A simple efficient GMM estimator of GARCH models

Jimmy Skoglund*
Department of Economic statistics,
Stockholm School of Economics
Box 6501, SE-113 83 Stockholm, Sweden

Stockholm School of Economics
SSE/EFI Working Paper Series in Economics and Finance
No. 434

February 2001

Abstract

This paper is concerned with efficient GMM estimation and inference in GARCH models. Sufficient conditions for the estimator to be consistent and asymptotically normal are established for the GARCH(1,1) conditional variance process. In addition efficiency results are obtained in the general framework of the GARCH(1,1)-M regression model.

Keywords: GARCH; GARCH-M; efficient GMM.
JEL codes: C12; C13; C22

*Jimmy.Skoglund@hhs.se

1 Introduction

Unconditional return distributions are characterized by the 'stylized facts' of excess kurtosis, high peakedness and are often skewed, see Mills (1999, ch 5). But the conditional distribution is also characterized by excess kurtosis and skewness when the ARCH model of Engle (1982) or its generalization to GARCH by Bollerslev (1986) are fitted to return series, see e.g. French, Schwert and Staumbaugh (1987), Engle and Gonzales-Rivera (1991).

GARCH models are typically estimated by the method of maximum likelihood applied to the normal density, regardless of whether the conditional distribution is assumed normal or not. This may of course result in a large loss of efficiency relative to the true but unknown maximum likelihood estimator. In response Engle and Gonzales-Rivera (1991) introduced the semi-parametric maximum likelihood estimator of GARCH models. The semi-parametric estimator is a two-step estimator. In the first step consistent estimates of the parameters are obtained and are used to estimate a non-parametric conditional density. The second step consists of using this non-parametric density to adapt the initial estimator. The method has not been applied much in the literature though. This may be because the estimator is not so simple to compute, furthermore a choice of density estimator is required and specification testing is not straightforward.

This paper is concerned with efficient GMM estimation of GARCH models. In particular we show that efficient GMM is a feasible alternative to the quasi-maximum likelihood and semi-parametric estimators. Compared to the semi-parametric estimator efficient GMM has the advantage of being simple to compute and specification testing is straightforward. As a competitor to the common quasi-maximum likelihood estimator efficient GMM is asymptotically efficient with the coefficient of skewness and excess kurtosis of the conditional density being important in explaining the differences.

The organization of the paper is as follows. In section 2 we define the estimator in case of a GARCH(1,1) conditional variance model and give sufficient conditions for the estimator to be consistent and asymptotically normal. That is, to have the CAN property. Asymptotic relative efficiency comparison to the quasi-maximum likelihood estimator shows that efficient GMM is asymptotically more efficient under asymmetry of the conditional density and a small Monte-Carlo experiment confirms that the finite-sample gain can be substantial. Section 3 is concerned with efficient GMM estimation of the GARCH(1,1)-M regression model of Engle, Lilien and Robins (1987).

It is shown that the introduction of a conditional mean makes the coefficient of excess kurtosis as well as of skewness important for explaining the relative efficiency gains of efficient GMM. In this section we also consider efficient GMM based specification tests. These tests are locally more powerful than the corresponding Bollerslev and Wooldridge (1992) robust classical tests whenever the efficient GMM estimator is asymptotically more efficient than quasi-maximum likelihood. Section 4 illustrates efficient GMM estimation and hypothesis testing with an application to the daily returns to the SP500 index, (1928-1991) and section 5 concludes. Proofs can be found in the appendix.

2 A GARCH(1,1) conditional variance process

2.1 The efficient GMM estimator

Consider the data generating process

$$\begin{aligned} \varepsilon_t &= z_t h_{0t},\; z_t \sim iid(0,1) \\ h_{0t}^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{0t-1}^2 \end{aligned} \qquad (1)$$

where the l:th conditional moment of the ε_t process is $E[\varepsilon_t^l | \digamma^{t-1}] = v_l h_{0t}^l$, $\digamma^{t-1}$ is the information generated up to $t-1$ and $v_l = E\left(z_t^l\right)$. In practice we observe a finite segment of the process (1) and the objective is to estimate the parameters $\theta_0 \in \Theta$ where $\theta_0 = (\alpha_0, \alpha_1, \beta)'$ and Θ is a compact parameter space. To this end, define the (raw) vector

$$\mathbf{r}_t = [\varepsilon_t, \left(\varepsilon_t^2 - h_t^2\right)]' \qquad (2)$$

and the generalized vector,

$$\mathbf{g}_t = \mathbf{F}_t' \mathbf{r}_t$$

where $\mathbf{F}_t$ is an instrumental variable function.

The GMM estimator of a parameter vector θ is the solution to (cf. Hansen (1982))

$$\text{argmin}_{\theta \in \Theta} \left[T^{-1} \sum_{t=1}^{T} \mathbf{g}_t \right]' \mathbf{W}_T \left[T^{-1} \sum_{t=1}^{T} \mathbf{g}_t \right]$$

with $\mathbf{W}_T = T^{-1} \sum_{t=1}^{T} \mathbf{W}_t$ an appropriate weighting matrix.

Efficient GMM corresponds to choosing $\mathbf{F}_t = \boldsymbol{\Sigma}_t^{-1}(\frac{\partial \mathbf{r}_t}{\partial \theta'})$ and $\mathbf{W}_t = (\frac{\partial \mathbf{r}_t'}{\partial \theta})\boldsymbol{\Sigma}_t^{-1} \times (\frac{\partial \mathbf{r}_t}{\partial \theta'})$ where $\boldsymbol{\Sigma}_t = \text{var}(\mathbf{r}_t | \mathcal{F}^{t-1})$ and $(\frac{\partial \mathbf{r}_t}{\partial \theta'})$ is the Jacobian matrix, see Newey and McFadden (1994). The objective function for an operational efficient GMM estimator is then given by

$$Q_T = T^{-1} \left[\sum_{t=1}^{T} \mathbf{g}_t \right]' \left(\sum_{t=1}^{T} \boldsymbol{\Lambda}_t \right)^{-1} \left[\sum_{t=1}^{T} \mathbf{g}_t \right] \tag{3}$$

where $\boldsymbol{\Lambda}_t = \mathbf{g}_t \mathbf{g}_t'$ is a parameter dependent weighting matrix. The elements of the generalized moment and the weighting matrix are given by

$$\begin{aligned} g_{it} &= \frac{1}{\Delta} \left(\frac{\partial h_t^2}{\partial \theta_i} \right) h_t^{-2} \left[\left(\frac{\varepsilon_t}{h_t} \right) v_3 - \left(\left(\frac{\varepsilon_t^2}{h_t^2} \right) - 1 \right) \right] \\ \Lambda_{ijt} &= \frac{1}{\Delta^2} \left(\frac{\partial h_t^2}{\partial \theta_i} \right) h_t^{-2} \left[\left(\frac{\varepsilon_t}{h_t} \right) v_3 - \left(\left(\frac{\varepsilon_t^2}{h_t^2} \right) - 1 \right) \right]^2 h_t^{-2} \left(\frac{\partial h_t^2}{\partial \theta_j} \right) \end{aligned}$$

with $\Delta = (v_4 - 1) - v_3^2$[1].

By construction the objective function (3) is exactly identified and it is well-known from the literature that the choice of weighting matrix above is sufficient but not necessary for asymptotic efficiency. In fact asymptotic theory does not discriminate between a parameter dependent weighting matrix and a weighting matrix based on an initial consistent estimator of θ_0, or for that matter the identity matrix. Simplicity suggests that the identity matrix might be a good choice in applications. However finite-sample evidence indicates that allowing the weighting matrix to be parameter dependent is preferred and since the weighting matrix in (3) is typically constructed for inference little additional effort is needed for this choice.

To put quasi-maximum likelihood in the GMM framework choose $\mathbf{W}_t = \mathbf{s}_t \mathbf{s}_t'$, where $\mathbf{s}_t = \frac{\partial l_t}{\partial \theta}$ with l_t the normal quasi-loglikelihood for observation t and $\mathbf{F}_t' \mathbf{r}_t = \mathbf{s}_t$. By noting that $\mathbf{g}_t = \frac{-2}{v_4 - 1} \mathbf{s}_t$ for $v_3 = 0$ the GMM objective function can be written

$$\left[T^{-1} \sum_{t=1}^{T} \mathbf{s}_t \right]' \mathbf{A}_T^{-1} \left[T^{-1} \sum_{t=1}^{T} \mathbf{s}_t \right]$$

where $\mathbf{A}_T = T^{-1} \sum_{t=1}^{T} \mathbf{s}_t \mathbf{s}_t'$ and the asymptotic equivalence of efficient GMM based on (2) and quasi-maximum likelihood follows in case of symmetric z_t.

[1] By definition $\boldsymbol{\Sigma}_t$ is positive definite which implies $(v_4 - 1) - v_3^2 > 0$.

Note however that in general efficient GMM requires an initial guess on the kurtosis and skewness of the rescaled variable. A guess on kurtosis can be based on the mean of $\frac{\varepsilon_t^4}{h_t^4(\widetilde{\theta}_T)}$, similarly the guess on skewness can be based on the mean of $\frac{\varepsilon_t^3}{h_t^3(\widetilde{\theta}_T)}$, where $\widetilde{\theta}_T$ is some initial estimator.

2.2 Asymptotic properties

All the asymptotic results below are derived with a parameter dependent weighting matrix. Compared to basing the weighting matrix on an initial consistent estimator of θ or simply the identity matrix no additional restrictions are needed. The latter estimators are of course special cases of the results given below. Furthermore allowing for a parameter dependent weighting matrix is unimportant for the asymptotic distribution.

The following assumptions are sufficient for the results

(a) $z_t \sim iid(0,1)$ and (z_t, h_t^2) stochastically independent

(b) $E\left(z_t^4\right) = v_4 < \infty$

(c) $\beta + \alpha_1 < 1$

(d) $\Theta \equiv \{\theta : 0 < a_{0l} \leq a_0 \leq a_{0u}, 0 < a_{1l} \leq a_1 \leq a_{1u}, 0 < b_l \leq b \leq b_u < 1\}$

(e) $\theta_0 \in int(\Theta)$

In contrast to Lee and Hansen (1994) and Lumsdaine (1996) we do not allow for IGARCH. Besides giving a simpler asymptotic theory this means that we do not have to consider further restrictions on the parameter space as is necessary in the case of IGARCH. In particular the present framework can be used to establish consistency and asymptotic normality of the quasi-maximum likelihood estimator of the GARCH(1,1) process without restrictions on the parameter space.

