This is to certify that the

dissertation entitled

Estimation of the GARCH Model:
Improving the Normal Quasi-MLE by Augmented GMM

presented by

Yi-Yi Chen

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Economics

Peter Schmidt
Major professor

Date Apr.12, 2000

ESTIMATION OF THE GARCH MODEL: IMPROVING THE NORMAL QUASI-MLE BY AUGMENTED GMM

By

Yi-Yi Chen

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

2000

ABSTRACT

ESTIMATION OF THE GARCH MODEL: IMPROVING THE NORMAL QUASI-MLE BY AUGMENTED GMM

By

Yi-Yi Chen

The standard estimator for ARCH models is the normal quasi maximum likelihood estimator (NQMLE). We interpret the NQMLE as a GMM estimator whose moment conditions are the normal score function, and we seek to improve it by adding more moment conditions based on autocorrelations of squares and on the score function for a rescaled t distribution. These augmented GMM estimators are asymptotically more efficient than the NQMLE when the data are non-normal. We evaluate the efficiency gain and find that it can be large, especially when the data are skewed. Simulations indicate that achieving these gains in practice will require a rather large sample size, such as 1,000 or more. Finally, we estimate a model for the DM/$ exchange rates, and find that the augmented GMM estimator performs largely as asymptotic theory and our simulations would predict.

ACKNOWLEDGEMENTS

Life is unpredictable. During these years of pursuing the Ph.D. degree, I have experienced some of the most dramatic moments in my life. Thank God for giving me the strength and courage to walk through them. I'll always remember my mother as a kind and loving woman who dedicated all her life to the family. I only wish she could share the enjoyment of my accomplishment with me.

For the completion of this dissertation, first I would like to thank my advisor Professor Peter Schmidt. The breadth of his research experience and the depth of his knowledge in economics have provided me with tremendous help throughout the course of writing this dissertation. His financial support in the last year and the editing on the dissertation are also gratefully acknowledged. I also thank Professor Jeffery Wooldridge who gave insightful comments on the dissertation and clarified some assumptions used in this work. The part of empirical work uses the data provided by Professors Richard Baillie, Tim Bollerslev and Charles Goodhart. This dissertation cannot be finished without their assistances.

Throughout these years of studying in the States, I never confronted financial problems. For this I sincerely thank my parents-in-law for their financial supports. My father's encouragement, both on the emotional and spiritual sides, are powerful support to me. In addition, I thank my sisters, Fang-Yi, Ang-Yi and Chu-En, for their unconditional love.

I would also like to thank my fellow classmates and friends for helping me in different ways. I specially thank Bih-Shiow, Te-Fen, Chiung-Ying, and Chung-Jung for sharing the happiness and life experiences with me, and also for driving me to shopping. At last, I thank my husband Hung-Jen; without his love and support, I cannot complete this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 4

Chapter 1
ARCH-TYPE MODELS AND ESTIMATION METHODS

1. Introduction

It is an empirical regularity that many data, such as stock returns, commodity prices, and foreign exchange rates, exhibit "volatility clustering" – periods of low volatility (variance) tend to cluster together followed by periods of high volatility. For example, French et al. (1987) show that in the period between 1928 and 1990, daily capital gains have a larger variance during the 1930's than during the 1960's. In the case of the Deutschmark/U.S. Dollar exchange rate during the period of 1981 to 1992, Baillie and Bollerslev (1989) also show that the daily exchange rates have periods of turbulence followed by periods of tranquillity.

This volatility clustering property was first documented by Mandelbrot (1963) and Fama (1965), and was modeled econometrically by Engle (1982) as an Autoregressive Conditional Heteroskedasticity (ARCH) process. The idea of ARCH modeling is to allow the conditional variance depend on the history of the series. This is very different from the traditional model in which the conditional variance is assumed to be independent of the past information.

The modeling of time varying variances has important implications for dynamic economic theory and modern finance theory. Recognizing the temporal pattern of time-varying heteroskedasticity can help to impose the accuracy of forecasts and of econometric inference. Furthermore, because risk and uncertainty play importance roles in finance theory, the ARCH model is useful to incorporate risk and uncertainty, as measured by variances and covariances, into the analysis of asset and option pricing.

2. ARCH and GARCH Models

Suppose that ϵ_t, t = 1,2,...,T is an observable series with $E(\epsilon_t) = 0$ and unconditional variance $Var(\epsilon_t) = \sigma^2$. Define Ψ_t as the information set at time t. We assert that $E(\epsilon_t|\Psi_{t-1}) = 0$. The conditional variance, h_t, can be expressed as $Var(\epsilon_t|\Psi_{t-1})$. How to model this conditional variance is what we are interested in.

The basic model of ARCH type is

$$\epsilon_t = h_t^{1/2} \cdot u_t, \quad \text{with } u_t \text{ iid } \ D(0,1), \tag{1}$$

where D is a specified distribution, such as the standard normal distribution or the standardized t distribution with a certain number of degrees of freedom, etc. The information set can be expressed as

$$\Psi_t = \{\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \cdots; h_t, h_{t-1}, \cdots\},$$

or in principle just as

$$\Psi_t = \{\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \cdots\},$$

since h_t ultimately is a function of past observations.

It may be noted that the representation in equation (1) is stronger than some possible definitions of an ARCH model. For example, we could define the model simply by the assumptions that $E(\epsilon_t|\Psi_{t-1}) = 0$, and $Var(\epsilon_t|\Psi_{t-1}) = h_t$. The representation in equation (1) implies these results for the first two conditional moments, but also places restrictions on the higher conditional moments. Specifically, it implies that $E(\epsilon_t^k|\Psi_{t-1}) = h_t^{k/2}\mu_k$ where $\mu_k \equiv E(u_t^k)$. The results in this thesis generally require the validity of the representation in equation (1), not just the correctness of the first two conditional moment assumptions.

Different specific models are defined depending on how h_t is related to Ψ_{t-1}. We discuss a number of the formulations below.

A. ARCH

A.1 First-order linear ARCH

In the simplest ARCH model, ARCH(1), which was introduced by Engle (1982), the conditional variance depends only on the latest past squared innovation,

$$h_t = \omega + \alpha_1 \epsilon_{t-1}^2, \quad \omega > 0,\ \alpha_1 \geq 0.$$

If $\alpha_1 = 0$, ϵ_t would be white noise. Otherwise, ϵ_t will be dependent through higher order moments.

A.2 General ARCH (ARCH(q))

The qth-order linear ARCH model is

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 = \omega + \alpha(L)\epsilon_t^2, \quad \text{with } \omega > 0,\ \alpha_i \geq 0 \text{ for each i},$$

where L is the lag operator so that $L\epsilon_t = \epsilon_{t-1}$ and $\alpha(L) = \sum_{i=1}^{q} \alpha_i L^i$. If and only if the sum of the α_i is less than one, the process is covariance stationary, in which case the unconditional variance is $\sigma^2 = \omega/(1 - \alpha_1 - \alpha_2 - \cdots - \alpha_q)$. Engle (1982) uses this specification to model the uncertainty of the inflation rate. Bodurtha and Mark

(1991) employ the ARCH(3) specification to model monthly NYSE stock returns, and the same formulation is adopted by Attanasio (1991) to model monthly excess returns on the S&P 500 index.

Although the ARCH model does imply volatility clustering, there are some difficulties in empirical applications. For example, without the restrictions on the lag structure, estimation may result in negative parameter estimates and fail the non-negative constraints. Also, one may need a large value of q in h_t in order to model the conditional variance correctly.

To avoid these two problems, Bollerslev (1986) proposed the Generalized Autoregressive Conditional Heteroskedasticity model, or GARCH model.

B. GARCH(p, q)

The GARCH process modifies the ARCH process by extending the AR process for ϵ_t^2 to an ARMA process, potentially permitting a more parsimonious parameterization.

In the GARCH(p, q) model of Bollerslev (1986), h_t is defined as follows:

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \tag{2}$$

where

$$p \geq 0, \quad q > 0,$$

$$\omega > 0, \ \alpha_i \geq 0, \quad i = 1, \ldots, q, \ \beta_i \geq 0, \ i = 1, \ldots, p.$$

The GARCH(p, q) process is covariance stationary if and only if $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$, in which case the unconditional variance is $\sigma^2 = \omega/(1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i)$. For $p = 0$ the process reduces to the ARCH(q) process.

The leading case of the GARCH model is the GARCH(1,1) model, with $p = q = 1$. The conditional variance for the GARCH(1,1) model is given by

$$h_t = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1},$$

where $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$. This model is covariance stationary if and only if $\alpha + \beta < 1$.

Many researchers have found that the GARCH(1,1) model performs well with high frequency data (Diebold, 1987). For example, Baillie and Bollerslev (1989) conclude "the conditional heteroskedasticity in daily spot rates is well represented by a GARCH(1,1) process with near unit roots" and the GARCH effects remain very significant for weekly data. Also, Hsieh (1988) demonstrates that the GARCH(1,1) model is suitable to describe daily data on nominal U.S. dollar return rates.

C. EGARCH

In the GARCH(p, q) model, the conditional variance only depends on the size of the past innovations and not their sign. Nelson (1991) points out that such a model cannot capture the "leverage effect" (the tendency for changes in stock prices to be negatively correlated with changes in stock volatility) which was first observed by Black (1976). Nelson (1991) proposed the Exponential GARCH(p, q) model, or EGARCH(p, q), which is related to the log-GARCH model proposed by Pantula (1986) and Geweke (1986). In the EGARCH model, h_t depends on both the sign and the magnitude of ϵ_t :

$$\ln(h_t) = \omega + \sum_{i=1}^{q} \alpha_i(\phi u_{t-i} + \gamma[|u_{t-i}| - E|u_{t-i}|]) + \sum_{i=1}^{p} \beta_i \ln(h_{t-i}).$$

There are no restrictions needed on α_i and β_i to ensure the nonnegativity of the conditional variance. If u_t is assumed iid normal, then ϵ_t is covariance stationary if $\sum_{i=1}^{p} \beta_i < 1$. Empirical evidence for stock returns supports this specification with $\alpha_i \phi < 0$, which implies that the conditional variance tends to increase (decrease) when ϵ_{t-i} is negative (positive), see Nelson (1991).

D. NARCH

As another variant of the ARCH family, Higgins and Bera (1992) propose the non-linear ARCH (NARCH) model:

$$h_t^{\gamma} = \omega + \sum_{i=1}^{q} \alpha_i (\epsilon_{t-i}^2)^{\gamma} + \sum_{i=1}^{p} \beta_i h_{t-i}^{\gamma},$$

where $\omega \geq 0, \alpha_i \geq 0, \beta_i \geq 0$, and $\gamma > 0$. The reason for proposing a nonlinear form for the conditional variance is that this specification is more flexible than the linear GARCH model. This formulation is equivalent to the general GARCH(p,q) model when $\gamma = 1$. With $\gamma = 1/2$, the conditional standard deviation $h_t^{1/2}$ is a distributed lag of absolute residuals as proposed by Taylor (1986) and Schwert (1989). Higgins and Bera (1992) suggest that the NARCH model does better for modeling the weekly exchange rate series than the linear GARCH model.

