3-1-1996

An Asymptotic Expansion in the Garch(1,1) Model

Oliver B. Linton

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1118

AN ASYMPTOTIC EXPANSION IN THE GARCH(1,1) MODEL

Oliver Linton

March 1996

AN ASYMPTOTIC EXPANSION IN THE GARCH(1,1) MODEL

By Oliver Linton

Yale University

Summary

We develop order T^{-1} asymptotic expansions for the quasi-maximum likelihood estimator (QMLE) and a two step approximate QMLE in the GARCH(1,1) model. We calculate the approximate mean and skewness and hence the Edgeworth-B distribution function. We suggest several methods of bias reduction based on these approximations.

1 Introduction

The first order asymptotic properties of estimators and test statistics in ARCH models are now well established. The Gaussian quasi-maximum likelihood estimator QMLE is asymptotically normal under quite general conditions, see for example Weiss (1986), Lumsdaine (1991), Lee and Hansen (1994), and Bollerslev and Wooldridge (1992). Indeed, Lumsdaine (1991) shows that this is true even when there is a unit (or even mildly explosive) root in the variance process.[1] The finite sample properties are less well known, although some simulation evidence has been presented: in an early paper Engle et al. (1985) found "direct evidence of a substantial negative bias in the coefficient" (estimate of the parameter on the lagged squared residual) in an ARCH model. They also found that sample standard deviations could be up to 30% below the limiting value for samples of size 50. Bollerslev and Wooldridge (1992) look at both estimates and tests and conclude that biases were relatively minor. In a recent Monte Carlo study, Lumsdaine (1995) found substantial evidence of skewness in estimates and test statistics, attributing some of this to the truncation of the parameter space imposed during estimation.

We investigate the small sample properties of the QMLE and a widely used two step estimator in the GARCH(1,1) and IGARCH(1,1) models, by developing asymptotic expansions and calculating the approximate bias and skewness from the truncated expansions. We then compute the Edgeworth approximate distribution derived from these approximate moments. This methodology has a long history of successful application in both econometrics and statistics. Bartlett (1953) gave the order T^{-1} bias and skewness of the MLE of a single parameter in i.i.d sampling, where T is sample size. Haldane and Smith (1956) gave the first four cumulants in the same setting. Shenton and Bowman in a series of papers (1963,1969,1977) developed these approximations further including higher orders of approximation (out even to order T^{-4} in some cases) to multivariate problems. Many of these results are discussed in McCullagh (1987). There has also been considerable work in the econometrics literature starting with Nagar (1959) who developed order T^{-1} second moment approximations for k-class estimators in a simultaneous system. Anderson (1974,1977) and Anderson and Sawa (1973,1979) give the expansions for estimators in simultaneous equation systems. Phillips (1977a) does the same for first order autoregression. Sargan (1974, 1976) and Phillips (1977b) establish the theoretical validity of the approximations for smooth functions of sample moments. Rothenberg (1984ab,1988) covers estimation and testing based on GLS procedures. Taniguchi (1991) develops expansions for many time series procedures. See Rothenberg (1986) for a review.

We apply these methods to our nonlinear time series setting. Many terms are involved in the expansions, in general. Nevertheless, several patterns emerge. Firstly, the order T^{-1} mean and skewness are nonlinear functions of the parameters. Secondly, the magnitude of the bias appears small except in the extreme case when the variance process approaches a deterministic difference equation. Thirdly, in some directions of approach to the unit root, the bias decreases, even when standardized by asymptotic standard deviation. This latter fact contrasts with the finding in linear autoregression, see for example Phillips (1977a), where the bias increases, and without limit on standardization by the asymptotic standard deviation, as the unit root is approached. This is entirely consistent with the known fact that for GARCH processes a unit root has no effect on the rate of convergence of parameter estimates as documented in Lumsdaine (1991). Our focus here is on estimation of the parameters of the variance process, although in passing we comment on the location estimator.

Some applications of our expansions include how to correct the estimator for bias, how to size-adjust test statistics, see Rothenberg (1988), and how to correct the certainty equivalent predictive likelihood function as in Cooley and Parke (1990). Finally, we confess that our expansions are purely formal.

In section 2 we introduce the model and procedures under investigation. The main theorem is stated in section 3, and in section 4 we investigate the approximations. In section 5 we give some extensions. We give our derivations in the appendix.

2 The model and estimators

The observed data $\{y_t\}_{t=1}^{T}$ are generated by the GARCH(1,1) process:

$$y_t = \beta + v_t^{1/2}\varepsilon_t$$

$$v_t = \theta_1 + \theta_2 v_{t-1} + \theta_3 (y_{t-1} - \beta)^2, \quad t = 1, 2, \ldots, T,$$

where ε_t are i.i.d., symmetric about zero, with variance one, fourth cumulant κ_4, and $\kappa_{23} = E\left\{(\varepsilon_t^2 - 1)^3\right\}$. We require ε_t to have finite J'th moments, for some large J> 6. We also assume that the true parameter values satisfy: $\theta_1 > 0$ and

$$E\left\{\ln(\theta_2 + \theta_3 \varepsilon_t^2)\right\} < \infty,$$

in which case v_t is strongly stationary and ergodic, see Nelson (1990). We assume that the initial observations are drawn from the stationary distribution.

We work throughout with estimators computed from the following conditional (on initial conditions) Gaussian Quasi-Likelihood:

$$\mathcal{L} = -\frac{1}{2}\sum_{t=1}^{T} h_t(\theta, \beta) - \frac{1}{2}\sum_{t=1}^{T} \varepsilon_t^2(\theta, \beta),$$

where $h_t = \ln v_t$ and $\theta = (\theta_1, \theta_2, \theta_3)'$. With regard to the family of processes induced by θ, we consider two alternative start-up rules documented further in the appendix:

(WS) Weakly Stationary Case. When $\theta_2 + \theta_3 < 1$, the process $v_t(\theta)$ is weakly stationary with unconditional variance $v = \theta_1/(1 - \theta_2 - \theta_3)$. In this case, the process $v_t(\theta)$ and its derivatives defined below are assumed to start from their unconditional mean.

(UR) Unit Root Case. When $\theta_2 + \theta_3 = 1$, the process $v_t(\theta)$ is not weakly stationary, i.e. the unconditional variance would not exist. In this case, the process $v_t(\theta)$ and its derivatives are assumed to start from some arbitrary value.

Let $\widehat{\beta}$ and $\widehat{\theta}$ be the maximizers of $\mathcal{L}(\theta, \beta)$ subject to the inequality restrictions $\theta_2, \theta_3 \geq 0$ and $\theta_1 > 0$ (in fact, we work with solutions of the quasi-score equations, $\partial\mathcal{L}(\widehat{\beta}, \widehat{\theta}) / \partial\phi = 0$, where $\phi = (\beta, \theta')')$. We also examine a convenient alternative to $\widehat{\theta}$: a maximizer $\widetilde{\theta}$ of $\mathcal{L}(\theta, \widetilde{\beta})$, where $\widetilde{\beta}$ is any $T^{1/2}$ consistent estimate of β, for example $\widetilde{\beta} = T^{-1} \sum_{t=1}^{T} y_t$. This method is widely used in practice. Note that under our symmetry assumption, the asymptotic distribution of $\widetilde{\theta}$ is independent of the choice of preliminary estimator $\widetilde{\beta}$.[2]

3 The main result

Our method is to expand the score equations in a power series in the likelihood derivatives about the true parameter value, and then to invert this expansion to yield an expansion for the standardized estimator. The moments of the truncated expansion are then found. Barndorff-Nielsen and Cox (1989) and Rothenberg (1986) review these techniques and the standard results. We first develop some notation. For β and for $i, j, k \in \{1, 2, 3\}$, the likelihood derivatives are:

$$\mathcal{L}_i = \tfrac{1}{2} \textstyle\sum_{t=1}^{T} (\varepsilon_t^2 - 1) h_{t;i},$$
$$\mathcal{L}_\beta = \tfrac{1}{2} \textstyle\sum_{t=1}^{T} (\varepsilon_t^2 - 1) h_{t;\beta} + \sum_{t=1}^{T} \varepsilon_t v_t^{-1/2}$$
$$\mathcal{L}_{ij} = \tfrac{1}{2} \textstyle\sum_{t=1}^{T} (\varepsilon_t^2 - 1) h_{t;ij} - \tfrac{1}{2} \sum_{t=1}^{T} \varepsilon_t^2 h_{t;i} h_{t;j},$$
$$\mathcal{L}_{\beta\beta} = \tfrac{1}{2} \textstyle\sum_{t=1}^{T} (\varepsilon_t^2 - 1) h_{t;\beta\beta} - \tfrac{1}{2} \sum_{t=1}^{T} \varepsilon_t^2 h_{t;\beta}^2 - 2 \sum_{t=1}^{T} \varepsilon_t v_t^{-1/2} h_{t;\beta} - \sum_{t=1}^{T} v_t^{-1}$$
$$\mathcal{L}_{i\beta} = \tfrac{1}{2} \textstyle\sum_{t=1}^{T} (\varepsilon_t^2 - 1) h_{t;i\beta} - \tfrac{1}{2} \sum_{t=1}^{T} \varepsilon_t^2 h_{t;i} h_{t;\beta} - \sum_{t=1}^{T} \varepsilon_t v_t^{-1/2} h_{t;i}$$
$$\mathcal{L}_{ijk} = \tfrac{1}{2} \textstyle\sum_{t=1}^{T} (\varepsilon_t^2 - 1) h_{t;ijk} - \tfrac{1}{2} \sum_{t=1}^{T} \varepsilon_t^2 \{h_{t;ij} h_{t;k} + h_{t;ik} h_{t;j} + h_{t;jk} h_{t;i} - h_{t;i} h_{t;j} h_{t;k}\},$$
$$\mathcal{L}_{\beta\beta\beta} = \tfrac{1}{2} \textstyle\sum_{t=1}^{T} (\varepsilon_t^2 - 1) h_{t;\beta\beta\beta} - \tfrac{1}{2} \sum_{t=1}^{T} \varepsilon_t^2 \{3 h_{t;\beta} h_{t;\beta\beta} - h_{t;\beta}^3\}$$

$$- \sum_{t=1}^{T} \varepsilon_t v_t^{-1/2} \{h_{t;\beta\beta} - 3h_{t;\beta}^2\} + 3 \sum_{t=1}^{T} v_t^{-1} h_{t;\beta},$$

$$\mathcal{L}_{ij\beta} = \tfrac{1}{2} \sum_{t=1}^{T} (\varepsilon_t^2 - 1) h_{t;ij\beta} - \tfrac{1}{2} \sum_{t=1}^{T} \varepsilon_t^2 \{h_{t;ij} h_{t;\beta} + h_{t;i\beta} h_{t;j} + h_{t;j\beta} h_{t;i} - h_{t;i} h_{t;j} h_{t;\beta}\}$$

$$- \sum_{t=1}^{T} \varepsilon_t v_t^{-1/2} \{h_{t;ij} - h_{t;i} h_{t;j}\}$$

$$\mathcal{L}_{i\beta\beta} = \tfrac{1}{2} \sum_{t=1}^{T} (\varepsilon_t^2 - 1) h_{t;i\beta\beta} - \tfrac{1}{2} \sum_{t=1}^{T} \varepsilon_t^2 \{h_{t;i} h_{t;\beta\beta} + 2h_{t;\beta} h_{t;\beta i} - h_{t;\beta}^2 h_{t;i}\}$$

$$+ \sum_{t=1}^{T} \varepsilon_t v_t^{-1/2} \{2h_{t;\beta} h_{t;i} - h_{t;\beta i}\} + \sum_{t=1}^{T} v_t^{-1} h_{t;i},$$

where $h_{t;i} = \partial h_t / \partial \theta_i$, for example. The logarithmic derivatives are related to the level derivatives of v_t by $h_{t;i} = v_t^{-1} v_{t;i}$ and $h_{t;ij} = v_t^{-1} v_{t;ij} - v_t^{-2} v_{t;i} v_{t;j}$, where, for example, $v_{t;i} = \partial v_t / \partial \theta_i$. Recursive expressions for the variance derivatives are given in the appendix. The scores are martingales relative to the natural filtration, while the Hessian and higher derivatives do not have this useful property.

