Finite Sample Theory and Bias Correction of Maximum Likelihood Estimators in the EGARCH Model*

Antonis Demos[†] and Dimitra Kyriakopoulou[‡]

Athens University of Economics and Business

and Université catholique de Louvain

February 23, 2018

Abstract

We derive analytical expressions of bias approximations for maximum likelihood (ML) and quasi-maximum likelihood (QML) estimators of the EGARCH$(1,1)$ parameters that enable us to correct after the bias of all estimators. The bias correction mechanism is constructed under the specification of two methods that are analytically described. We also evaluate the residual bootstrapped estimator as a measure of performance. Monte Carlo simulations indicate that, for given sets of parameters values, the bias corrections work satisfactory for all parameters. The proposed full-step estimator performs better than the classical one and is also faster than the bootstrap. The results can be also used to formulate the approximate Edgeworth distribution of the estimators.

Keywords: Exponential GARCH, maximum likelihood estimation, finite sample properties, bias approximations, bias correction, Edgeworth expansion, bootstrap.

JEL Classification: C13, C22

*This paper was previously circulated under the title "Bias Correction of ML and QML Estimators in the EGARCH(1,1) Model".

[†]Department of International European Economic Studies, Athens University of Economics and Business (AUEB), Athens, Greece, e-mail: demos@aueb.gr.

[‡]Corresponding author. Center for Operations Research and Econometrics (CORE), Université catholique de Louvain, Louvain-la-Neuve, Belgium, e-mail: dimitra.kyriakopoulou@uclouvain.be.

1 Introduction

In the econometric literature it is widely acknowledged that since the introduction of autoregressive conditional heteroskedasticity models, the so-called GARCH models, see Engle (1982) and Bollerslev (1986), the research on this class of models has been steadily increasing over the years. The study of statistical properties of the models is important from a theoretical perspective, but also for statistical inference and applied work. The asymptotic properties of estimators have been explored under many different model considerations, both in the univariate and multivariate frameworks, see for instance Weiss (1986), Lee and Hansen (1996), Lumsdaine (1996), Jensen and Rahbek (2004), Jeantheau (1998), Comte and Lieberman (2003) to cite a few of classical papers in the GARCH literature.

One of the most important extensions of GARCH models, that is also widely used in many financial applications, is the Exponential GARCH (EGARCH) model of Nelson (1991), which became immediately a classical GARCH model in financial econometrics. Comparing to the standard GARCH model, exponential-type models enable richer volatility dynamics and the distinctive advantage of the EGARCH is that the model captures the negative dynamic asymmetries noticed in many financial series, i.e. the so-called leverage effect. Also, their fitted values of volatility are guaranteed to be positive due to their log-linear form, which implies that non-negativity parameter restrictions are avoided. He *et al.* (2002) analyzed the theoretical properties of this model, and Hafner and Linton (2017) proposed an alternative estimator for the EGARCH model, which is available in a closed form and discussed its properties. The asymptotic properties of maximum likelihood estimators[1] in this model have not been fully explored until recently, see Wintenberger (2013), Kyriakopoulou (2015), Straumann (2005), and Straumann and Mikosch (2006).

Although the EGARCH model is in widespread use, there is no finite sample theory available for maximum likelihood (ML) and quasi-maximum likelihood (QML) estimators of its parameters. The existing results in the literature cover only the GARCH models and it seems that there is no direct extension to the case of the EGARCH model. Linton (1997) was the first to provide an asymptotic expansion in the first-order GARCH model. The small

[1] Zaffaroni (2009) estimated the EGARCH parameters with Whittle methods and the asymptotic distribution theory of these estimators was established.

sample estimation bias and properties of GARCH models and extension of that in the in-mean case have been studied by Iglesias and Phillips (2002, 2011, 2012). However, little is known about the small sample properties of the QML estimators in the EGARCH model. Only Deb (1996) examined the finite sample properties of the maximum likelihood and quasi-maximum likelihood estimators of the EGARCH(1,1) process using Monte Carlo methods. He used, however, response surface methodology in order to examine the finite sample bias and other properties of interest, by summarizing the results of a wide array of experiments. Perez and Zaffaroni (2008) compared the finite sample properties of the ML and Whittle estimators, in terms of bias and efficiency, in the EGARCH model and its long-memory version, confirming that maximum likelihood is more efficient.

In this paper we study the finite sample theory of maximum likelihood estimators of the EGARCH$(1,1)$ parameters and to do this we first derive the, up to $1/T$ order, Edgeworth type expansions, where T denotes the sample size. In fact, we derive the so-called Edgeworth-B coefficients, see Linton (1997). For a discussion about the Edgeworth expansion we refer to the monograph of Hall (1992). Specifically, we derive conditions on the parameter space so that the derivatives of the likelihood function are stationary. Based on the expansions we derive, we also provide bias approximations of the same order.

For a parameter vector φ and any consistent estimator $\widehat{\varphi}_T$ an approximate bias corrected estimator is the solution of

$$\min_{\varphi} \left\| \widehat{\varphi}_T - \varphi - B_T(\varphi) \right\|, \tag{1}$$

with $B_T(\varphi)$ the bias of $\widehat{\varphi}_T$, where $\varphi + B_T(\varphi)$ is a (potentially stochastic) approximation of $E_{\varphi}\widehat{\varphi}_T$ (see Arvanitis and Demos, 2015) and $\|\cdot\|$ denotes the Euclidean norm. The solution of the above minimization problem is called here as full-step bias corrected estimator. It is worth noticing that, according to Arvanitis and Demos (2015, 2016), the derived estimator in this way is an Indirect Inference estimator with the binding function being the identity (see e.g. Gourieroux *et al.*, 1993 for definitions of Indirect Inference estimators and also Arvanitis and Demos, 2015, Demos and Kyriakopoulou, 2013, and Gourieroux *et al.* 2000 for the bias properties of them).

Another possible solution of the above minimazation problem is

$$\widetilde{\varphi}_T = \widehat{\varphi}_T - B_T(\widehat{\varphi}_T),$$

where the term $B_T(\widehat{\varphi}_T)$ denotes the bias of $\widehat{\varphi}_T$ evaluated at $\widehat{\varphi}_T$. The estimator $\widetilde{\varphi}_T$ is called here as the first-step bias corrected one (see Arvanitis and Demos, 2015) and is the paradigm of a vast literature of approximate bias correction (see Cordeiro and McCullagh, 1991, Cordeiro and Klein, 1994, Cox and Hinkley, 1994, Fernandez-Val and Vella, 2011, Gourieroux *et al.,* 2000, Iglesias and Phillips, 2011, Linton, 1997, MacKinnon and Smith, 1998, and Rilstone *et al.*, 1996).

To asymptotically approximate $E_\varphi\widehat{\varphi}_T$, one could assume that $\widehat{\varphi}_T$ can be represented as a ratio of quadratic forms in normal (see Magnus, 1986) or non-normal (see Ullah and Srivastava, 1994) random variables and consequently, an asymptotic approximation of the expectation is given by the subsequent integral. Examples are provided by Phillips (2012), and Bao and Ullah (2007b) in the context of maximum likelihood estimator. Another procedure is to employ expansions of $\widehat{\varphi}_T$, as in e.g. Bao and Ullah (2007a), MacKinnon and Smith (1998), Newey and Smith (2004), or Rilstone *et al.* (1996), and then by employing Nagar (1956) type arguments (see Rothenberg, 1984), to approximate $E_\varphi\widehat{\varphi}_T$ by the expectation of the expansion. In fact, this is the way we follow in this paper, i.e. $E_\varphi\widehat{\varphi}_T$ is approximated by the expectation of the Edgeworth-B expansion of $\widehat{\varphi}_T$.

The approximation of $E_\varphi\widehat{\varphi}_T$ by Edgeworth-B means can yield analytically intractable functions of the parameter vector φ as (some of) the corresponding coefficients may depend on nuisance parameters, analytically intractable moments etc., especially for the QML estimator. In order to deal with such cases, we approximate them by employing the equivalent quantities using the standardized residuals. To this end, a Monte Carlo exercise is conducted and the results are presented and discussed. We provide two types of the bias correction mechanism, i.e. the first- and the full-step bias corrected, as they have been discribed above, in order to decide for the bias reduction in practice for the popular model of Nelson. Furthermore, we employ the residual bootstrap to approximately bias correct the (Q)ML estimators, so that we end up we three available approximations. We then compare the three approximately bias corrected estimators in terms of bias and also mean squared errors (MSE). This can be seen as the first time where analytically[2] the higher order biases appear in the GARCH literature

[2]Fiorentini, Calzolari and Panattoni (1996) argue that the computation of analytic derivatives of the log-likelihood is essential, as the computational benefit of their use is really substantial for estimation purposes.

for a nonlinear model like the EGARCH one and these results can now be used as to be incorporated into the relative analysis of other similar specifications, see e.g. Linton (1997), Iglesias and Linton (2007), Iglesias and Phillips (2012). We conclude that, for given sets of parameters values, the three approximately bias corrected estimators perform satisfactory. However, the full-step bias corrected one has relative advantages over the other two.

Notice also that one can employ these type of expansions, the Edgeworth-B distributions, to construct testing procedures or confidence sets with possibly better refined asymptotic properties that hold again independent of the choice of local alternatives (see e.g. Hoque *et al.*, 2006, Phillips and Park, 1988, and Rothenberg, 1984).

The plan of the paper is as follows. The EGARCH model and its estimation are presented in Section 2. The main results and our contributions are given in Sections 3-5. First, analytic derivatives and their expected values are presented. Second, conditions for stationarity of the log-variance derivatives are investigated. In the sequel, the theoretical bias approximations of ML and QML estimators are calculated and the simulation results for the bias correction of the estimators are presented. Finally, conclusions can be found in Section 6. Proofs are collected in the Appendix.

2 The model and estimators

We consider the following model, where the observed data $\{y_t\}$ are generated by the first-order EGARCH process of Nelson (1991), i.e. EGARCH$(1,1)$, in which the conditional variance of $\{u_t\}$, denoted by $\{h_t\}$, depends on both the size and the sign of the lagged residuals, i.e.

$$y_t = \mu + u_t = \mu + z_t\sqrt{h_t}, \quad t \in \mathbb{Z} \tag{2}$$

$$\ln(h_t) = \alpha + \theta z_{t-1} + \gamma |z_{t-1}| + \beta \ln(h_{t-1}), \tag{3}$$

where $\alpha \in \mathbb{R}$, $|\beta| < 1$, and $\{z_t\}$ is independently and identically distributed (i.i.d.) with zero mean and unit variance, allowing for the possibility of nonnormality in the conditional distribution of y_t.

The parameters $\theta, \gamma \in \mathbb{R}$ are for the asymmetries that the model captures. That is, θ shows the effect of the sign of u_t and, if it is negative, θ is the leverage effect parameter.

On the other hand, the term $\gamma |z_t|$ represents a magnitude effect and consequently, γ is the coefficient of the magnitude effect. Hence, the coefficients $(\theta + \gamma)$ and $(\theta - \gamma)$ (if $z_t \geq 0$ and $z_t < 0$, respectively) show the asymmetry in response to positive and negative u_t. Both terms together imply that $\theta z + \gamma |z| \geq 0$, for all $z \in \mathbb{R}$ (see Straumann, 2005) and it makes sense to impause the inequality $\gamma \geq |\theta|$, i.e. h_t as a function of z_{t-1} should be non-increasing on the negative real line and nondecreasing on the positive real line. It is worth mentioning that θ is expected to be negative to incorporate the negative correlation between current shocks and future conditional variance, the well known levarage effect in the stock market returns[3]. Furthermore, for financial data one expects that $0 < \beta < 1$, to incorporate the volatility clustering. However, for other data sets this is not the case (see e.g. Arvanitis and Demos, 2004, and McAleer *et al.*, 2007). Finally, one could consider the parameter α to be function of time, i.e. to have α_t, in order to accommodate the effect of any non-trading periods of forecastable effects (as in Nelson, 1991).