E. TARCH

The threshold ARCH model, or TARCH, is designed to take into account that the market's responses to good and bad news may be asymmetric. The conditional variance is defined as

$$h_t^{1/2} = \omega + \sum_{i=1}^{q} [\alpha_i^+ \ \mathrm{I}(\epsilon_{t-i} > 0)|\epsilon_{t-i}|^{\gamma} + \alpha_i^- \ \mathrm{I}(\epsilon_{t-i} \leq 0)|\epsilon_{t-i}|^{\gamma}] + \sum_{i=1}^{p} \beta_i h_{t-i}^{1/2},$$

where I(.) is the indicator function. Zakoian (1990) uses the model with $\gamma = 1$. Glosten, Jagannathan and Runkle (1993) use this model with $\gamma = 2$ for describing the nominal excess return on stocks. The model is attractive because it allows for more flexible responses of volatility to shocks of different signs and magnitude.

F. ARCH in mean

Consider a stochastic process, say y_t, where

$$y_t = f(\Psi_{t-1}; b) + \epsilon_t,$$

and $f(\Psi_{t-1}; b)$ is a function of Ψ_{t-1} and the parameter vector b.

In the ARCH-in-mean model, or ARCH-M, which was introduced by Engle, Lilien, and Robins (1987), the conditional mean is an explicit function of the conditional variance,

$$\mu_t = f(\Psi_{t-1}, h_t; b),$$

where μ_t is the mean of the stochastic process y_t. Some finance theories predict a tradeoff between the expected returns and the variance, or the covariance among the returns. This model is capable of explaining the relation between the conditional variance and the conditional mean provided the sign of the first derivative of $f(h_t, b)$ with respect to h_t is positive. The ARCH-M model has been applied to different stock index returns, such as the daily S&P index by French, Schwert, and Stambaugh (1987), and quarterly U.S. stock indices by Friedman and Kuttner (1988).

G. IGARCH

The Integrated GARCH model, or IGARCH, proposed by Engle and Bollerslev (1986), is the special case of the general GARCH(p, q) model in (2) with $\sum_{i=1}^{q} \alpha +$

$\sum_{i=1}^{p} \beta = 1$. It has the property that shocks to the conditional variance in the model persist permanently, so sometimes it is also characterized as having "persistent variance" or "integrated variance".

The unconditional variance for the IGARCH(p, q) model does not exist, since $\sum_{i=1}^{q} \alpha + \sum_{i=1}^{p} \beta = 1$. In the IGARCH model, the process is not covariance stationary, but Nelson (1990) shows that it is strictly stationary and ergodic.

3. MLE for the GARCH Process

To discuss maximum likelihood estimation, we presume that the distribution of u_t in equation (1) is known. That is, we presume that the distributional assumption made for u_t is correct. Let $f(u_t)$ denote the density function for $u_t \equiv \epsilon_t / h_t^{1/2}$, normalized to have mean zero and variance one. For the general model given in equation (1), the log likelihood function can be written as

$$L_T(\theta) = \sum_{t=1}^{T} l_t(\theta), \tag{3}$$

where the contribution of the tth observation is:

$$l_t(\theta) = -\frac{1}{2}\log\,[h_t(\theta)] + \log\,\{f[u_t(\theta)]\}. \tag{4}$$

Here the notation $h_t(\theta)$ indicates that h_t depends on some parameters θ, so that $u_t(\theta) = \epsilon_t / h_t(\theta)^{1/2}$, but for simplicity we will hereafter just use the notation h_t and u_t. The MLE of θ is obtained by maximizing $L_T(\theta)$, as given in equation (3), with respect to θ. For example, in the GARCH(p,q) model, $\theta = (\omega, \alpha_1, \alpha_2, \cdots, \alpha_q, \beta_1, \beta_2, \cdots, \beta_p)'$.

Now consider the commonly-assumed case in which the distribution of u_t is standard normal. Then the log-density function for the tth observation, apart from an irrelevant constant, is

$$l_t(\theta) = -\frac{1}{2}\log h_t - \frac{1}{2}\epsilon_t^2 h_t^{-1}. \tag{5}$$

The first and the second derivatives with respect to θ are

$$q_{1t}(\epsilon, \theta) = \frac{\partial l_t}{\partial \theta} = \frac{1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial \theta}\Big(\frac{\epsilon_t^2}{h_t} - 1\Big), \tag{6}$$

$$\frac{\partial^2 l_t}{\partial \theta \partial \theta'} = \Big(\frac{\epsilon_t^2}{h_t} - 1\Big)\frac{\partial}{\partial \theta'}\Big[\frac{1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial \theta}\Big] - \frac{1}{2}\frac{1}{h_t^2}\frac{\partial h_t}{\partial \theta}\frac{\partial h_t}{\partial \theta'}\frac{\epsilon_t^2}{h_t},$$

where

$$\epsilon = (\epsilon_0, \epsilon_1, \cdots, \epsilon_t)'.$$

In addition, if we consider the GARCH(p,q) model, we have:

$$\frac{\partial h_t}{\partial \theta} = z_t + \sum_{i=1}^{p} \beta_i \frac{\partial h_{t-i}}{\partial \theta},$$

$$\text{where} \quad z_t = (1, \epsilon_{t-1}^2, \cdots, \epsilon_{t-q}^2, h_{t-1}, \cdots, h_{t-p}).$$

Following Weiss (1986), the maximum likelihood estimators $\hat{\theta}_{MLE}$ of the parameters are consistent and asymptotically normal with mean θ_0 (the subscript 0 represents the vector of true parameters) and covariance matrix A^{-1}, where

$$A = -E[\partial^2 l_t / \partial\theta\partial\theta'] = E[(\partial l_t/\partial\theta)(\partial l_t/\partial\theta')].$$

This covariance matrix equals the Cramer-Rao lower bound.

It is difficult to establish asymptotic theory for the estimation of the IGARCH model. Hong (1987) provides Monte Carlo evidence for the IGARCH(1,1) model, and suggests that the sample size must be very large ($\approx$ 5,000 observations) for the asymptotic distributions to be good approximations. On the other hand, Lumsdaine (1996) proves that the Normal MLE of the parameters in the IGARCH(1,1) model is still consistent and asymptotically normal.

An interesting sidelight, not previously noted in the literature, is that the MLE will not be consistent for certain distributions of u_t. For example, let u_t be standard chi-square with ν degrees of freedom, so that $u_t = (\tilde{u}_t - \nu)/\sqrt{2\nu}$ where $\tilde{u}_t$ is χ^2_ν. Then $\tilde{u}_t \geq 0$ is equivalent to $u_t \geq -\nu/\sqrt{2\nu}$ and $\epsilon_t \geq -h_t^{1/2}\nu/\sqrt{2\nu}$, so that the range of ϵ_t depends on θ (through h_t). This violates one of the standard regularity conditions for the consistency of the MLE. More detail is given in Section A of Appendix 1. For an example of an article that misses this point, see Engle and Gonzalez-Rivera (1991).

4. QMLE for the GARCH Process

The likelihood function based on a distributional assumption provides a criterion function whose maximization defines an estimator. In the case that the distributional assumption is not correct, this estimator is called a quasi-maximum likelihood estimator, or QMLE. For example, the normal QMLE is simply the estimator that maximizes the normal log likelihood function.

The properties of the QMLE will depend upon the "assumed" distribution, which dictates the form of the estimator, and the "true" distribution, which is a characteristic of the data generating process. In general, when the assumed distribution is not the same as the true distribution, the QMLE will be inconsistent. However, for some assumed distributions, notably the normal, the QMLE is consistent and asymptotically normal, so long as the true distribution satisfies some regularity conditions. We now discuss two important types of QMLE.

A. Normal QMLE

The Normal QMLE was investigated by Weiss (1986) and Bollerslev and Wooldridge (1992). Bollerslev and Wooldridge use it in estimating a multivariate GARCH model. They find that even if normal distributional assumption does not hold, estimates based on the normal log-likelihood function are still consistent and asymptotically normal, provided that both the mean and variance equations are correctly specified and that some regularity conditions are satisfied.

The consistency and asymptotic normality of the Normal QMLE are given in the following theorem, which is proved by Bollerslev and Wooldridge (1992):

THEOREM(B&W) : If (1) The regularity conditions in Section B of Appendix 1 are satisfied, and (2) For some $\theta \in \text{int } \Theta$, $E(y_t|\Psi_{t-1}) = \mu_t(\theta_0)$ and $Var(y_t|\Psi_{t-1}) = \Omega_t(\theta_0)$, then

$$(A^{-1}BA^{-1})^{-1/2}\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow N(0, I),$$

where $\hat{\theta}_T$ is the QMLE

$$B = T^{-1}\sum_{t=1}^{T} E[s_t(\theta_0)s_t(\theta_0)'],$$

$s_t(\theta_0)$ is the score function of $l_t(\theta)$,

$$l_t = -1/2 \log|\Omega_t(\theta)| - 1/2 \; (y_t - \mu(\theta))'\Omega_t(\theta)^{-1}(y_t - \mu(\theta)),$$

and

$$A = T^{-1}\sum_{t=1}^{T} E[d_t(\theta_0)], \quad d_t(\theta_0) \text{ is the hessian function of } l_t(\theta),$$

In addition,

$$\hat{A}_T - A \to 0 \text{ and } \hat{B}_T - B \to 0,$$

where

$$\hat{A}_T = T^{-1}\sum_{t=1}^{T} d_t(\hat{\theta}_T), \text{ and } \hat{B}_T = T^{-1}\sum_{t=1}^{T} s_t(\hat{\theta}_T)s_t(\hat{\theta}_T)'.$$

The matrix $\hat{A}_T^{-1}\hat{B}_T\hat{A}_T^{-1}$ is the robust covariance matrix of White (1982). If the true distribution is normal, (or, more generally, if the third conditional moment is zero and the fourth conditional moment is three times the square of the conditional variance; that is, the first four conditional moments are the same as for the normal distribution), $A = B$, and therefore the asymptotic covariance matrix of the QMLE is simply A^{-1}.

