theta_3\right)y_{t-1}^2 + v_t - \theta_3 v_{t-1}, \quad \text{where} \quad v_t = y_t^2 - h_t,$$

and $v_t$ is a martingale difference sequence (see Bollerslev [6]).

Taking the ARMA$(1, 1)$ representation as an auxiliary model, define the auxiliary estimator as:

$$\beta_n = \left(\overline{y^2}, \widehat{\rho}_1, \frac{\widehat{\rho}_2}{\widehat{\rho}_1}\right)'$$

where $\overline{y^2} = \frac{1}{n}\sum_{t=1}^n y_t^2$, $\widehat{\rho}_i$ is the $i^{th}$ order sample autocorrelation of the squared $y's$, i.e. $\widehat{\rho}_i = \frac{\sum_{t=i+1}^n\left(y_t^2 - \overline{y^2}\right)\left(y_{t-i}^2 - \overline{y^2}\right)}{\sum_{t=i+1}^n\left(y_t^2 - \overline{y^2}\right)^2}$. It is easily seen that $\beta_n$ converges in probability to

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

The GMR1 estimator is given by

$$\theta_n = \begin{pmatrix} \overline{y^2} \\ \frac{1-(\widehat{r})^2-\sqrt{\left(1-(2\widehat{\rho_1}-\widehat{r})^2\right)\left(1-(\widehat{r})^2\right)}}{2(\widehat{\rho_1}-\widehat{r})} \\ \frac{-\left(1-2\widehat{\rho_1}\widehat{r}+(\widehat{r})^2\right)+\sqrt{\left(1-(2\widehat{\rho_1}-\widehat{r})^2\right)\left(1-(\widehat{r})^2\right)}}{2(\widehat{\rho_1}-\widehat{r})} \end{pmatrix},$$

where $\widehat{r} = \frac{\widehat{\rho_2}}{\widehat{\rho_1}}$ and equals the GT. As in the previous example, the evaluation of the GMR2 estimator needs the evaluation of the $E_\theta \beta_n$ (see definition D.3).

Now employing the GMR2 estimator, treating the GMR1 as an auxiliary one, we get the $1-$ GMR2 estimator. Again, the $E_\theta$ (GMR1) needs to be evaluated.

**Proposition 5** *$\beta_n$, GMR1, GT, GMR2 and $1-$GMR2 admit $5^{th}$ order valid Edgeworth expansions, uniformly over $\Theta$.*

Given the results of the previous proposition and those in section 3 the $2^{nd}$ 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 $2^{nd}$ order unbiased (see results in section 3.1). However, the $1-$ GMR2 is $3^{rd}$ 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 $\Theta^*$.

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

$$E_\theta \beta_n \simeq \left( \frac{1}{H} \sum_{h=1}^{H} \overline{y^2}^{(h}, \frac{1}{H} \sum_{h=1}^{H} \widehat{\rho}_1^{(h}, \frac{1}{H} \sum_{h=1}^{H} \frac{\widehat{\rho}_2^{(h}}{\widehat{\rho}_1^{(h}} \right)',$$

where $H = 60$, $\widehat{\rho}_i^{(h} = \frac{\sum_{t=i+1}^{n} \left( \left(y_t^{(h}\right)^2 - \overline{y^2} \right) \left( \left(y_{t-i}^{(h}\right)^2 - \overline{y^2}^{(h} \right)}{\sum_{t=i+1}^{n} \left( \left(y_t^{(h}\right)^2 - \overline{y^2}^{(h} \right)^2}$, $\overline{y^2}^{(h} = \frac{1}{n} \sum_{t=1}^{n} \left( y_t^{(h} \right)^2$ and $y_t^{(h}$ is as in equation (2), and similarly for the $E_\theta$ (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 $\beta_n$. We generate, with $n \in \{250, 500, 1000, 5000, 10000, 20000, 50000, 100000\}$, the $GARCH\,(1,1)$ model of equation (2), 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 $\beta_n$ estimator is always defined:

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

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 MSE$\,s$ 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 Edgeworth approximations for the three IE, 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 IE 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 reparameterized. 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.

Furthermore, by generalizing to multistep procedures we are able to provide recursive indirect estimators 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. [7]). This could also imply the first order asymptotic non-equivalence between the three IE. Second, an application of the Edgeworth approximations could lay in the derivation of higher order properties of indirect testing procedures. Third, the introduction of indirect estimators via the actual *use* of the Edgeworth approximations for the auxiliary one. For example, an indirect estimator could be defined by substituting $E_\theta \beta_n$ with $\mathcal{I}_{\varphi_{V(\theta)}}\left(k_{2_\beta}\right)$ in definition D.3. We leave all these questions for future work.

## 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] Bansal R.P., A. R. Gallant, R. Hussey and G. Tauchen (1995), "Non-parametric estimation of structural models for high-frequency currency market data" *Journal of Econometrics*, 66, 251-287.

[5] 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.

[6] Bollerslev T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* 31, 307-328.

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

[8] Chumacero R. A. (2001), "Estimating ARMA Models Efficiently", *Studies in Nonlinear Dynamics and Econometrics*, 5, 103–114.