Define $\widehat{\theta}_T$ as the sequence of minimizers of the objective function (3) and suppose some initial estimators $\widehat{v}_3, \widehat{v}_4$ are available and that these initial estimators only requires assumptions (a)-(e). We then have

Theorem 1 *Suppose assumptions (a)-(e) hold and that* $\widehat{v}_3 \xrightarrow{p} v_3^*, \widehat{v}_4 \xrightarrow{p} v_4^*$. *Then* $\widehat{\theta}_T \xrightarrow{p} \theta_0$ *on* Θ *regardless of* $v_3^* = v_3$ *or* $v_4^* = v_4$ *as long as both* v_3^*, v_4^* *are finite*

That is, $\widehat{\theta}_T$ is consistent for finite arbitrary guess on v_3, v_4. In practice we are of course interested in obtaining asymptotically valid inference about $\widehat{\theta}_T$ and for this purpose we need consistent initial estimators of v_3 and v_4. But this result has a useful consequence in terms of the asymptotic distribution of $\widehat{\theta}_T$. In particular we will be able to conclude that the asymptotic distribution of $\widehat{\theta}_T$ is the same regardless of wether v_3 and v_4 are known or estimated.

Let $\mathbf{\Lambda}_0 = E\mathbf{\Lambda}_t(\theta_0)$ and $\mathbf{G}_0 = E\mathbf{G}_t(\theta_0)$ where $\mathbf{G}_t(\theta) = \frac{\partial \mathbf{g}_t}{\partial \theta'}$

Theorem 2 *Suppose assumptions (a)-(e) holds and that* $\widehat{v}_3 \xrightarrow{p} v_3, \widehat{v}_4 \xrightarrow{p} v_4$ *or* v_3, v_4 *are known. Then* $T^{1/2}(\widehat{\theta}_T - \theta_0) \xrightarrow{d} N\left(\mathbf{0}, \left(\mathbf{G}_0 \mathbf{\Lambda}_0^{-1} \mathbf{G}_0\right)^{-1}\right) \cong N\left(\mathbf{0}, \mathbf{\Lambda}_0^{-1}\right)$ *where " $\cong$ " denotes equality in distribution*

The above result allows us to compare asymptotic variances of efficient GMM and quasi-maximum likelihood. Taking expectations of $\mathbf{\Lambda}_t$, $\left(\mathbf{B}_t \mathbf{A}_t^{-1} \mathbf{B}_t\right)$ evaluated at θ_0, where $\mathbf{B}_t = \frac{\partial \mathbf{s}_t}{\partial \theta}$ and $\mathbf{A}_t = \mathbf{s}_t \mathbf{s}_t'$ we obtain

$$\mathbf{V}_{GMM} = \left[E\mathbf{\Lambda}_t(\theta_0)\right]^{-1} = \left[(v_4 - 1) - v_3^2\right]\left[E\left(\frac{\partial h_{0t}^2}{\partial \theta}\frac{\partial h_{0t}^2}{\partial \theta'}\right)\right]^{-1} h_{0t}^4$$

and

$$\begin{aligned}\mathbf{V}_{QMLE} &= \left[\left(E\mathbf{B}_t(\theta_0)\left[E\mathbf{A}_t(\theta_0)\right]^{-1} E\mathbf{B}_t(\theta_0)\right)\right]^{-1} \\ &= \left[(v_4 - 1)\right]\left[E\left(\frac{\partial h_{0t}^2}{\partial \theta}\frac{\partial h_{0t}^2}{\partial \theta'}\right)\right]^{-1} h_{0t}^4\end{aligned}$$

The relative efficiency ratio is seen to depend only on $\frac{(v_4-1)-v_3^2}{(v_4-1)}$ and is strictly decreasing in v_3. That is, efficient GMM is strictly more efficient than the quasi-maximum likelihood estimator when the conditional innovations have a skewed distribution.

2.3 Finite-sample properties

A small Monte-Carlo experiment is conducted to evaluate the finite-sample properties of the estimators[2]. We generate data from the GARCH(1,1)

[2]Both estimators use the Newton algorithm. The efficient GMM estimator use Constrained Optimization (CO) module in GAUSS and quasi-maximum likelihood use the

process (1) with two sets of parameter combinations close to what one commonly encounters in applying GARCH(1,1) to real data, $\theta_{01} = (0.1, 0.2, 0.7)'$ and $\theta_{02} = (0.05, 0.05, 0.9)'$ where $\theta_0 = (\alpha_0, \alpha_1, \beta)'$. The sample sizes considered are $T = 1000$ and 5000. Initial values were chosen arbitrarily as $(0.1, 0.25, 0.67)'$ and $(0.05, 0.07, 0.87)'$ for each set of parameters.

All the efficient GMM estimations are performed with the parameter dependent weighting matrix. Efficient GMM estimators with the weighting matrix provided by the initial consistent estimator, $\widetilde{\theta}_T$ (where $\widetilde{\theta}_T$ is the quasi-maximum likelihood estimate) and the identity matrix respectively performed less well. The performance was especially poor with the identity matrix where we experienced serious convergence problems.

2.3.1 Symmetric densities

For many financial return series a t-distribution with a few degrees of freedom fits the empirical density of z_t quite well. The question is if we can improve on quasi-maximum likelihood with the efficient GMM estimator when the rescaled variable, z_t has a fat-tailed density. The rescaled variable is assumed to follow a t-distribution with 5 degrees of freedom which gives true kurtosis of 9. The $t(5)$ distributed random variable is generated as the ratio of a standard normal and $\sqrt{\chi_5^2/5}$ variate. To obtain a $(0,1)$ variable we divide by the standard deviation. Results are given in Table 1.

For the parameter combinations considered efficient GMM typically has a variance that is smaller than or equal to that of quasi-maximum likelihood, the gain is substantial for the case of θ_{02} and $T = 1000$. Furthermore the efficient GMM estimator has less bias than the quasi-maximum likelihood estimator in this case. The bias of the quasi-maximum likelihood estimator is small for θ_{01} but increases for the parameter vector θ_{02}. It appears that this tendency is not so strong for the efficient GMM estimator. For $T = 5000$ efficient GMM and quasi-maximum likelihood are equivalent for both θ_{01} and θ_{02}.

Constrained Maximum Likelihood (CML) module. In both cases analytical derivatives are used and constraints are imposed as $10^{-10} \leq a_0, a_1$ and $0 < b < 1$. All the results are based on 5000 replications and 100 initial values of the conditional variance process were discarded to avoid initialization effects. Quasi-maximum likelihood estimates are used to initialize a guess on v_3 and v_4 for the efficient GMM estimator.

Table 1 Finite sample comparison of efficient GMM (GMM) and quasi-maximum likelihood (QMLE), $z_t \sim t(5)$

	GMM		QMLE		
$T = 1000$	bias	std(1)	bias	std(2)	ratio (1)/(2)
θ_{01}					
a_0	0.010	0.045	0.011	0.046	0.978
a_1	0.005	0.066	0.005	0.068	0.970
b	−0.018	0.083	−0.018	0.085	0.976
θ_{02}					
a_0	0.011	0.051	0.027	0.081	0.629
a_1	0	0.030	0.005	0.032	0.938
b	−0.013	0.073	−0.034	0.104	0.702
$T = 5000$					
θ_{01}					
a_0	0.002	0.020	0.002	0.020	1
a_1	0.001	0.030	0.001	0.030	1
b	−0.005	0.037	−0.005	0.037	1
θ_{02}					
a_0	0.003	0.022	0.004	0.021	1.04
a_1	0.001	0.013	0.001	0.013	1
b	−0.005	0.030	−0.006	0.029	1.03

Table 2 Finite sample comparison of efficient GMM (GMM) and quasi-maximum likelihood (QMLE), $z_t \sim$ Gamma(2)

	GMM		QMLE		
$T = 1000$	bias	std(1)	bias	std(2)	ratio (1)/(2)
θ_{01}					
a_0	0.008	0.050	0.012	0.045	1.11
a_1	0.002	0.049	0.004	0.061	0.803
b	−0.013	0.074	−0.02	0.085	0.870
θ_{02}					
a_0	0.011	0.046	0.029	0.090	0.511
a_1	0.002	0.021	0.004	0.029	0.723
b	−0.015	0.059	−0.035	0.111	0.531
$T = 5000$					
θ_{01}					
a_0	0.001	0.013	0.002	0.018	0.722
a_1	0	0.020	0	0.026	0.789
b	−0.003	0.026	−0.005	0.035	0.743
θ_{02}					
a_0	0.002	0.013	0.004	0.018	0.722
a_1	0	0.008	0.001	0.011	0.727
b	−0.004	0.019	−0.005	0.026	0.730

2.3.2 Asymmetric densities

Unconditional return distributions are characterized by the 'stylized facts' of excess kurtosis, high peakedness and are often skewed. But the distribution of the rescaled variable as well is characterized by excess kurtosis and skewness when GARCH models are fitted to data see e.g. Engle and Gonzales-Rivera (1991). To consider asymmetric densities we generate z_t as Gamma-distributed with mean and variance parameter equal to 2. The Gamma(2) distributed random variable is obtained from *rndgam* in GAUSS and standardized by subtracting 2 from it and dividing by $2^{1/2}$. It has true kurtosis and skewness given by 6 and $2/\sqrt{2}$ respectively. Results are given in Table 2.

As for the case of symmetric z_t with $T = 1000$ the bias of the efficient

GMM estimator seems to be smaller and the gain is larger for θ_{02}. For $T = 5000$ the efficient GMM estimator has smaller variance than quasi-maximum likelihood for both θ_{01} and θ_{02} which is what we expect since it is asymptotically more efficient. In fact the estimated standard deviation ratios are quite close to the theoretical ratio of approximately 0.775.