Note from (3) that $\ln(h_t)$ constitutes a causal AR(1) process with mean $\alpha/(1-\beta)$ and error sequence $\theta z_{t-1} + \gamma |z_{t-1}|$. The unique stationary solution to (3), provided that $|\beta| < 1$, is given by its almost sure (a.s.) representation of an MA(∞), that is

$$\ln(h_t) = \alpha (1-\beta)^{-1} + \sum_{k=0}^{\infty} \beta^k \left(\theta z_{t-1-k} + \gamma |z_{t-1-k}|\right),$$

which further implies that the log-volatility is lower bounded as

$$\ln(h_t) \geq (\alpha + \gamma E|z_t|)(1-\beta)^{-1} \quad \text{a.s.}$$

The unconditional mean and variance of y_t are given by

$$E(y_t) = \mu,$$

and

$$Var(y_t) = \exp\left(\frac{\alpha}{1-\beta}\right) \prod_{i=0}^{\infty} E\left[\exp\left(\beta^i \left(\theta z_0 + \gamma |z_0|\right)\right)\right].$$

From Theorem 2 in He *et al.* (2002) and Proposition 1 in Demos (2002) we get that, under normality of the errors, the variance becomes

$$Var(y_t) = \exp\left(\frac{\alpha - \gamma\sqrt{\frac{2}{\pi}}}{1-\beta}\right) \prod_{i=0}^{\infty} \left[\exp\left(\frac{\beta^{2i}(\gamma^*)^2}{2}\right)\Phi\left(\beta^i \gamma^*\right) + \exp\left(\frac{\beta^{2i}\delta^2}{2}\right)\Phi\left(\beta^i \delta\right)\right],$$

[3] For an alternative definition of the leverage and asymmetry effect see McAleer (2014).

where $\gamma^* = \gamma + \theta$, $\delta = \gamma - \theta$ and $\Phi(k)$ is the value of the cumulative standard Normal distribution evaluated at k, i.e. $\Phi(k) = \int_{-\infty}^{k} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$.

By far, the most popular estimation method to estimate the parameters of an EGARCH model is the Maximum Likelihood (ML), assuming that the standardized errors z_t are independently and identically distributed (i.i.d.) standard normal random variables. However, for most applications, it is difficult to justify the normality assumption, and consequently the likelihood function may be misspecified. However, one can still obtain estimates by maximizing a Gaussian log-likelihood function even if the normality assumption is violated. The derived estimators are the so-called Gaussian Quasi Maximum Likelihood ones (QMLEs). This is the method employed here and is justified by empirical evidence that conditional distributions of asset returns are often thick tailed.

An important and interesting feature of the EGARCH model is that the assumption of the block diagonality of the information matrix no longer holds, even if the distribution of the standardized errors is symmetric. This is also the case for the ARCH-in-Mean model and the asymmetric Augmented ARCH model, see Bera and Higgins (1993), p. 34 and also Bollerslev *et al.* (1994), p. 2981. This fact implies that the off-diagonal blocks involving partial derivatives with respect to both mean and variance parameters are not null matrices, while this is the case in other GARCH-type models. Below we present analytic proofs of this argument in the context of the EGARCH$(1,1)$ model and these results disaccord with Malmsten (2004), even if the distribution of the innovations is symmetric, which implies that $Ez^3 = 0$.

In the EGARCH(1,1) model, the exact conditional log-likelihood function is not known, as it depends on unobserved initial values z_0 and h_0. Here we assume that $z_0 = 0$ and $\ln(h_0) = \frac{\alpha}{1-\beta}$ and consequently, we maximize an approximation of the Gaussian log-likelihood. Under the appropriate conditions, the filtered conditional variances converge to the true ones at an exponentially fast rate. This is the base for the almost sure convergence of the QML estimator (see Straumann, 2005, and Straumann and Mikosch, 2006). However, these conditions are difficult to check and almost imposible to impose. Hence, the approximate average conditional Gaussian log-likelihood function is given as follows

$$\ell_T\left(\alpha,\theta,\gamma,\beta,\mu\right| z_0,h_0) = -\frac{1}{2}\ln\left(2\pi\right) - \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left[\ln\left(h_t\right) + \frac{\left(y_t-\mu\right)^2}{h_t}\right]. \tag{4}$$

Notice that h_t and z_t are both functions of the parameters. We denote the parameter vector by φ, i.e. $\varphi = (\alpha,\theta,\gamma,\beta,\mu)'$. The first order conditions are recursive and consequently do not have explicit solutions.

The likelihood function is derived as though the errors are conditionally normal and is still maximized at the true parameters (see Straumann, 2005). Having specified the log-likelihood function, the QML estimator of φ is then defined as any measurable solution of

$$\widehat{\varphi}_T = \arg\max_{\varphi\in\Theta} \ell_T\left(\varphi\right). \tag{5}$$

The parameter space is of the form

$$\Theta = \mathbb{R}\times D\times[0,1)\times\mathbb{R},$$

where

$$D = \left\{(\theta,\gamma)' \in \mathbb{R}^2 \mid \theta\in\mathbb{R},\ \gamma\geq|\theta|\right\}.$$

3 Stationarity of the log-variance derivatives

In this section we investigate under which conditions the log-variance derivatives are stationary, needed for the existence and the evaluation of the log-likelihood derivatives, and hence in order to calculate the bias expressions of the QMLEs. The uniqueness of stationary and ergodic derivatives of the conditional variance is important when studying the validity of a Taylor series expansion of the first order log-likelihood derivatives. For the next Proposition we make the following assumption:

Assumption 1 *let us denote z_t by z, for simplicity. We assume that the following conditions hold true.*

1. $\left|\beta - \frac{1}{2}\gamma E\left|z\right|\right| < 1$

2. $\left|\beta^2 + \frac{1}{4}\theta^2 + \frac{1}{4}\gamma^2 - \gamma\beta E\left|z\right| + \frac{1}{2}\gamma\theta E\left(z\left|z\right|\right)\right| < 1$

3. $\left| \begin{array}{c} \beta^3 + \frac{3}{4}\beta\theta^2 + \frac{3}{4}\beta\gamma^2 - \frac{1}{8}\theta\left(\theta^2+3\gamma^2\right)E\left(z^3\right) - \frac{3}{2}\beta^2\gamma E\left|z\right| \\ +\frac{3}{2}\beta\theta\gamma E\left(z\left|z\right|\right) - \frac{1}{8}\gamma\left(\gamma^2+3\theta^2\right)E\left|z\right|^3 \end{array} \right| < 1$

Proposition 1 *If Assumption 1 is satisfied, then the second-order stationarity of all log-variance derivatives follows.*

Proof. The proof comes from the analytic results of the log-variance derivatives that are presented in the supplemented Technical Appendix (TA)[4]. ■

As we have already mentioned, it is reasonable to believe that γ and β are positive, at least for data that exhibit volatility clustering. Consequently, the conditions in the above proposition may seem to be restrictive, at least at a first glance. However, we might consider the following example that sheds some light on those conditions. Consider $(\beta, \gamma, \theta)' = (0.9, 0.7, -0.4)'$, which is the first set of parameters that we consider in the simulations section. Then, under the assumption of normality of errors we get that the conditions of Assumption 1 take the values of: 1. 0.621, 2. 0.392, and 3. 0.354, where for $(\beta, \gamma, \theta)' = (0.9, 0.6, -0.2)'$ the condition values are: 1. 0.661, 2. 0.422, and 3. 0.324.

Let us now illustrate the content of this section by considering the following example, where this notation applies: $h_{t;\circ} = \frac{\partial \ln(h_t)}{\partial \circ}$, $h_{t;\circ,\circ} = \frac{\partial^2 \ln(h_t)}{\partial \circ \partial \circ}$ etc.

$$\begin{aligned} h_{t;\alpha}h_{t;\alpha\alpha} &= \frac{1}{4}\left(\theta z_{t-1} + \gamma\left|z_{t-1}\right|\right)h_{t-1;\alpha}^2 + \frac{1}{4}\left(\theta z_{t-1} + \gamma\left|z_{t-1}\right|\right)\left(\beta - \frac{1}{2}\theta z_{t-1} - \frac{1}{2}\gamma\left|z_{t-1}\right|\right)h_{t-1;\alpha}^3 \\ &+ \left(\beta - \frac{1}{2}\theta z_{t-1} - \frac{1}{2}\gamma\left|z_{t-1}\right|\right)h_{t-1;\alpha,\alpha} \\ &+ \left(\beta - \frac{1}{2}\theta z_{t-1} - \frac{1}{2}\gamma\left|z_{t-1}\right|\right)^2 h_{t-1;\alpha}h_{t-1;\alpha,\alpha}. \end{aligned} \tag{6}$$

In order to calculate the expected value of the above expression, we first assume that $E\left(h_{t;\alpha}^2\right)$, $E\left(h_{t;\alpha}^3\right)$, and $E\left(h_{t;\alpha,\alpha}\right)$ exist. Next, define

$$\begin{aligned} A\left(z_{t-1}\right) &= \frac{1}{4}\left(\theta z_{t-1} + \gamma\left|z_{t-1}\right|\right)h_{t-1;\alpha}^2 \\ &+\frac{1}{4}\left(\theta z_{t-1} + \gamma\left|z_{t-1}\right|\right)\left(\beta - \frac{1}{2}\theta z_{t-1} - \frac{1}{2}\gamma\left|z_{t-1}\right|\right)h_{t-1;\alpha}^3 \\ &+\left(\beta - \frac{1}{2}\theta z_{t-1} - \frac{1}{2}\gamma\left|z_{t-1}\right|\right)h_{t-1;\alpha,\alpha}, \end{aligned}$$

[4] This is available from the authors if any reader is interested in those results.

and

$$B^2\left(z_{t-1}\right)=\left(\beta-\frac{1}{2}\theta z_{t-1}-\frac{1}{2}\gamma\left|z_{t-1}\right|\right)^2.$$

Then,

$$\begin{aligned} h_{t;\alpha}h_{t;\alpha\alpha} &= A\left(z_{t-1}\right)+B^2\left(z_{t-1}\right)h_{t-1;\alpha}h_{t-1;\alpha,\alpha}= \\ &= A\left(z_{t-1}\right)+\sum_{k=1}^{\infty}\prod_{i=0}^{k-1}B^2\left(z_{t-1-i}\right)A\left(z_{t-1-k}\right). \end{aligned}$$

The infinite sum converges almost surely. To see this, let

$$S_n=A\left(z_{t-1}\right)+\sum_{k=1}^{n}\prod_{i=0}^{k-1}B^2\left(z_{t-1-i}\right)A\left(z_{t-1-k}\right).$$

Then we have

$$\begin{aligned} E\left(S_n\right) &= E\left[A\left(z_{t-1}\right)\right]+\sum_{k=1}^{n}E\left[\prod_{i=0}^{k-1}B^2\left(z_{t-1-i}\right)\right]E\left[A\left(z_{t-1-k}\right)\right] \\ &= E\left[A\left(z_{t-1}\right)\right]\left[\sum_{k=0}^{n}\left\{E\left[B^2\left(z_{t-1-i}\right)\right]\right\}^k\right]. \end{aligned}$$

Thus, $E\left(\lim_{n\to\infty}S_n\right)=E\left[A\left(z_{t-1}\right)\right]\left\{1-E\left[B^2\left(z_{t-1-i}\right)\right]\right\}^{-1}<\infty$, providing that $E\left[A\left(z_{t-1}\right)\right]<\infty$. In order to ensure the existence of a stationary solution to the (6), we should impose the condition that

$$E\left[B^2\left(z_{t-1-i}\right)\right]<1.$$

In a similar manner, the rest stationarity conditions of all log-variance derivatives and products of them follow.

4 Finite sample properties and bias approximations

In this section we develop the bias approximations for ML and QML estimators in the EGARCH$(1,1)$ model[5]. One of the main advantages of developing the bias expressions is to employ them as a bias correction mechanism. This is one of the practical applications of the bias approximations. Moreover, these results help to analyse the consequences of introducing restrictions in the log-variance parameters. With these expressions, one can also

[5]Iglesias and Phillips (2002) developed theoretical bias approximations for the MLEs of the parameters in an ARCH(1) model.

compute the Edgeworth approximate distribution. It is also important to explore the theoretical properties of the estimators so that statistical inference can be possible.