Under the above conditions, the following moments converge[3] to finite limits as $T \to \infty$:

$$\tau_{ij} = T^{-1} E(\mathcal{L}_{ij}) = -\mu_{i,j}/2$$
$$\tau_{i,j} = T^{-1} E(\mathcal{L}_i \mathcal{L}_j) = (\kappa_4 + 2)\mu_{i,j}/4$$
$$\tau_{ijk} = T^{-1} E(\mathcal{L}_{ijk}) = -\{\mu_{ij,k} + \mu_{ik,j} + \mu_{jk,i} - \mu_{i,j,k}\}/2$$
$$\tau_{ij,k} = T^{-1} E(\mathcal{L}_{ij} \mathcal{L}_k) = -\{\mu_{k;i,j}^{\varepsilon\varepsilon} - (\kappa_4 + 2)(\mu_{ij,k} - \mu_{i,j,k})\}/4$$
$$\tau_{i,j,k} = T^{-1} E(\mathcal{L}_i \mathcal{L}_j \mathcal{L}_k) = \left\{\kappa_{23} \mu_{i,j,k} + (\kappa_4 + 2)(\mu_{k;i,j}^{\varepsilon\varepsilon} + \mu_{i;k,j}^{\varepsilon\varepsilon} + \mu_{j;i,k}^{\varepsilon\varepsilon})\right\}/8$$
$$\tau_{\beta\beta} = T^{-1} E(\mathcal{L}_{\beta\beta}) = -(\overline{\pi} + \mu_{\beta,\beta}/2)$$
$$\tau_{\beta,\beta} = T^{-1} E(\mathcal{L}_\beta^2) = \overline{\pi} + (\kappa_4 + 2)\mu_{\beta,\beta}/4$$
$$\tau_{i\beta\beta} = T^{-1} E(\mathcal{L}_{i\beta\beta}) = \overline{\pi}_i - \{\mu_{i,\beta\beta} + 2\mu_{\beta i,\beta} - \mu_{\beta,i,\beta}\}/2$$
$$\tau_{i\beta,\beta} = T^{-1} E(\mathcal{L}_{i\beta} \mathcal{L}_\beta) = -\{4\overline{\pi}_i + \mu_{\beta;i,\beta}^{\varepsilon\varepsilon} + 2\mu_{i,\beta}^{\varepsilon v} - (\kappa_4 + 2)(\mu_{i\beta,\beta} - \mu_{i,\beta,\beta})\}/4,$$

where $\mu_{i,j} = T^{-1}\sum_{t=1}^{T} E(h_{t;i}h_{t;j})$, which forms a nonsingular 3 by 3 matrix, while $\mu_{ij,k} = T^{-1}\sum_{t=1}^{T} E(h_{t;ij}h_{t;k})$ and $\mu_{i,j,k} = T^{-1}\sum_{t=1}^{T} E(h_{t;i}h_{t;j}h_{t;k})$. Here, $\overline{\pi} = T^{-1}\sum_{t=1}^{T} E(v_t^{-1})$, $\overline{v} = T^{-1}\sum_{t=1}^{T} E(v_t)$, and $\overline{\pi}_i = T^{-1}\sum_{t=1}^{T} E(v_t^{-1}h_{t;i})$, while

$$\mu^{\varepsilon\varepsilon}_{k;i,j} = T^{-1}\sum_{s<t}\sum E\left\{(\varepsilon_s^2-1)h_{s;k}h_{t;i}h_{t;j}\right\} \quad ; \quad \mu^{\varepsilon v}_{i,\beta} = T^{-1}\sum_{s<t}\sum E\left\{\varepsilon_s v_s^{-1/2}h_{t;i}h_{t;\beta}\right\}.$$

When the errors are symmetric, any moment involving an odd number of β subscripts is zero, so that $\tau_{i\beta}, \tau_{ij\beta}, \tau_{\beta\beta\beta} = 0$. When $\kappa_4 = 0$, the first two Bartlett identities hold; in particular,

$$\tau_{ij} + \tau_{i,j} = 0 \quad ; \quad \tau_{ijk} + \tau_{ij,k} + \tau_{ik,j} + \tau_{jk,i} + \tau_{i,j,k} = 0,$$

although more generally they do not.

Both $T^{1/2}(\widehat{\theta}-\theta)$ and $T^{1/2}(\widetilde{\theta}-\theta)$ are asymptotically normal with mean zero and the same covariance matrix, $\Omega = (\omega_{ij})_{i,j=1}^{3}$, whose $i, j'th$ element is

$$\omega_{ij} = \sum_{k,m=1}^{3} \tau^{ik}\tau_{k,m}\tau^{mj} = (\kappa_4+2)\mu^{i,j}, \tag{3.1}$$

where raising pairs of indices signifies matrix inversion, i.e. $\Omega^{-1} = (\omega^{ij})_{i,j=1}^{3}$. Similarly, $T^{1/2}(\widehat{\beta}-\beta)$ and $T^{1/2}(\widetilde{\beta}-\beta)$ are asymptotically normal. Let $\Phi_\chi(\cdot)$ be the c.d.f of an $N(0,\chi)$ random variable for any positive scalar χ, and let c be any 3 by 1 vector. In the appendix we establish the following:

THEOREM: *Our results below hold in both the weakly stationary case (1) and the unit root case (2). Firstly, $\widehat{\beta}$ is symmetrically distributed about β to order T^{-1}. Secondly, the asymptotic bias and skewness of $c'\widehat{\theta}$ are γ_1/T and γ_3/T, and of $c'\widetilde{\theta}$ are γ_1^*/T and γ_3^*/T, with*

$$\gamma_1 = \lambda_1 + \lambda_{10} - (\lambda_0 + \lambda_2) \quad ; \quad \gamma_3 = \lambda_4 - \lambda_3 - \lambda_5 \quad ;$$
$$\gamma_1^* = \lambda_1^* + \lambda_{10}^* - (\lambda_0 + \lambda_2) \quad ; \quad \gamma_3^* = \gamma_3,$$

where[4]

$$\lambda_0 = T^{-1} \sum\sum_{s<t} E\left\{(\varepsilon_s^2 - 1)c'\mathbf{M}^{-1}\mathbf{h}_{\theta t}\mathbf{h}_{\theta t}'\mathbf{M}^{-1}\mathbf{h}_{\theta s}\right\}$$
$$\lambda_1 = c'\mathbf{M}^{-1}T^{-1}\sum_{t=1}^T E\left\{\rho_1\mathbf{h}_{\theta t}h_{t;\beta\beta} + \rho_2\mathbf{h}_{\theta t}h_{t;\beta}^2 + \rho_3\mathbf{h}_{\theta\beta t}h_{t;\beta} + \rho_4 v_t^{-1}\mathbf{h}_{\theta t}\right\}$$
$$\lambda_{10} = \tau_{\beta\beta}^{-1}c'\mathbf{M}^{-1}T^{-1}\sum\sum_{s<t} E\left\{(\varepsilon_s^2 - 1)\mathbf{h}_{\theta t}h_{s;\beta}h_{t;\beta}\right\}/2 + E\left\{\varepsilon_s v_s^{-1/2}\mathbf{h}_{\theta t}h_{t;\beta}\right\}$$
$$\lambda_1^* = c'\mathbf{M}^{-1}T^{-1}\sum_{t=1}^T E\left\{\rho_1^*\mathbf{h}_{\theta t}h_{t;\beta\beta} + \rho_2^*\mathbf{h}_{\theta t}h_{t;\beta}^2 + \rho_3^*\mathbf{h}_{\theta\beta t}h_{t;\beta} + \rho_4^* v_t^{-1}\mathbf{h}_{\theta t} + \rho_5^*\mathbf{h}_{\theta t}\right\}$$
$$\lambda_{10}^* = -2c'\mathbf{M}^{-1}T^{-1}\sum\sum_{s<t} E\left\{\varepsilon_s v_s^{1/2}\mathbf{h}_{\theta t}h_{t;\beta}\right\}$$
$$\lambda_2 = (\kappa_4 + 2)T^{-1}\sum_{t=1}^T E\left\{\mathbf{tr}(\mathbf{M}^{-1}\mathbf{D}_t)c'\mathbf{M}^{-1}\mathbf{h}_{\theta t}\right\}/2$$
$$\lambda_3 = 3(\kappa_4 + 2)^2 T^{-1}\sum_{t=1}^T E\left\{c'\mathbf{M}^{-1}\mathbf{D}_t\mathbf{M}^{-1}cc'\mathbf{M}^{-1}\mathbf{h}_{\theta t}\right\}$$
$$\lambda_4 = \kappa_{23}T^{-1}\sum_{t=1}^T E\left\{(c'\mathbf{M}^{-1}\mathbf{h}_{\theta t})^3\right\}$$
$$\lambda_5 = 6(\kappa_4 + 2)T^{-1}\sum\sum_{s<t} E\left\{(c'\mathbf{M}^{-1}\mathbf{h}_{\theta t})^2 c'\mathbf{M}^{-1}\mathbf{h}_{\theta s}(\varepsilon_s^2 - 1)\right\},$$

with $\mathbf{h}_{\theta t} = (h_{t;1}, h_{t;2}, h_{t;3})'$, $\mathbf{D}_t = (v_t^{-1}v_{t;ij})_{i,j=1}^3$, $\mathbf{M} = (\mu_{i,j})_{i,j=1}^3$, *and* $\mathbf{h}_{\theta\beta t} = (\frac{\partial h_{t;1}}{\partial\beta}, \frac{\partial h_{t;2}}{\partial\beta}, \frac{\partial h_{t;3}}{\partial\beta})'$, *while* $\rho_1 = -\tau_{\beta\beta}^{-2}\tau_{\beta,\beta}/2$, $\rho_2 = \tau_{\beta\beta}^{-1}\left\{\frac{(\kappa_4+2)}{2} + \tau_{\beta,\beta}\tau_{\beta\beta}^{-1}/2\right\}$, $\rho_3 = -\tau_{\beta\beta}^{-1}\left\{\frac{(\kappa_4+2)}{2} - \tau_{\beta,\beta}\tau_{\beta\beta}^{-1}\right\}/2$, $\rho_4 = \tau_{\beta\beta}^{-1}\left\{2 + \tau_{\beta,\beta}\tau_{\beta\beta}^{-1}\right\}$, *and* $\rho_1^* = -\overline{v}$, $\rho_2^* = \overline{v}/2$, $\rho_3^* = -\overline{v}$, $\rho_4^* = \overline{v}$, $\rho_5^* = -2$.