3 Extension to models with a conditional mean

3.1 The GARCH(1,1)-M regression model

The GARCH(1,1) conditional variance process considered in section 2 may be somewhat restrictive in practice. Here we consider some practical details of estimating more general models with the efficient GMM estimator introduced in section 2.1. The model of interest is the GARCH(1,1)-M regression model introduced by Engle et al. (1987)

$$\begin{aligned} y_t &= \mathbf{X}_t'\boldsymbol{\mu} + \delta f(h_t^2) + \varepsilon_t \\ \varepsilon_t &= z_t h_t \\ h_t^2 &= a_0 + a_1\varepsilon_{t-1}^2 + b h_{t-1}^2 \end{aligned} \tag{4}$$

where $f(h_t^2)$ is a function of the conditional variance and in addition to section 2.1 we assume that z_t is independent of $\mathbf{X}_t$. Sufficient conditions for consistency and asymptotic normality are, to the authors knowledge, not known even for the quasi-maximum likelihood estimator. In what follows we simply assume that such conditions are satisfied. More specifically, we assume that the CAN property holds for both quasi-maximum likelihood and efficient GMM.

By using the (raw) vector (2) we can in analogy with section 2.1 define the efficient GMM estimator of the model (4) as a solution to

$$\operatorname{argmin}_{\theta\in\Theta} T^{-2}\left[\sum_{t=1}^{T}\mathbf{g}_t\right]'\boldsymbol{\Lambda}_T^{-1}\left[\sum_{t=1}^{T}\mathbf{g}_t\right] \tag{5}$$

where the parameter vector θ is given by $\theta = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)', \boldsymbol{\gamma}_1 = (a_0, a_1, b)'$ and $\boldsymbol{\gamma}_2 = (\delta, \mu')'$. The generalized moment, $\mathbf{g}_t$ is given by

$$\mathbf{g}_t = \frac{1}{\Delta}\left[\begin{array}{c} \left(\frac{\partial h_t^2}{\partial\theta}\right) h_t^{-2}\left[v_3\frac{\varepsilon_t}{h_t} - \left(\frac{\varepsilon_t^2}{h_t^2} - 1\right)\right] \\ +\left(\frac{\partial\varepsilon_t}{\partial\theta}\right) h_t^{-1}\left[\frac{\varepsilon_t}{h_t}(v_4 - 1) - v_3\left(\frac{\varepsilon_t^2}{h_t^2} - 1\right)\right] \end{array}\right] \tag{6}$$

with $\Delta = [(v_4 - 1) - v_3^2]$ and derivatives in (6) are computed recursively as

$$\frac{\partial h_t^2}{\partial \theta} = \mathbf{c}_{t-1} + \left(b - 2\delta a_1 \varepsilon_{t-1} \frac{\partial f(h_{t-1}^2)}{\partial h_{t-1}^2} \right) \frac{\partial h_{t-1}^2}{\partial \theta}$$

where $\mathbf{c}_t = (1, \varepsilon_t^2, h_t^2, -2a_1\varepsilon_t f(h_t^2), -2a_1\varepsilon_t \mathbf{X}_t')$, and

$$\frac{\partial \varepsilon_t}{\partial \theta} = \boldsymbol{\pi}_t - \delta \frac{\partial f(h_t^2)}{\partial h_t^2} \frac{\partial h_t^2}{\partial \theta}$$

with $\pi_t = (0, 0, 0, -f(h_t^2), -\mathbf{X}_t')$.

The equality of $E\boldsymbol{\Lambda}_t(\theta_0)$ and $E\mathbf{G}_t(\theta_0)$ is straightforward to verify from (6) using the iid property of z_t and hence the variance matrix of $T^{1/2}\widehat{\theta}_T$ can be consistently estimated by

$$\widehat{\boldsymbol{\Lambda}}_T^{-1} = \left(T^{-1} \sum_{t=1}^{T} \widehat{\boldsymbol{\Lambda}}_t \right)^{-1}$$

or

$$\widehat{\mathbf{G}}_T^{-1} = \left(T^{-1} \sum_{t=1}^{T} \frac{\partial \widehat{\mathbf{g}}_t}{\partial \theta} \right)^{-1}$$

where $\widehat{\boldsymbol{\Lambda}}_t = \boldsymbol{\Lambda}_t\left(\widehat{\theta}_T\right)$ and $\frac{\partial \widehat{\mathbf{g}}_t}{\partial \theta} = \frac{\partial \mathbf{g}_t\left(\widehat{\theta}_T\right)}{\partial \theta}$. Inference based on $\widehat{\boldsymbol{\Lambda}}_T$ only involves first derivatives of the conditional mean and the conditional variance function. This is useful since estimation of GARCH models frequently rely on numerical approximations to the analytical derivatives.

3.2 Asymptotic efficiency comparison

For the purpose of comparing the asymptotic variance matrices of efficient GMM and quasi-maximum likelihood we let

$$\mathbf{S} = \left[\begin{array}{cc} h_{0t}^{-1} \frac{\partial m_{0t}}{\partial \theta} & \frac{1}{2} h_{0t}^{-2} \frac{\partial h_{0t}^2}{\partial \theta} \end{array} \right]$$

where $h_{0t}^2 = h_t^2(\theta_0)$ and $m_{0t} = m_t(\theta_0)$ is the regression function and define the matrices

$$\mathbf{K} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$
$$\mathbf{L} = \begin{bmatrix} 1 & v_3 \\ v_3 & (v_4-1) \end{bmatrix}$$
$$\mathbf{M} = \frac{1}{\Delta}\begin{bmatrix} (v_4-1) & -2v_3 \\ -2v_3 & 4 \end{bmatrix}$$

with $\mathbf{L}, \mathbf{K}$ and $\mathbf{M}$ the 'weight matrices' for the $\mathbf{A}_t, \mathbf{B}_t$ and $\mathbf{\Lambda}_t$ matrices respectively. Define next

$$\mathbf{\Pi}_A = E(\mathbf{SDS}')$$

for arbitrary positive semi-definite matrix $\mathbf{D}$.

We can then write the difference between the inverses of the asymptotic variance matrices as

$$\mathbf{V}_{GMM}^{-1} - \mathbf{V}_{QMLE}^{-1} = \mathbf{\Pi}_M - \mathbf{\Pi}_K \mathbf{\Pi}_L^{-1} \mathbf{\Pi}_K \tag{7}$$

which is positive semi-definite for a general conditional density and the model (4)[3].

Note that $v_3 = 0$, $v_4 = 3$ implies $\mathbf{K} = \mathbf{L} = \mathbf{M}$ and hence $\mathbf{V}_{QMLE} = \mathbf{V}_{GMM}$ if the conditional density is normal. In the case of excess kurtosis and/or skewness of the conditional density, $\mathbf{K} \neq \mathbf{L} \neq \mathbf{M}$. However a formal proof that this is sufficient for efficiency gains is too difficult and we consider some simple cases where positive results can be obtained.

Consider first the GARCH(1,1) regression model ($\delta = 0$ in (4)). Imposing symmetry of the conditional density is not sufficient for $\mathbf{V}_{QMLE} = \mathbf{V}_{GMM}$ but for the conditional variance parameters, $\boldsymbol{\gamma}_1$ we have $\mathbf{V}(\boldsymbol{\gamma}_1)_{QMLE} = \mathbf{V}(\boldsymbol{\gamma}_1)_{GMM}$. Some tedious but straightforward algebra shows that

$$\begin{aligned} \mathbf{V}^{-1}(\boldsymbol{\mu})_{GMM} - \mathbf{V}^{-1}(\boldsymbol{\mu})_{QMLE} &= \mathbf{P}_{11} + \frac{4}{(v_4-1)}\mathbf{P}_{12} - [\mathbf{P}_{11} + 2\mathbf{P}_{12}] \\ &\quad \times [\mathbf{P}_{11} + (v_4-1)\mathbf{P}_{12}]^{-1} [\mathbf{P}_{11} + 2\mathbf{P}_{12}] \\ &= \frac{(v_4-3)^2}{(v_4-1)^2}\left[\mathbf{P}_{11}^{-1} + \frac{1}{(v_4-1)}\mathbf{P}_{12}^{-1}\right]^{-1} \end{aligned} \tag{8}$$

[3] The matrix difference on the right-hand side of (7) appears as part of the difference between the inverse asymptotic variance matrices of the semi-parametric and maximum likelihood estimators in Gonzalez-Rivera and Drost (1999). They prove that this difference is positive semi-definite for the model (4).

where $\mathbf{P}_{11} = E\mathbf{a}_{11,t}\mathbf{a}'_{11,t}$, $\mathbf{P}_{12} = E\mathbf{a}_{12,t}\mathbf{a}'_{12,t}$ with $\mathbf{a}_{11,t} = h_{0t}^{-1}\frac{\partial m_{0t}}{\partial \mu}$ and $\mathbf{a}_{12,t} = \frac{1}{2}h_{0t}^{-2}\frac{\partial h_{0t}^2}{\partial \mu}$. Note that (8) is positive definite for $v_4 > 3$ and increasing in v_4 implying there is efficiency gains for the conditional mean parameters in case of excess kurtosis.

A corresponding result for $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}_1$ under asymmetry of the conditional density is more difficult since the block-diagonal structure of $\mathbf{V}_{QMLE}$ and $\mathbf{V}_{GMM}$ is lost. We can however allow for $v_3 \neq 0$ if we set $\boldsymbol{\mu} = \mathbf{0}$. This gives the asymptotic relative efficiency result for the GARCH(1,1) conditional variance model, that is, $\frac{(v_4-1)-v_3^2}{v_4-1}$ obtained in section 2.2. In the GARCH(1,1)-M model (4) the block-diagonal structure of $\mathbf{V}_{QMLE}$ and $\mathbf{V}_{GMM}$ is lost even under symmetry of the conditional density. Thus in this case the algebra is too tedious to derive results such as (8) regardless of $v_3 = 0$ or not. However we conjecture that due to the absence of block-diagonal structure there is efficiency gains for *both* conditional mean and variance parameters under excess kurtosis. Similar considerations apply for an asymmetric conditional density.