We employ a McCullagh (1986) result for the standardized estimator having a stochastic expansion and taking expectations we end up with the asymptotic bias of the QML estimator. Our next step is to check our bias approximations through simulations. Note that McCullagh's expansion has already been applied in the literature to retrieve the bias in many nonlinear models, such as Linton (1997) and Iglesias and Phillips (2012). When dealing with nonlinear models, it is very common to have the bias expressions in terms of expectations and applying these expressions for bias correction. At this point, it is important to state briefly the main differences between our analysis and that of Linton (1997). We generalize the finite-sample analysis of heteroskedastic time series models considering a non-symmetric distribution of the errors. Furthermore, we show that the block-diagonality of the information matrix does not hold in our case, which implies that new terms appear in the bias expressions of the estimators. This means that we cannot use the results that appear in the literature from the analysis of the classical GARCH model, since our case seems not to be a direct extension.

Notation 1 *In what follows, for* $i,j,k=\{\alpha,\theta,\gamma,\beta,\mu\}$*,*

$$\begin{aligned}\tau_i &= \frac{1}{T}\sum_{t=1}^{T}E\left(h_{t;i}\right), \quad \tau_{i,j}=\frac{1}{T}\sum_{t=1}^{T}E\left(h_{t;i}h_{t;j}\right), \quad \tau_{ij,k}=\frac{1}{T}\sum_{t=1}^{T}E\left(h_{t;ij}h_{t;k}\right),\\ \tau_{i,j,k} &= \frac{1}{T}\sum_{t=1}^{T}E\left(h_{t;i}h_{t;j}h_{t;k}\right).\end{aligned}$$

Also,

$$\overline{\pi}=\frac{1}{T}\sum_{t=1}^{T}E\left(\frac{1}{h_t}\right), \text{ and } \overline{\pi_i}=\frac{1}{T}\sum_{t=1}^{T}E\left(\frac{1}{h_t}h_{t;i}\right).$$

Assumption 2 *We assume that the errors have bounded* J^{th} *moments, for some* $J>6$*, and we denote by* κ_3 *and* κ_4 *their third and fourth order cumulants, where the latter is given by:*

$$\kappa_4=E\left(z_t^4-3\right).$$

Under the Assumption 2, we are now able to present our next Theorem, which is useful for the evaluation of the bias approximations of all estimators and also to construct the Edgeworth expansions in this setting. This result may be viewed as generalization of that in

Linton (1997). Valid Edgeworth expansions in the case of the GARCH model are established by Corradi and Iglesias (2008). The next Theorem uses the same notation as in Linton (1997), i.e. $\mathcal{L}_i$, $\mathcal{L}_{ij}$, and $\mathcal{L}_{ijk}$ denote the derivatives with respect to i, j, k of the log-likelihood $\ell_T(\varphi)$.

Theorem 1 *Given that $z_t \sim iid(0,1)$ and non-symmetric, and for $i,j,k = \{\alpha, \theta, \gamma, \beta, \mu\}$, unless the parameter μ is used separately to underline the difference, the following moments of the log-likelihood derivatives (see Appendix A for analytical expressions) converge to finite limits as $T \to \infty$:*

$$c_{ij} = \tfrac{1}{T}E(\mathcal{L}_{ij}) = -\tfrac{1}{2}\tau_{i,j},$$

$$c_{ijk} = \tfrac{1}{T}E(\mathcal{L}_{ijk}) = -\tfrac{1}{2}(\tau_{ij,k} + \tau_{ik,j} + \tau_{jk,i} - \tau_{i,j,k}),$$

$$c_{ij,k} = \tfrac{1}{T}E(\mathcal{L}_{ij}\mathcal{L}_k) = -\tfrac{1}{4}\left[\tau^{zz}_{k;i,j} - (\kappa_4+2)(\tau_{ij,k} - \tau_{i,j,k})\right],$$

$$c_{\mu\mu} = \tfrac{1}{T}E(\mathcal{L}_{\mu\mu}) = -\left(\overline{\pi} + \tfrac{\tau_{\mu,\mu}}{2}\right),$$

$$c_{i\mu\mu} = \tfrac{1}{T}E(\mathcal{L}_{i\mu\mu}) = \overline{\pi_i} - \tfrac{1}{2}(\tau_{i,\mu\mu} + 2\tau_{\mu i,\mu} - \tau_{\mu,i,\mu}),$$

$$c_{\mu\mu\mu} = \tfrac{1}{T}E(\mathcal{L}_{\mu\mu\mu}) = -\tfrac{1}{2}\left(3\tau_{\mu\mu,\mu} - \tau^3_{\mu}\right) + 3\overline{\pi_{\mu}},$$

$$c_{i\mu,\mu} = \tfrac{1}{T}E(\mathcal{L}_{i\mu}\mathcal{L}_{\mu}) = -\tfrac{1}{4}\left\{\begin{array}{c} 4\overline{\pi_i} - (\kappa_4+2)(\tau_{i\mu,\mu} - \tau_{i,\mu,\mu}) \\ +\tau^{zz}_{\mu;i\mu} + 2\tau^{zh}_{i,\mu} + 2\kappa_3\left(2\tau^h_{i,\mu} - \tau^h_{i\mu}\right) \end{array}\right\},$$

$$c_{i\mu,j} = \tfrac{1}{T}E(\mathcal{L}_{i\mu}\mathcal{L}_j) = -\tfrac{1}{4}\left\{-(\kappa_4+2)(\tau_{i\mu,j} - \tau_{i,j,\mu}) + \tau^{zz}_{\mu;i\mu} + 2\kappa_3\tau^h_{ij}\right\},$$

$$c_{\mu\mu,i} = \tfrac{1}{T}E(\mathcal{L}_{\mu\mu}\mathcal{L}_i) = -\tfrac{1}{4}\left\{-(\kappa_4+2)(\tau_{\mu\mu,i} - \tau_{i,\mu,\mu}) + \tau^{zz}_{i;\mu\mu} + 4\kappa_3\tau^h_{i,\mu}\right\},$$

$$c_{ij,\mu} = \tfrac{1}{T}E(\mathcal{L}_{ij}\mathcal{L}_{\mu}) = -\tfrac{1}{4}\left\{\begin{array}{c} -(\kappa_4+2)(\tau_{ij,\mu} - \tau_{i,j,\mu}) + \tau^{zz}_{\mu;ij} + 2\tau^{zh}_{i,j} \\ +2\kappa_3\left(2\tau^h_{i,j} - \tau^h_{ij}\right) \end{array}\right\},$$

$$c_{\mu\mu,\mu} = \tfrac{1}{T}E(\mathcal{L}_{\mu\mu}\mathcal{L}_{\mu}) = -\tfrac{1}{4}\left\{\begin{array}{c} 8\overline{\pi_{\mu}} - (\kappa_4+2)(\tau_{\mu\mu,\mu} - \tau_{\mu,\mu,\mu}) \\ +\tau^{zz}_{\mu;\mu\mu} + 2\tau^{zh}_{\mu,\mu} + 2\kappa_3\left(3\tau^h_{\mu,\mu} - \tau^h_{\mu\mu}\right) \end{array}\right\},$$

where

$$\begin{aligned} \tau^{zz}_{k;i,j} &= \frac{1}{T}\sum_{s<t}\sum E\left[\left(z_s^2 - 1\right)h_{s;k}h_{t;i}h_{t;j}\right], \quad \tau^{zh}_{i,j} = \frac{1}{T}\sum_{s<t}\sum E\left(z_s\frac{1}{\sqrt{h_t}}h_{t;i}h_{t;j}\right), \\ \tau^h_{i,\mu} &= \frac{1}{T}\sum_{t=1}^{T} E\left(\frac{1}{\sqrt{h_t}}h_{t;i}h_{t;\mu}\right) \text{ and } \tau^h_{i\mu} = \frac{1}{T}\sum_{t=1}^{T} E\left(\frac{1}{\sqrt{h_t}}h_{t;i,\mu}\right). \end{aligned}$$

Proof. The proof comes immediately from the auxilliary Lemma 1 (in Appendix C) and the results of Appendix B. ■

In the following, we make use of the summation convention mechanism[6], that is $c^{ij}Z_j = \sum_j c^{ij}Z_j$, in which repeated indices in the expression are to be summed over. Note that c^{ij} is the i,j-element of the matrix inverse of $\{c_{ij}\}$. Let us first consider the case when the mean parameter is supposed to be equal to zero and not estimated. With techniques of McCullagh (1986), the standardized estimators, derived from choosing φ to solve $\mathcal{L}_i(\varphi) = 0$, for $i = \{\alpha, \theta, \gamma, \beta\}$, have the following stochastic expansion

$$\sqrt{T}\left\{\widehat{\varphi}_{T,i}(\mu) - \varphi_i\right\} = -c^{ij}Z_j + \frac{1}{\sqrt{T}}\left\{c^{ij}c^{kl}Z_{jk}Z_l - c^{ij}c^{kl}c^{mn}c_{j\,\mathrm{ln}}Z_kZ_m/2\right\} + O_P\left(T^{-1}\right), \quad (7)$$

where

$$Z_j = T^{-1/2}\mathcal{L}_j$$

and

$$Z_{jk} = T^{-1/2}\left\{\mathcal{L}_{jk} - E\left(\mathcal{L}_{jk}\right)\right\}$$

are evaluated at the true parameters and are jointly asymptotically normal. Raising pairs of indices signifies components from the matrix inversion. In order to derive the bias approximations, we need to find expressions for their following components: c^{ij}, c_{ijk} and $c_{jk,l}$.

Taking expectations of the right-hand side in (7), and up to order $O\left(T^{-1/2}\right)$, we get

$$E\left[\sqrt{T}\upsilon'\left\{\widehat{\varphi}_T(\mu) - \varphi\right\}\right] = \frac{1}{\sqrt{T}}\upsilon_i c^{ij}c^{kl}\left\{c_{jk,l} + c_{jkl}\left(\kappa_4 + 2\right)/4\right\},$$

where υ is the 4×1 choice parameter vector, i.e. if one wants the expectation for only the parameter γ then $\upsilon = (0,0,1,0,0)'$. It also allows to evaluate the expectation for a linear combination of the parameters of interest, e.g. if one is interested in the expectation of $\gamma + 2\theta$ then $\upsilon = (0,2,1,0,0)'$. If $\kappa_4 = 0$, QML equals ML and then the above formula equals the one of Cox and Snell (1968), i.e.

$$E\left[\sqrt{T}\upsilon'\left\{\widehat{\varphi}_T(\mu) - \varphi\right\}\right] = \frac{1}{\sqrt{T}}\upsilon_i c^{ij}c^{kl}\left\{c_{jk,l} + \frac{1}{2}c_{jkl}\right\}.$$

[6] This is a technique for simplifying expressions including summations of vectors, matrices, and general tensors.

Let us now consider the other case, where the mean parameter is unknown and estimated, i.e. $i = \{\alpha, \theta, \gamma, \beta, \mu\}$. Hence, if we incorporate the effects of estimating μ, the up to order $O_P\left(T^{-1/2}\right)$ stochastic expansions now take the following form

$$\sqrt{T}\left\{\widehat{\varphi}_{T,i} - \varphi_i\right\} - \sqrt{T}\left\{\widehat{\varphi}_{T,i}(\mu) - \varphi_i\right\} = \frac{1}{\sqrt{T}}\left\{c^{ij}c^{kl}Z_{jk}Z_l - c^{ij}c^{kl}c^{mn}c_{j\ln}Z_k Z_m/2\right\},$$

where now $i, j, k, l = \{\alpha, \theta, \gamma, \beta, \mu\}$. Taking expectations of the right-hand side, we find the asymptotic bias of the estimators.

Tables 1 and 2 show the evolution of the bias in absolute value of all parameters, for known values of $(\alpha, \theta, \gamma, \beta)$ which are $\alpha = 0.1, \beta = 0.9, \gamma = 0.7, \theta = -0.4$, under the assumptions of normality and mixture of normals, respectively. The results were checked through Monte Carlo simulations with 5000 replications. By far the constant α presents the highest bias. Further, the QMLEs are more biased than the MLEs, especially for smaller sample sizes.