In the GARCH(1,1) case, the (3×1) score function, $q_{1t}(\epsilon, \theta)$, can be represented as in equation (6), with $\theta = (\omega, \alpha, \beta)'$. The conditional expected value of the score function, $E(q_{1t}|\Psi_{t-1})$, is zero when the first two conditional moments are correctly specified as

$$E(\epsilon_t|\Psi_{t-1}) = 0,$$

$$Var(\epsilon_t|\Psi_{t-1}) = h_t = w + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}.$$

Lumsdaine (1996) provides different regularity conditions to prove asymptotic normality for the normal QMLE of the GARCH(1,1) model. There are two assumptions for the true parameters and the distribution of u_t made by Lumsdaine (1996); see Section C of Appendix 1.

B. Non-Gaussian QMLE

Non-Gaussian densities have become increasingly popular for the estimation of GARCH models. Examples include the student's t (Bollerslev (1987)) and the exponential power distribution (Nelson (1991)). The consistency and asymptotic normality of the QMLE based on non-Gaussian distributions has been investigated by Newey and Steigerwald (1997). Newey and Steigerwald show that consistency holds when a particular identification condition is satisfied. The identification condition is that there exists a unique maximum of the quasi-likelihood function at the true conditional mean and *relative* scale parameters. This condition is essential for the consistency of the QMLE. Newey and Steigerwald (1997) conclude that the identification condition holds if the conditional mean is identically zero, or a symmetry condition (the true and assumed densities are both unimodal and symmetric around zero) is satisfied. When the symmetry condition does not hold, one additional *location* parameter should be included to establish the identification condition for consistency.

According to Newey and Steigerwald's setup, the basic model can be represented as:

$$\epsilon_t = H_t^{1/2}(\gamma + \sigma_s u_t), \tag{7}$$

where γ is the *location* of the innovation distribution and σ_s is the scale parameter for the density of u_t. The conditional variance H_t is rescaled by a constant term. For example, in the GARCH(p,q) model, H_t is h_t in equation (2) rescaled by the constant term, ω. That is, $h_t = \omega H_t$ and correspondingly

$$H_t(\theta) = 1 + \sum_{i=1}^{q} \left(\frac{\alpha}{\omega}\right)_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i H_{t-i}, \tag{8}$$

where $(\frac{\alpha}{\omega})_i$ is the relative scale parameter and $\theta = ((\frac{\alpha}{\omega})_1, (\frac{\alpha}{\omega})_2, \cdots, (\frac{\alpha}{\omega})_q, \beta_1, \beta_2, \cdots, \beta_p, \varphi')'$, where φ is the parameter vector in mean function.

Thus we hope for consistent estimation of $\left(\alpha/\omega\right)_i$ and β_i, as opposed to the entire set of parameters $(\omega, \alpha_i, \beta_i)$. In particular, we generally do not obtain a consistent estimate of scale (ω).

Given the assumed density $a(u_t, \nu)$ with ν representing nuisance parameters in the density (e.g. degrees of freedom), the non-Gaussian QMLE is the value of $\delta = (\theta', \gamma, \sigma_s, \nu)'$ that maximizes the log-likelihood function,

$$L_T(\delta) = T^{-1} \sum_{t=1}^{T} l_t(\delta), \tag{9}$$

where $l_t(\delta)$ is given by:

$$l_t(\delta) = -\ln \sigma_s - \frac{1}{2}\ln H_t + \ln a([\sigma_s H_t^{1/2}(\theta)]^{-1}[y_t - f_t(\theta) - \gamma H_t^{1/2}(\theta)], \nu),$$

where $f_t(\theta)$ is the mean function. Notice that the additional parameter γ enters the conditional mean but not the conditional variance.

The identification condition for consistency as given by Newey and Steigerwald (1997) is stated in the following theorems. The first theorem applies to the case where the additional location parameter does not need to be included. The log-likelihood function is the same as equation (9) but with $\gamma \equiv 0$ and $\epsilon_t = H_t^{1/2}\sigma_s u_t$. The second theorem applies to the case where the location parameter is added to the model.

THEOREM (N&S, 1) : If Assumptions 1, 2, 3 in Section D of Appendix 1 are satisfied, and either Assumption 4 is satisfied or the conditional mean equals zero, then the expected log-likelihood $\bar{L}(\delta) = E[l_t(\delta)]$ has a unique maximum at some $\bar{\delta}$ with $\bar{\theta} = \theta_0$.

THEOREM (N&S, 2) : If Assumptions 1, 5, 6 in Section D of Appendix 1 are satisfied, then the expected log-likelihood $\bar{L}(\delta) = E[l_t(\delta)]$ has a unique maximum at some $\bar{\delta}$ with $\bar{\theta} = \theta_0$.

Also, for the ARCH and GARCH models, they prove that there is no asymptotic efficiency loss for the QMLE even if the location parameter is included in the case where the symmetry condition holds.

As previously stated, when the conditional mean is identically zero, the identification condition is satisfied without imposing the symmetry condition or the additional location parameter. The martingale-GARCH(1,1) model we used in this dissertation satisfies the condition that the conditional mean equals zero. Thus, the symmetry condition is not an issue in our model. The rescaled conditional variance of the GARCH(1,1) process, H_t, has the form of

$$H_t = 1 + \frac{\alpha}{\omega}\epsilon_{t-1}^2 + \beta H_{t-1}.$$

We will consider cases in which the student's t density is used. Many researchers find that the thick tail property of the student's t distribution describes financial data relatively well. The student's t QMLE, following the setup of Newey and Steigerwald

(1997), is the value of $\delta = (\sigma_s, \frac{\alpha}{\omega}, \beta)$ that maximizes equation (9), where $l_t(\delta)$ is

$$l_t(\delta) = \ln(\Gamma(\frac{\nu+1}{2})) - \ln(\Gamma(\frac{\nu}{2})) - \frac{1}{2}\ln(\pi(\nu-2)) - \ln\sigma_s - \frac{1}{2}\ln H_t - \frac{\nu+1}{2}\ln(1+\frac{\epsilon_t^2}{H_t\sigma_s^2(\nu-2)}). \tag{10}$$

Here ν, the number of degrees of freedom, is taken as given.

The score functions of the likelihood function contributions for observation t in equation (10) with respect to σ_s, $\frac{\alpha}{\omega}$, and β are

$$s_t^T(\sigma_s) = \frac{1}{\sigma_s}\left(\frac{\nu\epsilon_t^2 - H_t\sigma_s^2(\nu-2)}{\epsilon_t^2 + H_t\sigma_s^2(\nu-2)}\right), \tag{11}$$

$$s_t^T(\frac{\alpha}{\omega}) = \frac{1}{2}\frac{1}{H_t}\frac{\partial H_t}{\partial(\frac{\alpha}{\omega})}\left(\frac{\nu\epsilon_t^2 - H_t\sigma_s^2(\nu-2)}{\epsilon_t^2 + H_t\sigma_s^2(\nu-2)}\right), \tag{12}$$

$$s_t^T(\beta) = \frac{1}{2}\frac{1}{H_t}\frac{\partial H_t}{\partial\beta}\left(\frac{\nu\epsilon_t^2 - H_t\sigma_s^2(\nu-2)}{\epsilon_t^2 + H_t\sigma_s^2(\nu-2)}\right), \tag{13}$$

where

$$\frac{\partial H_t}{\partial(\frac{\alpha}{\omega})} = \epsilon_{t-1}^2 + \beta\frac{\partial H_{t-1}}{\partial(\frac{\alpha}{\omega})},$$
$$\frac{\partial H_t}{\partial\beta} = H_{t-1} + \beta\frac{\partial H_{t-1}}{\partial\beta}.$$

5. Improved Normal QMLE and Motivation

Although the Normal QMLE provides consistency and the asymptotic normality, its efficiency property is in doubt when the true distribution is far from normality. Engle and Gonzalez-Rivera (1991) show that the efficiency of NQMLE is low when the true density follows the student's t distribution with degree of freedom less than 12. One way to improve efficiency is to try to discover the true distribution of the innovations

u_t and use the MLE based on this true distribution. However, this procedure runs the risk of inconsistency if we get the "true" distribution wrong. So, instead, we may wish to improve upon the efficiency of the normal QMLE, but in such a way that we retain its robustness property.

The contribution of this thesis is to propose an augmented GMM estimator to improve the efficiency of the NQMLE. We interpret the normal QMLE as a GMM estimator, where the moment conditions it uses simply state that the score of the normal quasi-likelihood has expected value zero. Thus, the first-order conditions from the normal log-likelihood function are the base set of moment conditions with which we start. We then find other moment conditions to add on to the base set of moment conditions. In doing so we rely on the general result that adding extra valid moment conditions can never decrease the asymptotic efficiency of estimation. Of course, if the true density is Gaussian, the extra moments will be redundant. However, these extra moment conditions will generally improve efficiency if the true density is non-Gaussian.

There are two sets of extra moment conditions which we are particularly interested in. The first set of extra moments is derived from the relationship between the sample autocorrelations and the population autocorrelations of the squared observations. The use of such moments has been proposed by Baillie and Chung (1999). The second set of extra moment conditions is from the score of a non-Gaussian quasi log-likelihood function. Specifically, we consider the score function from the student's t quasi-likelihood function.

6. Plan of the Thesis

The thesis is organized as follows. Chapter 2 discusses in details the two extra sets of moment conditions that we will use. Combining all moment conditions, we show how to calculate the asymptotic variance of the augmented GMM estimators. To see the asymptotic efficiency gains from adding these extra moment conditions, we report asymptotic variances for specific parameter values and different assumptions on the true distribution. The results show that the augmented GMM estimators improve on the efficiency of the normal QMLE in a non-trivial way when the true density is far from normality.

Chapter 3 gives Monte Carlo simulation results for finite samples to investigate the finite sample properties of the normal QMLE and the augmented GMM estimators. We wish to see whether the asymptotic efficiency gains identified in Chapter 2 can actually be attained in reasonable sized samples. We find that rather large sample sizes, say T = 2,000, are required to do so. We also investigate the finite sample reliability of inferences based on asymptotic theory for the NQMLE and augmented GMM estimators.

Chapter 4 presents an empirical application to the high frequency data using the GARCH(1,1) model. The data we analyzed is the hourly exchange rate of the West German Deutschmark versus the U.S. dollar (DM/$) from 0:00 a.m. January 2, 1986 through 11:00 a.m. July 15, 1986. The same data has previously been analyzed by Baillie and Bollerslev (1990). Compared to the normal QMLE, the augmented GMM estimates appear to be considerably more precise, with conventionally calculated standard errors reduced by a factor of about two.

Chapter 5 gives our concluding remarks.