3.3 specification tests

As in the familiar maximum likelihood setting the classical LM (score), Wald and LR-tests are available and derived with expansions involving the first and second derivatives of the objective function (cf. Newey and West (1987)). Under the null hypothesis, say $H_0 : \mathbf{a}(\theta_0) = 0$, the LM (score) statistic is given by

$$\xi_{LM} = \sqrt{T}\widetilde{\mathbf{g}}'_T\widetilde{\boldsymbol{\Lambda}}_T^{-1}\widetilde{\mathbf{G}}_T\left[\widetilde{\mathbf{G}}'_T\widetilde{\boldsymbol{\Lambda}}_T^{-1}\widetilde{\mathbf{G}}_T\right]^{-1}\widetilde{\mathbf{G}}'_T\widetilde{\boldsymbol{\Lambda}}_T^{-1}\sqrt{T}\widetilde{\mathbf{g}}_T \tag{9}$$

where "$\sim$" denotes evaluated under the null hypothesis. Under our assumptions on the rescaled variable, z_t, we obtain the asymptotically equivalent form

$$\xi_{LM} = \sum_{t=1}^{T}\widetilde{\mathbf{g}}'_t\left(\sum_{t=1}^{T}\widetilde{\mathbf{g}}_t\widetilde{\mathbf{g}}'_t\right)^{-1}\sum_{t=1}^{T}\widetilde{\mathbf{g}}_t \tag{10}$$

which is simply TR_u^2 from the linear regression of 1 on $\widetilde{\mathbf{g}}'_t$. As an example of the TR^2 form of the LM test consider the efficient GMM counterpart of the Engle (1982) ARCH(m) classical LM test with no parameters in the conditional mean, we have

$$\xi_{LM} = \frac{1}{\Delta}\mathbf{W}_1'\mathbf{W}_2\left(\mathbf{W}_2\mathbf{W}_2'\right)^{-1}\mathbf{W}_2'\mathbf{W}_1$$

where $\mathbf{W}_1' = (w_{11}, ..., w_{1T})$, $\mathbf{W}_2' = (w_{21}, ..., w_{2T})$ with $w_{1t} = v_3\frac{\varepsilon_t}{\widetilde{h}_t} - \left(\frac{\varepsilon_t^2}{\widetilde{h}_t^2} - 1\right)$ and $w_{2t} = \widetilde{h}_t^{-2}\widetilde{\frac{\partial h_t^2}{\partial\theta}}$. Since

$$\text{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}\left[v_3\frac{\varepsilon_t}{\widetilde{h}_t} - \left(\frac{\varepsilon_t^2}{\widetilde{h}_t^2} - 1\right)\right]^2 = (v_4 - 1) - v_3^2$$

we can construct the asymptotically equivalent TR^2 form

$$\xi_{LM} = T\frac{\mathbf{W}_1'\mathbf{W}_2\left(\mathbf{W}_2\mathbf{W}_2'\right)^{-1}\mathbf{W}_2'\mathbf{W}_1}{\mathbf{W}_1'\mathbf{W}_1} = TR_u^2$$

where R_u^2 is the unadjusted squared multiple correlation coefficient from a regression of $\mathbf{W}_1$ on $\mathbf{W}_2$. Under the null hypothesis of no ARCH(m) $\widetilde{h}_t^2 = \sigma_\varepsilon^2$ and $\mathbf{W}_2 = \left(1, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, ..., \varepsilon_{t-m}^2\right)$ so that the LM test of the null of no ARCH(m) is computed by TR_u^2 from the regression of $\left[\widetilde{v}_3\frac{\varepsilon_t}{\sigma_\varepsilon} - \left(\frac{\varepsilon_t^2}{\sigma_\varepsilon^2} - 1\right)\right]$ on $\mathbf{W}_2$, where $\widetilde{v}_3$ is the skewness of y_t. That is, in contrast to the usual TR^2 form of the ARCH(m) LM test (which is computed by TR_u^2 from the regression of $\left(\frac{\varepsilon_t^2}{\sigma_\varepsilon^2} - 1\right)$ on $\mathbf{W}_2$) we have not imposed the auxiliary assumption of normality.

Similarly the Wald and quasi-LR statistics are available. The Wald statistic has its usual form and the quasi-LR test is given by

$$\xi_{QLR} = T\left[Q_T(\widetilde{\theta}_T) - Q_T(\widehat{\theta}_T)\right] \tag{11}$$

In the exactly identified case $Q_T(\widehat{\theta}_T) = 0$ and hence

$$\xi_{QLR} = T\left[Q_T(\widetilde{\theta}_T)\right] \tag{12}$$

which is computed holding the parameters in the null hypothesis fixed at their respective null hypothesis during the iterations. Note that $\xi_{QLR} = \xi_{LM}$ in this context.

As shown by Newey and West (1987) the LM, Wald and quasi-LR tests are asymptotically equivalent under the null and local alternatives. Under the alternative hypothesis $H_A : \mathbf{a}(\theta_0) = T^{-1/2}\boldsymbol{\lambda}$ these tests have the $\chi_q^2(\zeta_{GMM})$ distribution, where $\zeta_{GMM} = \boldsymbol{\lambda}' \left[\mathbf{A}(\theta_0)\mathbf{V}_{GMM}^{-1}\mathbf{A}(\theta_0)\right]^{-1}\boldsymbol{\lambda}$ is the non-centrality parameter and $\mathbf{A}(\theta_0)$ is the Jacobian matrix of the $q \times 1$ vector function $\mathbf{a}(\theta)$ of the $p \times 1$ vector θ with $q \leq p$. It is useful to compare this non-centrality parameter to the non-centrality parameter of the Bollerslev and Wooldridge (1992) robust classical tests which is given by $\zeta_{QMLE} = \boldsymbol{\lambda}' \left[\mathbf{A}(\theta_0)\mathbf{V}_{QMLE}^{-1}\mathbf{A}(\theta_0)\right]^{-1}\boldsymbol{\lambda}$. From the results of the previous section it is clear that the GMM based tests are more powerful in terms of local asymptotic power. In particular, whenever efficient GMM is asymptotically more efficient than quasi-maximum likelihood.

4 Empirical Illustration

In this section we apply the efficient GMM estimator to model the returns to the daily SP500 index, (1928-1991) as a GARCH-M model. Inclusion of a measure of volatility in the conditional mean of returns is an attempt to introduce a measure of risk. It is an implication of the 'mean-variance hypothesis' of the returns and under this hypothesis large returns are expected to be associated with high volatility.

The GARCH-M model has been applied by several researchers to model the relation between risk and return e.g. French et al. (1987) applied the GARCH-M model to subsets of the excess returns to the SP500 index and concluded that the data was consistent with a positive relation between conditional expected excess return and conditional variance. However Glosten, Jagannathan and Runkle (1993) obtained a significant negative relation between the conditional mean and the conditional variance of excess returns on stocks when the model was modified to allow positive and negative anticipated returns to have different impact on the conditional variance.

The distributional properties of the returns to the daily SP500 index has been analyzed in Mills (1999, ch 5) and Granger and Ding (1995) investigated the properties of absolute returns. Table 3 gives the estimates of the first four unconditional moments of the distribution of returns to the SP500 index.

The modelling strategy is to first specify the conditional mean and once satisfactory, tests for the conditional variance specification are performed. Only autocorrelation in the conditional mean is tested for and any possible

Table 3 Unconditional moments of returns to the SP500 index

Mean	0.01821
Variance	1.32377
Skewness	−0.48725
Kurtosis	25.4164

non-linearity of the conditional mean is disregarded. In tems of the conditional variance specification we restrict ourselves to the GARCH(1,1) case.

Fitting an AR(5) model to the returns and computing the LM (10) (quasi-LR (12)) test of the null hypothesis of no ARCH(4) gives a test-statistic of 55.67 with a corresponding p-value of 7.2×10^{-14} and hence we reject the null at any reasonable significance level. This leads us to estimate the following GARCH(1,1)-M model for the returns

$$\begin{aligned} y_t &= \mu_0 + \mu_1 y_{t-1} + \mu_2 y_{t-2} + \mu_3 y_{t-3} + \mu_4 y_{t-4} + \mu_5 y_{t-5} + \delta h_t + \varepsilon_t \\ h_t^2 &= a_0 + a_1 \varepsilon_{t-1}^2 + b h_{t-1}^2 \end{aligned}$$

where the choice of the conditional standard deviation specification of the GARCH(1,1)-M term is of course arbitrary.

Table 4 gives the estimation results together with Box-Pierce statistics of the levels and squares, $Q(\cdot)$ and $Q^2(\cdot)$ respectively (Box and Pierce (1970)) and Table 5 gives the first four moments of the rescaled residuals. Quasi-maximum likelihood estimates are given for comparison[4].

Inpsection of Table 4 shows that there is evidence for strong persistence in the conditional variance. The sum $a_1 + b$ is only slightly below unity for both efficient GMM and quasi-maximum likelihood. The estimates of the risk-premia term, δ are positive for both efficient GMM and quasi-maximum likelihood. Note however that the efficient GMM estimate is less than half the estimate of quasi-maximum likelihood and in contrast it is not significant at usual levels. Box-Pierce statistics of the levels and squares suggests that there might be some dynamics left in the data. If we follow the recommendation of

[4]Both estimators use the Newton algorithm with analytic first derivatives of the conditional mean and variance functions. The efficient GMM estimator use Constrained Optimization (CO) module in GAUSS and quasi-maximum likelihood use the Constrained Maximum Likelihood (CML) module. In both cases constraints are imposed which restricts a_0, a_1 from zero and $0 < b < 1$. Quasi-maximum likelihood estimates are used to initialize a guess on v_3 and v_4 for the efficient GMM estimator.

Table 4 GARCH(1,1)-M estimates of daily returns to the SP500, standard errors in parentheses based on weighting matrix for GMM and Bollerslev-Wooldridge robust standard errors for QMLE

	GMM	QMLE
μ_0	0.003194	−0.00172
	(0.01586)	(0.01778)
μ_1	0.140646	0.140532
	(0.00807)	(0.01015)
μ_2	−0.05300	−0.04540
	(0.00851)	(0.00982)
μ_3	0.01657	0.01715
	(0.00846)	(0.01053)
μ_4	0.01200	0.00854
	(0.00840)	(0.01012)
μ_5	0.01410	0.01850
	(0.00808)	(0.01034)
δ	0.02773	0.05700
	(0.02196)	(0.02404)
a_0	0.00754	0.00781
	(0.00190)	(0.00103)
a_1	0.08912	0.09082
	(0.00546)	(0.00889)
b	0.90782	0.90652
	(0.00627)	(0.00730)
$Q(12)$	18.53	16.91
$Q(24)$	35.15	33.87
$Q^2(12)$	21.93	20.65
$Q^2(24)$	32.10	30.88

Table 5 Conditional moments of rescaled residuals

	GMM	QMLE
Mean	−0.01129	−0.03466
Variance	1.00531	0.99850
Skewness	−0.51937	−0.51623
Kurtosis	7.40892	7.41753

Box and Pierce and compare with the χ^2_7 and χ^2_{19} distribution for the $Q(12)$ and $Q(24)$ statistics, respectively, they are significant at 5% level. The $Q^2(12)$ and $Q^2(24)$ statistics indicate that there might be some dynamics left in the conditional variance as well.