[9] Chung, C. and G. Tauchen (2001), "Testing target-zone models using efficient method of moments" *Journal of Business and Economic Statistics*, 19, 255–269.

[10] Davidson J. (1994), "Stochastic Limit Theory" Oxford University Press.

[11] Demos A. and D. Kyriakopoulou (2008), "Edgeworth expansions for the MLE and MM estimators of an MA(1) model", *Communications in Statistics-Theory and Methods*, forthcoming.

[12] Dridi R., and E. Renault (2000),"Semi-parametric Indirect Inference", mimeo, LSE and CREST-INSEE and Universite de Paris IX Dauphine.

[13] Gallant R.A., D. Hsieh and G. Tauchen (1997), "Estimation of stochastic volatility models with diagnostics", *Journal of Econometrics*, 81, 159-192.

[14] Gallant R.A. and J.R. Long (1997) "Estimating stochastic differential equations efficiently by Minimum Chi- Squared", *Biometrika*, 84, 125-141.

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

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

[17] Ghysels I., L. Khalaf and C. Vodounou (2003), "Simulation based inference in moving average models" Annales d' Econommie et de Statistique, 69, 85-99.

[18] Götze F., and C. Hipp (1983), "Asymptotic expansions for sums of weakly dependent random vectors", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandie Gebiete*, 64, pp. 211-239.

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

[20] Gourieroux C.,and A. Monfort (1996), "*Simulation-based Econometric Methods*", CORE Lectures, Ox. Un. Press.

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

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

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

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

[25] Hall, P. and J.L. Horowitz (1996), "Bootstrap critical values for tests based on generalized-method-of-moment estimators", *Econometrica* 64, 891-916.

[26] 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.

[27] 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.

[28] Linton, O. (1997), "An asymptotic expansion in the GARCH(1,1) model", *Econometric Theory* 13, 558-581.

[29] Magdalinos, M.A. (1992), "Stochastic Expansions and Asymptotic Approximations", *Econometric Theory*, Vol. 8, No. 3, pp. 343-367.

[30] McCullagh, P. (1987), "*Tensor Methods in Statistics*" Chapman and Hall, London, New York.

[31] 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.

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

[33] 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.

[34] Phillips P.C.B. (2011), "Folklore theorems, implicit maps and Indirect Inference" *Econometrica*, forthcoming.

[35] Robinson, P.M. (1988) "The stochastic difference between econometric statistics", *Econometrica* 56, 531-548.

[36] Sargan, J.D. (1976) "Econometric estimators and the Edgeworth approximation", *Econometrica* 44, 421-448.

[37] 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.

[38] Skovgaard I.M. (1981), "Transformation of an Edgeworth expansion by a sequence of smooth functions", *Scandinavian Journal* of Statistics 8, 207-217.

[39] Smith A.A. (1993), "Estimating nonlinear time-series models using Simulated Vector Autoregressions", *Journal of Applied Econometrics* 8, 63-84.

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

[41] Sweeting T.J. "Speeds of convergence for the multidimensional central limit theorem", *The Annals of Statistics* 5, 28-41.

[42] van der Vaart, A.W. and J.A. Wellner, (2000) "*Weak convergence and empirical processes*" Springer.

## Appendix-Proofs

**Proof of Lemma 2.1.** i). Assumptions A.6, A.7a) and the triangle inequality imply that for any $\varepsilon > 0$

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

This along with assumption A.6 implies that for any $\varepsilon > 0$

$$\sup_{\theta^*\in\Theta} P_{\theta^*} \left(\sup_{\theta\in\Theta} \left| \left\|\beta_n - b\left(\theta\right)\right\|_{W_n^*\left(\theta_n^*\right)} - \left\|b\left(\theta^*\right) - b\left(\theta\right)\right\|_{E_{\theta^*} W^*\left(\theta^*\right)}\right| > \varepsilon\right) = o\left(n^{-a^*}\right)$$

which in turn along with A.2 and A.3 implies the uniform consistency of GMR1. ii). Given i), we have that GMR1 $\in \mathcal{O}_{\varepsilon'}\left(\theta_0\right) \supset \overline{\mathcal{O}}_\varepsilon\left(\theta_0\right)$ with probability $1 - o\left(n^{-a^*}\right)$ that is independent of $\theta$. Hence due to the mean value theorem submultiplicativity and assumption A.9

$$\begin{aligned} & \sup_{\theta\in\Theta} P_\theta \left(\sqrt{n}\left\|\text{GMR1} - \theta\right\| > M \ln^{1/2} n\right) \\ \leq\ & \sup_{\theta\in\Theta} P_\theta \left(\sqrt{n}\left\|\beta_n - b\left(\theta\right)\right\| > M \ln^{1/2} n\right) + o\left(n^{-a^*}\right) \end{aligned}$$