Appendix 1

A. MLE under Asymmetric distribution

Suppose that $\tilde{u}_t$ follows a χ^2 distribution with degrees of freedom ν:

$$f(\tilde{u}_t) = \frac{1}{\Gamma(\frac{\nu}{2})}\Big(\frac{1}{2}\Big)^{\frac{\nu}{2}}\big(\tilde{u}_t\big)^{(\frac{\nu}{2}-1)}exp\Big(-\frac{\tilde{u}_t}{2}\Big), \tilde{u}_t \geq 0,$$

This χ^2 variable has mean ν and variance 2ν. Now we standardize $\tilde{u}_t$ to obtain u_t with mean zero and variance one: $u_t = (\tilde{u}_t - \nu)/\sqrt{2\nu}$. After this standardization, the density of u_t is

$$f(u_t) = \frac{\nu^{1/2}}{\Gamma(\frac{\nu}{2})}\Big(\frac{1}{2}\Big)^{\frac{\nu-1}{2}}\big(\sqrt{2\nu}u_t + v\big)^{(\frac{\nu}{2}-1)}exp\Big(-\frac{\sqrt{2\nu}u_t + \nu}{2}\Big), u_t \geq -\frac{\nu}{\sqrt{2\nu}}.$$

Transforming to the conditional distribution of ϵ_t ($\epsilon_t = h_t^{1/2}u_t$),

$$f(\epsilon_t|\psi_{t-1}) = \frac{v^{1/2}}{\Gamma(\frac{\nu}{2})h_t^{\frac{1}{2}}}\Big(\frac{1}{2}\Big)^{\frac{\nu-1}{2}}\Big(\sqrt{2\nu}\frac{\epsilon_t}{h_t^{\frac{1}{2}}} + \nu\Big)^{(\frac{\nu}{2}-1)}exp\Big(-\frac{1}{2}(\sqrt{2\nu}\frac{\epsilon_t}{h_t^{\frac{1}{2}}} + \nu)\Big),$$

where $\epsilon_t \geq -\frac{\nu}{\sqrt{2\nu}}h_t^{\frac{1}{2}}$.

We note that the natural constraint $\tilde{u}_t \geq 0$ corresponds to $u_t \geq -\nu/\sqrt{2\nu}$ and $\epsilon_t \geq -h_t^{1/2}\nu/\sqrt{2\nu}$. Thus the range of ϵ_t depends on the parameter θ, because h_t depends on θ. This dependence of the range of ϵ_t on θ violates one of the standard regularity conditions for the consistency and asymptotic normality of the MLE[1],and suggests that the MLE will be inconsistent.

More specially, let $s_T(\epsilon, \theta) = \frac{\partial L_T(\theta)}{\partial \theta}$ be the score function. The fundamental condition for the consistency of the MLE is that $E[s_T(\epsilon, \theta_0)] = 0$, where θ_0 is the

[1]The conditions for asymptotic normality of MLE are stated as Theorem 5.2 of Wooldridge (1994).

true value of θ. This condition fails for the case that u_t is standardized χ^2. In Table A.1 we evaluate the expected value of the score function, with T =2,000 and r $\equiv$ number of replications = 500, for the χ_1^2 and χ_5^2 cases and two different parameter values. In each case the mean appears to be nonzero. More formally, the hypothesis of zero mean is rejected by the usual asymptotic t test, where the standard error is calculated with the usual Newey-West formula with m = number of lags = 50.

B. Regularity Conditions – Bollerslev and Wooldridge

1. Θ is compact and has a nonempty interior.

2. The conditional mean and the conditional variance are measurable for all θ and twice continuously differentiable on Θ.

3. (a)$\{l_t(\epsilon_t, \theta), t = 1, 2, ...\}$ satisfies the uniform weak law of large numbers. (b) θ_0 is the identifiably unique maximizer of $E[\sum_{t=1}^{T} l_t(\theta)]$.

4. (a) $\partial l_t^2/\partial\theta'\partial\theta$ and $E[\partial l_t^2/\partial\theta'\partial\theta]$ satisfy the uniform weak law of large numbers. (b) $A = T^{-1}\sum_{t=1}^{T} E[a_t(\theta_0)]$ is uniform positive definite.

5. $s_t(\theta)s_t(\theta)'$ satisfies the uniform weak law of large numbers.

6. (a) $B = T^{-1}\sum_{t=1}^{T} E[s_t(\theta_0)s_t(\theta_0)']$ is uniformly positive definite.
   (b) $B^{-1/2}T^{-1/2}\sum_{t=1}^{T} s_t(\theta) \to N(0, I)$.

C. Regularity Conditions – Lumsdaine

1. The true parameter vector θ_0 is in the interior of Θ, a compact, convex parameter space. For any vector $(\omega, \alpha, \beta) \in \Theta$, assume that $m \leq \omega \leq M$,

$m \leq \alpha \leq (1-m)$, and $m \leq \beta \leq (1-m)$ for some constant $m > 0$, and $\alpha_0 + \beta_0 \leq 0$.

2. $u_t$is iid drawn from a symmetric, unimodal density, bounded in a neighborhood of 0, with mean 0, variance 1, and $E(u_t^{32}) < \infty$. In addition, assume that h_{0t} is independent of $\{u_t, u_{t+1}, \cdots\}$.

D. Identification conditions for consistency – Newey and Steigerwald

1. $E[|l_t(\delta)|] < \infty$ for all $\theta \in \Phi$, $\nu \in N$ and $\sigma_s > 0$, where Φ and N are feasible sets for θ and ν respectively.

2. The function $H_t^{1/2}(\theta_0) > 0$, and if $\theta \neq \theta_0$ then *either* $H_t^{1/2}(\theta)/H_t^{1/2}(\theta_0)$ is not constant *or* $P[f_t(\theta) \neq f_t(\theta_0)] > 0$.

3. The function $Q(\sigma_s, \nu) = -\ln\, \sigma_s + E[\ln\, a(\sigma_{s0}u_t/\sigma_s, \nu)]$ has a unique maximum at some σ_s and ν over $\sigma_s > 0$ and $\nu \in N$.

4. The innovation u_t is symmetrically distributed around zero with unimodal density $k(u)$ satisfying $k(u_1) \leq k(u_2)$ for $|u_1| \geq |u_2|$. For each ν, $a(u, \nu)$ is symmetric around zero and $a(u_1, \nu) < a(u_2, \nu)$ for $|u_1| > |u_2|$.

5. The function $H_t^{1/2}(\theta_0) > 0$, and if $\theta \neq \theta_0$ then either $H_t^{1/2}(\theta)/H_t^{1/2}(\theta_0)$ or $[f_t(\theta) - f_t(\theta_0)]/H_t^{1/2}(\theta_0)$ is not constant.

6. The function $Q(\gamma, \sigma_s, \nu) = -\ln\, \sigma_s + E[\ln\, a((\sigma_{s0}u_t + \gamma)/\sigma_s, \nu)]$ has a unique maximum in $(\gamma, \sigma_s, \nu)'$.

Table A1

Test for the expected value of the score function of the standardized chi-square distribution

T=2,000, r = 500

$\omega = 0.2$, $\alpha = 0.15$, $\beta = 0.65$

True distribution: Standardized chi-square with degree of freedom 1

	mean	s.e.(m = 50)	t-value
ω	-0.54	0.18	-2.93
α	-0.30	0.10	-3.06
β	-0.46	0.15	-3.04

True distribution: Standardized chi-square with degree of freedom 5

	mean	s.e.(m = 50)	t-value
ω	-0.35	0.11	-3.27
α	-0.25	0.09	-2.88
β	-0.32	0.11	-2.78

T=2,000, r = 500

$\omega = 0.1$, $\alpha = 0.1$, $\beta = 0.8$

True distribution: Standardized chi-square with degree of freedom 1

	mean	s.e.(m = 50)	t-value
ω	-0.96	0.33	-2.87
α	-0.56	0.18	-3.06
β	-0.81	0.27	-3.02

True distribution: Standardized chi-square with degree of freedom 5

	mean	s.e.(m = 50)	t-value
ω	-0.60	0.23	-2.59
α	-0.46	0.16	-2.90
β	-0.55	0.20	-2.76

Chapter 2

EXTRA MOMENT CONDITIONS AND ASYMPTOTIC ANALYSIS

1. Introduction

The normal QMLE is widely used in estimating ARCH type models. As we mentioned in chapter 1, the normal QMLE is consistent and asymptotically normal, but the loss in efficiency may be large when the true distribution is far from normality. The normal QMLE is a GMM estimator, based on the fact that the expectation of the score function of the normal log likelihood equals zero, whether the true distribution of the innovations is normal or not. We want to improve the normal QMLE by adding other moment conditions that also do not require normality to be valid.

In this chapter we give details of two extra sets of moment conditions; one is based on the autocorrelations of ϵ_t^2 and the other is based on the rescaled student's t distribution. Then we show how to calculate the asymptotic variance of the augmented GMM estimator. The asymptotic variances for specific parameter values and different true distributions are reported and compared to see how much efficiency we can gain from adding these extra moment conditions.

2. Moment Conditions Based on Autocorrelations of ϵ_t^2

Baillie and Chung (1999) apply a Minimum Distance Estimator (MDE) to the GARCH model. This estimator minimizes the distance between the population and sample autocorrelations of the squared observation (ϵ_t^2). They perform Monte Carlo simulations and find that for non-normal innovations, especially in the asymmetric case, the MDE can be more efficient than the normal QMLE.

The MDE is defined by solving the following minimization problem:

$$\text{Min} \ \ [\hat{\rho} - \rho(\theta)]' W [\hat{\rho} - \rho(\theta)], \tag{1}$$

where W is a positive definite weighting matrix, and $\hat{\rho}$ and $\rho(\theta)$ are $(g \times 1)$ vectors, which contain g sample autocorrelations and g population autocorrelations respectively. We briefly explain how to calculate the three components in the objective function, $\hat{\rho}$, $\rho(\theta)$, and W. First, the sample autocorrelation $\hat{\rho}_k, k = 1, \cdots, g$, is defined by

$$\hat{\rho}_k = \sum_{t=k+1}^{T} (\epsilon_t^2 - \bar{\epsilon^2})(\epsilon_{t-k}^2 - \bar{\epsilon^2}) / \sum_{t=k+1}^{T} (\epsilon_t^2 - \bar{\epsilon^2})^2,$$

where $\bar{\epsilon^2}$ is the sample mean of the ϵ_t^2. Second, the population autocorrelation function of ϵ_t^2, $\rho(\theta)$, depends on the model used in describing the series. We will consider the case of the GARCH(1,1) model. For this case an explicit formula for the autocorrelations is available. Specifically, the autocorrelation function of ϵ_t^2 for the GARCH(1,1) process has been derived by Bollerslev (1988) and Ding and Granger (1996). If $\alpha + \beta < 1$ and the fourth moment of ϵ_t exists[2], the autocorrelation function for the GARCH(1,1) process is

$$\rho_1 = \left(\alpha + \frac{\alpha^2 \beta}{1 - 2\alpha\beta - \beta^2}\right),$$

and

[2]That is, $\kappa\alpha^2 + 2\alpha\beta + \beta^2 < 1$, where κ is 3 plus the coefficient of excess. In the normal case, κ is equal to 3.

$$\rho_k = \left(\alpha + \frac{\alpha^2\beta}{1-2\alpha\beta-\beta^2}\right)(\alpha+\beta)^{k-1}, \quad \text{for} \quad k \geq 2.$$

Finally, the optimal weighting matrix W is the inverse of the asymptotic covariance matrix of $\hat{\rho}$, a result given by Chiang (1965) and Ferguson (1958).