In terms of the mean squared error, from (7) we have up to $O\left(T^{-1}\right)$

$$E\left[\sqrt{T}\upsilon'\left\{\widehat{\varphi}_T - \varphi\right\}\right]^2 = -\upsilon_i c^{ij}\left(\kappa_4 + 2\right)/2, \tag{8}$$

which is the asymptotic variance. If we let the remainder to be of order $O\left(T^{-3/2}\right)$, then the mean squared error is again evaluated by (8), with the difference now that there would be added terms of order $O\left(T^{-1}\right)$. Of course, as $T \to \infty$, the mean squared error approaches the asymptotic variance.

5 Simulations

In this section we make a simulation exercise in order to check the adequacy of our theoretical results and be able to proceed with the bias correction of the estimators. We draw random samples of size $500, 750, 1500, 3000, 5000, 10000, 25000, 50000$ observations and 500 observations for initialization, under first the assumption of normality and second the assumption of mixture of normals. We make 5000 replications for all sample sizes. The mean parameter μ is supposed to be equal to zero and hence is not estimated, so the parameter vector is $\varphi = (\alpha, \theta, \gamma, \beta)'$. We check the performance of the bias correction mechanism for different sets of parameter values and we will present the results for three sets, i.e. $(0.1, -0.4, 0.7, 0.9)'$,

$(-0.1, -0.2, 0.6, 0.9)'$ and $(0.5, -0.5, 0.8, 0.5)'$. The first two sets include values for the parameters that are close to what is observed from financial data. We multiply the bias by T and not $\sqrt{T}$, i.e. $E\left(T\left\|\widehat{\varphi}_T - \varphi\right\|\right)$, as in this way we keep a constant term in the bias expressions that is important to distinguish what happens when we increase the sample size, as the next terms in the expressions will tend to zero, as $T \rightarrow \infty$.

The bias correction mechanism is constructed, first, under the specification of two methods. The first one, called first-step correction, is the classical one, in which we estimate the model and we retrieve the estimated parameters. Next, we compute the bias expressions by using the estimates and we are then able to correct the bias of the estimators with the corresponding values of the bias, i.e.

$$\widetilde{\varphi}_T = \widehat{\varphi}_T - \frac{1}{T} bias\left(\widehat{\varphi}_T\right),$$

where $bias\left(\widehat{\varphi}_T\right)$ is the $\frac{1}{T}$ term in the expansion of $E\left(\widehat{\varphi}_T\right)$ evaluated at $\widehat{\varphi}_T$, i.e. all c^{ij}, $c_{ij,k}$, and c_{ijk}, *for* $i, j, k = \{\alpha, \theta, \gamma, \beta\}$, are evaluated at $\widehat{\varphi}_T$ and the nuisance parameters are estimated from the standardized residuals. Further, notice that there is nothing to prevent the case of $\widetilde{\varphi}_T$ being outside the admissible area (see also Linton, 1997 as well as Iglesias and Linton, 2007). This will affect both the estimated bias and the MSE of the first-step bias corrected estimator.

The second method that we employ, called full-step correction, is a method proposed by Arvanitis and Demos (2015), in which we solve the optimization problem in (1), and the full-step correction estimator, denoted by $\widehat{\varphi}_T^F$, is given by

$$\widehat{\varphi}_T^F = \arg\min_{\varphi \in \Theta} \left\| \widehat{\varphi}_T - \varphi - \frac{1}{T} bias\left(\varphi\right) \right\|,$$

where $bias\left(\varphi\right)$ is the $\frac{1}{T}$ term in the expansion of $E\left(\widehat{\varphi}_T\right)$ as a function of φ and the nuisance parameters are, again, estimated from the standardized residuals. In this respect, this method is a multi-step maximization procedure, using numerical derivatives. This justifies the name of the first method, which is the first step of the multi-step optimization problem. In this way, the second method incorporates the constraints that are imposed on the coefficients and as a consequence the corrected estimate of the EGARCH parameters cannot lie outside the admissible region, i.e. the corrected β will be less than one in absolute value.

Furthermore, as a third bias corrected estimator we evaluate the bootstrapped one[7]. For each Monte Carlo experiment we perform $H = 5000$ bootstraps to the standardized residuals, estimated employing the (Q)MLE. For each bootstrap sample we then evaluate the (Q)MLEs and form the bias corrected bootstraped estimator as:

$$\widehat{\varphi}_T^B = 2\widehat{\varphi}_T - \frac{1}{H}\sum_{i=1}^{H} \widehat{\varphi}_T^i.$$

Figures 1 and 2 represent the bias correction performance under the normality assumption, of the first two sets of parameters. To conserve space, the results for the third set of parameters is not presented, as the are qualitatively the same. In both figures we present $\frac{T}{S}\sum_{i=1}^{S} \left\|\widehat{\varphi}_{T,i} - \varphi\right\|$, an estimator of $E\left(T\left\|\widehat{\varphi}_T - \varphi\right\|\right)$, for $S = 5000$, the number of Monte Carlo experiments and $\widehat{\varphi}_{T,i}$ any of the four estimators, i.e. the MLE $(\widehat{\varphi}_T)$, the first-step bias corrected $(\widetilde{\varphi}_T)$, the full-step bias corrected $\left(\widehat{\varphi}_T^F\right)$, and the bootstrap corrected one $\left(\widehat{\varphi}_T^B\right)$. For the first set of parameter values (Figure 1) we see that the bias correction works in all cases and the approximate corrected bias estimators have estimated biases almost 25% of the estimated MLE. The same applies for the second set of parameters (Figure 2). Hence, we can conclude that all three approximate bias corrected estimators, i.e. first-step, full-step, and bootstrap corrected, reduce the bias of the MLE by 75%, at least for the set of parameters considered here. It is worth mentioning that the estimated biases of the MLE is close to the theoretical ones, i.e. for the first set of parameter values the theoretical bias is 21.75, whereas for the second one is 19.86,

When dropping the normality assumption, we run the simulations under the hypothesis of mixture of normals for standardized random variables (see Figure 3 and Figure 4). In fact, the errors are drawn from a normal distribution with mean 0.01 and variance 9, with probability 0.1, and with probability 0.9 they are drawn from a normal distribution with mean -0.001 and variance 0.111. In this way, the theoretical mean and variance of the distribution are 0 and 1, respectively. Notice that with these hyperparameter values the theoretical skewness and kurtosis of the random errors are 0.0266 and 24.334 respectively, approximately matching the sample counterparts of most high frequency financial data.

From Figures 3 and 4 it seems that the estimated biases of the corrected estimators are

[7] For asymptotics, see Bickel and Freedman (1981).

almost 50% lower the ones of the QMLEs. The only exception is the first-step corrected estimator for the first set of parameter values. This is attributed to the fact that in more than 10 Monte Carlo cases, the bias corrected estimator of β is outside the admissible interval, i.e. the estimated values of β are greater than 1. Again, the estimated biases of the QMLEs are close to the theoretical ones which are 50.23 and 39.38, for the two set of parameter values, respectively.

It seems that the non-normality of the errors increases, almost doubles, the bias of the QMLEs, for both set of parameter values. in which we have selected different values of the coefficient β, i.e. low (0.5) and high (0.9). Figure 1 (under normality) and Figure 4 (under mixture of normals) are constructed under the same set of parameter values and it is interesting to compare between the two cases. As in the case of normality, we see that in Figure 4 the bias correction of the estimators works in most cases and the results are satisfactory. In Figure 3, the corrected bias is again under the bias of the MLEs, indicating that the theoretical results correct the bias, under the assumptions made.

In terms of mean squared error, in Figure 5 we present the estimated MSEs, multiplied by the number of observations T, of the four estimators for the first set of parameter values, whereas Figure 6 present the ones for the second set, both under normality. It seems that all estimators have the same, more or less, MSEs.

Dropping normality, it seems that the estimated MSEs are double, as compared to the ones under normality, for both set of parameters values (see Figures 7 and 8). The exception is again the MSE of the first-step bias corrected one, due to the few cases over correction for the parameter β.

In all, it seems that, in terms of bias and MSE, the full-step and bootstrap approximate bias corrected estimators perform better than the classical first-step one, with litle difference between them. However, there is a big difference between the two in terms of time efficiency. Table 3 presents the average CPU time per Monte Carlo iteration for the first set of parameters, under normality[8]. It is obvious that the procedure to get the full-step approximate bias corrected estimator is from 178 to 293 times faster than the boostrap one. Of course,

[8]To solve the optimization problems the same routine is employed, i.e. the NAG routine EO4JBQ in Fortran.

one could improve the time performance of the boostrap estimator by decreasing the number of bootstraps. This, however, would increase the bootstrap error and thereby increasing the MSE of the bootstrap corrected estimator. To investigate the optimal number of bootstraps, in terms of time-MSE, is outside the scope of the present paper.

6 Conclusions

In this paper we study the finite sample properties of ML and QML estimators in the EGARCH$(1,1)$ model of Nelson (1991). We present analytic derivatives of both the log-likelihood and the log-variance functions. We further develop theoretical bias approximations for the estimators of the model parameters, up tp order $O\left(T^{-1}\right)$, and we derive conditions for the second-order stationarity of the log-variance derivatives. We employ the provided formulae to produce approximate bias corrected estimators, for both ML and QML estimators. As these estimators are solutions of multistep minimization procedures are called full-step bias corrected estimators. In fact these estimators are Indirect Inference ones, where the binding function is the identity.

In a simulation exercise, we compare the performance of the suggested approximate bias corrected estimator with the commonly employed "feasibly biased corrected" one, called here first-step, as it is an one-computational step approximation of the suggested estimator and the bootstrap corrected one. The results suggest that the full-step estimator avoids the overcorrection of the first-step one. Further, its bias parformance is more or less the same as the bootstrap corrected one and they share the same MSEs. However, the suggested method is considerably faster than the bootstrap one.

An interesting idea for future research would be the investigation of necessary and sufficient conditions for the existence and validity of the Edgeworth approximations in this context. Also, one might consider the case of the EGARCH-in-Mean model and employ the results presented here. This model examines the relation between the level of market risk and required return and its estimation theory and asymptotic properties of the QMLE have been studied by Hafner and Kyriakopoulou (2017).

Acknowledgements

We are grateful to Stelios Arvanitis, Luc Bauwens, Gabriele Fiorentini, Christian Hafner, the participants at Spring Meeting of Young Economists, Istanbul, the 3rd International Conference on Computational and Financial Econometrics, Limassol, the Rimini Conference in Economics and Finance, Rimini, the 10th International Vilnius Conference on Probability Theory and Mathematical Statistics, Vilnius, the 9th Conference on Research on Economic Theory & Econometrics (CRETE), Tinos, the Conference in Honour of Sir David F. Hendry, St. Andrews, Scotland, the European Meeting of Statisticians (EMS), Piraeus, the 4th International Conference on Computational and Financial Econometrics (CFE), London and the seminar participants at University of Piraeus and Athens University of Economics and Business. Financial support from the Basic Research Funding Program (PEVE 2) is gratefully acknowledged. All remaining errors are solely our responsibility.

References

[1] Arvanitis, S., Demos, A. (2004). Time dependence and moments of a family of time-varying parameter GARCH in mean models. *Journal of Time Series Analysis* 25: 1-25.

[2] Arvanitis, S., Demos, A. (2015). A class of indirect inference estimators: higher-order asymptotics and approximate bias correction. *Econometrics Journal* 18: 200-241.

[3] Arvanitis, S., Demos, A. (2016). On the Validity of Edgeworth Expansions and Moment Approximations for Three Indirect Inference Estimators. *Journal of Econometric Methods*, forthcoming, DOI: https://doi.org/10.1515/jem-2015-0009.

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

[5] Bao, Y., Ullah, A. (2007b). Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics* 140: 396–413.

[6] Bera, A. K., Higgins, M. L. (1993). ARCH models: properties, estimation and testing. *Journal of Economic Surveys* 7: 305-362.

[7] P. J. Bickel, P. J., Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics* 9: 1196–1217.

[8] Bollerslev, T. (1986). A generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307-327.