Therefore, we have the formal Edgeworth-B approximation: for any $x \in \mathbb{R}$,

$$\Pr\left\{T^{1/2}c'(\widehat{\theta} - \theta) \leq x\right\} = \Phi_{c'\Omega c}\left\{x - \frac{6\gamma_1 + \gamma_3(x^2 - 1)}{6T^{1/2}}\right\} + o(T^{-1/2}). \tag{3.2}$$

Likewise for $T^{1/2}c'(\widetilde{\theta} - \theta)$ *with* γ_1^* *and* γ_3^* *replacing* γ_1 *and* γ_3 *in (3.2).*

Unfortunately, we cannot in general obtain simple expressions for γ_1 and γ_3, as one can in the AR(1) model[5], and therefore present our results in the above form which is suggestive of how one might compute them. We use numerical integration to compute the bias, see section 4 below. The formulae simplify in some special cases:

When β is known and not estimated: the bias terms λ_1, λ_{10}, λ_1^*, and λ_{10}^* are all zero.

When $\theta_2 = 0$ is known and only β, θ_1, and θ_3 are estimated: both λ_2 and λ_3 are zero because then all second derivatives $v_{t;ij}$ are zero.

Under Gaussianity: $\rho_1, \rho_2, \rho_3 = \tau_{\beta\beta}^{-1}/2$, and $\rho_4 = \tau_{\beta\beta}^{-1}$.

When $v_t = v$ is constant and only v and β are estimated (the QMLE of v here is the sample variance $\widehat{v} = T^{-1}\sum_{t=1}^{T}(y_t - \overline{y})^2)$. In this case, $\mathbf{h}_{\theta t} = v^{-1}$ and $\mathbf{M} = v^{-2}$, so that $\lambda_0, \lambda_{10}, \lambda_1^*, \lambda_{10}^*, \lambda_2, \lambda_3, \lambda_5 = 0$ – only λ_1 and λ_4 are no zero. In fact, $\lambda_1 = \rho_4 = -v$ and $\lambda_4 = v^3\kappa_{23}$.[6] Thus the bias and skewness of $\widehat{v}$ are $-vT^{-1}$ and $v^3\kappa_{23}T^{-1}$ respectively, which agrees with the known moments of the sample variance.

Our results cast doubt on the relevance of some of the calculations carried out in Engle et al. (1985, p91): specifically, the bias is generally affected by how β is estimated, witness the difference between λ_1 and λ_1^*. However, the skewness, although generally nonzero, is unaffected by whether or how location is estimated. This accords with the well-known result, see Pfanzagl (1980, p32) and Akahira and Takeuchi (1981), that first order efficiency implies second order efficiency, i.e. the skewness of all first order efficient estimators is the same. Note, however, that neither of our estimators is necessarily efficient yet still this result appears to hold.

A Special Case: Local Power of a Wald Test

We now consider an important special case where $\theta_2 = 0$ and $\theta_3 = \overline{\theta}_3/T^{1/2}$ for some fixed $\overline{\theta}_3$, and the parameters $\beta, \theta_1, \theta_3$ are estimated by maximizing $\mathcal{L}$. In this case, we can approximate the precision process by Taylor expansion and calculate explicitly the required moments for the Edgeworth approximation. We have

$$\frac{1}{v_t} = \frac{1}{\theta_1} - \frac{\overline{\theta}_3}{\theta_1^2 T^{1/2}}(y_{t-1} - \beta)^2 + O_p(T^{-1}),$$

while $v_{t;1} = 1$ and $v_{t;3} = (y_{t-1} - \beta)^2$. Therefore,

$$\mathbf{M} = \begin{pmatrix} \frac{1}{\theta_1^2} & \frac{1}{\theta_1} \\ \frac{1}{\theta_1} & E(\varepsilon_1^4) \end{pmatrix} + \frac{\overline{\theta}_3}{T^{1/2}} \begin{pmatrix} \frac{-2}{\theta_1^2} & \frac{1-2E(\varepsilon_1^4)}{\theta_1} \\ \frac{1-2E(\varepsilon_1^4)}{\theta_1} & 2\left\{E(\varepsilon_1^4) - E(\varepsilon_1^6)\right\} \end{pmatrix} = \mathbf{M}_0 + \frac{1}{T^{1/2}}\mathbf{M}_1,$$

and thus

$$\mathbf{M}^{-1} = \mathbf{M}_0^{-1} - \frac{1}{T^{1/2}}\mathbf{M}_0^{-1}\mathbf{M}_1\mathbf{M}_0^{-1}$$

to order $T^{-1/2}$, where

$$\mathbf{M}_0^{-1} = \frac{\theta_1^2}{E(\varepsilon_1^4) - 1} \begin{pmatrix} E(\varepsilon_1^4) & \frac{-1}{\theta_1} \\ \frac{-1}{\theta_1} & \frac{1}{\theta_1^2} \end{pmatrix}.$$

We have written $\mathbf{M}$ in two parts: one that occurs under homoskedasticity and a correction factor due to $\overline{\theta}_3 \neq 0$.

As for the bias and skewness, note that: $\lambda_2, \lambda_3 = 0$ because we are not estimating θ_2, while $\lambda_{10}, \lambda_{10}^* = 0$, $\lambda_1 = c'\mathbf{M}^{-1}T^{-1}\sum_{t=1}^T E(\rho_4 v_t^{-1}\mathbf{h}_{\theta t})$ and $\lambda_1^* = c'\mathbf{M}^{-1}T^{-1}\sum_{t=1}^T E\left\{\rho_4^* v_t^{-1}\mathbf{h}_{\theta t} + \rho_5^*\mathbf{h}_{\theta t}\right\}$ because $v_{t;\beta} = 0$, finally, we can replace $\mathbf{M}^{-1}$ by $\mathbf{M}_0^{-1}$, $\mathbf{h}_{\theta t}$ by $(1, \varepsilon_{t-1}^2 v_{t-1})'/\theta_1$, and $\mathbf{h}_{\theta 1}$ by $(1, \theta_1)'/\theta_1$. We have

$$E\left\{(\varepsilon_1^2-1)\mathbf{h}_{\theta 2}\mathbf{h}'_{\theta 2}\right\} = \begin{pmatrix} 0 & \{E(\varepsilon_1^4)-1\}/\theta_1 \\ \{E(\varepsilon_1^4)-1\}/\theta_1 & \{E(\varepsilon_1^6)-E(\varepsilon_1^4)\} \end{pmatrix},$$

while

$$E\left\{(\varepsilon_1^2-1)\mathbf{h}_{\theta t}\mathbf{h}'_{\theta t}\right\} = O(T^{-(t-2)/2}), \quad t=3,\ldots,T.$$

Therefore, to the required order of magnitude

$$\begin{aligned}
\lambda_0 &= c'\mathbf{M}_0^{-1}E\left\{(\varepsilon_1^2-1)\mathbf{h}_{\boldsymbol{\theta}\mathbf{2}}\mathbf{h}'_{\theta 2}\right\}\mathbf{M}_0^{-1}\mathbf{h}_{\theta 1} \\
&= \frac{\theta_1^3}{\left\{E(\varepsilon_1^4)-1\right\}^2}c'\begin{pmatrix} E(\varepsilon_1^4) & \frac{-1}{\theta_1} \\ \frac{-1}{\theta_1} & \frac{1}{\theta_1^2} \end{pmatrix}\begin{pmatrix} 0 & \{E(\varepsilon_1^4)-1\}/\theta_1 \\ \{E(\varepsilon_1^4)-1\}/\theta_1 & E(\varepsilon_1^6)-E(\varepsilon_1^4) \end{pmatrix}\begin{pmatrix} E(\varepsilon_1^4)-1 \\ 0 \end{pmatrix} \\
&= c'\begin{pmatrix} -\theta_1 \\ 1 \end{pmatrix}.
\end{aligned}$$

Also, $E(v_2^{-1}\mathbf{h}_{\theta 2}) = (1,\theta_1)'/\theta_1^2$ and $\rho_4 = -\theta_1$, so that $\lambda_1 = -\theta_1 c'\mathbf{M}_0^{-1}E(v_t^{-1}\mathbf{h}_{\theta t}) = -c'(\theta_1, 0)'$. Therefore,

$$\gamma_1 = \begin{cases} 0 & \text{if } c=(1,0)' \\ -1 & \text{if } c=(0,1)' \end{cases}.$$ [7]

Our results for $\widehat{\theta}_3$ agree exactly with those presented in Engle et al. (1985) even though their estimation method and setting were slightly different. Specifically, they worked with the special case that $\theta_1 = 1, \theta_3 = 0$, and β known. Their estimation method was to use the score function and the conditional mean of the Hessian $-$ $\left(\frac{-1}{2}\sum_{t=1}^{T} h_{t;i}h_{t;j}\right)_{i,j}$ $-$ to take one step starting from the true values $\theta_1 = 1, \theta_3 = 0$. Our results for $\widehat{\theta}_1$, however, indicate a difference between the maximum likelihood estimator (MLE) and the Engle et al. (1985) approximation: the bias of the MLE is zero to this order, while for their procedure there is a positive bias of magnitude θ_1/T.