Because we want to use standard GMM results to evaluate potential efficiency gains, we want to consider a GMM estimator rather than a MDE estimator. To do so, we note that

$$T^{1/2}(\hat{\rho} - \rho(\theta)) = T^{-1/2} \sum_{t=g+1}^{T} Z_t/\hat{\gamma}_0,$$

where

$$\hat{\gamma}_0 = T^{-1} \sum_{t=k+1}^{T} (\epsilon_t^2 - \bar{\epsilon^2})^2,$$

$$Z_t = \begin{pmatrix} (\epsilon_t^2 - \bar{\epsilon^2})(\epsilon_{t-1}^2 - \bar{\epsilon^2}) - \rho_1(\theta)(\epsilon_t^2 - \bar{\epsilon^2})^2 \\ \vdots \\ (\epsilon_t^2 - \bar{\epsilon^2})(\epsilon_{t-g}^2 - \bar{\epsilon^2}) - \rho_g(\theta)(\epsilon_t^2 - \bar{\epsilon^2})^2 \end{pmatrix}. \tag{2}$$

Thus the MDE should be asymptotically equivalent to a GMM estimator based on the moment conditions $E(Z_t) = 0$.

The augmented GMM estimator combines these moment conditions with those based on the score of the normal likelihood. Explicitly, it is based on the moment conditions: $E[q_t(\theta_0)] = 0$, where

$$q_t(\theta) = [q_{1t}(\theta)', q_{2t}(\theta)']', \tag{3}$$

where $q_{1t}(\theta)$ is given in equation (6) of chapter 1, and where $q_{2t}(\theta) = Z_t$ as in equation (2) above.

A standard result of GMM estimation is that adding more valid moment conditions cannot decrease the asymptotic efficiency of estimation. Thus the augmented GMM estimator is asymptotically at least as efficient as the normal quasi-MLE. If the data are normal, the moment conditions based on q_{2t} must be redundant, whereas otherwise the augmented GMM estimator would be expected to be strictly more efficient than the normal quasi-MLE.

A relevant detail of estimation is that, whereas q_{1t} is uncorrelated over t, q_{2t} is correlated over different t. The estimation (or evaluation) of the weighting matrix needs to recognize this fact. Thus we would use a Newey-West type estimator:

$$\hat{\Omega} = \hat{\Gamma}_0 + \sum_{j=1}^{m}(1 - \frac{j}{1+m})(\hat{\Gamma}_j + \hat{\Gamma}_j'), \tag{4}$$

where $\hat{\Gamma}_j = \frac{1}{T}\sum_t q_t(\hat{\theta})q_{t-j}(\hat{\theta})'$, where $\hat{\theta}$ is an initial consistent estimator of θ , and where m grows with T at a suitable rate. Then $W = \hat{\Omega}^{-1}$.

3. Moment Conditions Based on the Score Function from the Rescaled Student's t Distribution

In section 4.B of Chapter 1, we discussed the fact, proved by Newey and Steigerwald (1997), that the QMLE based on the rescaled student's t distribution yields a consistent estimator for the parameters $(\sigma_s, \alpha/\omega, \beta)$ of the GARCH model. Since we are interested in the parameters (α, β, ω) in the natural GARCH parameterization, we simply will use the two moment conditions that correspond to the scores with respect to α/ω and β. That is, we consider

$$q_{3t}(\theta) = [s_t^T(\alpha/\omega), s_t^T(\beta)]', \tag{5}$$

where $s_t^T(\alpha/\omega)$ and $s_t^T(\beta)$ are given in equations (12) and (13) of chapter 1. Then our set of extra moment conditions is :$E[q_{3t}(\theta_0)] = 0$. We can then use these conditions to augment the moment conditions based on the score of the normal likelihood function.

4. Asymptotic Variance for the Augmented-GMM Estimator

Putting together the moment conditions based on the score functions from NQMLE, the autocorrelations of ϵ_t^2, and the score functions from the rescaled t distribution, there are $g + 5$ moment conditions, where g is the number of autocorrelations of ϵ_t^2 that are considered. Explicitly, we have

$$E\ [q_t(\theta_0)] \equiv E\ [\begin{pmatrix} q_{1t}(\theta_0) \\ q_{2t}(\theta_0) \\ q_{3t}(\theta_0) \end{pmatrix}] = 0. \tag{6}$$

Here q_{1t}, q_{2t} and q_{3t}, are as defined in the preceding two sections.

The augmented-GMM estimator, $\hat{\theta}$, minimizes

$$\bar{q}_T(\theta)'W\bar{q}_T(\theta),$$

where $\bar{q}_T(\theta) = T^{-1}\sum_t q_t(\theta)$ and W is the weighting matrix. Under some standard regularity conditions, $\hat{\theta}$ is consistent and has the following asymptotic distribution:

$$T^{\frac{1}{2}}(\hat{\theta} - \theta) \rightarrow N[0, (D'WD)^{-1}D'W\Omega WD(D'WD)^{-1}],$$

where $D = E[\partial q_t(\epsilon, \theta)/\partial\theta']$ and $\sqrt{T}\bar{q}_T(\theta_0) \rightarrow N(0, \Omega)$. Since q_{2t} is correlated over time, Ω not only includes the expected square of $q_t(\theta)$, but also involves the cross products over time of $q_t(\theta)$. Thus

$$\Omega = \lim_{T\rightarrow\infty} Var(\sqrt{T}\bar{q}_T) = \Gamma_0 + \sum_{l=1}^{\infty}(\Gamma_l + \Gamma_l'),$$

where $\Gamma_0 = E[q_t(\theta)q_t(\theta)']$, and $\Gamma_l = E[q_t(\theta)q_{t-l}(\theta)']$.

The optimal GMM estimator is obtained when W is a consistent estimator of Ω^{-1}, as shown in Hansen (1982). In this, the asymptotic variance simplifies to $(D'\Omega^{-1}D)^{-1}$.

In practice, given an initial consistent estimator $\hat{\theta}$, D can be consistently estimated by

$$\hat{D} = T^{-1}\sum_{t=1}^{T} \partial q_t(\epsilon,\theta)/\partial\theta'_{|\theta=\hat{\theta}} \ .$$

Ω can be consistently estimated by the method of Newey and West (1987), as in equation (4) above.

5. Evaluation and Comparison of Asymptotic Variances

In order to assess the relative asymptotic efficiency gain of the augmented-GMM estimator over the normal QMLE, we simply wish to calculate and compare the asymptotic variances of the estimators. These depend on expectations that we cannot take analytically. Thus, we will rely on simulation to give us numerical results. That is, the asymptotic variance is of the form $(D'\Omega^{-1}D)^{-1}$, and we use simulation to evaluate D and Ω. We stress that this is still an evaluation of the *asymptotic* variance of the estimates. We do not calculate any estimates of θ; we simply use simulation to replace the expectations in the definitions of D and Ω (evaluated at the "true value" θ_0).

The model is a GARCH(1,1) process,

$$\epsilon_t = h_t^{\frac{1}{2}} u_t,$$

where u_t is drawn from a distribution with mean 0 and variance 1. Thus

$$h_t = w + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}, \quad t = 1, \ldots, T,$$

where $\omega > 0, \alpha \geq 0, \beta \geq 0$. For sample size equal to T and number of replications equal to R, the asymptotic variance is evaluated as $(\tilde{D}'\tilde{\Omega}^{-1}\tilde{D})^{-1}$, where

$$\tilde{D} = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial q_t^{(r)}(\epsilon,\theta)}{\partial\theta'},$$

and

$$\tilde{\Omega} = \frac{1}{R}\sum_{r=1}^{R}[\tilde{\Gamma}_0 + \sum_{l=1}^{m}(1 - \frac{l}{1+m})(\tilde{\Gamma}_l + \tilde{\Gamma}_l')], \quad \text{with}$$

$$\tilde{\Gamma}_0 = \sum_{t=1}^{T} q_t^{(r)}(\epsilon,\theta)q_t^{(r)}(\epsilon,\theta)',$$

$$\tilde{\Gamma}_l = \sum_{t=l+1}^{T} q_t^{(r)}(\epsilon,\theta)q_{t-l}^{(r)}(\epsilon,\theta)',$$

where $r = 1, 2, \cdots, R$, and $t = 1, 2, \cdots, T$. All evaluations are for $\theta = \theta_0$.

For large R and T, and for m picked suitably, $(\tilde{D}'\tilde{\Omega}^{-1}\tilde{D})^{-1}$ should be close to the asymptotic variance $(D'\Omega^{-1}D)^{-1}$. We calculate our results with $R = 500$ and $T = 2,000$, so that we are averaging over 1,000,000 different observations to evaluate D and Ω. After some experimentation, we picked m $= 50$ as the number of lags to use in evaluating Ω.

The parameter values we choose satisfy the following constraints. For stationarity of the process, $\alpha + \beta < 1$ must hold. In addition, the fourth-order moment of ϵ_t must exist in order for asymptotic theory to hold for the moment conditions based on the autocorrelations of ϵ_t^2. In Table 2.1, we tabulate the value of $\kappa\alpha^2 + 2\alpha\beta + \beta^2$, which must be less than one for the fourth moment to exist, for various values of κ, α and β.

Since different distributions have different κ, Table 2.1 shows all the cases which we will analyze in this paper. The variance of the unconditional density, $\omega/(1-\alpha-\beta)$, is chosen to equal to 10. Also, to minimize the effect of startup problems, we generate a sample of size equal to $T+200$, and discard the first 200 data points; the remaining T data points are then used in simulation.

We first report the results with $\alpha = 0.15$ and $\beta = 0.7$, in Table 2.2. We compare results from five different GMM estimators based on different sets of moment conditions. The **NQMLE** uses the moment conditions q_{1t}, the score from the normal log-likelihood function. The **MDE** is based on the moment conditions q_{2t}, which equate sample and population autocorrelations of ϵ_t^2. The estimator using moment conditions q_{1t} and q_{2t} is called **GMM1**. **GMM2** refers to the estimator that uses the moment conditions q_{1t} and q_{3t}, where q_{3t} is based on the score for the rescaled student's t distribution. Finally, **GMM3** uses all $g+5$ moment conditions: q_{1t}, q_{2t}, and q_{3t}.