[9] Bollerslev, T., Engle, R. F., Nelson, D. B. (1994). ARCH models. In R. F. Engle and D. L. McFadden (eds.), *The Handbook of Econometrics*, Volume 4.

[10] Comte, F., Lieberman, O. (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84: 61-84.

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

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

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

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

[15] Cox, D., Snell, E. (1968). A general definition of residuals. *Journal of the Royal Statistical Society, Series, B* 30: 248-275.

[16] Deb, P. (1996). Finite sample properties of maximum likelihood and quasi-maximum likelihood estimators of EGARCH models. *Econometric Reviews* 15: 51-68.

[17] Demos, A. (2002). Moments and dynamic structure of a time-varying parameter stochastic volatility in mean model. *Econometric Journal* 5: 345-357.

[18] Demos, A., Kyriakopoulou, D. (2013). Edgeworth and moment approximations: the case of MM and QML estimators for the MA(1) models. *Communications in Statistics: Theory and Methods* 42: 1713-1747.

[19] Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50: 987-1008.

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

[21] Fiorentini, G., Calzolari, G., Panattoni, L. (1996). Analytic derivatives and the computation of GARCH estimates. *Journal of Applied Econometrics* 11: 399-417.

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

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

[24] Hafner, C. M., Kyriakopoulou, D. (2017). Exponential-type GARCH models with linear-in-variance risk premium. *Working Paper.*

[25] Hafner, C. M., Linton, O. (2017). An almost closed form estimator for the EGARCH model. *Econometric Theory* 33: 1013-1038.

[26] Hall, P. (1992). *The bootstrap and Edgeworth expansion.* Springer-Verlag, New York.

[27] He, C., Terasvirta, T., Malmsten, H. (2002). Moment structure of a family of first-order exponential GARCH models. *Econometric Theory* 18: 868-885.

[28] Hoque, Z., Khan, S., Billah, B. (2006). Edgeworth size corrected W, LR and LM tests in the formation of the preliminary test estimator. *Journal of Statistical Research* 37: 43-55.

[29] Iglesias, E. M., Linton, O. (2007). Higher order asymptotic theory when a parameter is on a boundary with an application to GARCH models. *Econometric Theory* 23: 1136-1161.

[30] Iglesias, E. M., Phillips, G. D. A. (2002). Small sample estimation bias in ARCH models. Cardiff University.

[31] Iglesias, E. M., Phillips, G. D. A. (2011). Small sample estimation bias in GARCH models with any number of exogenous variables in the mean equation. *Econometric Reviews* 30: 303–336.

[32] Iglesias, E. M., Phillips, G. D. A. (2012). Estimation, testing, and finite sample properties of quasi-maximum likelihood estimators in GARCH-M models. *Econometric Reviews* 31: 532-557.

[33] Jeantheau, T. (1998). Strong consistency of estimators of multivariate ARCH models. *Econometric Theory* 14: 70-86.

[34] Jensen, S. T., Rahbek, A. (2004). Asymptotic normality of the QML estimator of ARCH in the nonstationary case. *Econometrica* 72: 641-646.

[35] Kyriakopoulou, D. (2015). Asymptotic normality of the QML estimator of the EGARCH$(1, 1)$ model, Working Paper.

[36] Lee, S-W., Hansen, B. E. (1994). Asymptotic theory for the GARCH(1,1) quasi maximum likelihood estimator. *Econometric Theory* 10: 29-52.

[37] Linton, O. (1997). An asymptotic expansion in the GARCH(1,1) model. *Econometric theory* 13: 558-581.

[38] Lumsdaine, R. L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64: 575-596.

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

[40] Magnus, J. R. (1986). The exact moments of a ratio of quadratic forms in normal variables. *Annales d'Economie et de Statistique* 4: 95-109.

[41] Malmsten, H. (2004). Evaluating exponential GARCH models. *SSE/EFI Working Paper Series in Economics and Finance*, no. 564, Stockholm School of Economics.

[42] McAleer, M. (2014). Asymmetry and leverage in conditional volatility models. *Econometrics* 2: 145-150.

[43] McAleer, M., Chan, F., Marinova, D. (2007). An econometric analysis of asymmetric volatility: theory and application to patents. *Journal of Econometrics* 139: 259-284.

[44] McCullagh, P. (1986). *Tensor methods in statistics.* Monographs on Statistics and Applied Probability, Chapman and Hall.

[45] Nagar, A. L. (1959). The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica* 27: 573-595.

[46] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59: 347-370.

[47] Newey, W. K., Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72: 219-255.

[48] Perez, A., Zaffaroni, P. (2008). Finite-sample properties of maximum likelihood and Whittle estimators in EGARCH and FIEGARCH models. *Quantitative and Qualitative Analysis in Social Sciences* 2: 78-97.

[49] Phillips, P.C.B. (2012). Folklore theorems, implicit maps and indirect inference. *Econometrica* 80: 425-454.

[50] Phillips, P. C. B., Park, J. Y. (1988). On the formulation of Wald tests of nonlinear restrictions. *Econometrica* 56: 1065-83.

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

[52] Rothenberg, T. J. (1984). Approximating the distributions of econometric estimators and test statistics. In Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics* Volume 2, New York, North-Holland.

[53] Straumann, D. (2005). *Estimation in Conditionally Heteroskedastic Time Series Models.* Lecture notes in Statistics, Springer.

[54] Straumann, D., Mikosch, T. (2006). Quasi-maximum likelihood estimation in conditionally heteroskedastic time series: a stochastic recurrence equations approach. *Annals of Statistics* 34: 2449-2495.

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

[56] Weiss, A. A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory* 2: 107-131.

[57] Wintenberger, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH(1,1) model. *Scandinavian Journal of Statistics* 40: 846-867.

[58] Zaffaroni, P. (2009). Whittle estimation of EGARCH and other exponential volatility models. *Journal of Econometrics* 151: 190-200.

Appendix A Analytic derivatives of the log-likelihood function

In this Appendix we present analytic derivatives of the log-likelihood function, which are needed to evaluate the asymptotic bias of the QMLEs and to calculate the cumulants of the Edgeworth distribution. It is of great importance to mention here that there are no such analytic results in the related literature of the finite sample theory of the conditional heteroskedastic models, and it is especially this feature that makes this analysis to differ from the previous one, that of Linton (1997), who studied the case of the GARCH$(1, 1)$ model. Let us proceed with the derivatives of the log-likelihood function and their analytic representation.

In the following, $h_{t;\circ}$ denotes the first derivative of the log-variance, $h_{t;\circ,\circ}$ the second derivative and so on. The derivatives of the log-likelihood function with respect to all parameters are stated below

$$
\begin{aligned}
\frac{\partial}{\partial \mu}\ell_T(\varphi) &\equiv \mathcal{L}_\mu = \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\mu} + \sum_{t=1}^{T}\frac{z_t}{\sqrt{h_t}}, \\
\frac{\partial^2}{\partial \mu^2}\ell_T(\varphi) &\equiv \mathcal{L}_{\mu\mu} = \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\mu,\mu} - \sum_{t=1}^{T}\left(\frac{1}{h_t} + 2\frac{z_t}{\sqrt{h_t}}h_{t;\mu} + \frac{1}{2}z_t^2 h_{t;\mu}^2\right), \\
\frac{\partial^3}{\partial \mu^3}\ell_T(\varphi) &\equiv \mathcal{L}_{\mu\mu\mu} = \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\mu,\mu,\mu} + 3\sum_{t=1}^{T}\frac{1}{h_t}h_{t;\mu} \\
&\quad -3\sum_{t=1}^{T}\frac{z_t}{\sqrt{h_t}}\left(h_{t;\mu,\mu} - h_{t;\mu}^2\right) - \frac{1}{2}\sum_{t=1}^{T}z_t^2\left(3h_{t;\mu}h_{t;\mu,\mu} - h_{t;\mu}^3\right)
\end{aligned}
$$

while for $i,j,k = \{\alpha,\theta,\gamma,\beta\}$ the derivatives are

$$
\begin{aligned}
\frac{\partial}{\partial i}\ell_T(\varphi) &\equiv \mathcal{L}_i = \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;i}, \\
\frac{\partial^2}{\partial i\partial j}\ell_T(\varphi) &\equiv \mathcal{L}_{ij} = \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;i,j} - \frac{1}{2}\sum_{t=1}^{T}z_t^2 h_{t;i}h_{t;j}, \\
\frac{\partial^3}{\partial i\partial j\partial k}\ell_T(\varphi) &\equiv \mathcal{L}_{ijk} = \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;i,j,k} - \frac{1}{2}\sum_{t=1}^{T}z_t^2\left(3h_{t;i}h_{t;j,k} - h_{t;i}h_{t;j}h_{t;k}\right).
\end{aligned}
$$

The cross derivatives are given by the following expressions

$$
\begin{aligned}
\frac{\partial^2}{\partial i\partial \mu}\ell_T(\varphi) &\equiv \mathcal{L}_{i\mu} = \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;i,\mu} - \frac{1}{2}\sum_{t=1}^{T}z_t^2 h_{t;i}h_{t;\mu} - \sum_{t=1}^{T}\frac{z_t}{\sqrt{h_t}}h_{t;i}, \\
\frac{\partial^3}{\partial i\partial^2 \mu}\ell_T(\varphi) &\equiv \mathcal{L}_{i\mu\mu} = \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;i,\mu,\mu} - 2\sum_{t=1}^{T}\frac{z_t}{\sqrt{h_t}}\left(h_{t;i,\mu} - h_{t;i}h_{t;\mu}\right) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}z_t^2\left(2h_{t;\mu}h_{t;i,\mu} - h_{t;i}h_{t;\mu}^2 + h_{t;i}h_{t;\mu,\mu}\right) + \sum_{t=1}^{T}\frac{1}{h_t}h_{t;i}, \\
\frac{\partial^3}{\partial i\partial j\partial \mu}\ell_T(\varphi) &\equiv \mathcal{L}_{ij\mu} = \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;i,j,\mu} - \sum_{t=1}^{T}\frac{z_t}{\sqrt{h_t}}\left(h_{t;i,j} - h_{t;i}h_{t;j}\right) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}z_t^2\left(h_{t;j}h_{t;i,\mu} - h_{t;j}h_{t;i}h_{t;\mu} + h_{t;i,j}h_{t;\mu} + h_{t;i}h_{t;j,\mu}\right).
\end{aligned}
$$

Note that the log-likelihood derivatives are expressions of the log-variance derivatives, $h_{t;\circ}$, where the latter are given in the Appendix. The expected values of the log-likelihood derivatives are also given in the Appendix.