which is $o\left(n^{-a^*}\right)$ due to A.8 (see also the second part of the proof of lemma 3.1). Assumptions A.7b), and A.9 enable a Taylor expansion of order $s^*$ around $\left(\theta_n - \theta, m_n\right)$ of the f.o.c.'s that the estimator satisfies. Moreover assumption A.7b) implies that $D^i W_n^*\left(\theta\right)$ is asymptotically equi-Lipschitz on $\overline{\mathcal{O}}_\varepsilon\left(\theta_0\right)$ with probability $1 - o\left(n^{-a^*}\right)$ independent of $\theta$, for $i = 0, \ldots, s$. Analogously assumption A.9 implies that $D^i b\left(\theta\right)$ is Lipschitz on $\overline{\mathcal{O}}_\varepsilon\left(\theta_0\right)$, for $i = 0, \ldots, s+1$. These along with A.8 imply that the remainder (say) $R_n\left(\theta\right)$ satisfies

$$\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon\left(\theta_0\right)} P_\theta\left(R_n\left(\theta\right) > \gamma_n\right) = o\left(n^{-a^*}\right)$$

with $\gamma_n = o\left(n^{-a^*}\right)$ independent of $\theta$. Also due to A.9 an inversion of the expansion implies that

$$\sqrt{n}\left(\theta_n - \theta\right) = f_n\left(\sqrt{n} m_n\right) + R_n^*\left(\theta\right)$$

where $f_n$ is polynomial with coefficients satisfying the conditions (3.1.I-II) of Skovgaard [38] uniformly on $\overline{\mathcal{O}}_\varepsilon\left(\theta_0\right)$. Hence a generalization of lemma 4.6 of Skovgaard [38] implies that $f_n\left(\sqrt{n} m_n\right)$ admits an Edgeworth expansion of order $s^*$ uniformly on $\overline{\mathcal{O}}_\varepsilon\left(\theta_0\right)$. Due to the behavior of $R_n\left(\theta\right)$ and A.7b), A.9

$$\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon\left(\theta_0\right)} P_\theta\left(R_n^*\left(\theta\right) > \gamma_n^*\right) = o\left(n^{-a^*}\right)$$

with $\gamma_n^* = o\left(n^{-a^*}\right)$ independent of $\theta$. Hence $\sqrt{n}\,(\theta_n - \theta)$ admits the same Edgeworth expansion with $f_n\left(\sqrt{n}m_n\right)$. ■

**Proof of Lemma 2.2.** i). The proof is almost identical to the proof of 2.1.i) except for the fact that A.6 along with A.2 imply that

$$\sup_{\theta\in\Theta}\|E_\theta\beta_n - b(\theta)\| = o(1)$$

which in turn implies that for any $\varepsilon > 0$

$$\sup_{\theta^*\in\Theta} P_{\theta^*}\left(\sup_{\theta\in\Theta}\left|\|\beta_n - E_\theta\beta_n\|_{W_n^*(\theta_n^*)} - \|b(\theta^*) - b(\theta)\|_{E_{\theta^*}W^*(\theta^*)}\right| > \varepsilon\right) = o\left(n^{-a^*}\right)$$

ii). Notice that assumptions A.9, A.10 due to remark R.3 imply that for $n$ large enough rank $E_\theta\beta_n = p$, for all $\theta$ in $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ and that $D^i E_\theta\beta_n$ is equi-Lipschitz on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$ on $\overline{\mathcal{O}}_\varepsilon(\theta_0)$. The rest follow in an analogous manner to the proof of 2.1.ii). ■

**Proof of Lemma 2.3.** We obtain that

$$\begin{aligned}
&\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta\left(\left\|\sqrt{n}\left(E_{\theta_n^*}\beta_n - E_{\theta_0}\beta_n\right) - A_n(\theta)\right\| > \gamma_n\right)\\
\leq\;& \sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta\left(\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)}\sqrt{n}\left\|E_\theta\beta_n - b(\theta) - \textstyle\sum_{i=1}^s \frac{1}{n^{\frac{i}{2}}}\mathcal{I}_V(k_i(\theta))\right\| > \frac{\gamma_n}{6}\right)\\
&+\textstyle\sum_{i=1}^s \sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta\left(\frac{1}{n^{\frac{i-1}{2}}}\|B_n(\theta)\| > \frac{\gamma_n}{3s}\right)\\
&+\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta\left(\sqrt{n}\left\|b(\theta_n^*) - b(\theta) - \textstyle\sum_{i=1}^s \frac{1}{i!}D^i b(\theta)\left((\theta_n^* - \theta)^i\right)\right\| > \frac{\gamma_n}{3}\right) + o\left(n^{-\frac{s^*-1}{2}}\right)
\end{aligned}$$

where

$$B_n(\theta) = \mathcal{I}_V(k_i(\theta_n^*)) - \mathcal{I}_V(k_i(\theta)) - \textstyle\sum_{j=1}^{s-i}\frac{1}{i!}D^j\mathcal{I}_V(k_i(\theta))\left((\theta_n^* - \theta)^j\right)$$