When the true distribution is normal, the asymptotic standard errors are reported in column 1 of Table 2.2. The NQMLE is the MLE in this case and is therefore efficient. Adding more moment conditions should not improve efficiency. The MDE should be inefficient relative to NQMLE, and it is. All of the other GMM estimators (GMM2 and all variants of GMM1 and GMM3) should have asymptotic standard errors equal to that of the NQMLE, and this is a check on the accuracy of the calculations. There are only very minor differences, usually no more than 3%, and this indicates that our calculations are reasonably accurate.

Column 2 of Table 2.2 gives the asymptotic standard errors when the true distribution is a standardized student's t distribution with the degree of freedom 5. We consider the same estimators as before, but we also consider the MLE based on the t_5 distribution (TMLE), which should be efficient. The asymptotic standard errors for NQMLE are larger than those for TMLE by about 50%. The MDE still has larger asymptotic standard errors than those of NQMLE, even for g as large as 20. For GMM1 with g = 5, the asymptotic standard error of α is smaller than for NQMLE by about 20%, but there are only very small differences for ω and β. Increasing g reduces the asymptotic standard errors in all three parameters, with larger decreases for ω and β than for α. With g = 20, the GMM1 asymptotic standard errors are about 20% smaller for α and about 10% smaller for ω and β, compared to NQMLE. The efficiency gains from GMM2 are of comparable magnitude. It is perhaps surprising that GMM2 is not more efficient, since it uses the score function from the rescaled student's t distribution; but it uses only two moment conditions based on student's t, not three. Finally, GMM3 results in further efficiency gains. For estimation of α it essentially reaches the TMLE lower bound, whereas for ω and β its standard errors are roughly midway between those for NQMLE and TMLE.

The results for the χ^2 distribution with two degrees of freedom are given in column 3 of Table 2.2. The MDE does better here than previously. For MDE with g = 10, the asymptotic standard error of α is smaller than that of NQMLE, and when g increases to 20, the asymptotic standard error of β is close to that of NQMLE. For GMM1, the asymptotic standard error of α is almost 30% less than that of NQMLE. The improvements for ω and β are still minor. For GMM2, the efficiency improvement for α is comparable to those for GMM1, while for β and ω the results are similar to those for GMM1 with g = 5 but not as good as for GMM1 with g larger. GMM3 is not much better than GMM1.

In column 4 of Table 2.2, the true distribution is the standardized gamma distribution with two degrees of freedom . These numbers show roughly the same pattern as the standardized t distribution with the degree of freedom 5.

We also analyze three different sets of values for α and β. Tables 2.3, 2.4 and 2.5 correspond to $\{ \alpha = 0.1, \beta = 0.75 \}$, $\{ \alpha = 0.1, \beta = 0.8 \}$ and $\{ \alpha = 0.2, \beta = 0.6 \}$ respectively. We will not discuss these results in detail because they are fairly similar to those in Table 2.2. The augmented GMM estimators improve on the NQMLE, in terms of asymptotic standard error, by an amount that varies over parameters and distributions, but is perhaps typically in the range of 10% $\sim$ 20%.

6. Conclusions

When the true distribution is Gaussian, using the first set of moment conditions q_{1t} based on the normal score function is enough to obtain the Cramer-Rao lower bound. The extra two sets of moment conditions are redundant. The augmented GMM has no gain in efficiency when the data series actually follows the normal distribution.

When the true distribution is non-Gaussian, such as student's t distribution, χ^2 distribution, or gamma distribution, we gain asymptotic efficiency by using extra moment conditions. GMM1 uses g moment conditions based on the autocorrelations of the ϵ_t^2, while GMM2 uses the score for the rescaled student's t distribution. GMM3 uses all of these and must be most efficient. The efficiency gains that are achieved typically amount to reduction of the asymptotic standard error by $10\% \sim 20\%$.

This benefit does not come without some costs. The augmented GMM estimators are computationally more complicated than the NQMLE. Furthermore, GMM1 and GMM3 especially are potentially heavily overidentified and there may be worries about their finite-sample properties. We will present Monte Carlo evidence on this point in the next chapter.

Table 2.1

$\kappa\alpha^2+2\alpha\beta+\beta^2$

(κ = "degree of excess" plus 3. Fourth moment exists if $\kappa\alpha^2+2\alpha\beta+\beta^2 < 1$).

$\kappa = 3$

standard normal distribution

		α					
		0.1	0.15	0.2	0.25	0.3	0.35
	0.4	0.2700	0.3475	0.4400	0.5475	0.6700	0.8075
	0.45	0.3225	0.4050	0.5025	0.6150	0.7425	0.8850
	0.5	0.3800	0.4675	0.5700	0.6875	0.8200	0.9675
	0.55	0.4425	0.5350	0.6425	0.7650	0.9025	1.0550
β	0.6	0.5100	0.6075	0.7200	0.8475	0.9900	1.1475
	0.65	0.5825	0.6850	0.8025	0.9350	1.0825	1.2450
	0.7	0.6600	0.7675	0.8900	1.0275	1.1800	1.3475
	0.75	0.7425	0.8550	0.9825	1.1250	1.2825	1.4550
	0.8	0.8300	0.9475	1.0800	1.2275	1.3900	1.5675

$\kappa = 6$

standardized gamma distribution with degree of freedom 2

		α					
		0.1	0.15	0.2	0.25	0.3	0.35
	0.4	0.3000	0.4150	0.5600	0.7350	0.9400	1.1750
	0.45	0.3525	0.4725	0.6225	0.8025	1.0125	1.2525
	0.5	0.4100	0.5350	0.6900	0.8750	1.0900	1.3350
	0.55	0.4725	0.6025	0.7625	0.9525	1.1725	1.4225
β	0.6	0.5400	0.6750	0.8400	1.0350	1.2600	1.5150
	0.65	0.6125	0.7525	0.9225	1.1225	1.3525	1.6125
	0.7	0.6900	0.8350	1.0100	1.2150	1.4500	1.7150
	0.75	0.7725	0.9225	1.1025	1.3125	1.5525	1.8225
	0.8	0.8600	1.0150	1.2000	1.4150	1.6600	1.9350

$\kappa = 9$

standardized t with degree of freedom 5

standardized chisquare with degree of freedom 2

		α					
		0.1	0.15	0.2	0.25	0.3	0.35
	0.4	0.3300	0.4825	0.6800	0.9225	1.2100	1.5425
	0.45	0.3825	0.5400	0.7425	0.9900	1.2825	1.6200
	0.5	0.4400	0.6025	0.8100	1.0625	1.3600	1.7025
	0.55	0.5025	0.6700	0.8825	1.1400	1.4425	1.7900
β	0.6	0.5700	0.7425	0.9600	1.2225	1.5300	1.8825
	0.65	0.6425	0.8200	1.0425	1.3100	1.6225	1.9800
	0.7	0.7200	0.9025	1.1300	1.4025	1.7200	2.0825
	0.75	0.8025	0.9900	1.2225	1.5000	1.8225	2.1900
	0.8	0.8900	1.0825	1.3200	1.6025	1.9300	2.3025

Note:

standard normal distribution	$\kappa = 3$
standardized t with degree of freedom v1,	$\kappa = (6/(\upsilon 1-4)) + 3$
standardized chisquare with degree of freedom v2,	$\kappa = (12/\upsilon 2) + 3$
standardized gamma distribution with degree of freedom v3,	$\kappa = (6/\upsilon 3) + 3$

Table 2.2

Asymptotic Standard Error

$\omega = 1.5, \alpha = 0.15, \beta = 0.7$

True distribution	standard normal distribution			standardized t(5)			standardized chisquare(2)			standardized gamma(2)		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β
NQMLE	0.3571	0.0249	0.0510	0.5356	0.0428	0.0762	0.5233	0.0479	0.0816	0.4642	0.0389	0.0703
MDE												
g = 5		0.0386	0.1022		0.0623	0.1636		0.0579	0.1615		0.0559	0.1548
g = 10		0.0345	0.0720		0.0479	0.0998		0.0456	0.0986		0.0435	0.0934
g = 20		0.0339	0.0681		0.0452	0.0854		0.0431	0.0838		0.0409	0.0813
GMM1												
g = 5	0.3553	0.0245	0.0507	0.5260	0.0337	0.0749	0.5011	0.0349	0.0765	0.4584	0.0313	0.0686
g =10	0.3533	0.0242	0.0509	0.4938	0.0335	0.0713	0.4763	0.0347	0.0734	0.4391	0.0312	0.0666
g = 20	0.3508	0.0242	0.0507	0.4649	0.0333	0.0680	0.4527	0.0345	0.0705	0.4227	0.0311	0.0646
GMM2	0.3539	0.0250	0.0507	0.5112	0.0338	0.0738	0.5134	0.0343	0.0764	0.4615	0.0312	0.0678
GMM3												
g = 5	0.3529	0.0244	0.0506	0.5018	0.0308	0.0727	0.4989	0.0313	0.0744	0.4561	0.0287	0.0671
g = 10	0.3508	0.0242	0.0505	0.4720	0.0306	0.0691	0.4713	0.0310	0.0707	0.4349	0.0287	0.0647
g = 20	0.3483	0.0241	0.0503	0.4437	0.0304	0.0656	0.4440	0.0308	0.0673	0.4163	0.0286	0.0624
TMLE				0.3633	0.0300	0.0547						

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 2.3

Asymptotic standard error

$\omega = 1.5, \alpha = 0.1, \beta = 0.75$

True distribution	standard normal distribution			standardized t(5)			standardized chisquare(2)			standardized gamma(2)		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β
NQMLE	0.4686	0.0219	0.0606	0.6536	0.0356	0.0849	0.6449	0.0383	0.0885	0.5804	0.0321	0.0781
MDE												
g = 5		0.0308	0.1172		0.0514	0.1862		0.0477	0.1880		0.0446	0.1734
g = 10		0.0272	0.0757		0.0393	0.1080		0.0378	0.1060		0.0357	0.1009
g = 20		0.0266	0.0693		0.0369	0.0889		0.0361	0.0897		0.0336	0.0847
GMM1												
g = 5	0.4671	0.0218	0.0605	0.6490	0.0298	0.0838	0.6219	0.0315	0.0841	0.5750	0.0282	0.0767
g =10	0.4657	0.0217	0.0604	0.6196	0.0296	0.0807	0.6010	0.0314	0.0818	0.5608	0.0281	0.0753
g = 20	0.4628	0.0217	0.0602	0.5804	0.0295	0.0767	0.5799	0.0313	0.0795	0.5424	0.0280	0.0734
GMM2	0.4640	0.0219	0.0601	0.6443	0.0290	0.0840	0.6289	0.0290	0.0839	0.5761	0.0268	0.0763
GMM3												
g = 5	0.4634	0.0218	0.0600	0.6351	0.0275	0.0830	0.6165	0.0274	0.0821	0.5722	0.0252	0.0756
g = 10	0.4620	0.0217	0.0599	0.6075	0.0273	0.0795	0.5953	0.0272	0.0794	0.5571	0.0255	0.0740
g = 20	0.4589	0.0216	0.0597	0.5647	0.0271	0.0748	0.5719	0.0271	0.0767	0.5361	0.0254	0.0716
TMLE				0.4557	0.0249	0.0611						