To entertain the possibility that we need to differentiate between bad and good news we consider a GJR-GARCH(1,1) specification for the conditional variance (Glosten et al. (1993)). The GJR-GARCH(1,1) specification is

$$h_t^2 = a_0 + \left(a_1 + a_2 S_{t-1}^-\right) \varepsilon_{t-1}^2 + b h_{t-1}^2 \tag{13}$$

where S_{t-1}^- is an indicator function which takes the value 1 when $\varepsilon_t < 0$ and value 0 when $\varepsilon_t > 0$.

An LM test of the null of no asymmetry against the alternative of the GJR-GARCH(1,1) specification gives a test-statistic of 31.55 with a corresponding p-value of 1.94×10^{-8} suggesting that there is a need to differentiate beween good and bad news.

Results from re-estimation with the GJR-GARCH(1,1) specification (13) for the conditional variance yield only marginally different results from Table 4 and are not reproduced here. In summary the Box-Pierce statistics have been reduced considerably and the GJR parameter is positive and strongly significant but δ is now close to zero and insignificant at usual levels for both efficient GMM and quasi-maximum likelihood. These results offer no evidence for a GARCH(1,1)-M formulation of the returns to the SP500 index but suggests that it is important to allow for asymmetric GARCH.

5 Final remarks

This paper has introduced a feasible alternative to the common quasi-maximum likelihood estimator and the semi-parametric estimator of GARCH models.

It has been shown that the efficient GMM estimator is simple to compute and asymptotically efficient relative to quasi-maximum likelihood. Hence in practice there is little reason not to prefer it over the common quasi-maximum likelihood estimator. Compared to the semi-parametric estimator efficient GMM has the advantage of being simple to compute and specification testing is also straightforward. One expects that the efficient GMM estimator will find its use in applications.

A Proofs

We first give a series of lemmas that will be useful in the proofs of the theorems. For a random variable X_t let X_T denote $T^{-1}\sum_{t=1}^{T} X_t$, $||X_t||_p$ the L^p-norm of X_t and by $||X_t||$ the ordinary Euclidean norm.

Define the unobserved variance process, which is obtained by extending the observed process into the infinite past history

$$h_t^{2u} = \frac{a_0}{1-b} + a_1 \sum_{k=0}^{\infty} b^k \varepsilon_{t-1-k}^2$$

Lumsdaine (1996) lemma 1 show that $\left|\frac{\partial h_t^{2u}}{\partial a_0} h_t^{-2u}\right|$, $\left|\frac{\partial h_t^{2u}}{\partial a_1} h_t^{-2u}\right|$ are naturally bounded. The lemma below deals with the term $\left|\frac{\partial h_t^{2u}}{\partial b} h_t^{-2u}\right|$ which is more difficult

Lemma A.1 $E\left|\frac{\partial h_t^{2u}}{\partial b} h_t^{-2u}\right|^q < \infty$ *uniformly in* $\theta \in \Theta$ for all $1 \leq q < \infty$

Proof. By Minkowskis inequality

$$\left|\left|\frac{\partial h_t^{2u}}{\partial b} h_t^{-2u}\right|\right|_q^q = \left|\left|\sum_{k=0}^{\infty} b^k h_{t-k-1}^{2u} h_t^{-2u}\right|\right|_q^q \leq \left(\sum_{k=0}^{\infty} b^k \left|\left|\frac{h_{t-k-1}^{2u}}{h_t^{2u}}\right|\right|_q\right)^q$$

Write

$$h_t^{2u} = \frac{a_0}{1-b}\left(1-b^{k+1}\right) + a_1 \sum_{j=0}^{k} b^j \varepsilon_{t-j-1}^2 + b^{k+1} h_{t-k-1}^{2u}$$

and

$$\begin{aligned}
\frac{h_{t-k-1}^{2u}}{h_t^{2u}} &= \frac{h_{t-k-1}^{2u}}{\frac{a_0}{1-b}\left(1-b^{k+1}\right) + a_1 \sum_{j=0}^{k} b^j \varepsilon_{t-j-1}^2 + b^{k+1} h_{t-k-1}^{2u}} \\
&= \left(\frac{1}{b^{k+1}}\right) \frac{b^{k+1} h_{t-k-1}^{2u}}{\frac{a_0}{1-b}\left(1-b^{k+1}\right) + a_1 \sum_{j=0}^{k} b^j \varepsilon_{t-j-1}^2 + b^{k+1} h_{t-k-1}^{2u}} \\
&\leq \left(\frac{1}{b^{k+1}}\right) \frac{b^{k+1} h_{t-k-1}^{2u}}{\frac{a_0}{1-b}\left(1-b^{k+1}\right) + b^{k+1} h_{t-k-1}^{2u}}
\end{aligned}$$

We then have, denoting $v = \frac{a_0}{1-b}\left(b^{-k} - b\right)$ and for c_1 a strictly positive constant

$$
\begin{aligned}
\left(\sum_{k=0}^{\infty} b^k \left\|\frac{h_{t-k-1}^{2u}}{h_t^{2u}}\right\|_q\right)^q &\leq \left(\sum_{k=0}^{\infty}\left(E\left|\frac{1}{\frac{v}{h_{t-k-1}^{2u}} + b}\right|^q\right)^{1/q}\right)^q \\
&\leq \left(\frac{1}{b}\sum_{k=0}^{\infty}\left(E\left|\frac{h_{t-k-1}^{2u}}{v + h_{t-k-1}^{2u}}\right|\right)^{1/q}\right)^q \\
&\leq \left(\frac{1}{b}\sum_{k=0}^{\infty}\left(\frac{1}{c_1\left(b^{-k} - b\right) + 1}\right)^{1/q}\right)^q
\end{aligned}
$$

where the first inequality follows from above, the second since $0 < b < 1$ and $q \geq 1$ and the third by Jensens inequality. Taking the limit, as $q \rightarrow \infty$, of the last term raised to the power of $\frac{1}{q}$ shows that it is not uniformly bounded. But for $q < \infty$ we have

$$
\left(\frac{1}{b}\sum_{k=0}^{\infty}\left(\frac{1}{c_1\left(b^{-k} - b\right) + 1}\right)^{1/q}\right)^q < \infty
$$

uniformly in $\theta \in \Theta$ and hence all finite-order moments exist ■

The following lemma bounds the expectation of the ratio $\left|\frac{h_t^{2u}(\theta_0)}{h_t^{2u}(\theta)}\right|$ uniformly in θ

Lemma A.2 $E\left|\frac{h_{0t}^2}{h_t^{2u}}\right|^q < \infty$ *uniformly in* $\theta \in \Theta$ for all $1 \leq q < \infty$

Proof. We have for $b \geq \beta$

$$
\left|\frac{h_{0t}^2}{h_t^{2u}}\right|^q \leq \left|\frac{\frac{\alpha_0}{1-\beta}}{\frac{a_0}{1-b}} + \frac{\alpha_1}{a_1}\right|^q
$$

and for $b \leq \beta$

$$
\begin{aligned}
\left\|\frac{h_{0t}^2}{h_t^{2u}}\right\|_q^q &\leq \left\|\frac{\frac{\alpha_0}{1-\beta}}{\frac{a_0}{1-b}} + \frac{\alpha_1 \sum_{k=0}^{\infty}\beta^k \varepsilon_{t-k-1}^2}{h_t^{2u}}\right\|_q^q \\
&\leq \left|\frac{\frac{\alpha_0}{1-\beta}}{\frac{a_0}{1-b}}\right|^q + \left|\frac{\alpha_1}{a_1}\right|^q \left\|\sum_{k=0}^{\infty}\left(\frac{\beta}{b}\right)^k \frac{b^k h_{t-k}^{2u}}{\frac{a_0}{1-b}\left(1 - b^k\right) + b^k h_{t-k}^{2u}}\right\|_q^q
\end{aligned}
$$

By a similar argument to that in lemma A.1

$$\begin{aligned}
&\left|\left|\sum_{k=0}^{\infty}\left(\frac{\beta}{b}\right)^{k}\frac{h_{t-k}^{2u}}{\frac{a_0}{1-b}\left(b^{-k}-1\right)+h_{t-k}^{2u}}\right|\right|_{q}^{q} \\
\leq\ & \left(\sum_{k=0}^{\infty}\left(E\left|\left(\frac{\beta}{b}\right)^{k}\frac{h_{t-k}^{2u}}{\frac{a_0}{1-b}\left(b^{-k}-1\right)+h_{t-k}^{2u}}\right|^{q}\right)^{1/q}\right)^{q} \\
\leq\ & \left(\sum_{k=0}^{\infty}\left(\frac{\beta}{b}\right)^{k}\left(E\left|\frac{1}{\frac{\frac{a_0}{1-b}\left(b^{-k}-1\right)}{h_{t-k}^{2u}}+1}\right|\right)^{1/q}\right)^{q}
\end{aligned}$$

and using Jensens inequality

$$\begin{aligned}
&\left(\sum_{k=0}^{\infty}\left(\frac{\beta}{b}\right)^{k}\left(E\left|\frac{x}{\frac{a_0}{1-b}\left(b^{-k}-1\right)+x}\right|\right)^{1/q}\right)^{q} \qquad (14)\\
\leq\ & \left(\sum_{k=0}^{\infty}\left(\left(\frac{\beta}{b}\right)^{kq}\frac{1}{c_2\left(b^{-k}-1\right)+1}\right)^{1/q}\right)^{q} \\
=\ & \left(\frac{\beta}{b}\sum_{k=0}^{\infty}\left(\frac{\beta^{k}}{c_2\left(1-b^{k}\right)+b^{k}}\right)^{1/q}\right)^{q}
\end{aligned}$$

where $c_2 > 0$ and the last term is convergent uniformly in $\theta \in \Theta$ for all $q < \infty$ ■

The lemma below concerns the convergence of the unobserved objective function, Q_T^u based on h_t^{2u} to the limiting objective function, Q

Lemma A.3 $\sup_{\theta\in\Theta}|Q_T^u - Q| \xrightarrow{p} 0$

Proof. Applying the triangle and Cauchy-Schwarz inequalities to $|Q_T^u - Q|$ as in Newey and McFadden (1994) theorem 2.1 we need to show uniform convergence of $||\mathbf{g}_T^u - \mathbf{g}||$ and $||\mathbf{\Lambda}_T^u - \mathbf{\Lambda}||$ to zero. The method chosen here is to first establish a law of large numbers for all $\theta \in \Theta$. Uniform convergence (and continuity of the limiting function) will follow if we can establish stochastic equicontinuity, see Andrews (1992, theorem 3 (a)). Under assumption (c) Nelson (1990) show that h_{0t}^2 is strictly stationary (and covariance-stationary) and ergodic, hence $\mathbf{g}_T^u$ and $\mathbf{\Lambda}_T^u$ are strictly stationary and ergodic since they are measurable functions of h_{0t}.