Now we have that

$$a_n = \sqrt{n}\left\|E_\theta\beta_n - b(\theta) - \textstyle\sum_{i=1}^s \frac{1}{n^{\frac{i}{2}}}\mathcal{I}_V k_i(\theta)\right\| = o\left(n^{-a}\right)$$

independent of $\theta$, due to lemma 3.1 and similarly due to Taylor's theorem

$$\begin{aligned}
&\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta\left(\frac{1}{n^{\frac{i-1}{2}}}\|B_n(\theta)\| > \frac{\gamma_n}{3s}\right)\\
\leq\;& \sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta\left(\frac{1}{n^{\frac{i-1}{2}}}\frac{1}{(s-i+1)!}\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)}\left\|D^{s-i+1}\mathcal{I}_V(k_i(\theta))\right\|\|\theta_n^* - \theta\|^{s-i+1} > \frac{\gamma_n}{3s}\right)
\end{aligned}$$

which due to the continuity of $D^{s-i+1}\mathcal{I}_{\varphi_{V(\theta)}}(k_i)$ and the assumed behavior of $\theta_n^*$ the latter bound is less than or equal to

$$\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta\left(\frac{\ln^{\frac{s-i+1}{2}} n}{n^{\frac{s}{2}}}\frac{\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)}\left\|D^{s-i+1}\mathcal{I}_V\left(k_i\left(\theta\right)\right)\right\|}{(s-i+1)!} > \frac{\gamma_n}{3s}\right) + o\left(n^{-a^*}\right)$$

and the displayed bound is zero when $\gamma_n \geq \frac{\ln^{\frac{s-i+1}{2}} n}{n^{\frac{s}{2}}}\frac{3s\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)}\left\|D^{s-i+1}\mathcal{I}_V(k_i(\theta))\right\|}{(s-i+1)!}$. Furthermore using the same reasoning as above

$$\begin{aligned}
&\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta\left(\sqrt{n}\left\|b\left(\theta_n^*\right)-b\left(\theta_0\right)-\sum_{i=1}^s\frac{1}{i!}D^i b\left(\theta_0\right)\left(\left(\theta_n^*-\theta_0\right)^i\right)\right\| > \frac{\gamma_n}{3}\right)\\
\leq\ &\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta\left(\sqrt{n}\left\|\theta_n^*-\theta_0\right\|^{s+1} > \frac{(s+1)!\gamma_n}{3\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)}\left\|D^{s+1}b\left(\theta\right)\right\|}\right)+o\left(n^{-a^*}\right)\\
\leq\ &\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta\left(\frac{\ln^{\frac{s+1}{2}} n}{n^{\frac{s}{2}}} > \frac{(s+1)!\gamma_n}{3\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)}\left\|D^{s+1}b\left(\theta\right)\right\|}\right)+o\left(n^{-a^*}\right)
\end{aligned}$$

the last probability is zero when $\gamma_n \geq \frac{3\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)}\left\|D^{s+1}b(\theta)\right\|}{(s+1)!}\frac{\ln^{\frac{s+1}{2}} n}{n^{\frac{s}{2}}}$. Hence for $\gamma_n = \max\left(\frac{3\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)}\left\|D^{s+1}b(\theta)\right\|}{(s+1)!}\frac{\ln^{\frac{s+1}{2}} n}{n^{\frac{s}{2}}}, 6a_n, \frac{\ln^{\frac{s-i+1}{2}} n}{n^{\frac{s}{2}}}\frac{3s\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)}\left\|D^{s-i+1}\mathcal{I}_V(k_i(\theta))\right\|}{(s-i+1)!}, i=1,\ldots,s\right)$ the result follows for large enough $n$. ■

**Proof of Lemma 2.4.** Notice that the uniform consistency and its subsequences follow for the GMR1 and GMR2 due to the first parts of lemmas 2.1, 2.2. Then for i) we have that remark R.3 implies that for $n$ large enough rank $E_\theta\beta_n = p$, for all $\theta$ in $\overline{\mathcal{O}}_\varepsilon(\theta_0)$, hence $\beta_n = E_{\mathrm{GMR2}}\beta_n$. Then due to 2.1

$$\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} P_\theta\left(\sqrt{n}\left\|\mathrm{GMR1}-\theta\right\| > M\ln^{1/2} n\right) = o\left(n^{-a^*}\right)$$

Hence with probability $1-o\left(n^{-a^*}\right)$ independent of $\theta$, due to the mean value theorem and A.9

$$\begin{aligned}
\left\|\mathrm{GMR1}-\mathrm{GMR2}\right\| &\leq M\left\|\beta_n - b\left(\mathrm{GMR2}\right)\right\|\\
&\leq M\left(\left\|\beta_n - E_{\mathrm{GMR2}}\beta_n\right\| + \left\|E_{\mathrm{GMR2}}\beta_n - b\left(\mathrm{GMR2}\right)\right\|\right)
\end{aligned}$$

and the first term in the last display is zero. Due to lemma 3.1 the second term is $O\left(\frac{1}{n}\right)$. Hence

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

Therefore due to lemmas 2.3, 3.1 we obtain that

$$\sup_{\theta\in\overline{\mathcal{O}}_{\varepsilon}(\theta_0)} P_\theta\left(\left\|\sqrt{n}\left(\beta_n - E_{\mathrm{GMR2}}\beta_n\right) - \Gamma_n(\theta)\right\| > \gamma_n\right) = o\left(n^{-a^*}\right)$$