Now for the skewness. We have

$$\begin{aligned} E\left\{(c'\mathbf{M}_0^{-1}\mathbf{h}_{\theta 2})^3\right\} &= \frac{\theta_1^3}{\{E(\varepsilon_1^4)-1\}^3} E\left\{c'\begin{pmatrix} E(\varepsilon_1^4) & \frac{-1}{\theta_1} \\ \frac{-1}{\theta_1} & \frac{1}{\theta_1^2}\end{pmatrix}\begin{pmatrix} 1 \\ \varepsilon_1^2\theta_1\end{pmatrix}\right\}^3 \\ &= \frac{\theta_1^3}{\{E(\varepsilon_1^4)-1\}^3} E\left\{c'\begin{pmatrix} E(\varepsilon_1^4)-\varepsilon_1^2 \\ (\varepsilon_1^2-1)/\theta_1\end{pmatrix}\right\}^3, \end{aligned}$$

and so

$$\lambda_4 = \begin{cases} \theta_1^3\kappa_{23}\frac{E\left\{E(\varepsilon_1^4)-\varepsilon_1^2\right\}^3}{\left\{E(\varepsilon_1^4)-1\right\}^3} & \text{if } c=(1,0)' \\ \\ \frac{\kappa_{23}^2}{\left\{E(\varepsilon_1^4)-1\right\}^3} & \text{if } c=(0,1)'. \end{cases}$$

Also, $E\left\{(c'\mathbf{M}_0^{-1}\mathbf{h}_{\theta 2})^2(\varepsilon_1^2-1)\right\}$

$$= \frac{\theta_1^2}{\left\{E(\varepsilon_1^4)-1\right\}^2}c'\begin{pmatrix} E(\varepsilon_1^4) & \frac{-1}{\theta_1} \\ \frac{-1}{\theta_1} & \frac{1}{\theta_1^2}\end{pmatrix}\begin{pmatrix} 0 & \{E(\varepsilon_1^4)-1\}\theta_1 \\ \{E(\varepsilon_1^4)-1\}\theta_1 & \{E(\varepsilon_1^6)-E(\varepsilon_1^4)\}\theta_1^2\end{pmatrix}\begin{pmatrix} E(\varepsilon_1^4) & \frac{-1}{\theta_1} \\ \frac{-1}{\theta_1} & \frac{1}{\theta_1^2}\end{pmatrix}c$$

$$= \frac{\theta_1^2}{\left\{E(\varepsilon_1^4)-1\right\}^2}c'\begin{pmatrix} E(\varepsilon_1^6)-2E^2(\varepsilon_1^4)+E(\varepsilon_1^4)+1 & \{E^2(\varepsilon_1^4)+E(\varepsilon_1^4)-E(\varepsilon_1^6)-1\}/\theta_1 \\ \{E^2(\varepsilon_1^4)+E(\varepsilon_1^4)-E(\varepsilon_1^6)-1\}/\theta_1 & \{E(\varepsilon_1^6)-3E(\varepsilon_1^4)+1\}/\theta_1^2\end{pmatrix}c,$$

while

$$E\left\{(c'\mathbf{M}_0^{-1}\mathbf{h}_{\theta t})^2(\varepsilon_1^2-1)\right\} = O(T^{-(t-2)/2}), \quad t=3,\ldots,T.$$

Multiplying by $c'\mathbf{M}_0^{-1}\mathbf{h}_{\theta 1} = c'(\theta_1, 0)'$, we find that

$$\lambda_5 = \begin{cases} 6(\kappa_4+2)\theta_1^3 \frac{E(\varepsilon_1^6)-2E^2(\varepsilon_1^4)+E(\varepsilon_1^4)+1}{\left\{E(\varepsilon_1^4)-1\right\}^2} & \text{if } c=(1,0)' \\ \\ 0 & \text{if } c=(0,1)'. \end{cases}$$

Finally, γ_3 is the difference between λ_4 and λ_5. In the Gaussian case, $\gamma_3 = 9\theta_1^3$ when $c=(1,0)'$, while $\gamma_3 = 8$ when $c=(0,1)'$.

As an application consider the following non-robust Wald test for the presence of ARCH. Suppose that the errors are Gaussian and that $\widehat{\beta}$ and $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_3)'$ are estimated by the maximum likelihood method. Our test is to

$$\text{reject if} \quad T^{1/2}\widehat{\theta}_3 > z_\alpha,$$

where $\Phi_1(-z_\alpha) = \alpha$. By the above arguments

$$\Pr\left\{T^{1/2}\widehat{\theta}_3 \le x\right\} = \Phi_1\left\{x - \overline{\theta}_3 - \frac{8(x-\overline{\theta}_3)^2 - 12\overline{\theta}_3(x-\overline{\theta}_3) - 14}{6T^{1/2}}\right\} + o(T^{-1/2}),$$

so that under the null hypothesis that $\theta_3 = 0$,

$$\Pr\left\{T^{1/2}\widehat{\theta}_3 \le x\right\} = \Phi_1\left\{x - \frac{8x^2-14}{6T^{1/2}}\right\} + o(T^{-1/2}).$$

For the usual 5% level with $z_{0.05} = 1.645$ the second order rejection frequency is predicted to be 6.4% for $T=100$, i.e. over-rejection is expected. This agrees somewhat with Lumsdaine (1995, Table 2).

4 Magnitude of the second order effect in the GARCH(1,1) case

We now investigate the magnitude of the asymptotic bias in the general setting of the GARCH(1,1) model. We calculate the asymptotic bias by numerical integration. In fact, we use simulation techniques to avoid the high dimensional integration. This technique is now widely established in other areas of econometrics, see Hajivassiliou and Ruud (1994) for discussion and references. By using sufficiently many random draws one can achieve any given degree of accuracy; we use over a million random variables in our computations and achieve third significant figure accuracy.

We work only with a normal error distribution and with the situation where the location parameter is known and not estimated in which case the bias is $-(\lambda_0+\lambda_2)/T$ for both $c'\widehat{\theta}$ and $c'\widetilde{\theta}$. For computational purposes it is convenient to rewrite the λ's in terms of matrix notation, i.e. $\lambda_2 = c'\mathbf{M}^{-1}\mathbf{\Theta}_2\text{vec}(\mathbf{M}^{-1})$ and $\lambda_0 = c'\mathbf{M}^{-1}\mathbf{\Theta}_0\text{vec}(\mathbf{M}^{-1})$, where $\mathbf{\Theta}_2 = T^{-1}\sum_{t=1}^{T} E\{\mathbf{h}_{\theta t}\text{vec}'(\mathbf{D}_t)\}$ and $\mathbf{\Theta}_0 = T^{-1}\sum\sum_{s<t} E\left\{(\varepsilon_s^2-1)\mathbf{h}_{\theta s}\otimes\mathbf{h}_{\theta t}\mathbf{h}_{\theta t}'\right\}$. We generate n samples of size R using the recursions given in the appendix, and estimate $\mathbf{M}$ by $\overline{\mathbf{M}} = n^{-1}\sum_{i=1}^{n} R^{-1}\sum_{t=1}^{R}\mathbf{h}_{\theta t}^{i}\mathbf{h}_{\theta t}^{i\prime}$, $\mathbf{\Theta}_2$ by $\overline{\mathbf{\Theta}}_2 = n^{-1}R^{-1}\sum_{t=1}^{R}\sum_{i=1}^{n}\mathbf{h}_{\theta t}^{i} vec'(\mathbf{D}_t^{i})$, and $\mathbf{\Theta}_0$ by $\overline{\mathbf{\Theta}}_0 = n^{-1}\sum_{i=1}^{n} R^{-1}\sum_{s<t}\sum_{t=1}^{R} E\left\{(\varepsilon_s^{i2}-1)\mathbf{h}_{\theta s}^{i}\otimes\mathbf{h}_{\theta t}^{i}\mathbf{h}_{\theta t}^{i\prime}\right\}$, where $\mathbf{h}_{\theta t}^{i}$ is the $i'th$ realization at time t. Then we estimate λ_2 by $c'\overline{\mathbf{M}}^{-1}\overline{\mathbf{\Theta}}_2\text{vec}(\overline{\mathbf{M}}^{-1})$ and λ_0 by $c'\overline{\mathbf{M}}^{-1}\overline{\mathbf{\Theta}}_0\text{vec}(\overline{\mathbf{M}}^{-1})$. We take $n = 10,000$ and $R = 100$.

As Lumsdaine (1995) shows, $\widehat{\theta}_2$ and $\widehat{\theta}_3$ are invariant to changes in θ_1, while $\widehat{\theta}_1$ varies proportionately with θ_1^2. Therefore, we restrict attention to $\theta_1 = 0.1$ and vary the other parameters. A stationary start-up was used, see the appendix. In Figure 1 below, we give the raw biases, γ_1/T, for each parameter estimate computed at a total of 171 equally spaced points inside the unit simplex (assuming a sample size of $T = 100$).[8]

*** FIGURE 1 ***

The bias of $\widehat{\theta}_1$ is positive and tends to increase with θ_2 but decrease with θ_3, although a dramatic U-shaped pattern is evident for $\theta_3 \approx 0$ (which corresponds to an essentially deterministic variance process).

*** FIGURE 2 ***

The bias of $\widehat{\theta}_2$ is generally negative, and displays considerable nonlinearity with respect to both θ_2 and θ_3 throughout, although, especially, a dramatic U-shape is evident for $\theta_3 \approx 0$, as above.

*** FIGURE 3 ***

The bias of $\widehat{\theta}_3$ trends upward from right to left, taking negative values for $\theta_2 \approx 0$ and $\theta_3 \approx 1$ and positive values for the reverse position.

Figures 1-3 make evident that the bias can decrease as the unit root region is approached. Note also that the magnitude of the bias is relatively small in each case except when $\theta_3 \approx 0$.

We now try and get some idea of the average effect of each parameter on the biases of $\widehat{\theta}_j$, $j = 1, 2, 3$. One way of displaying the multivariate function $b(\theta_2, \theta_3) = \gamma_1(\theta_2, \theta_3)/T$ is to report the one-dimensional slices $b_2(\theta_2) = \int b(\theta_2, \theta_3) dQ_3(\theta_3)$ and $b_3(\theta_3) = \int b(\theta_2, \theta_3) dQ_2(\theta_2)$, where Q_2 and Q_3 are probability measures, see Linton and Nielsen (1995). When b is either additive or multiplicative, i.e. $b(\theta_2, \theta_3) = g_2(\theta_2) + g_3(\theta_3)$ or $b(\theta_2, \theta_3) = g_2(\theta_2) g_3(\theta_3)$ for some functions g_2 and g_3, then b_2 and b_3 give the component functions up to a constant. More generally, they measure an average individual effect. We use empirical weighting and report b_2 and b_3 below in Figure 4.

*** FIGURE 4 ***

These pictures confirm the general impression given by Figures 1-3; for example, the U-shaped relationship between the bias of $\widehat{\theta}_2$ and θ_2.

Finally, we look at the unit root case. We take start-up $v_0 = 1$ and, as before, $\theta_1 = 0.1$. In Figure 5 we give $\gamma_1(\theta_2, \theta_3)/T$ for a grid of points on the unit simplex, $\{(0.95, 0.05), (0.90, 0.10), \ldots\}$, plotted against θ_2.

*** Figure 5 ***

The relationship is far from linear. Note that the biases are uniformly small in this case.

We suspect that when the affects of estimating β are factored in a more complicated picture might emerge, but that the main qualitative conclusions stated above remain.

5 Extensions

1. Nonlinear Functions of θ

In many cases, nonlinear functions of the parameters are also of interest. For example, the *cumulated impulse response* (CIR) function, $CIR = 1/(1 - \theta_2 - \theta_3)$, and the *half life of a unit shock* (HLS), $HLS = \ln 2/\ln(\theta_2 + \theta_3)$, which measure the long run response of the conditional variance to a unit shock and the length of time until the impulse response of a unit shock is half its initial value, respectively. The MSE optimal predictor of v_{t+k}, which is, in the weakly stationary case, $v^e_{t+k} = v + (\theta_2 + \theta_3)^{k-1}(v_{t+1} - v)$, see Baillie and Bollerslev (1992), and tests of hypotheses about θ based on the Wald, LM, or LR principles are also of interest here.