Note that (ignoring some constants)

$$|g_{it}^u| \leq \left|\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\right| + \left|\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\left(\frac{\varepsilon_t}{h_t^u}\right)\right| + \left|\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)\right|$$

From lemma A.1 we have for the first term

$$E\left|\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\right| < \infty$$

Applying lemma A.1 and A.2 to the second and third terms respectively, using Hölders inequality and the independence of z_t we have $E\left|g_{it}^u\right| < \infty$.

Consider next the cross-product

$$\begin{aligned}
\left|g_{it}^u g_{jt}^{u\prime}\right| \leq\ & \left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j}\right)' h_t^{-4u}\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)^2\right| \\
& + \left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j}\right)' h_t^{-4u}\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)\right| \\
& + \left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j}\right)'\right| + 2\left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j}\right)' h_t^{-4u}\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)^{3/2}\right| \\
& +2\left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j}\right)' h_t^{-4u}\left(\frac{\varepsilon_t}{h_t^u}\right)\right|
\end{aligned}$$

and again using lemma A.1, A.2, Hölders inequality and the independence of z_t gives $E\left|g_{it}^u g_{jt}^{u\prime}\right| < \infty$. This establishes a pointwise law of large numbers. To show convergence is uniform and continuity of the limiting function we establish that (a) $\sup_{\theta\in\Theta} E\left|\frac{\partial g_{it}^u}{\partial \theta_j}\right| < \infty$ and (b) $\sup_{\theta\in\Theta} E\left|\frac{\partial \mathbf{\Lambda}_{ijt}^u}{\partial \theta_k}\right| < \infty$ for all i, j and k. Consider (a)

$$\begin{aligned}
\left|\frac{\partial g_{it}^u}{\partial \theta_j}\right| \leq\ & \left|\frac{\partial^2 h_t^{2u}}{\partial \theta_i \partial \theta_j} h_t^{-2u}\left(\frac{\varepsilon_t}{h_t^u}\right)\right| + \left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j} h_t^{-2u}\right)\left(\frac{\varepsilon_t}{h_t^u}\right)\right| \\
& + \left|\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\frac{\partial h_t^{u}}{\partial \theta_j} h_t^{-u}\left(\frac{\varepsilon_t}{h_t^u}\right)\right| + \left|\frac{\partial^2 h_t^{2u}}{\partial \theta_i \partial \theta_j} h_t^{-2u}\left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1\right)\right| \\
& + \left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j} h_t^{-2u}\right)\left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1\right)\right| \\
& + \left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j} h_t^{-2u}\right)\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)\right|
\end{aligned}$$

and hence we need to find convenient expressions for the second derivative $\left|\frac{\partial^2 h_t^{2u}}{\partial\theta_i\partial\theta_j}h_t^{-2u}\right|$. Differentiating h_t^{2u} twice

$$\begin{aligned}
\frac{\partial h_t^{2u}}{\partial a_0 \partial b}h_t^{-2u} &= \sum_{k=0}^{\infty} b^k \frac{\partial h_{t-k-1}^{2u}}{\partial a_0}h_t^{-2u} \leq \frac{1}{1-b_u}\frac{1-b_l}{a_{0l}} \\
\frac{\partial h_t^{2u}}{\partial a_1 \partial b}h_t^{-2u} &= \sum_{k=0}^{\infty} b^k \frac{\partial h_{t-k-1}^{2u}}{\partial a_1}h_t^{-2u} \leq \sum_{k=0}^{\infty} b^k \varepsilon_{t-k-2}^2 h_t^{-2u} \\
&\leq \frac{1}{a_{1l}}\sum_{k=0}^{\infty} b^k h_{t-k-1}^{2u} h_t^{-2u}
\end{aligned}$$

where we used that $\varepsilon_{t-k-2}^2 \leq \frac{1}{a_{1l}}h_{t-k-1}^{2u}$. For the most demanding derivative with respect to b

$$\begin{aligned}
\frac{\partial^2 h_t^{2u}}{\partial b \partial b}h_t^{-2u} &= 2\sum_{k=0}^{\infty} b^k \frac{\partial h_{t-k-1}^{2u}}{\partial b}h_t^{-2u} = 2\sum_{k=0}^{\infty} b^k \left(\sum_{k=0}^{\infty} b^k \frac{h_{t-k-2}^{2u}}{h_{t-k-1}^{2u}}\right)\frac{h_{t-k-1}^{2u}}{h_t^{2u}} \\
&\leq 2\left(\sum_{k=0}^{\infty} b^k \left|\frac{h_{t-k-1}^{2u}}{h_t^{2u}}\right|\right)^2
\end{aligned}$$

which shows that $\sup_{\theta\in\Theta} E\left|\frac{\partial g_{it}^u}{\partial\theta_j}\right| < \infty$ for all i, j by application of lemma A.1 and A.2. Next for (b) we have

$$\begin{aligned}
\frac{\partial \Lambda_{ijt}^u}{\partial\theta_k} \leq & \left|\frac{\partial h_t^{2u}}{\partial\theta_i}h_t^{-2u}\frac{\partial^2 h_t^{2u}}{\partial\theta_j\partial\theta_k}h_t^{-2u}\left(\frac{\varepsilon_t}{h_t^u} - (\frac{\varepsilon_t^2}{h_t^{2u}} - 1)\right)^2\right| \\
& + \left|\left(\frac{\partial h_t^{2u}}{\partial\theta_i}h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial\theta_j}h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial\theta_k}h_t^{-2u}\right)\left(\frac{\varepsilon_t}{h_t^u} - (\frac{\varepsilon_t^2}{h_t^{2u}} - 1)\right)^2\right| \\
& + \left|\begin{array}{c}\left(\frac{\partial h_t^{2u}}{\partial\theta_i}h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial\theta_j}h_t^{-2u}\right)\frac{\partial h_t^{u}}{\partial\theta_k}h_t^{-u} \\ \left(\frac{\varepsilon_t}{h_t^u}\right)\left(\left(\frac{\varepsilon_t}{h_t^u}\right) - (\frac{\varepsilon_t^2}{h_t^{2u}} - 1)\right)\end{array}\right| \\
& + \left|\begin{array}{c}\left(\frac{\partial h_t^{2u}}{\partial\theta_i}h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial\theta_j}h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial\theta_k}h_t^{-2u}\right) \\ \left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)\left(\left(\frac{\varepsilon_t}{h_t^u}\right) - (\frac{\varepsilon_t^2}{h_t^{2u}} - 1)\right)\end{array}\right|
\end{aligned}$$

and we need to consider $\left|\frac{\partial h_t^{2u}}{\partial\theta_i}h_t^{-2u}\frac{\partial^2 h_t^{2u}}{\partial\theta_j\partial\theta_k}h_t^{-2u}\right|$. For the most demanding derivative

$$\left|\frac{\partial h_t^{2u}}{\partial b} h_t^{-2u} \frac{\partial^2 h_t^{2u}}{\partial b^2} h_t^{-2u}\right| \leq 2\left(\sum_{k=0}^{\infty} b^k \left|\frac{h_{t-k-1}^{2u}}{h_t^{2u}}\right|\right)^3$$

such that $\sup_{\theta\in\Theta} E\left|\frac{\partial \mathbf{\Lambda}_{ijt}^u}{\partial \theta_k}\right| < \infty$ for all i, j and k. Hence the sequences $||\mathbf{g}_T^u - \mathbf{g}||$, $||\mathbf{\Lambda}_T^u - \mathbf{\Lambda}||$ are stochastically equicontinuous ■

The next and final lemma is concerned with the convergence of the unobserved objective function, Q_T^u to the corresponding observed objective function, Q_T

Lemma A.4 $\sup_{\theta\in\Theta} |Q_T^u - Q_T| \xrightarrow{p} 0$

Proof. By the same argument as in lemma A.3 we need to show that $\sup_{\theta\in\Theta} ||\mathbf{g}_T^u - \mathbf{g}_T|| \xrightarrow{p} 0$ and $\sup_{\theta\in\Theta} ||\mathbf{\Lambda}_T^u - \mathbf{\Lambda}_T|| \xrightarrow{p} 0$. First we observe some properties of the conditional variance process

$$\begin{aligned} h_t^{2u} &= h_t^2 + b^{t-1}\left(a_1 \sum_{k=0}^{\infty} b^k \varepsilon_{-k}^2 + a_0 \sum_{k=0}^{\infty} b^k\right) \\ &= h_t^2 + b^{t-1} h_1 \end{aligned}$$

where $h_1 = h_1(\theta)$ is the initial condition. Next

$$E \sup_{\theta\in\Theta} h_t^{2u} \leq \frac{a_{0u}}{1-b_u} + a_{1u} \sum_{k=0}^{\infty} b_u^k E \sup_{\theta\in\Theta} \varepsilon_{t-1-k}^2 < \infty$$

since $E\sup_{\theta\in\Theta} \varepsilon_t^2 < \infty$ under assumption (c) and does not depend on t, which in turn imply that $\sup_{\theta\in\Theta} \varepsilon_t^2 < \infty$. We also have

$$E \sup_{\theta\in\Theta} \frac{\varepsilon_t^2}{h_t^{2u}} \leq \frac{1-b_l}{a_{0l}} E \sup_{\theta\in\Theta} \varepsilon_t^2 < \infty$$