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

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

This along with A.9 implies the result as in the last part of the proof of 2.2. ii) follows the same way as i) except now $\left\|\beta_n - E_{\mathrm{GMR2}}\beta_n\right\|$ is zero with probability $1 - o\left(n^{-a^*}\right)$ independent of $\theta^*$. ■

**Proof of Lemma 2.5.** It is easy to see that this special assumption implies that GMR1 = GT with probability $1 - o\left(n^{-a^*}\right)$ independent of $\theta$. The rest are trivial consequences of lemma 2.1. ■

**Proof of Lemma 2.6.** i). Assumptions A.6, A.7a) and the triangle inequality imply that for any $\varepsilon > 0$

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

Assumption A.12 implies that

$$\sup_{\theta\in\Theta}\left\|E_\theta c_n(\beta) - c(\theta,\beta)\right\| = o(1)$$

and

$$\left\|E_\theta c_n(\beta) - E_\theta c_n(\beta')\right\| \leq M\left\|\beta - \beta'\right\|, \text{ for all } \beta, \beta'$$

which in turn along with A.6 imply that for all $\varepsilon > 0$

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

which along with the asymptotic behavior of the weighting matrix imply that for all $\varepsilon > 0$

$$\sup_{\theta^*\in\Theta} P_{\theta^*}\left(\left|\left\|E_\theta c_n(\beta_n)\right\|_{W_n^{**}(\theta_n^*)} - \left\|c(\theta, b(\theta^*))\right\|_{E_{\theta^*}W^{**}(\theta^*)}\right| > \varepsilon\right) = o\left(n^{-a^*}\right)$$

which along with A.12 and A.3 implies the uniform consistency of GT. ii). The proof follows analogously to the proof of 2.1 by noting that assumption A.13 implies that $D^i E_\theta c_n(\beta)$ is equi-Lipschitz on $\overline{\mathcal{O}}_\eta(\varphi_0)$, for $i = 0, \ldots, s+1$. ■

**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

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

Due to the hypothesis for the support of $P_n$

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

Hence

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

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

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

and $n^{a+\frac{m}{2}}\rho^mq^mP\left(\|\zeta_n\|>c(\ln n)^\epsilon\right)=o(1)$ if $\epsilon\geq\frac{1}{2}$ and $c\geq\sqrt{2a+m+1}$ by lemma 2 of Magdalinos [29]. 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 [24] formula 8.357. For the uniform case first notice that

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

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

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

where $\lambda_{\max}\left(\theta\right)$ denotes the maximum absolute eigenvalue of $V^{1/2}\left(\theta\right)$ and $\theta_i^*\in\overline{\mathcal{O}}_{\varepsilon}\left(\theta_0\right)$ exist for all $i=1,\ldots,s^*$ due to the continuity and are independent of $z$ due to the positivity and the fact that $\pi_i$ are polynomials in $x$, and $\theta^*$ exists due to continuity of $V$ and the compactness of $\overline{\mathcal{O}}_{\varepsilon}\left(\theta_0\right)$. For $M\geq s^*\lambda_{\max}\left(\theta^*\right)$ the result follows from lemma 2 of Magdalinos [29]. The rest follows in the same spirit of the first part. ■

**Proof of Lemma 3.2.** The assumptions and lemma 3.1 ensure the validity of the mean approximation. Following the procedure described in the paragraph immediately after lemma 3.1 we obtain that the relevant inversion is

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

Integrating with respect to $\left(1+\frac{\pi_1(z,\theta)}{\sqrt{n}}\right)\varphi_{V(\theta)}\left(z\right)$, and noting that $k_{1_\beta}\left(z,\theta\right)=z$, $k_{2_\beta}\left(z,\theta\right)=z\pi_1\left(z,\theta\right)$ we obtain the result. ■

**Proof of Lemma 3.3.** Argue as in the proof of the previous lemma and use lemma 2.3 in order to obtain that the relevant inversion is given by

$$\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W^*\left(k_{1_\beta}-\frac{\mathcal{I}_{\varphi_V}\left(k_{2_\beta}\right)}{\sqrt{n}}\right)-\frac{1}{\sqrt{n}}\mathcal{C}^{-1}\left[k'_{1_{\theta^*}}\frac{\partial^2 b_j}{\partial\theta\partial\theta'}\right]_{j=1,\dots,l}W^*\frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W^*k_{1_\beta}$$
$$-\frac{1}{2\sqrt{n}}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W^*\left[\frac{\partial b'}{\partial\theta}k'_{1_\beta}W^*\mathcal{C}^{-1}\frac{\partial b_j}{\partial\theta\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W^*k_{1_\beta}\right]_{j=1,\dots,q}$$
$$-\frac{1}{\sqrt{n}}\mathcal{C}^{-1}\left(\frac{\partial b'}{\partial\theta}\frac{1}{\sqrt{n}}k_{1_{w^*}}+\left[\frac{\partial}{\partial\theta'}W^*_{j,j'}k_{1_{\theta^*}}\right]_{j,j'=1,\dots,l}\right)\frac{\partial b}{\partial\theta'}\mathcal{C}^{-1}\frac{\partial b'}{\partial\theta}W^*k_{1_\beta}$$