Let g be a scalar valued three times continuously differentiable function of θ, and let $\mathbf{g} = (\frac{\partial g}{\partial \theta_1}, \frac{\partial g}{\partial \theta_2}, \frac{\partial g}{\partial \theta_3})'$ and $\mathbf{G} = (\frac{\partial^2 g}{\partial \theta_i \partial \theta_j})_{i,j=1}^3$. Then, $T^{1/2}\left\{g(\widehat{\theta}) - g(\theta)\right\}$ is asymptotically normal with mean 0 and variance $\mathbf{g}'\boldsymbol{\Omega}\mathbf{g}$. Furthermore, the asymptotic mean and skewness of $g(\widehat{\theta}) - g(\theta)$ are approximately

$$T^{-1}\{\gamma_1(\mathbf{g}'\widehat{\theta}) + \mathbf{tr}(\mathbf{G}\boldsymbol{\Omega})/2\} \ ; \ T^{-1}\{\gamma_3(\mathbf{g}'\widehat{\theta}) + 3\mathbf{g}'\boldsymbol{\Omega}^{-1}\mathbf{G}\boldsymbol{\Omega}^{-1}\mathbf{g}\},$$

respectively. For the CIR function,

$$\mathbf{g} = \frac{-1}{(1-\theta_2-\theta_3)^2}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \ ; \ \mathbf{G} = \frac{-2}{(1-\theta_2-\theta_3)^3}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Although there is no apparent unit root effect for $\widehat{\theta}$, there is one for $\widehat{CIR}$, for the obvious reason that $CIR = \infty$ when $\theta_2 + \theta_3 = 1$. Clearly, if $\theta_2 + \theta_3$ is close to 1, the moments of $\widehat{CIR}$ could be very large. The asymptotic standard deviation of $g(\widehat{\theta}) - g(\theta)$ is $\{\mathbf{g}'\boldsymbol{\Omega}\mathbf{g}\}^{1/2}T^{-1} = O([1-\theta_2-\theta_3]^{-2}T^{-1})$ as $\theta_2 + \theta_3 \to 1$, since $\boldsymbol{\Omega}$ is finite for the unit root case, while the asymptotic bias and skewness are $O([1-\theta_2-\theta_3]^{-3}T^{-1})$ and $O([1-\theta_2-\theta_3]^{-7}T^{-1})$ respectively.

2. Bias Reduction

One practical application of our theorem is to bias reduction.[9] Andrews (1993) defines an exact median bias correction for least squares estimators in an autoregressive model. His procedure is based on inverting the relationship between the bias (computed by simulation) and the single parameter being estimated. This method is not well suited to the multiple parameter situation.[10] An alternative bias reduction method based on the asymptotic approximations of section 3 is now given. These approximate methods, at least the additive versions, are somewhat standard – see Cox and Hinkley (1974, p310) for a discussion – but have not been applied in this context before. A major advantage of our method is that it can be implemented without detailed knowledge of the bias function $\gamma_1(\theta)$, all that is required is that one can estimate it consistently. This method also does not require knowledge of the error distribution.[11] Concentrating on mean bias correction,[12] let $\widehat{\gamma}_1 = \gamma_1(\widehat{\theta})$ be the estimated bias, where the function $\gamma_1(\cdot)$ is given in the theorem. Then, $\widehat{\gamma}_1$ consistently estimates γ_1. With this quantity we define the additive and multiplicative bias-corrected estimates[13]:

(1) $c'\widehat{\theta}^{abr} = c'\widehat{\theta} - T^{-1}\widehat{\gamma}_1$

(2) $c'\widehat{\theta}^{mbr} = c'\widehat{\theta}/(1 + \widehat{\gamma}_1/c'\widehat{\theta}T)$

(3) $c'\widehat{\theta}^{ebr} = c'\widehat{\theta}\exp(-\widehat{\gamma}_1/c'\widehat{\theta}T)$.

Then, $c'\widehat{\theta}^{abr}$, $c'\widehat{\theta}^{mbr}$, and $c'\widehat{\theta}^{ebr}$ are unbiased to order T^{-1}. Bias-corrected estimates of $g(\theta)$ are obtained by replacing $\widehat{\theta}$ by $g(\widehat{\theta})$ and $\widehat{\gamma}_1$ by $T^{-1}\{\widehat{\gamma}_1(\widehat{\mathbf{g}}'\widehat{\theta})+(\widehat{\kappa}_4+2)\mathbf{tr}(\widehat{\mathbf{G}}\widehat{\mathbf{M}}^{-1})/2\}$ in (1), (2) and (3). One advantage of the multiplicative corrections $\widehat{\theta}^{mbr}$ and $\widehat{\theta}^{ebr}$ is that they ensure positivity for the corrected estimate.

6 Concluding Remarks

The asymptotic bias of the QMLE seems to vary with the parameters in a nonlinear fashion, which would suggest that the linear response surface approach used by Engle et al. (1985) is flawed and certainly cannot work well over the entire parameter space. Rather surprisingly, the bias does not necessarily increase when the unit root is approached. Some caution is due when interpreting these results literally, but they are at least suggestive of some major differences in behavior compared with linear autoregression. In this sense, our results confirm the implications of Lumsdaine (1991), although our arguments have nothing to do with the "pile-up" phenomenon investigated in Lumsdaine (1995).

We derived explicitly the second order approximation to the distribution of a certain Wald test for the presence of ARCH effects under a sequence of local alternatives to homoskedasticity. The test should over-reject in small samples, by about 30% for sample size 100.

Finally, our calculations extend to linear regression model with exogenous regressors with only minor changes to the formulae by incorporating the moments of the regressors. The estimate of the regression parameters are symmetrically distributed as before. If a lagged dependent variable is included, however, we can expect the estimate of its parameter to be mean and median biased as it is without ARCH affects.

7 Appendix

The variance derivatives satisfy the following recursions:

$$
\begin{aligned}
v_{t;1} &= 1 + \theta_2 v_{t-1;1} = \theta_2^t v_{0;1} + \textstyle\sum_{s=0}^{t-1} \theta_2^s \\
v_{t;2} &= v_{t-1} + \theta_2 v_{t-1;2} = \theta_2^t v_{0;2} + \textstyle\sum_{s=0}^{t-1} \theta_2^s v_{t-1-s} \\
v_{t;3} &= \theta_2 v_{t-1;3} + (y_{t-1} - \beta)^2 = \theta_2^t v_{0;3} + \textstyle\sum_{s=0}^{t-1} \theta_2^s (y_{t-1-s} - \beta)^2 \\
v_{t;22} &= 2v_{t-1;2} + \theta_2 v_{t-1;22} = \theta_2^t v_{0;22} + 2\textstyle\sum_{s=0}^{t-1} \theta_2^s v_{t-1-s;2} \\
v_{t;12} &= v_{t-1;1} + \theta_2 v_{t-1;12} = \theta_2^t v_{0;12} + \textstyle\sum_{s=0}^{t-1} \theta_2^s v_{t-1-s;1}
\end{aligned}
$$

$$v_{t;23} = v_{t-1;3} + \theta_2 v_{t-1;23} = \theta_2^t v_{0;23} + \sum_{s=0}^{t-1} \theta_2^s v_{t-1-s;3}$$
$$v_{t;\beta} = \theta_2 v_{t-1;\beta} - 2\theta_3(y_{t-1} - \beta) = \theta_2^t v_{0;\beta} - 2\theta_3 \sum_{s=0}^{t-1} \theta_2^s (y_{t-1-s} - \beta)$$
$$v_{t;\beta\beta} = \theta_2 v_{t-1;\beta\beta} + 2\theta_3 = \theta_2^t v_{0;\beta\beta} + 2\theta_3 \sum_{s=0}^{t-1} \theta_2^s$$
$$v_{t;\beta 2} = v_{t-1;\beta} + \theta_2 v_{t-1;\beta 2} = \theta_2^t v_{0;\beta 2} + \sum_{s=0}^{t-1} \theta_2^s v_{t-1-s;\beta}$$
$$v_{t;\beta 3} = \theta_2 v_{t-1;\beta 3} - 2(y_{t-1} - \beta) = \theta_2^t v_{0;\beta 3} - 2\sum_{s=0}^{t-1} \theta_2^s (y_{t-1-s} - \beta)$$

while the remaining second derivatives are all zero. For our purposes, the third derivatives are not required. In the ARCH process, $\theta_2 = 0$, all second derivatives are zero. As far as the theoretical calculations are concerned, how the model is started up is irrelevant. In our computations we take all start-ups (observations dated 0) to be fixed numbers, either derived from the stationary start-up where relevant, in which case

$$v_0 = \theta_1/(1 - \theta_2 - \theta_3)$$
$$v_{0;1} = 1/(1 - \theta_2)$$
$$v_{0;2}, v_{0;3} = v_0/(1 - \theta_2)$$
$$v_{0;22} = 2v_0/(1 - \theta_2)^2$$
$$v_{0;12} = 1/(1 - \theta_2)^2$$
$$v_{0;23} = v_0/(1 - \theta_2)^2$$
$$v_{0;\beta\beta} = 2\theta_3/(1 - \theta_2),$$

with all other start-ups zero, or some arbitrary quantity.

Proof of Theorem. We first derive the asymptotic properties of $\widehat{\theta}(\beta)$, the pure GARCH procedure where β is known and not estimated. We then examine the QMLE $\widehat{\theta}(\widehat{\beta})$, incorporating the effects of estimating β. Finally, we consider $\widetilde{\theta}(\widetilde{\beta})$.

We use throughout the summation convention that repeated indices in an expression are to be summed over, e.g. $\tau^{ij}\tau^{jl} = \sum_j \tau^{ij}\tau^{jl}$.

1. *Properties of* $\widehat{\theta}(\beta)$. Ignoring the inequality restrictions[14], the standardized estimators, derived from choosing θ to solve $\mathcal{L}_i(\theta, \beta) = 0$, $i = 1, 2, 3$, have the following stochastic expansions (see McCullagh (1987, p209)):

$$T^{1/2}\left\{\widehat{\theta}_i(\beta) - \theta_i\right\} \approx -\tau^{ij}\mathcal{Z}_j + T^{-1/2}\left\{\tau^{ij}\tau^{kl}\mathcal{Z}_{jk}\mathcal{Z}_l - \tau^{ij}\tau^{kl}\tau^{mn}\tau_{jln}\mathcal{Z}_k\mathcal{Z}_m/2\right\}, \quad (A1.1)$$

where $\mathcal{Z}_j = T^{-1/2}\mathcal{L}_j$ and $\mathcal{Z}_{jk} = T^{-1/2}\left\{\mathcal{L}_{jk} - E(\mathcal{L}_{jk})\right\}$ are evaluated at the true parameters and are jointly asymptotically normal. The truncation error in (A1.1) is $O_P(T^{-1})$.