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 2.4

Asymptotic standard error

$\omega = 1.0, \alpha = 0.1, \beta = 0.8$

True distribution	standard normal distribution			standardized t(5)			standardized chisquare(2)			standardized gamma(2)		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β
NQMLE	0.2971	0.0197	0.0425	0.4060	0.0335	0.0616	0.4117	0.0345	0.0638	0.3664	0.0288	0.0554
MDE												
g = 5		0.0347	0.1117		0.0562	0.1780		0.0554	0.1824		0.0514	0.1695
g = 10		0.0274	0.0646		0.0370	0.0914		0.0372	0.0931		0.0349	0.0867
g = 20		0.0253	0.0534		0.0322	0.0659		0.0325	0.0678		0.0302	0.0646
GMM1												
g = 5	0.2956	0.0195	0.0425	0.4030	0.0263	0.0596	0.4037	0.0274	0.0615	0.3677	0.0245	0.0549
g =10	0.2948	0.0193	0.0425	0.3917	0.0261	0.0582	0.3920	0.0273	0.0603	0.3589	0.0244	0.0541
g = 20	0.2914	0.0192	0.0422	0.3620	0.0259	0.0546	0.3697	0.0272	0.0576	0.3444	0.0243	0.0525
GMM2	0.2941	0.0197	0.0422	0.4005	0.0260	0.0585	0.4001	0.0261	0.0594	0.3663	0.0242	0.0541
GMM3												
g = 5	0.2933	0.0194	0.0422	0.3967	0.0239	0.0579	0.3946	0.0241	0.0585	0.3645	0.0225	0.0537
g = 10	0.2924	0.0192	0.0422	0.3846	0.0238	0.0565	0.3830	0.0240	0.0571	0.3550	0.0225	0.0527
g = 20	0.2889	0.0191	0.0419	0.3529	0.0235	0.0526	0.3589	0.0239	0.0542	0.3384	0.0224	0.0508
TMLE				0.2891	0.0225	0.0435						

GMM1–the score from Normal distribution & MDE

GMM2– the scores from Normal distribution & rescaled t distribution

GMM3– the scores from Normal distribution & rescaled t distribution and MDE

Table 2.5

Asymptotic standard error

$\omega = 2.0, \alpha = 0.2, \beta = 0.6$

True distribution	standard normal distribution			standardized t(5)			standardized chisquare(2)			standardized gamma(2)		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β
NQMLE	0.4004	0.0294	0.0568	0.6419	0.0521	0.0897	0.6165	0.0604	0.0959	0.5429	0.0480	0.0819
MDE												
g = 5		0.0453	0.1040		0.0702	0.1580		0.0644	0.1546		0.0623	0.1503
g = 10		0.0428	0.0842		0.0601	0.1123		0.0554	0.1081		0.0532	0.1042
g = 20		0.0427	0.0829		0.0587	0.1049		0.0542	0.0998		0.0520	0.0981
GMM1												
g = 5	0.3994	0.0286	0.0567	0.6170	0.0408	0.0875	0.5833	0.0412	0.0884	0.5297	0.0370	0.0792
g =10	0.3989	0.0285	0.0567	0.5696	0.0404	0.0818	0.5472	0.0407	0.0837	0.5019	0.0369	0.0762
g = 20	0.3983	0.0285	0.0566	0.5513	0.0403	0.0796	0.5275	0.0406	0.0813	0.4889	0.0368	0.0745
GMM2	0.3967	0.0294	0.0563	0.5918	0.0407	0.0848	0.5120	0.0397	0.0767	0.5097	0.0367	0.0745
GMM3												
g = 5	0.3949	0.0284	0.0562	0.5718	0.0365	0.0824	0.4974	0.0340	0.0741	0.5014	0.0329	0.0735
g = 10	0.3920	0.0282	0.0561	0.5310	0.0361	0.0773	0.4759	0.0388	0.0714	0.4767	0.0329	0.0707
g = 20	0.3905	0.0281	0.0560	0.5145	0.0360	0.0752	0.4614	0.0338	0.0696	0.4645	0.0329	0.0692
TMLE				0.4210	0.0367	0.0630						

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Chapter 3
FINITE-SAMPLE PROPERTIES

1. Introduction

In Chapter 2 we found that, for cases in which the true distribution is far away from the gaussian, augmented GMM gives a non-trivial gain in asymptotic efficiency. The gains were on the order of $10\% \sim 20\%$ for α and smaller percentages for ω and β in terms of asymptotic standard error. In this section, we try to investigate whether these efficiency gains can be realized in finite samples.

An issue of concern in finite samples is the problem of weak instruments. Some studies find that when the instruments are weak, i.e. the partial correlation between the instruments and the included endogenous variable is low, the conventional asymptotic results fail even if the sample size is quite large. In applications of two-stage least squares (2SLS), Nelson and Startz (1990) and Bound, Jaeger and Baker (1995) find that the 2SLS estimator is biased away from the true value and in the direction of the ordinary least squares (OLS) estimator. Also, the asymptotic distribution is a poor approximation to the true distribution in finite samples when the instruments are weak, so that inference based on asymptotics is inaccurate.

Since some of the augmented GMM estimators employ many moment conditions, we have a high degree of overidentification. A highly overidentified problem is particularly vulnerable to the problem of weak instruments. In such cases, the estimates may be biased and the asymptotic standard errors may be overly optimistic in finite samples. Correspondingly the efficiency gain may not be as large as it is asymptotically, and in fact may not even exist. Moreover, inference may not be accurate, e.g. t tests may overreject the null hypothesis. Monte Carlo experiments will be carried out to check the quality of the asymptotic approximation to the finite sample distributions, and to see whether augmented GMM is really better than the normal QMLE in finite samples.

2. Design of the Experiment

We generate data from the GARCH(1,1) model in the same way as in section 2.5. An experiment is defined by a choice of the sample size (T), the parameter value (ω, α, β), and the true distribution (e.g. normal, t_5). For each experiment, we estimate the model by NQMLE, MDE, GMM1, GMM2 and GMM3. There are three sets of results for MDE, GMM1, and GMM3 corresponding to g = 5, 10, 20. In addition, the non-Gaussian MLE based on the true distribution will be reported when the MLE is available.

For the estimators other than NQMLE, we use the estimated weighting matrix proposed by Newey and West (1987), with the lag length determined by an automatic-lag selection criterion suggested by Newey and West (1994)[3]. We use NQMLE as the initial consistent estimator for estimation of the weighting matrix.

For each experiment and estimator, we obtain the following statistics; mean, standard error, root mean square error (RMSE) and the average estimated asymptotic

[3] $m = \hat{v}T^{\frac{1}{3}}$, where $\hat{v} = 1.1447\{\hat{s}^{(1)}/\hat{s}^{(0)}\}^{2/3}, \hat{s}^{(1)} = 2\sum_{j=1}^{n} j\hat{\Omega}_j, \hat{s}^{(0)} = \hat{\Omega}_0 + 2\sum_{j=1}^{n}\hat{\Omega}_j$, and $n = [4(T/100)^{2/9}]$. $\hat{\Omega}_j$ is defined as equation (4) in chapter 2.

standard error (using the usual GMM formula)[4]. We want to check (i) whether the bias is significant, (ii) whether the standard error and RMSE are lower for the augmented GMM estimators than for the NQMLE, and (iii) how large is the difference between the finite sample standard error and the average estimated asymptotic standard error.

The Monte Carlo simulations are performed using Gauss (Windows version 3.2.32) on a Pentium II 350 PC. We use the CML (Constrained Maximum Likelihood) and CO (Constrained Optimization) modules to do the optimization. The NEWTON (Newton-Raphson) method is the main optimization algorithm, but in a few cases we also used the BFGS (Broyden, Fletcher, Goldfarb, Shanno) algorithm because of convergence problems using NEWTON. We use 500 replications ($R = 500$).

We organize our experiments around a "base case" with $T = 2,000$, $\alpha = 0.15$, $\beta = 0.7$, unconditional variance = 10, and data generated from the t_5 distribution. We then perform four sets of experiments. First, holding all other features of the DGP constant, we consider different distributions: standard normal, student's t distribution with five degrees of freedom, standardized chi-square with two degrees of freedom, and standardized gamma with two degrees of freedom[5]. In the optimization process, we found that some replications did not satisfy the constraint that α must be greater than or equal to zero. Such results would cause the asymptotic standard error to be incorrectly calculated. This happens only in the smallest sample size ($T = 500$), and

[4]For each replication, we apply asymptotic theory to calculate the standard error evaluated at the parameter estimates. These standard errors are then averaged over all replications.

[5]The density of the standardized gamma distribution, u_t, with ν degrees of freedom is $f(u_t) = [(\nu/2)^{1/2}/\Gamma(\nu/2)] \cdot ((\nu/2)^{1/2}u_t + \nu/2)^{(\nu-2)/2} \cdot exp(-(\nu/2)^{1/2}u_t - \nu/2)$, where $u_t > -(\nu/2)^{1/2}$. Its mean and variance are equal to ν.

happens in less than 10 out of every 500 replications. We discard these replications. Keeping them would have little effect on bias or RMSE. Second, again holding all other features of the DGP as in the base case, we vary T: we consider $T = 500$ and 1,000 in addition to the base value of 2,000. Third, we verify that changing ω does not change the results, in a sense to be made precise later. Fourth, again holding all other features of the GDP as in the base case, we vary α and β.

3. Results of the Experiments

3.1 Different Distributions

In this set of experiments we consider different distributions. In all cases we use the base case parameter values: $T = 2,000$, $\alpha = 0.15$, $\beta = 0.7$, unconditional variance $=$ 10 ($\omega = 1.5$).

A. The True Distribution is Gaussian

Table 3.1 shows the results for the case in which the true distribution is standard normal. Asymptotically, augmentation makes no difference but intuition suggests that redundant moment conditions may be harmful in finite samples, so that NQMLE might be best. This is generally true but the differences are not too large. The results for GMM2 are very similar in all respects to those for NQMLE. The GMM1 estimates of ω are severely biased. Apart from that, however, GMM2 and GMM3 show some bias but not very much, and their RMSE is larger but not much larger than that of NQMLE. The main disadvantage of GMM1 and GMM3 is that the asymptotic standard errors understate the finite sample standard errors, especially when g is large.