Integrating the above w.r.t. $\left(1+\frac{\pi_1(z,\theta)}{\sqrt{n}}\right)\varphi_{V(\theta)}(z)$ we get the result. ■

**Proof of Lemma 3.5.** It follows directly by lemmas 2.5 and 3.2. ■

**Proof of Lemma 3.4.** The assumptions and lemma 3.1 ensure the validity of the mean approximation. Following the procedure described in the paragraph immediately after lemma 3.1 we obtain that the relevant inversion is

$$\mathcal{D}^{-1}\mathcal{E}\frac{\partial c(\theta,b)}{\partial\beta'}k_{1_\beta}$$
$$+\frac{1}{2\sqrt{n}}\mathcal{D}^{-1}\mathcal{E}\left[k'_{1_\beta}\frac{\partial^2 c_j(\theta,b)}{\partial\beta\partial\beta'}k_{1_\beta}\right]_{j=1,\dots,l}-\frac{1}{\sqrt{n}}\mathcal{D}^{-1}\mathcal{E}\left[q'_{1_\beta}\frac{\partial b'}{\partial\theta}\frac{\partial^2 c_j(\theta,b)}{\partial\beta\partial\beta'}k_{1_\beta}\right]_{j=1,\dots,l}$$
$$+\frac{1}{2\sqrt{n}}\mathcal{D}^{-1}\mathcal{E}\left[q'_{1_\beta}\mathcal{H}_j q_{1_\beta}\right]_{j=1,\dots,l}$$
$$+\frac{1}{\sqrt{n}}\mathcal{D}^{-1}\left[q'_{1_\beta}\mathcal{H}_j-k'_{1_\beta}\frac{\partial b'}{\partial\theta}\frac{\partial^2 c_j(\theta,b)}{\partial\beta\partial\beta'}\right]_{j=1,\dots,l}W^{**}(\theta)\mathcal{J}^*k_{1_\beta}$$
$$-\frac{1}{\sqrt{n}}\mathcal{D}^{-1}\frac{\partial b'}{\partial\theta}\frac{\partial c'(\theta,b)}{\partial\beta}\mathcal{J}\mathcal{J}^*k_{1_\beta}$$

Integrating the above w.r.t. $\left(1+\frac{\pi_1(z,\theta)}{\sqrt{n}}\right)\varphi_{V(\theta)}(z)$ we get the result. ■

**Proof of Lemma 3.6.** The assumptions and lemma 3.1 ensure the validity of the second moment approximation. Following the procedure described in the paragraph immediately after lemma 3.1 the result is obtained when the outer products of the formulae in the proofs of lemmas 3.2 and 3.3 are computed in the particular case and the results are integrated w.r.t. to $\left(1+\frac{\pi_1(z,\theta)}{\sqrt{n}}\right)\varphi_{V(\theta)}(z)$. ■

**Proof of Lemma 3.7.** The assumptions and lemma 3.1 ensure the validity of the second moment approximation. Following the procedure described in the paragraph immediately after lemma 3.1 the result is obtained when the outer product of the formula in the proof of lemma 3.4 is computed and the result is integrated w.r.t. $\left(1+\frac{\pi_1(z,\theta)}{\sqrt{n}}\right)\varphi_{V(\theta)}(z)$. ■

**Proof of Lemma 4.1.** First notice that in any step of the procedure the binding function is the identity. Next the $o\left(n^{-a^*}\right)$ uniform consistency of $\theta_n^{(0}$ ensures the analogous for any step of the recursion. Then condition $\limsup_n \sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} E_\theta \left\|\sqrt{n}\bar{l}_n(\theta)\right\|^2 < \infty$ ensures that $\sup_{\theta\in\overline{\mathcal{O}}_\varepsilon(\theta_0)} \left\|D^2 E_\theta \theta_n^{(h}\right\| < M$ for any $h = 0, \ldots, \zeta$ and the same holds for condition A.11. These imply that in any step of the recursion

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

and that accordingly admits a locally uniform Edgeworth expansion of order $s^*$. The proof for the moment approximations for the case $h = 1$ follows easily. Using induction if these hold for some $h$, then notice that the Taylor inversion approximating appropriately $\sqrt{n}\left(\theta_n^{(h+1} - \theta\right)$ is

$$k_{1_{\theta(h}} - \sum_{i=2h+2}^{2h+3} \frac{\mathcal{I}_V\left(k_{i_{\theta(h}}\right)}{n^{\frac{i-1}{2}}} - \frac{D\mathcal{I}_V\left(k_{2h+2_{\theta(h}}\right)\left(k_{1_{\theta(h}}\right)}{n^{\frac{2h+2}{2}}}$$

The result follows by integrating this and its exterior product with respect to the Edgeworth distribution of $\sqrt{n}\left(\theta_n^{(h} - \theta\right)$ that has density of the form $\left(1 + \sum_{i=1}^{2s+2} \frac{\pi_i(z,\theta)}{n^{i/2}}\right)\varphi_{V(\theta)}(z)$ and by holding the relevant terms. ■

**Proof of Proposition 4.** Let $s^* = 5$, $y_0^2$ is observed and for $t = 1, \ldots, n$, $x_t = (y_t, y_t y_{t-1})$ and $c(x_t, \beta) = \begin{pmatrix} y_{i-1}^2 - \beta_1 \\ y_i y_{i-1} - \beta_2 \end{pmatrix}$.