1a. *Asymptotic bias of* $\widehat{\theta}(\beta)$. Taking expectations of the right hand side of (A1.1), and using (3.1) and symmetry of τ^{lk}, we get

$$E\left[T^{1/2}c'\left\{\widehat{\theta}(\beta) - \theta\right\}\right] \approx T^{-1/2}c_i\tau^{ij}\tau^{kl}\{\tau_{jk,l} + \tau_{jkl}(\kappa_4 + 2)/4\}, \quad (A1.2)$$

by a cancellation of $\tau^{kl}\tau_{km}\tau^{mn}$. We substitute for the $\tau's$ in terms of the $\mu's$, using, for example, that $\mu^{i,j}\mu^{k,l}\mu_{jk,l} = \mu^{i,j}\mu^{k,l}\mu_{jl,k}$ for any i, by symmetry of $\mu^{k,l}$. Let $a_0 = c_i\mu^{i,j}\mu^{k,l}\mu^{\varepsilon\varepsilon}_{l:k,j}$, $a_1 = c_i\mu^{i,j}\mu^{k,l}\mu_{jl,k}$, $a_2 = c_i\mu^{i,j}\mu^{k,l}\mu_{kl,j}$ and $a_3 = c_i\mu^{i,j}\mu^{k,l}\mu_{j,k,l}$. Then

$$c_i\tau^{ij}\tau^{kl}\tau_{jk,l} = (\kappa_4+2)(a_1 - a_3) - a_0 \quad ; \quad c_i\tau^{ij}\tau^{kl}\tau_{jkl}\frac{\kappa_4+2}{4} = -\frac{\kappa_4+2}{2}(2a_1 + a_2 - a_3).$$

We get a cancellation of the terms involving a_1, so that the asymptotic bias of $c'\widehat{\theta}$ is approximately $-T^{-1}\left\{\frac{\kappa_4+2}{2}(a_2 + a_3) + a_0\right\}$. We now switch to matrix notation, so that $a_0 = T^{-1}\sum\sum_{s<t} E\left\{(\varepsilon_s^2 - 1)c'\mathbf{M}^{-1}\mathbf{h}_{\theta t}\mathbf{h}'_{\theta t}\mathbf{M}^{-1}\mathbf{h}_{\theta s}\right\}$, $a_2 = T^{-1}\sum_{t=1}^T E\left\{\mathbf{tr}(\mathbf{M}^{-1}\mathbf{H}_{\theta\theta t})c'\mathbf{M}^{-1}\mathbf{h}_{\theta t}\right\}$

and $a_3 = T^{-1}\sum_{t=1}^{T} E\left\{\mathbf{tr}(\mathbf{M}^{-1}\mathbf{h}_{\theta t}\mathbf{h}'_{\theta t})c'\mathbf{M}^{-1}\mathbf{h}_{\theta t}\right\}$, where $\mathbf{H}_{\theta\theta t} = (h_{t;ij})_{i,j=1}^{3}$. Then, since $\mathbf{H}_{\theta\theta t} = \mathbf{D}_t - \mathbf{h}_{\theta t}\mathbf{h}'_{\theta t}$,

$$a_2 + a_3 = T^{-1}\sum_{t=1}^{T} E\left\{\mathrm{tr}(\mathbf{M}^{-1}\mathbf{D}_t)c'\mathbf{M}^{-1}\mathbf{h}_{\theta t}\right\}.$$

Therefore, the bias of $c'\widehat{\theta}(\beta)$ is $-(\lambda_0 + \lambda_2)/T$.

1b. *Skewness of* $\widehat{\theta}(\beta)$. We first recentre the standardized estimator. Let $E_\infty\{T^{1/2}[\widehat{\theta}_i(\beta) - \theta_i]\}$ be the expectation of the truncated standardized estimator given by the right hand side of (A1.1), and let

$$\mathbb{P}_i = T^{1/2}\left\{\widehat{\theta}_i(\beta) - \theta_i\right\} - E_\infty\left[T^{1/2}\left\{\widehat{\theta}_i(\beta) - \theta_i)\right\}\right] = \mathbb{A}_i + T^{-1/2}\mathbb{B}_i,$$

where $\mathbb{A}_i = -\tau^{ij}\mathcal{Z}_j$ and $\mathbb{B}_i = \mathbb{C}_i + \mathbb{D}_i$, with $\mathbb{C}_i = \tau^{ij}\tau^{kl}\{\mathcal{Z}_{jk}\mathcal{Z}_l - E(\mathcal{Z}_{jk}\mathcal{Z}_l)\}$ and $\mathbb{D}_i = -\tau^{ij}\tau^{kl}\tau^{mn}\tau_{jln}\{\mathcal{Z}_k\mathcal{Z}_m - E(\mathcal{Z}_k\mathcal{Z}_m)\}/2$. Thus

$$E(\mathbb{P}_{i_1}\mathbb{P}_{i_2}\mathbb{P}_{i_3}) \approx E(\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{A}_{i_3}) + T^{-1/2}\{E(\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{B}_{i_3}) + E(\mathbb{B}_{i_1}\mathbb{A}_{i_2}\mathbb{A}_{i_3}) + E(\mathbb{A}_{i_1}\mathbb{B}_{i_2}\mathbb{A}_{i_3})\},$$

where

$$\begin{aligned}
E(\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{A}_{i_3}) &= -\tau^{i_1j_1}\tau^{i_2j_2}\tau^{i_3j_3}E(\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\mathcal{Z}_{j_3}) \\
E(\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{C}_{i_3}) &= \tau^{i_1j_1}\tau^{i_2j_2}\tau^{i_3j_3}\tau^{k_3l_3}E\left[\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\left\{\mathcal{Z}_{j_3k_3}\mathcal{Z}_{l_3} - E(\mathcal{Z}_{j_3k_3}\mathcal{Z}_{l_3})\right\}\right] \\
E(\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{D}_{i_3}) &= -\tfrac{1}{2}\tau^{i_1j_1}\tau^{i_2j_2}\tau^{i_3j_3}\tau^{k_3l_3}\tau^{m_3n_3}\tau_{j_3l_3n_3}E\left[\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\left\{\mathcal{Z}_{k_3}\mathcal{Z}_{m_3} - E(\mathcal{Z}_{k_3}\mathcal{Z}_{m_3})\right\}\right].
\end{aligned}$$

Now,

$$E(\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\mathcal{Z}_{j_3}) = T^{-1/2}\kappa_{23}\mu_{j_1,j_2,j_3}/8,$$

by direct calculation. Next, using the joint asymptotic normality of $\mathcal{Z}_{j_3k_3}$ and $\mathcal{Z}_{l_3}$,

$$E\left[\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\left\{\mathcal{Z}_{j_3k_3}\mathcal{Z}_{l_3} - E(\mathcal{Z}_{j_3k_3}\mathcal{Z}_{l_3})\right\}\right] = \tau_{j_1,l_3}\tau_{j_3k_3,j_2} + \tau_{j_1,j_3k_3}\tau_{j_2,l_3} + o(1)$$

$$E\left[\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\left\{\mathcal{Z}_{k_3}\mathcal{Z}_{m_3} - E(\mathcal{Z}_{k_3}\mathcal{Z}_{m_3})\right\}\right] = \tau_{j_1,k_3}\tau_{j_2,m_3} + \tau_{j_1,m_3}\tau_{j_2,k_3} + o(1).$$

Therefore, substituting $\mu's$ for $\tau's$, and switching to matrix notation,

$$\begin{aligned} E(c_{i_1}c_{i_2}c_{i_3}\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{A}_{i_3}) &= T^{-1/2}\kappa_{23}c_{i_1}c_{i_2}c_{i_3}\mu^{i_1,j_1}\mu^{i_2,j_2}\mu^{i_3,j_3}\mu_{j_1,j_2,j_3} \\ &= T^{-3/2}\textstyle\sum_{t=1}^{T}\kappa_{23}E\left\{(c'\mathbf{M}^{-1}\mathbf{h}_{\theta t})^3\right\} \end{aligned}$$

while

$$\begin{aligned} E(c_{i_1}c_{i_2}c_{i_3}\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{D}_{i_3}) &= -(\kappa_4+2)^2 c_{i_1}c_{i_2}c_{i_3}\mu^{i_1,l_3}\mu^{i_3,j_3}\mu^{i_2,n_3}\times \\ &\quad \{\mu_{j_3l_3,n_3} + \mu_{j_3n_3,l_3} + \mu_{l_3n_3,j_3} - \mu_{j_3,l_3,n_3}\} \\ &= -(\kappa_4+2)^2 T^{-1}\textstyle\sum_{t=1}^{T}E\left\{3c'\mathbf{M}^{-1}\mathbf{h}_{\theta t}\mathbf{M}^{-1}cc'\mathbf{M}^{-1}\mathbf{h}_{\theta t} - (c'\mathbf{M}^{-1}\mathbf{h}_{\theta t})^3\right\}, \end{aligned}$$

and

$$\begin{aligned} E(c_{i_1}c_{i_2}c_{i_3}\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{C}_{i_3}) &= (\kappa_4+2)^2 c_{i_1}c_{i_2}c_{i_3}\{\mu^{i_1,k_3}\mu^{i_3,j_3}\mu^{j_2,i_2}\mu_{j_3k_3,j_2} - \mu^{i_3,j_3}\mu^{i_1,k_3}\mu^{i_2,j_2}\mu_{j_3,k_3,j_2} + \\ &\quad \mu^{i_3,j_3}\mu^{i_2,k_3}\mu^{i_1,j_1}\mu_{j_3k_3,j_1} - \mu^{i_3,j_3}\mu^{i_1,j_1}\mu^{i_2,k_3}\mu_{j_3,k_3,j_1}\} - \\ &\quad 2c_{i_1}c_{i_2}c_{i_3}\{\mu^{i_3,j_3}\mu^{i_2,j_2}\mu^{i_1,k_3}\mu^{\varepsilon\varepsilon}_{j_2:k_3,j_3} + \mu^{i_3,j_3}\mu^{i_2,k_3}\mu^{i_1,j_1}\mu^{\varepsilon\varepsilon}_{j_1:k_3,j_3}\} \\ &= 2(\kappa_4+2)^2 T^{-1}\textstyle\sum_{t=1}^{T}E\left\{c'\mathbf{M}^{-1}\mathbf{h}_{\theta t}\mathbf{M}^{-1}cc'\mathbf{M}^{-1}\mathbf{h}_{\theta t} - (c'\mathbf{M}^{-1}\mathbf{h}_{\theta t})^3\right\} - \\ &\quad 2(\kappa_4+2)T^{-1}\sum_{s<t}\sum E\left\{(c'\mathbf{M}^{-1}\mathbf{h}_{\theta t})^2 c'\mathbf{M}^{-1}\mathbf{h}_{\theta s}(\varepsilon_s^2 - 1)\right\}. \end{aligned}$$

Again we get a cancellation of terms involving $(c'\mathbf{M}^{-1}\mathbf{h}_{\theta t})^3$, and the skewness of $c'\widehat{\theta}(\beta)$ is approximately $T^{-1}\left\{\lambda_4 - \lambda_3 - \lambda_5\right\}$.

2. *Properties of* $\widehat{\theta}(\widehat{\beta})$. We have

$$T^{1/2}\left\{\widehat{\theta}_i(\widehat{\beta})-\theta_i\right\}-T^{1/2}\left\{\widehat{\theta}_i(\beta)-\theta_i\right\}\approx T^{-1/2}\left\{\tau^{ij}\tau^{\beta\beta}\mathcal{Z}_{j\beta}\mathcal{Z}_{\beta}-\tau^{ij}\tau^{\beta\beta}\tau^{\beta\beta}\tau_{j\beta\beta}\mathcal{Z}_{\beta}^2/2\right\}, \tag{A1.3}$$

because $\tau_{k\beta}, \tau_{jl\beta}, \tau_{j\beta n} = 0$.