The cross-products of the log-likelihood derivatives for $i,j=\{\alpha,\theta,\gamma,\beta\}$ are

$$
\begin{aligned}
\mathcal{L}_i\mathcal{L}_{ij} &= \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;i}\left(\frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;i,j}-\frac{1}{2}\sum_{t=1}^{T}z_t^2h_{t;i}^2\right),\\
\mathcal{L}_i\mathcal{L}_{j\mu} &= \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;i}\left(\frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;j,\mu}-\frac{1}{2}\sum_{t=1}^{T}z_t^2h_{t;j}h_{t;\mu}-\sum_{t=1}^{T}\frac{z_t}{\sqrt{h_t}}h_{t;i}\right),\\
\mathcal{L}_i\mathcal{L}_{\mu\mu} &= \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;i}\left[\frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\mu,\mu}-\sum_{t=1}^{T}\left(\frac{1}{h_t}+2\frac{z_t}{\sqrt{h_t}}h_{t;\mu}+\frac{1}{2}z_t^2h_{t;\mu}^2\right)\right],\\
\mathcal{L}_\mu\mathcal{L}_{ij} &= \left(\frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\mu}+\sum_{t=1}^{T}\frac{z_t}{\sqrt{h_t}}\right)\left(\frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;i,j}-\frac{1}{2}\sum_{t=1}^{T}z_t^2h_{t;i}^2\right),\\
\mathcal{L}_\mu\mathcal{L}_{j\mu} &= \left(\frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\mu}+\sum_{t=1}^{T}\frac{z_t}{\sqrt{h_t}}\right)\left(\begin{array}{c}\frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;j,\mu}-\frac{1}{2}\sum_{t=1}^{T}z_t^2h_{t;j}h_{t;\mu}\\ -\sum_{t=1}^{T}\frac{z_t}{\sqrt{h_t}}h_{t;i}\end{array}\right),\\
\mathcal{L}_\mu\mathcal{L}_{\mu\mu} &= \left(\frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\mu}+\sum_{t=1}^{T}\frac{z_t}{\sqrt{h_t}}\right)\left[\begin{array}{c}\frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\mu,\mu}\\ -\sum_{t=1}^{T}\left(\frac{1}{h_t}+2\frac{z_t}{\sqrt{h_t}}h_{t;\mu}+\frac{1}{2}z_t^2h_{t;\mu}^2\right)\end{array}\right].
\end{aligned}
$$

Appendix B Expected values of cross products of the log-likelihood derivatives

Here we present the expected values of cross-products of the log-likelihood derivatives. To conserve space, we present only some indicative. That is,

1. $\frac{1}{T}E\left(\mathcal{L}_\alpha\mathcal{L}_{\alpha\alpha}\right)=-\frac{1}{4}\left[\sum\sum_{s<t}E\left[\left(z_s^2-1\right)h_{s;\alpha}h_{t;\alpha}h_{t;\alpha}\right]-\left(\kappa_4+2\right)\sum_{t=1}^{T}E\left(h_{t;\alpha}h_{t;\alpha,\alpha}-h_{t;\alpha}^3\right)\right]$

2. $\frac{1}{T}E\left(\mathcal{L}_\alpha\mathcal{L}_{\alpha\mu}\right)=-\frac{1}{4}\left[\begin{array}{c}\sum\sum_{s<t}E\left[\left(z_s^2-1\right)h_{s;\mu}h_{t;\alpha}h_{t;\mu}\right]-\left(\kappa_4+2\right)\sum_{t=1}^{T}E\left(h_{t;\alpha}h_{t;\alpha,\mu}-h_{t;\mu}h_{t;\alpha}^2\right)\\ +2\kappa_3\sum_{t=1}^{T}E\left(\frac{1}{\sqrt{h_t}}h_{t;\alpha}^2\right)\end{array}\right]$

3. $\frac{1}{T}E\left(\mathcal{L}_\alpha\mathcal{L}_{\mu\mu}\right)=-\frac{1}{4}\left[\begin{array}{c}\sum\sum_{s<t}E\left[\left(z_s^2-1\right)h_{s;\alpha}h_{t;\mu}^2\right]-\left(\kappa_4+2\right)\sum_{t=1}^{T}E\left(h_{t;\alpha}h_{t;\mu,\mu}-h_{t;\alpha}h_{t;\mu}^2\right)\\ +4\kappa_3\sum_{t=1}^{T}E\left(\frac{1}{\sqrt{h_t}}h_{t;\alpha}h_{t;\mu}\right)\end{array}\right]$

4. $\frac{1}{T}E\left(\mathcal{L}_\mu\mathcal{L}_{\alpha\alpha}\right)=-\frac{1}{4}\left[\begin{array}{c}\sum\sum_{s<t}E\left[\left(z_s^2-1\right)h_{s;\mu}h_{t;\alpha}^2\right]-\left(\kappa_4+2\right)\sum_{t=1}^{T}E\left(h_{t;\mu}h_{t;\alpha,\alpha}-h_{t;\mu}h_{t;\alpha}^2\right)\\ +2\sum\sum_{s<t}E\left(z_s\frac{1}{\sqrt{h_t}}h_{t;\alpha}^2\right)+2\kappa_3\sum_{t=1}^{T}E\left(\frac{1}{\sqrt{h_t}}h_{t;\alpha}^2-\frac{1}{\sqrt{h_t}}h_{t;\alpha,\alpha}\right)\end{array}\right]$

5. $\frac{1}{T}E\left(\mathcal{L}_{\mu}\mathcal{L}_{\alpha\mu}\right)=-\frac{1}{4}\left[\begin{array}{c}\sum\sum_{s<t}E\left[\left(z_s^2-1\right)h_{s;\mu}h_{t;\alpha}h_{t;\mu}\right]-\left(\kappa_4+2\right)\sum_{t=1}^{T}E\left(h_{t;\mu}h_{t;\alpha,\mu}-h_{t;\alpha}h_{t;\mu}^2\right)\\+2\sum\sum_{s<t}E\left(z_s\frac{1}{\sqrt{h_t}}h_{t;\alpha}h_{t;\mu}\right)+2\kappa_3\sum_{t=1}^{T}E\left(\frac{2}{\sqrt{h_t}}h_{t;\alpha}h_{t;\mu}-\frac{1}{\sqrt{h_t}}h_{t;\alpha,\mu}\right)\\+4\sum_{t=1}^{T}E\left(\frac{1}{h_t}h_{t;\alpha}\right)\end{array}\right]$

6. $\frac{1}{T}E\left(\mathcal{L}_{\mu}\mathcal{L}_{\mu\mu}\right)=-\frac{1}{4}\left[\begin{array}{c}\sum\sum_{s<t}E\left[\left(z_s^2-1\right)h_{s;\mu}h_{t;\mu}^2\right]-\left(\kappa_4+2\right)\sum_{t=1}^{T}E\left(h_{t;\mu}h_{t;\mu,\mu}-h_{t;\mu}^3\right)\\+2\sum\sum_{s<t}E\left(z_s\frac{1}{\sqrt{h_t}}h_{t;\mu}^2\right)+2\kappa_3\sum_{t=1}^{T}E\left(\frac{3}{\sqrt{h_t}}h_{t;\mu}^2-\frac{1}{\sqrt{h_t}}h_{t;\mu,\mu}\right)\\+8\sum_{t=1}^{T}E\left(\frac{1}{h_t}h_{t;\mu}\right)\end{array}\right]$

At this point, we should note that these results differ from those in the paper of Linton (2007), due to the fact that we assume non-symmetric distribution of the errors and also none of these expressions are zero, since the block-diagonality of the information matrix in the case of the EGARCH$(1,1)$ model does not hold.

Analytic proof of the first result is given as follows

$$\begin{aligned}\mathcal{L}_{\alpha}\mathcal{L}_{\alpha\alpha} &= \frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\alpha}\left(\frac{1}{2}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\alpha,\alpha}-\frac{1}{2}\sum_{t=1}^{T}z_t^2\left(h_{t;\alpha}\right)^2\right)\\ &= \frac{1}{4}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\alpha}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\alpha,\alpha}-\frac{1}{4}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\alpha}\sum_{t=1}^{T}z_t^2h_{t;\alpha}^2\\ &= \frac{1}{4}\sum_{t=1}^{T}\left(z_t^2-1\right)^2h_{t;\alpha}h_{t;\alpha,\alpha}+\frac{1}{4}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\alpha}\sum_{s=t+1}^{T}\left(z_s^2-1\right)h_{s;\alpha,\alpha}\\ &\quad+\frac{1}{4}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\alpha}\sum_{s=1}^{t-1}\left(z_s^2-1\right)h_{s;\alpha,\alpha}\\ &\quad-\frac{1}{4}\sum_{t=1}^{T}\left(z_t^2-1\right)z_t^2h_{t;\alpha}^3-\frac{1}{4}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\alpha}\sum_{s=t+1}^{T}z_s^2h_{s;\alpha}^2-\frac{1}{4}\sum_{t=1}^{T}\left(z_t^2-1\right)h_{t;\alpha}\sum_{s=1}^{t-1}z_s^2h_{s;\alpha}^2.\end{aligned}$$

Hence

$$E\left(\mathcal{L}_{\alpha}\mathcal{L}_{\alpha\alpha}\right)=\frac{T\left(\kappa_4+2\right)}{4}\left[E\left(h_{t;\alpha}h_{t;\alpha,\alpha}\right)-E\left(h_{t;\alpha}^3\right)\right]-\frac{1}{4}E\sum\sum_{s<t}\left(z_s^2-1\right)z_t^2h_{t;\alpha}^2h_{s;\alpha},$$

where

$$h_{t;\alpha}=1+\left(\beta-\frac{1}{2}\theta z_{t-1}-\frac{1}{2}\gamma\left|z_{t-1}\right|\right)h_{t-1;\alpha},$$

and

$$h_{t;\alpha}^2 = 1 + 2\left(\beta - \frac{1}{2}\theta z_{t-1} - \frac{1}{2}\gamma\left|z_{t-1}\right|\right) h_{t-1;\alpha} + \left(\beta - \frac{1}{2}\theta z_{t-1} - \frac{1}{2}\gamma\left|z_{t-1}\right|\right)^2 h_{t-1;\alpha}^2.$$

Let

$$h_{t+k;\alpha} = 1 + \left(\beta - \frac{1}{2}\theta z_{t+k-1} - \frac{1}{2}\gamma\left|z_{t+k-1}\right|\right) h_{t+k-1;\alpha},$$

and

$$h_{t+k;\alpha}^2 = 1 + 2\left(\beta - \frac{1}{2}\theta z_{t+k-1} - \frac{1}{2}\gamma\left|z_{t+k-1}\right|\right) h_{t+k-1;\alpha} + \left(\beta - \frac{1}{2}\theta z_{t+k-1} - \frac{1}{2}\gamma\left|z_{t+k-1}\right|\right)^2 h_{t+k-1;\alpha}^2.$$

Hence,

$$\begin{aligned} E\left[\left(z_t^2 - 1\right) h_{t+k;\alpha}^2 h_{t;\alpha}\right] &= E\left[\left(z_t^2 - 1\right)\left[\begin{array}{c} h_{t;\alpha} + 2\left(\beta - \frac{1}{2}\theta z_{t+k-1} - \frac{1}{2}\gamma\left|z_{t+k-1}\right|\right) h_{t+k-1;\alpha} h_{t;\alpha} \\ + \left(\beta - \frac{1}{2}\theta z_{t+k-1} - \frac{1}{2}\gamma\left|z_{t+k-1}\right|\right)^2 h_{t+k-1;\alpha}^2 h_{t;\alpha} \end{array}\right]\right] \\ &\overset{k \geq 1}{=} 2\left(\beta - \frac{1}{2}\gamma E\left|z\right|\right) E\left(z_t^2 - 1\right) h_{t+k-1;\alpha} h_{t;\alpha} \\ &\quad + \left[\beta^2 + \frac{1}{4}\left(\theta^2 + \gamma^2\right) - \beta\gamma E\left|z\right| + \frac{1}{2}\theta\gamma E\left(z\left|z\right|\right)\right] E\left(z_t^2 - 1\right) h_{t+k-1;\alpha}^2 h_{t;\alpha}, \end{aligned}$$

while for $k = 1$

$$\begin{aligned} E\left[\left(z_t^2 - 1\right) h_{t+1;\alpha}^2 h_{t;\alpha}\right] &= E\left[\left(z_t^2 - 1\right)\left[\begin{array}{c} h_{t;\alpha} + 2\left(\beta - \frac{1}{2}\theta z_t - \frac{1}{2}\gamma\left|z_t\right|\right) h_{t;\alpha}^2 \\ + \left(\beta - \frac{1}{2}\theta z_t - \frac{1}{2}\gamma\left|z_t\right|\right)^2 h_{t;\alpha}^3 \end{array}\right]\right] \\ &= 2E\left[\left(z_t^2 - 1\right)\left(\beta - \frac{1}{2}\theta z_t - \frac{1}{2}\gamma\left|z_t\right|\right)\right] E h_{t;\alpha}^2 \\ &\quad + E\left[\left(z_t^2 - 1\right)\left(\beta - \frac{1}{2}\theta z_t - \frac{1}{2}\gamma\left|z_t\right|\right)^2\right] E h_{t;\alpha}^3. \end{aligned}$$