For some arbitrary i, j

$$\begin{aligned} \left|T^{-1}\sum_{t=1}^{T}\left(\Lambda_{ijt}^u - \Lambda_{ijt}\right)\right| &\leq T^{-1}\sum_{t=1}^{T}\left[|g_{it}^u| + |g_{it}|\right] \\ &\left\{\begin{array}{c} \left|\begin{array}{c} \frac{\partial h_t^2}{\partial \theta_j}\frac{\varepsilon_t}{h_t} - \frac{\partial h_t^{2u}}{\partial \theta_j}\frac{\varepsilon_t}{h_t^u} - \frac{\partial h_t^2}{\partial \theta_j}\frac{\varepsilon_t^2}{h_t^2} - \\ \frac{\partial h_t^{2u}}{\partial \theta_j}\frac{\varepsilon_t^2}{h_t^{2u}} + \frac{\partial h_t^2}{\partial \theta_j} - \frac{\partial h_t^{2u}}{\partial \theta_j} \end{array}\right| h_t^{-2} - \\ \left|\frac{\partial h_t^{2u}}{\partial \theta_j}\left(\frac{\varepsilon_t}{h_t^u} - \left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1\right)\right)\right| \left|h_t^{-2u} - h_t^{-2}\right| \end{array}\right\} \end{aligned}$$

since $\sup_{\theta\in\Theta}|g_{it}^u|$, $\sup_{\theta\in\Theta}|g_{it}|$ are finite and $h_t^{-2} \leq \frac{1-b_l}{a_{0l}}$, we must show that

$$(i)\ T^{-1}\sum_{t=1}^{T}\left|\frac{\partial h_t^2}{\partial\theta_j}\left(\frac{\varepsilon_t^r}{h_t^r}\right) - \frac{\partial h_t^{2u}}{\partial\theta_j}\left(\frac{\varepsilon_t^r}{h_t^{ru}}\right)\right| \xrightarrow{p} 0$$

$$(ii)\ T^{-1}\sum_{t=1}^{T}\left|\frac{\partial h_t^{2u}}{\partial\theta_j}\left(\frac{\varepsilon_t}{h_t^u} - \left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1\right)\right)\right|\left|h_t^{-2u} - h_t^{-2}\right| \xrightarrow{p} 0$$

uniformly in θ for $r = 0, 1$ or 2. For (i) we have

$$\left|\frac{\partial h_t^2}{\partial\theta_j}\left(\frac{\varepsilon_t^r}{h_t^r}\right) - \frac{\partial h_t^{2u}}{\partial\theta_j}\left(\frac{\varepsilon_t^r}{h_t^{ru}}\right)\right| \leq \left|\begin{array}{c} \frac{\partial h_t^2}{\partial\theta_j}\left[\left(\frac{\varepsilon_t^r}{h_t^{ru}}\right) - \left(\frac{\varepsilon_t^r}{h_t^r}\right)\right] \\ +\frac{\partial}{\partial\theta_j}b^{t-1}h_1\left(\frac{\varepsilon_t^r}{h_t^{ru}}\right) \end{array}\right|$$

By lemma 3 of Lee and Hansen (1994)

$$T^{-1}\sum_{t=1}^{T}\left[\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right) - \left(\frac{\varepsilon_t^2}{h_t^2}\right)\right] \xrightarrow{p} 0$$

uniformly in θ, hence for $r = 1$ and trivially for $r = 0$. Note that

$$\sup_{\theta\in\Theta}\left|\left|\frac{\partial h_t^2}{\partial\theta}\right|\right| < \infty$$

since

$$\begin{aligned}
\frac{\partial h_t^2}{\partial a_0} &= 1 + b\frac{\partial h_{t-1}^2}{\partial a_0} \leq \sum_{k=0}^{\infty} b^k \leq \frac{1}{1-b_u} \\
\frac{\partial h_t^2}{\partial a_1} &= \varepsilon_{t-1}^2 + b\frac{\partial h_{t-1}^2}{\partial a_0} \leq \sum_{i=0}^{\infty} b^k\varepsilon_{t-k-1}^2 \\
\frac{\partial h_t^2}{\partial b} &= h_{t-1}^2 + b\frac{\partial h_{t-1}^2}{\partial b} \leq \sum_{k=0}^{\infty} b^k h_{t-k-1}^2
\end{aligned}$$

Furthermore $\sup_{\theta\in\Theta}\left(\frac{\varepsilon_t^r}{h_t^{ru}}\right) < \infty$ holds from above, and

$$\begin{aligned}
\sum_{t=1}^{T}\left|\frac{\partial}{\partial\theta_i}b^{t-1}h_1\right| &\leq \sum_{t=1}^{T} b^{t-1}\left|\frac{\partial h_1}{\partial\theta_i}\right| + \sum_{t=1}^{T}(t-1)b^{t-2}h_1 \\
&\leq \frac{1}{1-b_u}\left|\frac{\partial h_1}{\partial\theta_i}\right| + \frac{b_u}{(1-b_u)^2}h_1
\end{aligned}$$

hence

$$T^{-1}\left[\frac{1}{1-b_u}\left|\frac{\partial h_1}{\partial \theta_i}\right| + \frac{b_u}{(1-b_u)^2}h_1\right] \xrightarrow{p} 0$$

which shows (i). Next for (ii) $T^{-1}\sum_{t=1}^{T}\left|h_t^{-2u}-h_t^{-2}\right| \xrightarrow{p} 0$ uniformly in θ by lemma 6 (c) of Lumsdaine (1996), and

$$\left|\frac{\partial h_t^{2u}}{\partial \theta_j}\left(\frac{\varepsilon_t}{h_t^u} - \left(\frac{\varepsilon_t^2}{h_t^{2u}}-1\right)\right)\right|$$

is bounded in probability uniformly in θ which shows (ii). Hence $\sup_{\theta\in\Theta}$ $||\mathbf{\Lambda}_T^u - \mathbf{\Lambda}_T|| \xrightarrow{p} 0$ holds. Finally $||\mathbf{g}_T^u - \mathbf{g}_T|| \xrightarrow{p} 0$ uniformly in θ follows from (i) above ■

We are now ready to give the proofs of the theorems in the text.

Proof theorem 1. By theorem 2.1 (Consistency theorem for extremum estimators) in Newey and McFadden (1994) we need to establish that (i) $\sup_{\theta\in\Theta}|Q_T - Q| \xrightarrow{p} 0$, (ii) Q is continuous, (iii) Θ is compact (which holds by assumption) and (iv) Q is uniquely minimized at θ_0. By the triangle inequality

$$\sup_{\theta\in\Theta}|Q_T - Q| \le \sup_{\theta\in\Theta}|Q_T^u - Q| + \sup_{\theta\in\Theta}|Q_T - Q_T^u|$$

where $\sup_{\theta\in\Theta}|Q_T^u - Q| \to 0$ was established in lemma A.3 and $\sup_{\theta\in\Theta}|Q_T - Q_T^u| \to 0$ was established in lemma A.4. This proves (i) and (ii) follows from lemma A.3 and continuity of multiplication. To finally show (iv) we note that Lumsdaine (1996) lemma 5 prove that $E\left[\frac{\partial h_t^{2u}}{\partial \theta}\frac{\partial h_t^{2u}}{\partial \theta'}h_t^{-4u}\right]$ is a positive definite matrix for all $\theta \in \Theta$. It follows that $E\mathbf{\Lambda}_t^u$ is a positive definite matrix for $v_4 - v_3^2 > 1$, since

$$E\mathbf{\Lambda}_t^u = \frac{1}{(v_4-1)-v_3^2}E\frac{\partial h_t^{2u}}{\partial \theta}\frac{\partial h_t^{2u}}{\partial \theta'}h_t^{-4u}$$

Identification holds if θ_0 is unique solution to $\lim_{T\to\infty}E\mathbf{g}_T^u = 0$. By application of dominated convergence theorem $\lim_{T\to\infty}E\mathbf{g}_T^u = E\mathbf{g}_0$, where $\mathbf{g}_0 = \mathbf{g}(\theta_0)$ and since $E\left(\frac{\varepsilon_t}{h_{0t}}v_3 - \left(\frac{\varepsilon_t^2}{h_{0t}^2}-1\right)\right) = 0$ a consistent root exists at θ_0. In addition this root exists for finite arbitrary initial guess on v_3, v_4 ■

Proof theorem 2. First we note that the fact that a root exists at θ_0 for finite arbitrary guess on v_3, v_4 ensures we can apply theorem 6.2 of Newey and McFadden (1994) to obtain that the asymptotic distribution of the estimator is independent of the guess on v_3, v_4.