2a. *Asymptotic bias of* $\widehat{\theta}(\widehat{\beta})$. We take expectations of the right hand side of (A1.3), and substitute for the τ's, finding

$$\begin{aligned}\tau^{ij}\tau^{\beta\beta}\tau_{j\beta,\beta} &= \tau_{\beta\beta}^{-1}\mu^{i,j}\{4\overline{\pi}_j+\mu^{\varepsilon\varepsilon}_{\beta:j,\beta}+2\mu^{\varepsilon v}_{j,\beta}-(\kappa_4+2)(\mu_{j\beta,\beta}-\mu_{j,\beta,\beta})\}/2,\\ -\tau^{ij}\tau^{\beta\beta}\tau^{\beta\beta}\tau_{j\beta\beta}\tau_{\beta,\beta}/2 &= \tau_{\beta\beta}^{-2}\tau_{\beta,\beta}\mu^{i,j}\{2\overline{\pi}_j-(\mu_{j,\beta\beta}+2\mu_{j\beta,\beta}-\mu_{j,\beta,\beta})\}/2.\end{aligned}$$

We now collect terms, finding the coefficients on

$\overline{\pi}_j$:	$\tau_{\beta\beta}^{-1}\{2+\tau_{\beta\beta}^{-1}\tau_{\beta,\beta}\}$;	ρ_4
$\mu^{\varepsilon\varepsilon}_{\beta:j,\beta}$:	$\tau_{\beta\beta}^{-1}/2$	
$\mu^{\varepsilon v}_{j,\beta}$:	$\tau_{\beta\beta}^{-1}$	
$\mu_{j\beta,\beta}$:	$\tau_{\beta\beta}^{-1}\{-\frac{\kappa_4+2}{2}+\tau_{\beta\beta}^{-1}\tau_{\beta,\beta}\}/2$;	ρ_3
$\mu_{j,\beta,\beta}$:	$\tau_{\beta\beta}^{-1}\{\frac{\kappa_4+2}{2}-\tau_{\beta\beta}^{-1}\tau_{\beta,\beta}/2\}$;	ρ_2
$\mu_{j,\beta\beta}$:	$-\tau_{\beta\beta}^{-2}\tau_{\beta,\beta}/2$;	ρ_1

along with their common factor of $T^{-1/2}\mu^{i,j}$. Combining with the pure GARCH bias, we get the required result for the bias of $c'\widehat{\theta}(\widehat{\beta})$.

2b. *Skewness of* $\widehat{\theta}(\widehat{\beta})$. We have to calculate terms like $E(\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{E}_{i_3})$, where

$$\mathbb{E}_{i_3}=T^{-1/2}\tau^{i_3j}\tau^{\beta\beta}\left\{\mathcal{Z}_{j\beta}\mathcal{Z}_{\beta}-E(\mathcal{Z}_{j\beta}\mathcal{Z}_{\beta})\right\}-T^{-1/2}\tau^{i_3j}\tau^{\beta\beta}\tau^{\beta\beta}\tau_{j\beta\beta}\left\{\mathcal{Z}_{\beta}^2-E(\mathcal{Z}_{\beta}^2)\right\}/2.$$

By asymptotic normality

$$E\left[\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\left\{\mathcal{Z}_{j_3\beta}\mathcal{Z}_\beta - E(\mathcal{Z}_{j_3\beta}\mathcal{Z}_\beta)\right\}\right] = \tau_{j_2,\beta}\tau_{j_3\beta,j_1} + \tau_{j_1,\beta}\tau_{j_1,j_3\beta} + o(1)$$

$$E\left[\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\left\{\mathcal{Z}_\beta^2 - E(\mathcal{Z}_\beta^2)\right\}\right] = 2\tau_{j_1,\beta}\tau_{j_2,\beta} + o(1),$$

but, because $E(\mathcal{Z}_j\mathcal{Z}_\beta) = 0$, the right hand sides are both zero. Therefore, the skewness of $\widehat{\theta}(\widehat{\beta})$ is the same as that of $\widehat{\theta}(\beta)$ to order T^{-1}.

3. *The properties of* $\widetilde{\theta}$. In this case, the truncated stochastic expansion is

$$T^{1/2}c'(\widetilde{\theta}-\theta) \approx -c_i\widetilde{\tau}^{ij}\widetilde{\mathcal{Z}}_j + T^{-1/2}\left\{c_i\widetilde{\tau}^{ij}\widetilde{\tau}^{kl}\widetilde{\mathcal{Z}}_{jk}\widetilde{\mathcal{Z}}_l - c_i\widetilde{\tau}^{ij}\widetilde{\tau}^{kl}\widetilde{\tau}^{mn}\widetilde{\tau}_{jln}\widetilde{\mathcal{Z}}_k\widetilde{\mathcal{Z}}_m/2\right\},$$

with approximation error of order T^{-1}. Here, tildes denote dependency on $\widetilde{\beta}$. In the order $T^{-1/2}$ terms we can replace $\widetilde{\beta}$ by β with error that is of order T^{-1}. However, we must further expand the leading term in a second order Taylor series in $\widetilde{\beta}-\beta$. We obtain

$$T^{1/2}c'(\widetilde{\theta}-\theta) - T^{1/2}c'\left\{\widehat{\theta}(\beta)-\theta\right\} \approx -T^{-1/2}c_i\tau^{ij}\mathcal{Z}_{j\beta}\mathcal{W} - T^{-1/2}c_i\tau^{ij}\tau_{j\beta\beta}\mathcal{W}^2/2, \quad (A1.4)$$

where $\mathcal{Z}_{j\beta} = T^{-1/2}\mathcal{L}_{j\beta}$ and $\mathcal{W} = T^{1/2}(\widetilde{\beta}-\beta) = T^{-1/2}\sum_{t=1}^T \varepsilon_t v_t^{1/2}$ are both $O_p(1)$. The other terms: $T^{-1/2}c_i\tau^{ij}\tau_{jk\beta}\tau^{kl}\mathcal{Z}_l\mathcal{W}$, $T^{-1}c_i\tau^{ij}\tau_{jk\beta}\tau^{kl}\tau_{lm\beta}\tau^{mn}\mathcal{Z}_n\mathcal{W}^2$, $T^{-1}c_i\tau^{ij}\tau_{jk\beta\beta}\tau^{kl}\mathcal{Z}_l\mathcal{W}^2/2$ and $T^{-1}c_i\tau^{ij}\tau_{jk\beta}\tau^{kl}\mathcal{Z}_{l\beta}\mathcal{W}^2$ are $o_P(T^{-1/2})$ because $\tau_{j\beta}, \tau_{jk\beta} = 0$, by symmetry. Furthermore, $E(\mathcal{Z}_j\mathcal{W}) = 0$.

3a. *Asymptotic bias of* $\widetilde{\theta}$. Firstly, we have that $E(\mathcal{W}^2) = \overline{v}$ and $E(\mathcal{Z}_{j\beta}\mathcal{W}) = -\mu_j - \frac{1}{2}T^{-1}\sum\sum_{s<t} E\left\{h_{t;j}h_{t;\beta}\varepsilon_s v_s^{1/2}\right\}$. We take expectations of the right hand side of (A1.4), and substitute for the τ's, finding

$$-\tau^{ij}\mathcal{Z}_{j\beta}\mathcal{W} = -2\mu^{i,j}\left[\mu_j + \tfrac{1}{2}T^{-1}\sum\sum_{s<t} E\left\{h_{t;j}h_{t;\beta}\varepsilon_s v_s^{1/2}\right\}\right]$$
$$-\tau^{ij}\tau_{j\beta\beta}\mathcal{W}^2/2 = \mu^{i,j}\overline{v}\left[2\overline{\pi}_j - \left\{\mu_{j,\beta\beta} + 2\mu_{\beta j,\beta} - \mu_{\beta,j,\beta}\right\}/2\right].$$

We have to add both these terms to the bias of $c'\widehat{\theta}(\beta)$, in order to get the bias of $c'\widetilde{\theta}$.

3b. *Skewness of* $\widetilde{\theta}$. We have to calculate terms like $T^{-1/2}E(\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{G}_{i_3})$, $T^{-1/2}E(\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{H}_{i_3})$, and $T^{-1/2}E(\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{K}_{i_3})$,where $\mathbb{G}_i = \sum_{j,k,l}\tau^{ij}\tau_{jk\beta\beta}\tau^{kl}\mathcal{Z}_l\mathcal{W}/2$ (note that $E(\mathcal{Z}_l\mathcal{W}) = 0$), $\mathbb{H}_i = -\sum_j \tau^{ij}\{\mathcal{Z}_{j\beta}\mathcal{W} - E(\mathcal{Z}_{j\beta}\mathcal{W})\}$, and $\mathbb{K}_i = -\sum_j \tau^{ij}\tau_{j\beta\beta}\{\mathcal{W}^2 - E(\mathcal{W}^2)\}/2$. We have

$$E(\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{G}_{i_3}) = \tau^{i_1j_1}\tau^{i_2j_2}\tau^{i_3j_3}\tau_{j_3k_3\beta\beta}\tau^{k_3l_3}E(\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\mathcal{Z}_{l_3}\mathcal{W})$$
$$E(\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{H}_{i_3}) = -\tau^{i_1j_1}\tau^{i_2j_2}\tau^{i_3j_3}E[\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\{\mathcal{Z}_{j_3\beta}\mathcal{W} - E(\mathcal{Z}_{j_3\beta}\mathcal{W})\}]$$
$$E(\mathbb{A}_{i_1}\mathbb{A}_{i_2}\mathbb{K}_{i_3}) = -\tau^{i_1j_1}\tau^{i_2j_2}\tau^{i_3j_3}\tau_{j_3\beta\beta}E[\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\{\mathcal{W}^2 - E(\mathcal{W}^2)\}]/2.$$

Now,

$$\begin{aligned} E(\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\mathcal{Z}_{l_3}\mathcal{W}) &= o(1) \\ E\left[\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\left\{\mathcal{Z}_{j_3\beta}\mathcal{W} - E(\mathcal{Z}_{j_3\beta}\mathcal{W})\right\}\right] &= o(1) \\ E\left[\mathcal{Z}_{j_1}\mathcal{Z}_{j_2}\left\{\mathcal{W}^2 - E(\mathcal{W}^2)\right\}\right] &= o(1), \end{aligned}$$

by joint asymptotic normality, and that $\mathcal{W}$ is uncorrelated with $\mathcal{Z}_j$, for $j = 1, 2, 3$. Therefore, the skewness of $\widetilde{\theta}$ is the same as that of $\widehat{\theta}(\beta)$.

4. *The properties of* $\widehat{\beta}$. In this case,

$$T^{1/2}(\widehat{\beta}-\beta) = -\tau^{\beta\beta}\mathcal{Z}_\beta + T^{-1/2}\left\{\tau^{\beta\beta}\tau^{kl}\mathcal{Z}_{\beta k}\mathcal{Z}_l - \tau^{\beta\beta}\tau^{\beta\beta}\tau^{kl}\tau_{\beta\beta l}\mathcal{Z}_\beta\mathcal{Z}_k\right\},$$

because $\tau_{k\beta}, \tau_{\beta kl}, \tau_{\beta\beta\beta} = 0$. Therefore, both the mean and skewness of $T^{1/2}(\widehat{\beta}-\beta)$ are zero to order T^{-1}, since also $\tau_{\beta k,l} = 0$. ∎

References

Akahira, M., and K. Takeuchi (1981): *Asymptotic efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency.* Springer-Verlag, Berlin.