Hence,

$$\begin{aligned} E\left[\left(z_t^2 - 1\right) h_{t+k;\alpha}^2 h_{t;\alpha}\right] &\overset{k \geq 1}{=} -\left(\beta - \frac{1}{2}\gamma E\left|z\right|\right)\left[\theta E z^3 + \gamma\left(E\left|z\right|^3 - E\left|z\right|\right)\right]\left(\beta - \frac{1}{2}\gamma E\left|z\right|\right)^{k-2} E h_{t;\alpha}^2 \\ &+ \left[\beta^2 + \frac{1}{4}\left(\theta^2 + \gamma^2\right) - \beta\gamma E\left|z\right| + \frac{1}{2}\theta\gamma E\left(z\left|z\right|\right)\right] E\left(z_t^2 - 1\right) h_{t+k-1;\alpha}^2 h_{t;\alpha}. \end{aligned}$$

Set

$$
\begin{aligned}
A &= -\left(\beta - \frac{1}{2}\gamma E\left|z\right|\right)\left[\theta E z^3 + \gamma\left(E\left|z\right|^3 - E\left|z\right|\right)\right] E h_{t;\alpha}^2, \\
C &= \beta^2 + \frac{1}{4}\left(\theta^2+\gamma^2\right) - \beta\gamma E\left|z\right| + \frac{1}{2}\theta\gamma E\left(z\left|z\right|\right).
\end{aligned}
$$

We have that

$$
E\left[\left(z_t^2-1\right)h_{t+k;\alpha}^2 h_{t;\alpha}\right] \overset{k\geq 1}{=} A\left(\beta - \frac{1}{2}\gamma E\left|z\right|\right)^{k-2} + CE\left(z_t^2-1\right)h_{t+k-1;\alpha}^2 h_{t;\alpha}.
$$

By repeating substitution $E\left[\left(z_t^2-1\right)h_{t+k;\alpha}^2 h_{t;\alpha}\right] \overset{k\geq 1}{=}$

$$
\begin{aligned}
&A\left[\left(\beta - \frac{1}{2}\gamma E\left|z\right|\right)^{k-2} + C\left(\beta - \frac{1}{2}\gamma E\left|z\right|\right)^{k-3} + \ldots + C^{k-2}\right] + C^{k-1}E\left(z_t^2-1\right)h_{t+1;\alpha}^2 h_{t;\alpha} \\
&\overset{k\geq 1}{=} A\frac{C^{k-1} - \left(\beta - \frac{1}{2}\gamma E\left|z\right|\right)^{k-1}}{C - \left(\beta - \frac{1}{2}\gamma E\left|z\right|\right)} + C^{k-1}E\left(z_t^2-1\right)h_{t+1;\alpha}^2 h_{t;\alpha}.
\end{aligned}
$$

Consequently, for $k \geq 1$, $E\left[\left(z_t^2-1\right)h_{t+k;\alpha}^2 h_{t;\alpha}\right]$ is given by

$$
\begin{aligned}
&A\frac{C^{k-1} - \left(\beta - \frac{1}{2}\gamma E\left|z\right|\right)^{k-1}}{C - \left(\beta - \frac{1}{2}\gamma E\left|z\right|\right)} + \\
&C^{k-1}\left[2E\left[\left(z_t^2-1\right)\left(\beta - \frac{1}{2}\theta z_t - \frac{1}{2}\gamma\left|z_t\right|\right)\right]Eh_{t;\alpha}^2 + E\left[\left(z_t^2-1\right)\left(\beta - \frac{1}{2}\theta z_t - \frac{1}{2}\gamma\left|z_t\right|\right)^2\right]Eh_{t;\alpha}^3\right],
\end{aligned}
$$

where

$$
2E\left[\left(z_t^2-1\right)\left(\beta - \frac{1}{2}\theta z_t - \frac{1}{2}\gamma\left|z_t\right|\right)\right]Eh_{t;\alpha}^2 = -\left(\theta E z^3 + \gamma\left(E\left(\left|z\right|^3\right) - E\left|z\right|\right)\right)Eh_{t;\alpha}^2
$$

and

$$
E\left[\left(z_t^2-1\right)\left(\beta - \frac{1}{2}\theta z_t - \frac{1}{2}\gamma\left|z_t\right|\right)^2\right]Eh_{t;\alpha}^3 = \left[\begin{array}{c} \frac{1}{4}\left(\theta^2+\gamma^2\right)\left(Ez^4-1\right) \\ -\beta\theta E z^3 + \beta\gamma\left(E\left|z\right| - E\left|z\right|^3\right) \\ +\frac{1}{2}\theta\gamma\left(E\left(z^3\left|z\right|\right) - E\left(z\left|z\right|\right)\right) \end{array}\right]Eh_{t;\alpha}^3.
$$

Hence we have

$$
\begin{aligned}
E\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left(z_s^2-1\right)z_t^2 h_{t;\alpha}^2 h_{s;\alpha} &= E\sum_{t=1}^{T-1}\sum_{k=1}^{T-t}\left(z_t^2-1\right)h_{t+k;\alpha}^2 h_{t;\alpha} \\
&= \sum_{t=1}^{T-1}\sum_{k=1}^{T-t} A\frac{C^{k-1} - \left(\beta - \frac{1}{2}\gamma E\left|z\right|\right)^{k-1}}{C - \left(\beta - \frac{1}{2}\gamma E\left|z\right|\right)} + C^{k-1}\Lambda,
\end{aligned}
$$

where

$$
\begin{aligned}
\Lambda &= -\left(\theta E z^3+\gamma\left(E\left(|z|^3\right)-E|z|\right)\right) E h_{t;\alpha}^2 \\
&+\left[\begin{array}{c}
\frac{1}{4}\left(\theta^2+\gamma^2\right)\left(E z^4-1\right)-\beta\theta E z^3 \\
+\beta\gamma\left(E|z|-E|z|^3\right)+\frac{1}{2}\theta\gamma\left(E\left(z^3|z|\right)-E(z|z|)\right)
\end{array}\right] E h_{t;\alpha}^3 .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left(z_s^2-1\right) z_t^2 h_{t;\alpha}^2 h_{s;\alpha} &= \left(\frac{A}{C-\left(\beta-\frac{1}{2}\gamma E|z|\right)}+\Lambda\right)\sum_{t=1}^{T-1}\sum_{k=1}^{T-t} C^{k-1} \\
&\quad -\frac{A}{C-\left(\beta-\frac{1}{2}\gamma E|z|\right)}\sum_{t=1}^{T-1}\sum_{k=1}^{T-t}\left(\beta-\frac{1}{2}\gamma E|z|\right)^{k-1}
\end{aligned}
$$

and keeping only terms of order T, we have $E\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left(z_s^2-1\right) z_t^2 h_{t;\alpha}^2 h_{s;\alpha}=$

$$
\left(\frac{A}{C-\left(\beta-\frac{1}{2}\gamma E|z|\right)}+\Lambda\right)\frac{T}{1-C}-\frac{A}{C-\left(\beta-\frac{1}{2}\gamma E|z|\right)}\frac{T}{1-\left(\beta-\frac{1}{2}\gamma E|z|\right)},
$$

provided that $|C|<1$ and $\left|\beta-\frac{1}{2}\gamma E|z|\right|<1$. Hence,

$$
\begin{aligned}
E\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left(z_s^2-1\right) z_t^2 h_{t;\alpha}^2 h_{s;\alpha} &= T\left(\begin{array}{c}
\frac{A}{C-\left(\beta-\frac{1}{2}\gamma E|z|\right)}\frac{1}{1-C}+\Lambda\frac{1}{1-C} \\
-\frac{A}{C-\left(\beta-\frac{1}{2}\gamma E|z|\right)}\frac{1}{1-\left(\beta-\frac{1}{2}\gamma E|z|\right)}
\end{array}\right)+O(1) \\
&= T\frac{A}{(1-C)\left(1-\left(\beta-\frac{1}{2}\gamma E|z|\right)\right)}+T\frac{\Lambda}{1-C}+O(1)
\end{aligned}
$$

where

$$
\begin{aligned}
A &= -\left(\beta-\frac{1}{2}\gamma E|z|\right)\left[\theta E z^3+\gamma\left(E|z|^3-E|z|\right)\right] E h_{t;\alpha}^2, \\
C &= \beta^2+\frac{1}{4}\left(\theta^2+\gamma^2\right)-\beta\gamma E|z|+\frac{1}{2}\theta\gamma E(z|z|), \\
\Lambda &= -\left(\theta E z^3+\gamma\left(E\left(|z|^3\right)-E|z|\right)\right) E h_{t;\alpha}^2 \\
&+\left[\begin{array}{c}
\frac{1}{4}\left(\theta^2+\gamma^2\right)\left(E z^4-1\right)-\beta\theta E z^3 \\
+\beta\gamma\left(E|z|-E|z|^3\right)+\frac{1}{2}\theta\gamma\left(E\left(z^3|z|\right)-E(z|z|)\right)
\end{array}\right] E h_{t;\alpha}^3 .
\end{aligned}
$$

Appendix C Auxilliary Lemma

Lemma 1 (log-likelihood derivatives) *The expected values of the second and third log-likelihood derivatives are given by the following expressions.*

For $i, j = \{\alpha, \theta, \gamma, \beta\}$,

$$\begin{aligned} E(\mathcal{L}_{ij}) &= -\frac{T}{2} E(h_{t;i} h_{t;j}), \\ E(\mathcal{L}_{\mu j}) &= -\frac{T}{2} E(h_{t;\mu} h_{t;j}), \\ E(\mathcal{L}_{\mu\mu}) &= -TE\left(\frac{1}{h_t}\right) - \frac{T}{2} E\left(h_{t;\mu}^2\right), \end{aligned}$$

and

$$E(\mathcal{L}_{iii}) = -\frac{T}{2} E\left(3h_{t;i} h_{t;i,i} - h_{t;i}^3\right),$$

for $i = \{\alpha, \theta, \gamma, \beta\}$, *and* $j = \{\alpha, \theta, \gamma, \beta, \mu\}$,

$$E(\mathcal{L}_{iij}) = -\frac{T}{2} E\left(h_{t;j} h_{t;i,i} - h_{t;i}^2 h_{t;j} + 2h_{t;i} h_{t;i,j}\right),$$

for $i, j = \{\alpha, \theta, \gamma, \beta\}$, *and* $k = \{\alpha, \theta, \gamma, \beta, \mu\}$,

$$E(\mathcal{L}_{ijk}) = -\frac{T}{2} E(h_{t;j} h_{t;i,k} + h_{t;k} h_{t;i,j} + h_{t;i} h_{t;j,k} - h_{t;j} h_{t;i} h_{t;k},),$$

for $i = \{\alpha, \theta, \gamma, \beta\}$, *and* $j = \mu$,

$$E(\mathcal{L}_{ijj}) = -\frac{T}{2} E\left(h_{t;i} h_{t;j,j} + 2h_{t;j} h_{t;i,j} - h_{t;i} (h_{t;j})^2\right) + TE\left(\frac{1}{h_t} h_{t;i}\right),$$

for $j = \mu$,

$$E(\mathcal{L}_{jjj}) = -\frac{T}{2} E\left(3h_{t;j} h_{t;j,j} - h_{t;j}^3\right) + TE\left(3\frac{1}{h_t} h_{t;j}\right).$$

Proof. The results come immediately from applying the expectation to the terms of the log-likelihood derivatives which are given in the Appendix A. Analytic results of these expectations to what they are equal to are given in the Technical Appendix (TA). ■

T	$\widehat{\alpha}$ bias	$\widehat{\beta}$ bias	$\widehat{\gamma}$ bias	$\widehat{\theta}$ bias
500	0.0389	−0.0048	−0.0134	−0.0046
750	0.0257	−0.0029	−0.0101	−0.0031
1500	0.0123	−0.0014	−0.0043	−0.0013
3000	0.0063	−0.0007	−0.0028	−0.0006
5000	0.0036	−0.0004	−0.0017	−0.0003
10000	0.0020	−0.0003	−0.0008	−0.0001
25000	0.0009	−0.0001	−0.0005	−0.0001
50000	0.0003	0.0000	−0.0001	−0.0001

Table 1: Biases of ML estimators under normality with $\alpha = 0.1, \beta = 0.9, \gamma = 0.7, \theta = -0.4$.