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 $W^* = \mathrm{Id}_2$. Notice that $b(\theta) = E_\theta\begin{pmatrix} y_t y_{t-1} \\ y_{t-1}^2 \end{pmatrix} = \begin{pmatrix} \theta \\ 1+\theta^2 \end{pmatrix}$, $B = b(\Theta)$ compact and assumption A.9. Now for $E\left(\varepsilon_i^{14}\right) < \infty$ assumption (2) of Götze and Hipp [19] is satisfied uniformly over $\Theta$. Further, their assumption (3) is satisfied, uniformly over $\Theta$, due to the $1-$ Dependence of the MA(1) model. For the same reason conditions (i) and (ii) of their lemma 2.3 are satisfied, uniformly over $\Theta$, while for their condition (iii), the choice of $l_1 = 0$ and $l_2 = 1$ (in their notation) implies that the relevant derivative equals a linear combination of $\varepsilon_{-2}$, $\varepsilon_{-1}$, $\varepsilon_0$, $\varepsilon_1$ and $\varepsilon_2$, with coefficients that are not zero for any choice of $\theta$ in $\Theta$, and the condition follows due to independence. Consequently, $\sqrt{n}\begin{pmatrix} \frac{1}{n}\sum_i y_i y_{i-1} - \theta \\ \frac{1}{n}\sum_i y_{i-1}^2 - 1 - \theta^2 \end{pmatrix}$ has a $5^{th}$ order Edgeworth expansion uniformly over $\Theta$.

For the auxiliary estimator, $\beta_n = \frac{\sum_{t=1}^n y_t y_{t-1}}{\sum_{t=1}^n y_{t-1}^2}$, consider the function $f(x, y) = \frac{x}{y}$. All derivatives, up to $5^{th}$ order, of $f$ evaluated at $x = \theta$ and $y = 1 + \theta^2$ are of the form either $\frac{\theta}{\left(1+\theta^2\right)^k}$ or $\frac{1}{\left(1+\theta^2\right)^k}$. Hence, the derivatives of $f(x, y)$ satisfy assumption 3.1 I and II of Skovgaard [38], uniformly with respect to $\theta$ (see the second part of remark R.2). Assumption A.11 also follows from differentiability of $E_\theta y_i^r$, for $r = 1, ..., 14$, the differentiability of $f(x, y)$ and dominated convergence. Consequently, $\sqrt{n}\left(\beta_n - \frac{\theta}{1+\theta^2}\right)$ has a $5^{th}$ order Edgeworth expansion, uniformly over $\Theta$ implying also assumption A.6. This implies that GMR1, and GMR2 are uniformly consistent with rate $o\left(n^{-2}\right)$. Assumption A.9 applies for the function $b(\theta) = \frac{\theta}{1+\theta^2}$ and therefore due to lemma 2.1 $\sqrt{n}\,(\text{GMR1} - \theta)$ has a $5^{th}$ order Edgeworth expansion, uniformly over $\Theta$. This implies also that the $1 - \text{GMR2}$ is uniformly consistent with rate $o\left(n^{-2}\right)$. The same moment condition implies that $\sqrt{n}\bar{l}_n(\theta)$ (where $\bar{l}_n(\theta)$ as in lemma 4.1) has an Edgeworth expansion of third order uniformly over $\Theta$ and therefore $DE_\theta\beta_n$ is non degenerate for all $\theta$ for large enough $n$. Hence due to lemma 2.4.i) GMR2 has a $5^{th}$ order Edgeworth expansion, uniformly over $\Theta$ which is also true for $1 - \text{GMR2}$ due to lemma 4.1. ■

**Proof of Proposition 5.** Proceed as in the proof of proposition 4 where now $y_{-1}^2$ and $y_0^2$ are observed and for $i = 1, \ldots, n$, $x_i = \left(y_i^2, y_i^2 y_{i-1}^2, y_i^2 y_{i-2}^2\right)$,

$$c(x_i, \beta^*) = \begin{pmatrix} y_i^2 - \beta_1^* \\ y_i^4 - \beta_2^* \\ y_i^2 y_{i-1}^2 - \beta_3^* \\ y_i^2 y_{i-2}^2 - \beta_4^* \end{pmatrix} \text{ and and } W^* = \text{Id}_4. \text{ Let } \beta^*(\theta) = E_\theta \begin{pmatrix} y_i^2 \\ y_i^4 \\ y_i^2 y_{i-1}^2 \\ y_i^2 y_{i-2}^2 \end{pmatrix}.$$