Consider the gradient of the efficient GMM objective function

$$\begin{aligned}\frac{\partial}{\partial\theta}Q_T &= \frac{\partial \mathbf{g}_T'}{\partial\theta}\left(\boldsymbol{\Lambda}_T^{-1}+\boldsymbol{\Lambda}_T^{-1}\right)\mathbf{g}_T + \left[\left(\mathbf{g}_T'\otimes\mathbf{g}_T'\right)\frac{\partial \operatorname{vec}\boldsymbol{\Lambda}_T^{-1}}{\partial\theta'}\right]' \\ &= 2\mathbf{G}_T'\boldsymbol{\Lambda}_T^{-1}\mathbf{g}_T - \left[\left(\mathbf{g}_T'\otimes\mathbf{g}_T'\right)\left(\boldsymbol{\Lambda}_T^{-1}\otimes\boldsymbol{\Lambda}_T^{-1}\right)\frac{\partial \operatorname{vec}\boldsymbol{\Lambda}_T}{\partial\theta'}\right]'\end{aligned} \tag{15}$$

with $\mathbf{G}_T = T^{-1}\sum_{t=1}^{T}\frac{\partial \mathbf{g}_t}{\partial\theta'}$, and the second derivative

$$\begin{aligned}\frac{\partial Q_T^2(\theta)}{\partial\theta\partial\theta'} &= \frac{\partial \mathbf{g}_T'}{\partial\theta}\left(\boldsymbol{\Lambda}_T^{-1}+\boldsymbol{\Lambda}_T^{-1}\right)\frac{\partial \mathbf{g}_T}{\partial\theta'} \\ &\quad + \left[\mathbf{g}_T'\left(\boldsymbol{\Lambda}_T^{-1}+\boldsymbol{\Lambda}_T^{-1}\right)\otimes\mathbf{I}_p\right]\frac{\partial}{\partial\theta'}\left[\operatorname{vec}\left(\frac{\partial \mathbf{g}_T'}{\partial\theta}\right)\right] + o_p(1) \\ &= 2\mathbf{G}_T'\boldsymbol{\Lambda}_T^{-1}\mathbf{G}_T \\ &\quad + \left[\mathbf{g}_T'\left(\boldsymbol{\Lambda}_T^{-1}+\boldsymbol{\Lambda}_T^{-1}\right)\otimes\mathbf{I}_p\right]\frac{\partial \operatorname{vec}\mathbf{G}_T'}{\partial\theta'} + o_p(1)\end{aligned} \tag{16}$$

where the $o_p(1)$ term in (16)comes from the derivative of the second term in the gradient. It assumes that $\frac{\partial}{\partial\theta}\left[\frac{\partial \operatorname{vec}\boldsymbol{\Lambda}_T}{\partial\theta'}\right]' = o_p(T)$ uniformly in $\theta\in\Theta$ which can be shown. The proof is complete if we can show that the last terms of (15, 16) are negligible asymptotically and the conditions of theorem 3.2 of Newey and McFadden (1994) holds (Asymptotic normality of minimum distance estimators). Asymptotic normality holds (given consistency) if (i) $\sup_{\theta\in\Theta}||\boldsymbol{\Lambda}_T-\boldsymbol{\Lambda}|| \xrightarrow{p} 0$ and $\boldsymbol{\Lambda}$ is non-singular (which holds from above) (ii) $\sup_{\theta\in\Theta}\left|\left|\frac{\partial \mathbf{g}_T^u}{\partial\theta'}-\mathbf{G}\right|\right| \xrightarrow{p} 0$ and $\mathbf{G} = E\frac{\partial \mathbf{g}_T^u}{\partial\theta'}$ is non-singular (iii) $\theta_0\in int(\Theta)$ (which holds by assumption) and (iv) asymptotic normality of $\sqrt{T}\mathbf{g}_T$.

To prove (ii) note that we have already shown that $\left|\left|\frac{\partial \mathbf{g}_T^u}{\partial\theta'}-\mathbf{G}\right|\right| \xrightarrow{p} 0$ in the proof of theorem 1. As in lemma A.3 uniform convergence and continuity of the limiting function follows if $\sup_{\theta\in\Theta}E\left|\frac{\partial \mathbf{G}_{ijt}^u}{\partial\theta_k}\right| < \infty$ for all i,j and k.

Differentiating $\frac{\partial g_{it}^u}{\partial \theta_j}$ once more

$$\begin{aligned}
\left|\frac{\partial^2 g_{it}^u}{\partial \theta_j \partial \theta_k}\right| &\leq \left|\frac{\partial^3 h_t^{2u}}{\partial \theta_i \partial \theta_j \partial \theta_k} h_t^{-2u}\left(\left(\frac{\varepsilon_t}{h_t^u}\right)-\left(\frac{\varepsilon_t^2}{h_t^{2u}}-1\right)\right)\right| \\
&+\left|\begin{array}{c} \frac{\partial^2 h_t^{2u}}{\partial \theta_i \partial \theta_j} h_t^{-2u} \frac{\partial h_t^{2u}}{\partial \theta_k} h_t^{-2u} \\ \left[3\left(\left(\frac{\varepsilon_t}{h_t^u}\right)-\left(\frac{\varepsilon_t^2}{h_t^{2u}}-1\right)\right)-\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)^{3/2}\right] \end{array}\right| \\
&+\left|\begin{array}{c} \left[\frac{\partial^2 h_t^{2u}}{\partial \theta_i \partial \theta_j} h_t^{-2u} \frac{\partial h_t^u}{\partial \theta_k} h_t^{-u}-\left(\frac{\partial h_t^{2u}}{\partial \theta_j} h_t^{-2u}\right)^2 \frac{\partial h_t^u}{\partial \theta_k} h_t^{-u}\right] \\ \left(\frac{\varepsilon_t}{h_t^u}\right)\left(\frac{\varepsilon_t^2}{h_t^{2u}}-1\right) \end{array}\right| \\
&+\left|2\left(\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_k} h_t^{-2u}\right)\left(\frac{\varepsilon_t}{h_t^u}\right)\right| \\
&+\left|\begin{array}{c} 2\left(\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_k} h_t^{-2u}\right) \\ \left(-\left(\frac{\varepsilon_t^2}{h_t^{2u}}-1\right)-\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)^{3/2}\right) \end{array}\right|
\end{aligned}$$

and in view of lemma A.1 and A.2 we need to consider the third derivative of h_t^{2u}. For the derivative with respect to b we have

$$\frac{\partial^3 h_t^{2u}}{\partial b^3} h_t^{-2u} \leq 3 \sum_{k=0}^{\infty} b^k \frac{\partial^2 h_{t-k-1}^{2u}}{\partial b^2} h_t^{-2u} \leq 6\left(\sum_{k=0}^{\infty} b^k\left|\frac{h_{t-k-1}^{2u}}{h_t^{2u}}\right|\right)^3$$

and hence applying lemma A.1 and A.2 using Hölders inequality and the independence of z_t we conclude that convergence is uniform and $\mathbf{G}$ is continuous on Θ. Then $\sup_{\theta \in \Theta}||\mathbf{G}_T-\mathbf{G}|| \xrightarrow{p} 0$ holds if $\sup_{\theta \in \Theta}||\mathbf{G}_T^u-\mathbf{G}_T|| \rightarrow 0$ which can be shown. It is straightforward to verify that $E\mathbf{G}_{ijt}(\theta_0)=E\mathbf{\Lambda}_{ijt}(\theta_0)$ for all i, j which shows the nonsingularity of $\mathbf{G}$ as well. Next (iv) follows since $\mathbf{g}_t$ is a stationary ergodic martingale difference sequence with finite variance and hence $T^{\frac{1}{2}}\mathbf{g}_T \xrightarrow{d} N(\mathbf{0}, \mathbf{\Lambda})$. It remains to show that the last terms in (15, 16) are unimportant for the asymptotic distribution. We have from (15)

$$\frac{\partial}{\partial \theta} Q_T=2\mathbf{G}_T' \mathbf{\Lambda}_T^{-1} \mathbf{g}_T-\left[\left(\mathbf{g}_T' \otimes \mathbf{g}_T'\right)\left(\mathbf{\Lambda}_T^{-1} \otimes \mathbf{\Lambda}_T^{-1}\right) \frac{\partial \operatorname{vec} \mathbf{\Lambda}_T}{\partial \theta'}\right]'$$

and scaling by $T^{1/2}$, since $T^{1/2}\mathbf{g}_T$ is bounded in probability, $\sup_{\theta \in \Theta}||\mathbf{\Lambda}_T-\mathbf{\Lambda}|| \xrightarrow{p} 0$ and $\frac{\partial \operatorname{vec} \mathbf{\Lambda}_T}{\partial \theta'}=O_p(1)$ uniformly in $\theta \in \Theta$ we have the result. A similar argument applied to the second term of (16) establishes the result here as well since $\frac{\partial \operatorname{vec} \mathbf{G}_T}{\partial \theta'} = O_p(1)$ uniformly in $\theta \in \Theta$. Applying a standard mean-value expansion of the gradient vector as in Newey and McFadden (1994) theorem 3.2 then obtains the distributional result given in theorem 2 ■

References

Andrews, D. W. K. (1992), 'Generic uniform convergence', *Econometric Theory* **8**, 241–257.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Bollerslev, T. and Wooldridge, J. M. (1992), 'Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances', *Econometric Reviews* **11**, 143–172.

Box, G. E. P. and Pierce, D. A. (1970), 'Distribution of residual autocorrelation in autoregressive-integrated moving average time series models', *Journal of The American Statistical Association* **65**, 1509–1526.

Engle, R. F. (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation', *Econometrica* **50**, 987–1007.

Engle, R. F. and Gonzales-Rivera, G. (1991), 'Semiparametric ARCH models', *Journal of Business and Economic Statistics* **9/4**, 345–359.

Engle, R. F., Lilien, D. M. and Robins, R. P. (1987), 'Estimating time-varying risk premia in the term structure', *Econometrica* **55/2**, 391–407.

French, K. R., Schwert, W. G. and Staumbaugh, R. F. (1987), 'Expected stock returns and volatility', *Journal of Financial Economics* **19**, 391–407.

Glosten, L., Jagannathan, R. and Runkle, D. (1993), 'On the relation between expected values and the volatility of the nominal excess return on stocks', *Journal of Finance* **48**, 1779–1801.

Gonzalez-Rivera, G. and Drost, F. C. (1999), 'Efficiency comparisons of maximum-likelihood based estimators in GARCH models', *Journal of Econometrics* **93**, 93–111.

Granger, C. W. J. and Ding, Z. (1995), 'Some properties of the absolute returns: An alternative measurement of risk', *Annales d'Economie et de Statistique* **40**, 67–91.

Hansen, L. P. (1982), 'Large sample properties of generalized method of moment estimators', *Econometrica* **50**, 1029–1054.

Lee, S. W. and Hansen, B. (1994), 'Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator', *Econometric Theory* **10**, 29–52.

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.

Mills, T. C. (1999), *The Econometric Modelling of Financial Time Series*, Cambridge University Press.

Nelson, D. D. (1990), 'Stationarity and persistence in the GARCH(1,1) models', *Econometric Theory* **6**, 318–334.

Newey, W. K. and McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* R. F. Engle and D. McFadden, eds, 'Handbook of Econometrics', Vol. 4, Amsterdam North Holland, chapter 36, pp. 2111–2245.

Newey, W. K. and West, K. (1987), 'Hypothesis testing with efficient method of moments estimation', *International Economic Review* **28**, 777–787.