T	$\widehat{\alpha}$ bias	$\widehat{\beta}$ bias	$\widehat{\gamma}$ bias	$\widehat{\theta}$ bias
500	0.0720	−0.0163	−0.0560	−0.0139
750	0.0492	−0.0107	−0.0401	−0.0073
1500	0.0194	−0.0036	−0.0229	−0.0008
3000	0.0089	−0.00017	−0.0106	−0.0001
5000	0.0056	−0.0010	−0.0085	−0.0001
10000	0.0030	−0.0006	−0.0034	−0.0002
25000	0.0003	−0.0001	−0.0014	0.0001
50000	0.0001	0.0000	−0.0008	−0.0001

Table 2: Biases of QML estimators under mixture of normals with $\alpha = 0.1, \beta = 0.9, \gamma = 0.7, \theta = -0.4$.

T	Full-Step	Bootstrap
500	0.20	52.23
750	0.22	64.65
1500	0.30	95.31
3000	0.54	148.26
5000	0.73	163.44
10000	1.43	254.67
25000	3.11	683.49
50000	5.23	1202.35

Table 3: Average CPU time/MC iteration, first set of parameter values, under normality

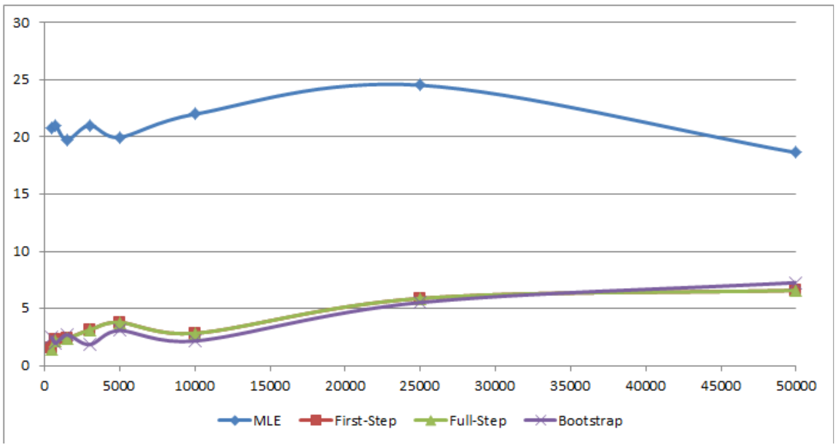

Figure 1: Bias of all estimators under normality with $\alpha = 0.1, \beta = 0.9, \gamma = 0.7, \theta = -0.4$.

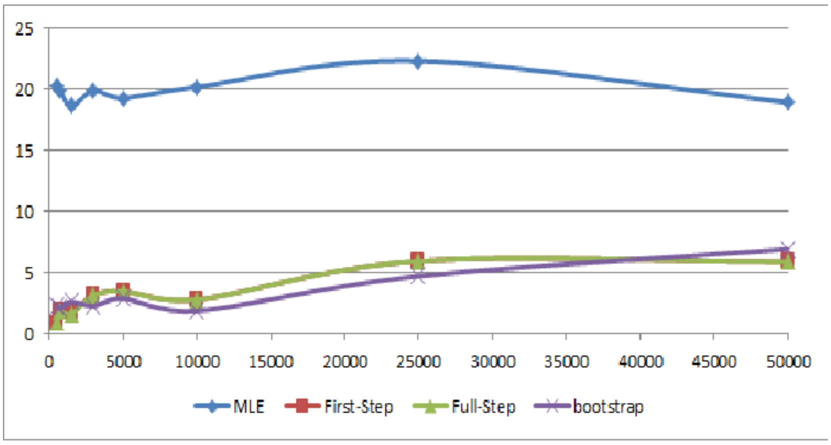

Figure 2: Bias of all estimators under normality with $\alpha = -0.1, \beta = 0.9, \gamma = 0.6, \theta = -0.2$.

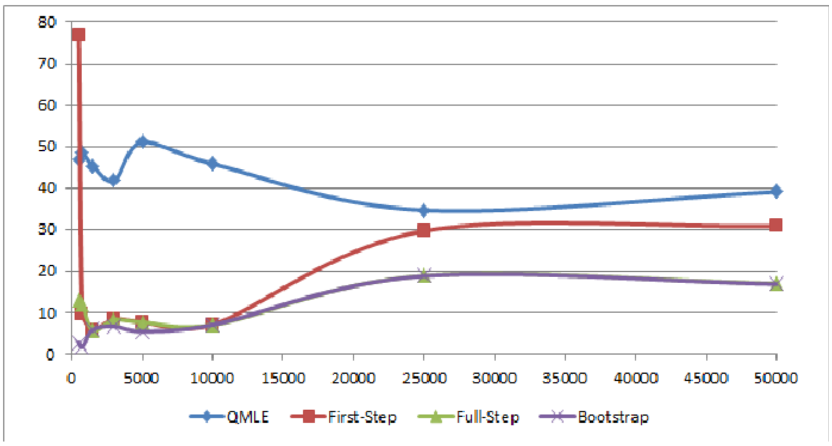

Figure 3: Bias of all estimators under mixture of normals with

$\alpha = 0.1, \beta = 0.9, \gamma = 0.7, \theta = -0.4$.

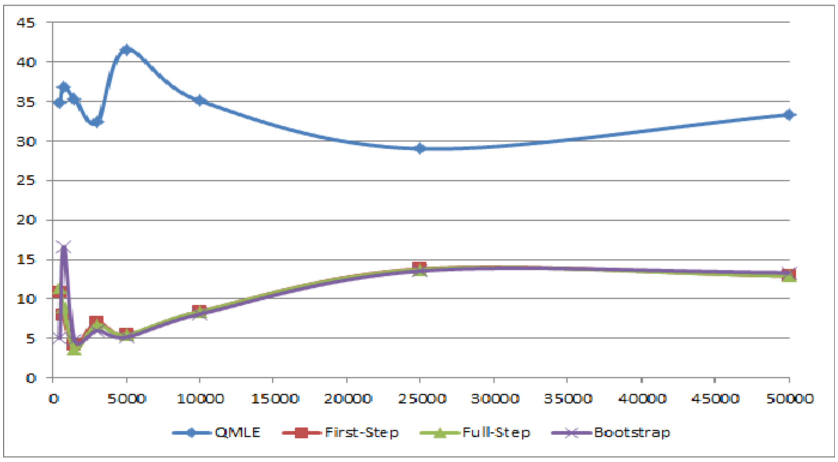

Figure 4: Bias of all estimators under mixture of normals with $\alpha = -0.1, \beta = 0.9, \gamma = 0.6, \theta = -0.2$.

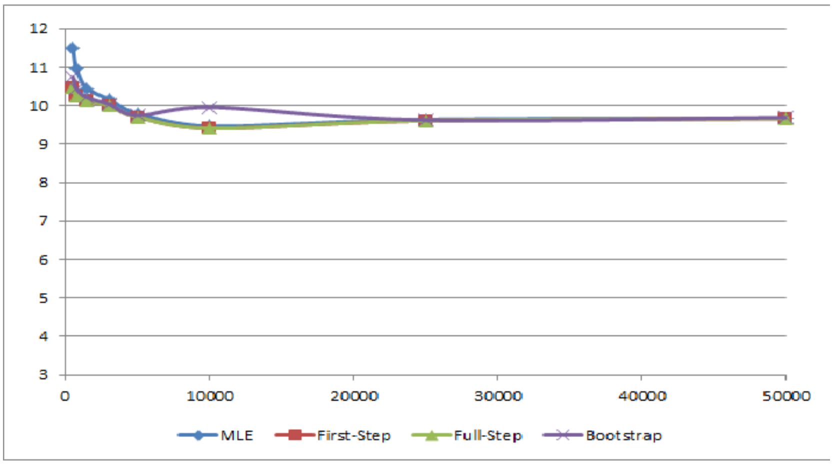

Figure 5: MSEs of all estimators under normality with $\alpha = 0.1, \beta = 0.9, \gamma = 0.7, \theta = -0.4$.

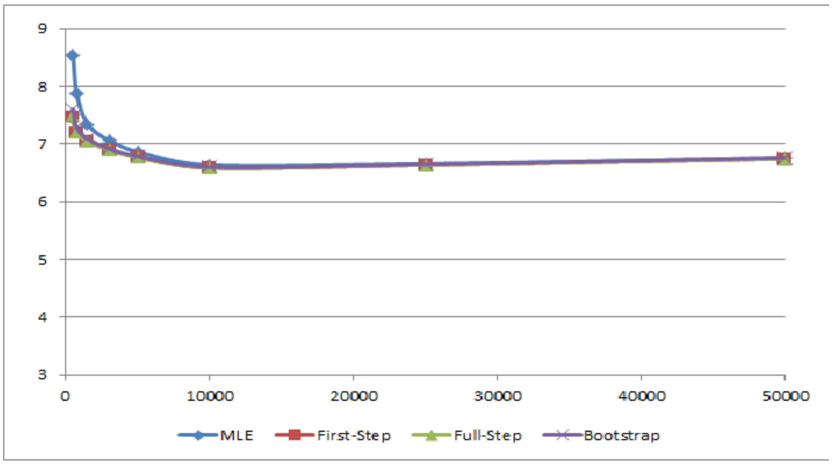

Figure 6: MSEs of all estimators under normality with $\alpha = -0.1, \beta = 0.9, \gamma = 0.6, \theta = -0.2$.

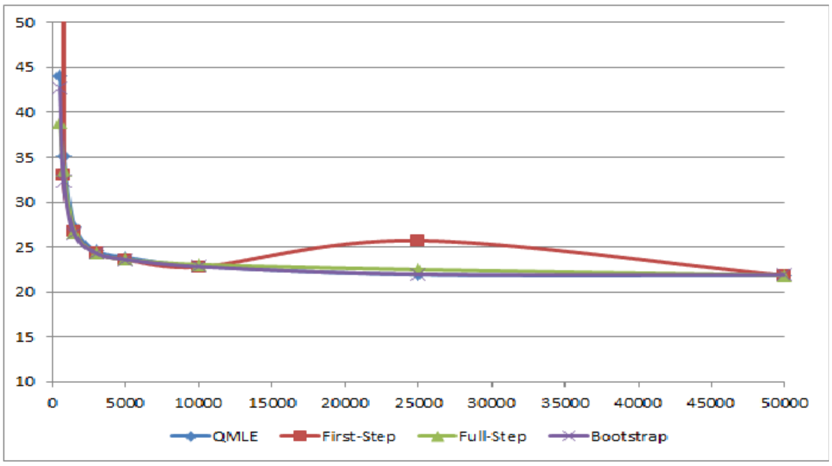

Figure 7: MSEs of all estimators under mixture of normals with $\alpha = 0.1, \beta = 0.9, \gamma = 0.7, \theta = -0.4$.

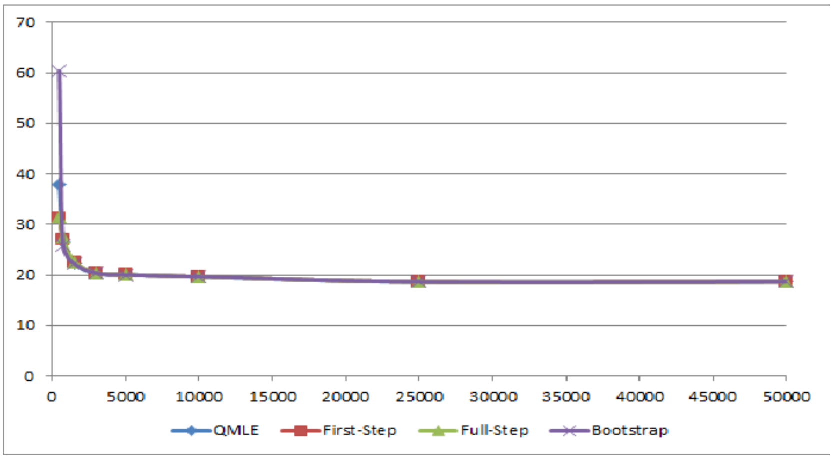

Figure 8: Bias of all estimators under mixture of normals with $\alpha = -0.1, \beta = 0.9, \gamma = 0.6, \theta = -0.2$.