The uniform versions of the Götze and Hipp conditions (see remark R.2) can be verified due to the moment condition employed and the weak-dependence of the ARMA$(1, 1)$ representation of the model. Hence $\sqrt{n}\left(\beta_n^* - \beta^*(\theta)\right)$ admits a fifth order Edgeworth expansion uniformly over $\Theta$. Considering the auxiliary estimator $\beta_n = \left(\beta_{1_n}^*, \frac{\beta_{3_n}^*}{\beta_{2_n}^*}, \frac{\beta_{4_n}^*}{\beta_{2_n}^*}\right)'$ proceed analogously to the proof of the previous proposition using properties of $f(x, y, z, w) = \left(x, \frac{z}{y}, \frac{w}{y}\right)'$ to obtain the needed results. ■

## Figures

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

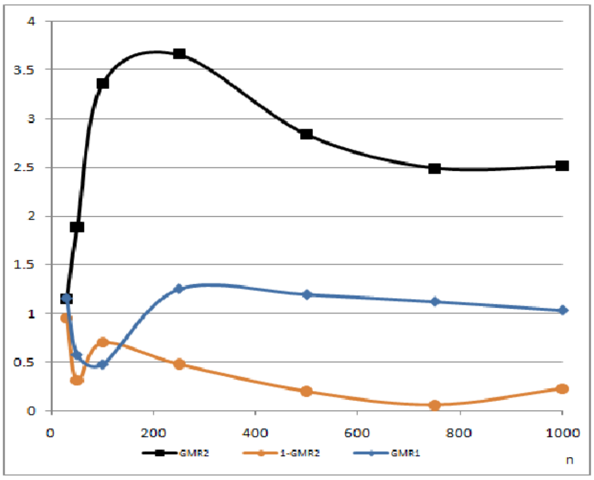


Figure 1: $n \times \left|\widehat{\text{Bias}}\right|$.

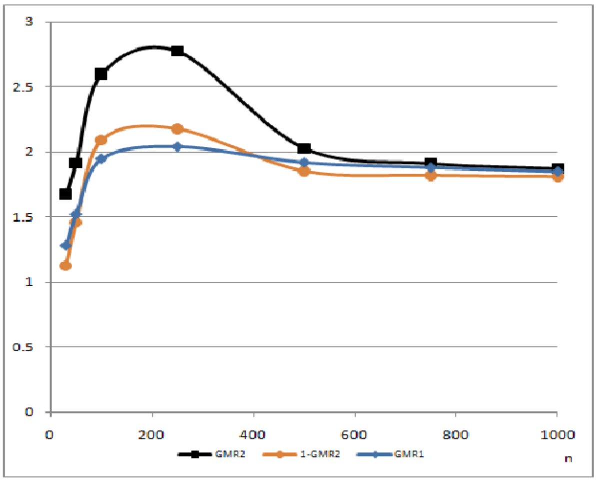


Figure 2: $n \times \widehat{\text{MSE}}$.

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

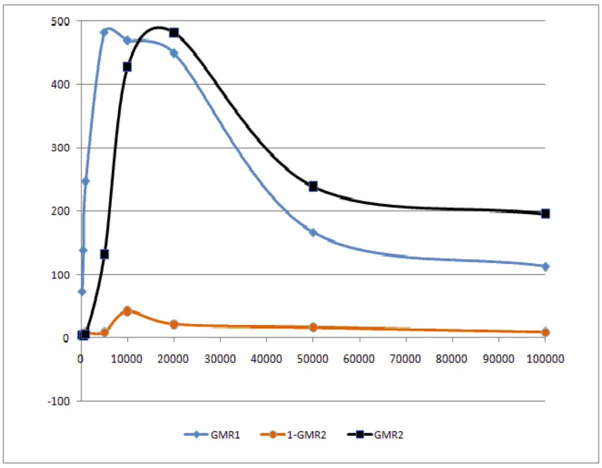


Figure 3: $n \times \left\|\widehat{\text{Bias}}\right\|$.

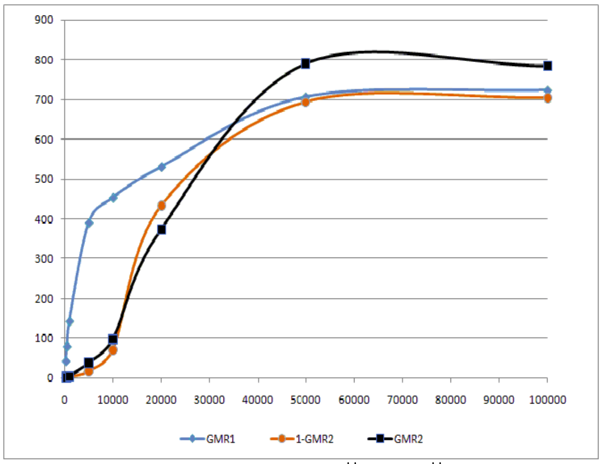


Figure 4: $n \times \left\|\widehat{\text{MSE}}\right\|$.