Anderson, T.W. (1974): "An asymptotic expansion of the distribution of the limited information maximum likelihood estimate of a coefficient in a simultaneous equation system," *Journal of the American Statistical Association*, **69**, 565-73.

Anderson, T.W. (1977): "Asymptotic expansions of the distributions of estimates in simultaneous equations for alternative parameter sequences," *Econometrica*, **45**, 509–18.

Anderson, T.W. and T. Sawa (1973): "Distributions of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansion," *Econometrica*, **41**, 683-714.

Anderson, T.W. and T. Sawa (1979): "Evaluation of the distribution function of the two stage least squares estimate," *Econometrica* **47**, 163-82.

Andrews, D.W.K., (1993): "Exactly median unbiased estimation of first order autoregressive/unit root models." *Econometrica* **61**, 139-166.

Baillie, R.T. and T. Bollerslev (1992): "Prediction in dynamic models with time-dependent conditional variances." *Journal of Econometrics.* **52**, 91-113.

Barndorff-Nielsen, O.E., and D.R. Cox (1989): *Asymptotic techniques for use in statistics.* Chapman and Hall.

Bartlett, M.S. (1953): "Approximate confidence intervals II," *Biometrika* **40**, 306-317.

Bollerslev, T., R.F. Engle, and D.B. Nelson (1993): "ARCH Models," in *The Handbook of Econometrics*, Vol. IV. Eds D.F. McFadden and R.F. Engle III. North-Holland.

Bollerslev, T., and J.M. Wooldridge (1992): "Quasi Maximum Likelihood Estimation and inference in dynamic models with time varying covariates." *Econometric Reviews.* **11**, 143-172.

Cooley, T.F. and W.R. Parke (1990): "Asymptotic likelihood-based prediction functions." *Econometrica* **58**, 1215-1234.

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

Engle, R.F., D.F. Hendry, and D. Trumble (1985): "Small sample properties of ARCH estimators and tests." *Canadian Journal of Economics.* **18**, 66-93.

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

Ghosh, J.K., B.K. Sinha, and H.S. Wieand (1980): "Second order efficiency of the MLE with respect to any bowl shaped loss function." *Annals of Statistics.* **8**, 506-521.

Hajivassiliou, V.A., and P.A. Ruud (1994): "Classical Estimation methods for LDV Models using Simulation," in *The Handbook of Econometrics*, Vol. IV. eds. D.F. McFadden and R.F. Engle III. North Holland. p2384-2441.

Haldane J.B.S., and S.M. Smith (1956): "The sampling distribution of a maximum likelihood estimate," *Biometrika*, **63**, 96-103.

Härdle, W., and O.B. Linton (1994): "Applied nonparametric methods." in *The Handbook of Econometrics*, Vol. IV. eds. D.F. McFadden and R.F. Engle III. North Holland. p2297-2339.

Hendry, D.F. (1986): Discussion of Engle, R.F. and T. Bollerslev, " Modelling the persistence of conditional variances." *Econometric Reviews* **5**, 1-50.

Lee, S-W., and B.E. Hansen (1994): "Asymptotic theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator." *Econometric Theory* **10**, 29-52.

Linton, O.B., and J.P. Nielsen (1995): "Kernel Estimation of Structured Nonparametric Regression Models," *Biometrika* **82**, 93-100.

Lumsdaine, R.L. (1991): "Consistency and asymptotic normality of the Quasi-Maximum Likelihood estimator in GARCH(1,1) and IGARCH(1,1) models." Manuscript, Princeton University.

Lumsdaine, R.L. (1995): "Finite sample properties of the Maximum Likelihood estimator in GARCH(1,1) and IGARCH(1,1) models: a monte carlo investigation." *Journal of Business and Economic Statistics* **13**, 1-10.

McCullagh, P., (1987): *Tensor methods in statistics.* Chapman and Hall.

Malinovskii, V.K. (1994): "Asymptotic expansions in sequential estimation for the first-order random coefficient autoregressive model: a regenerative approach." *Acta Applicandae Mathematicae.* **34**, 261-281.

Nagar, A.L. (1959): "The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations," *Econometrica*, **27**, 573-95.

Nelson, D.B. (1990): "Stationarity and persistence in the GARCH(1,1) models." *Econometric Theory.* **6**, 318-334.

Nelson, D.B. (1991): "Conditional heteroskedasticity in asset returns: A new approach." *Econometrica.* **59**, 347-470.

Pfanzagl, J. (1980): "Asymptotic expansions in parametric statistical theory," in P.R. Krishnaiah, ed., *Developments in Statistics*, vol 3. New York: Academic Press.

Phillips, P.C.B., (1977a): "Approximations to some finite sample distributions associated with a first-order stochastic difference equation." *Econometrica.* **45**, 463-485.

Phillips, P.C.B., (1977b): "A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators." *Econometrica.* **45**, 1517-34.

Robinson, P.M. (1988): "Root-n consistent semiparametric regression." *Econometrica.* **56**, 931-954.

Rothenberg, T.J., (1984a): "Approximate normality of Generalized Least Squares Estimates," *Econometrica* **52**, 811-825.

Rothenberg, T.J., (1984b): "Hypothesis testing in Linear Models when the Error Covariance Matrix is Nonscalar," *Econometrica* **52**, 827-842.

Rothenberg, T.J., (1986): "Approximating the distributions of econometric estimators and test statistics." In *The Handbook of Econometrics*, Volume II. North-Holland.

Rothenberg, T.J., (1988): "Approximate Power Functions for some Robust Tests of Regression Coefficients." *Econometrica.* **56**, 997-1019.

Sargan, J.D., (1974): "On the validity of Nagar's expansion for the moments of econometric estimators," *Econometrica*, **42**, 169-76

Sargan, J.D., (1976): "Econometric estimators and the Edgeworth expansion," *Econometrica*, **44**, 421-448; erratum, **45**, 272.

Shenton, L.R., and K. Bowman (1963): "Higher moments of a maximum-likelihood estimate," *Journal of the Royal Statistical Society B*, **25**, 305-17.

Shenton, L.R., and K. Bowman (1969): "Maximum likelihood estimate moments for the two-parameter gamma distribution," *Sankha B*, **31**, 379–96.

Shenton, L.R., and K. Bowman (1977): *Maximum Likelihood Estimation in Small Samples.* London: Griffin.

Taniguchi, M., and K. Maekawa (1990): "Asymptotic expansions of the distribution of statistics related to the spectral density matrix in multivariate time series and their applications," *Econometric Theory* **6**, 75-96.

Taniguchi, M. (1991): *Higher order asymptotic theory for time series analysis*, Springer Verlag, Berlin.

Weiss, A.A. (1986): "Asymptotic theory for ARCH models: Estimation and testing." *Econometric Theory.* **2**, 107-131.

NOTES

0. I would like to thank R. Lumsdaine for helpful comments on an earlier version of this paper and T. Bollerslev for some useful references. Thanks also to Vsevolod Malinowskii for sending me copies of some of his work. Thanks are also due to a referee, the co-Editor, and the Editor himself for correcting some mistakes. Financial support from the National Science Foundation is gratefully acknowledged. Address correspondence to: Oliver Linton, Cowles Foundation for Research in Economics, Yale University, Yale Station Box 208281, New Haven, CT 06520-8281, USA. Phone: (203) 432-3699. Fax: (203) 432-6167. http://www.econ.yale.edu/~linton.

1. In this case, the variance process is strongly stationary (Nelson (1990)), but the unconditional variance is infinite.

2. Under Gaussianity, the information matrix is block diagonal between θ and β, and $\widetilde{\theta}$ procedure is fully efficient. When ε_t is not Gaussian, $\widetilde{\theta}$ and $\widehat{\theta}$ are both consistent but not fully efficient.

3. See Lemmas 1-5 of Lumsdaine (1991).

4. The double summation terms, $\lambda_0, \lambda_{10}, \lambda_{10}^*$, and λ_5, arise from the fact that the centred Hessian is not a martingale. That these quantities are finite should follow from the asymptotic independence of the relevant stochastic processes.

5. In the stationary AR(1) model, $y_t = \rho y_{t-1} + \varepsilon_t$, where $|\rho| < 1$, the asymptotic bias of the least squares estimate of ρ is $-2\rho T^{-1}$ (relative to the asymptotic standard deviation it is $-2T^{-1}\rho/\sqrt{1-\rho^2}$), see Phillips (1977a).

6. Or rather, are not present.

7. Also, note that $\gamma_1^* = \gamma_1$ here.

8. The biases standardized by the asymptotic standard deviations $\omega_{ii}^{1/2} = \sqrt{2\mu^{i,i}}$ gives essentially the same pattern.

9. Although not directly relevant to the usual optimality criteria in regular parametric problems, since biased estimates can have smaller MSE than unbiased ones, there are some advantages of reduced bias. Impartiality itself may be desirable, for example in the context of testing for unit roots, see Andrews (1993). In situations with large numbers of nuisance parameters – for example, when estimating the error variance in fixed effect panel data regression models with large n but small T – it may be necessary to correct the MLE for bias otherwise it can be inconsistent. In nonparametric and semiparametric estimation problems, bias reduction can improve convergence rates, see Härdle and Linton (1994) and Robinson (1988).

10. Computationally, the Andrews' method is also very demanding here because of the iterative procedure used to estimate the parameters.

11. When the error distribution is Gaussian, they possess some large sample optimality: for example, Ghosh, Sinha and Wieand (1980) prove that bias-reduced MLE's are second order MSE optimal amongst all second order unbiased estimators. If one insists on having an unbiased estimator then the adjusted MLE is best.

12. The median bias of $c'\widehat{\theta}$ is approximately $(6\gamma_1 - \gamma_3)/6T$.

13. Firth (1993) recommends bias correction by modifying the score function and even the likelihood itself directly.

14. The inequality restrictions can be ignored because the event that they are violated is of exponentially small order of magnitude for the Edgeworth measure that serves as an $O(T^{-1/2})$ approximation to the distribution of the unrestricted estimate.

FIGURES 1-3. Bias of $\widehat{\theta}_1, \widehat{\theta}_2$, and $\widehat{\theta}_3$ against θ_2 and θ_3, with: $\theta_1 = 0.1$, $T = 100$, weakly stationary design, Gaussian errors, location not estimated.

FIGURE 4. Univariate effects on biases of $\widehat{\theta}_1, \widehat{\theta}_2$, and $\widehat{\theta}_3$ of θ_2 and θ_3. Same design as before.

FIGURE 5. Unit root design. Bias of $\widehat{\theta}_1, \widehat{\theta}_2$, and $\widehat{\theta}_3$ against θ_2 and θ_3, with: $v_0 = 1$ and rest as before.