B. The True Distribution is Standardized t_5

For the student's t distribution with ν degrees of freedom, the mean is zero and the variance is $\nu/(\nu-2)$ for $\nu > 2$. The fourth moment is equal to $3\nu^2[(\nu-2)(\nu-4)]^{-1}$ for $\nu > 4$. When $\nu = 5$, the degree of excess[6] equals 6, which indicates that the t distribution has a thicker tail than the normal distribution (whose degree of excess equals zero).

Table 3.2 provides the Monte Carlo simulation results when the true distribution is student's t with five degrees of freedom. We first note that the NQMLE performs adequately, in the sense that there is little bias and the average asymptotic standard error is only a little smaller than the finite sample standard error. GMM2 also shows little bias, and its RMSE is considerably smaller than that of the NQMLE, for ω and β at least. So GMM2 does achieve finite sample efficiency gains over NQMLE. However, its asymptotic standard errors are somewhat less reliable than NQMLE's. The highly overidentified estimators (MDE, GMM1, and GMM3) all have noticeable bias for at least some of the parameters, and their asymptotic standard errors are not very reliable, especially when g is large. However, GMM3 does have smaller MSE than NQMLE, and the differences are non-trivial, especially when g is not too large.

TMLE is the MLE with the standardized student's t likelihood function. The

[6]The degree of excess is measured by the fourth central moment normalized by the squared variance minus 3.

RMSE for all three parameters is quite small: The RMSE of NQMLE is more than 35% larger than that of TMLE. So, if the true distribution were known, there would be a fairly large efficiency gain from using the true MLE.

C. The True Distribution is Standardized χ_2^2

The χ^2 distribution with ν degree of freedom is asymmetric, and from Patel, Kapadia, and Owen (1976) the mean and the variance are equal to ν and 2ν. Its coefficient of skewness and coefficient of excess are equal to $2^{3/2}/\nu^{1/2}$ and $12/\nu$, respectively. The χ_2^2 distribution has a fat tail and a "long tail" in the right direction; this is seen from the fact that the coefficient of excess (6) and the coefficient of skewness (2) are both larger than zero.

Table 3.3 presents the Monte Carlo simulation results when the true distribution follows the standardized χ_2^2 distribution. These results show many of the same patterns as in the case of the standardized t_5 distribution, but they are much more favorable for the augmented GMM estimators. The NQMLE performs adequately in the same sense as before – little bias, and relatively reliable asymptotic standard errors. In terms of RMSE, GMM2 and GMM3 are better than NQMLE, while GMM1 and even MDE are sometimes better. MDE, GMM1 and GMM3 give biased estimates, especially of α, but GMM2 is essentially unbiased. Its asymptotic standard errors are somewhat less reliable than those of NQMLE, but much more reliable than those of GMM1 and GMM3. Overall, GMM2 seems best, since it achieves considerable finite sample efficiency gains relative to NQMLE, without being badly biased and without having very unreliable asymptotic standard errors.

As mentioned before, MLE is not valid in this case. The efficiency loss of the NQMLE cannot be shown through the comparison with the χ_2^2 MLE.

D. The True Distribution is Standardized Gamma Distribution with Two Degrees of freedom

The gamma distribution with ν degrees of freedom has mean and variance both equal to ν, while the coefficient of skewness equals $2\nu^{-1/2}$ and the coefficient of excess is $6/\nu$. With two degrees of freedom, the gamma distribution has coefficient of skewness equal to 1.414, and coefficient of excess equal to 3. This means that the gamma distribution has a fat tail and is asymmetric.

The Monte Carlo simulation results for this distribution are shown in Table 3.4. We will not discuss these results in detail because they are relatively similar to those for the χ_2^2 case. Both NQMLE and GMM2 are essentially unbiased, and both have fairly reliable asymptotic standard errors, but GMM2 is better in the sense of smaller RMSE. GMM1 and GMM3 are generally good in terms of RMSE, but they are biased and their asymptotic standard errors are not reliable, especially when g is large. Overall GMM2 seems like the best choice.

3.2 Different Values of Sample Size

We now consider the effects of changing the sample size (T). Our base case (Table 3.2) had $T = 2,000$, with the standardized t_5 distribution and with $\alpha = 0.15$, $\beta = 0.7$, $\omega = 1.5$. We now give results for the same parameter values and distribution, but for

$T = 500$ (Table 3.5) and $T = 1,000$ (Table 3.6). We wish to check our intuition that the augmented GMM estimators will do better relative to NQMLE when T is larger, and worse when it is smaller.

This intuition is supported by our results. When $T = 500$, GMM2 offers little efficiency gain over NQMLE, while it is more biased and its asymptotic standard errors are less reliable. GMM1 and GMM3 do offer reductions in RMSE over NQMLE, but their asymptotic standard errors are quite unreliable. When $T = 1,000$, the results are (unsurprisingly) between those for $T = 500$ and $T = 2,000$, and in particular the comparison between NQMLE and GMM2 depends on which parameter (α, β or ω) you look at.

3.3. ω Doesn't Matter

Up to now, we assume the unconditional variance is equal to 10. In this section we verify that its value does not affect any of our substantive results (e.g. comparisons of estimators).

In fact, a change in the unconditional variance only rescales the whole series. That is, the series with the unconditional variance ϕ is the same as the series with the unconditional variance 1 times $\sqrt{\phi}$, if the same basic random numbers are used. Since these two data series are identical in the above sense, the simulation results for estimates of α and β should be the same. For ω, the estimates (and their standard errors and RMSE) should all change by the proportion ϕ.

Table 3.7 gives the results for a case that is the same as the base case except that the unconditional variance equals 1 ($\omega = 0.15$) instead of 10 ($\omega = 1.5$). Comparing these results to those for the base case (Table 3.2), we see that the results do match in the sense described above: for α and β the results are the same, while for ω they decrease by a factor of 10. This correspondence is exact (to the number of decimal places reported) for NQMLE, TMLE and GMM1, while there are some very minor differences for GMM2 and GMM3. Apart from these minor computational differences, the results verify that the choice of ω doesn't matter.

3.4 Effects of Changing α and β

We have assumed $\alpha = 0.15$ and $\beta = 0.7$ in the previous sections. In this section, we analyze the effects of changing the parameter values of α and β while we maintain the assumptions that $T = 2,000$, and that the true distribution is a student's t distribution with 5 degrees of freedom.

Table 3.8, 3.9 and 3.10 give the results when $\{\alpha = 0.1,\ \beta = 0.75\}$, $\{\alpha = 0.1,\ \beta = 0.8\}$ and $\{\alpha = 0.2,\ \beta = 0.6\}$ respectively. If we compare these results to those for the base case $\{\alpha = 0.15,\ \beta = 0.7\}$ in Table 3.2, we see that the new results are broadly similar. NQMLE and GMM2 are fairly similar, but GMM2 generally has smaller RMSE. The more highly overidentified estimators often are reasonably good in terms of RMSE, but bad in terms of bias and the reliability of their asymptotic standard errors.

3.5 Accuracy of Inference

In previous sections we have compared the finite sample standard error to the average estimated asymptotic standard error, to see whether the estimated asymptotic standard errors are reliable. For example, when the asymptotic standard error is on average smaller than the finite sample standard error, we would expect that tests based on the asymptotic standard errors would overreject. In this section we provide direct evidence on this point, by giving the rejection frequency (true size) of the normal 5% Wald tests of the hypotheses that ω, α and β equal their true values.

These results are given in Table 3.11. We assume the base case except that we consider four different choices of the true distribution, as in section 3.1 above.

For the NQMLE, the simulation size is less than 0.05 when the true distribution is standard normal; for the other (non-Gaussian) cases the simulation size is larger than 0.05. For GMM2, the size for ω and β are below 0.05, while the size for α is over 0.05 for all distributions. Size is roughly in the range from 0.02 to 0.10. In that sense inference is relatively reliable.

For the MDE, GMM1, and GMM3 estimators, the size distortions are quite serious, especially for α. The frequency of rejection for α is often over 30%. The overrejection problem is generally worst for the highly overidentified estimators (MDE, GMM1 and GMM3 with large values of g). This pattern is exactly what we would expect from our earlier comparisons of finite sample and asymptotic standard errors.

4. Conclusions

Recall that the results in Chapter 2 indicate that, in large samples, the augmented GMM estimators provide an asymptotic efficiency gain when the distribution of the innovations is not normal, and the only price to pay for this gain is computational time and effort. In this chapter, we turn our attention to the finite sample properties of NQMLE, MDE and the augmented GMM estimators, to see whether an efficiency gain can be realized in samples of reasonable size.

The simulation results show that, although for some estimators there is an efficiency gain in terms of smaller MSE, the gain comes with non-trivial cost: the estimates may be biased, and inference based on the asymptotic distribution of the estimates may be very inaccurate. These problems are particularly serious for MDE, GMM1, and GMM3, even for rather large sample sizes. Therefore, doubts can be raised regarding these estimators' usefulness in empirical studies. Fortunately, not all of the augmented GMM estimators perform badly. GMM2 does not suffer from a serious bias problem, and inference based on its asymptotic distribution is reasonably reliable if the sample size is large enough. Also, if the sample size is large enough, GMM2 does provide non-trivial efficiency gains over NQMLE. This is especially true when the data are asymmetric.

In practice, our results support the use of GMM2 if the sample size is as large as 1,000 and there is evidence of asymmetry, or if the sample size is as large as 2,000 and the data are symmetric but with fat tails. For smaller sample sizes, NQMLE would remain the preferred method.

Table 3.1

Monte Carlo simulation result

Ture distribution: standard normal distribution

T = 2,000, R = 500

$w = 1.5$, $\alpha = 0.15$, $\beta = 0.7$

	ω				α				β			
	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg. asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.
NQMLE	1.5540	0.3241	0.3286	0.3699	0.1510	0.0242	0.0242	0.0248	0.6931	0.0461	0.0466	0.0524
MDE												
g = 5					0.1360	0.0356	0.0383	0.0339	0.6934	0.0937	0.0940	0.0974
g = 10					0.1370	0.0276	0.0305	0.0229	0.6846	0.0671	0.0688	0.0613
g = 20					0.1386	0.0256	0.0280	0.0194	0.6853	0.0590	0.0608	0.0500
GMM1												
g = 5	1.8321	0.5243	0.6206	0.4239	0.1427	0.0234	0.0245	0.0206	0.6648	0.0642	0.0733	0.0579
g =10	1.7451	0.4206	0.4868	0.3633	0.1410	0.0229	0.0246	0.0187	0.6729	0.0563	0.0625	0.0506
g = 20	1.6675	0.3850	0.4199	0.3099	0.1411	0.0232	0.0248	0.0166	0.6790	0.0537	0.0577	0.0436
GMM2	1.5201	0.3499	0.3504	0.3616	0.1499	0.0246	0.0246	0.0247	0.6974	0.0502	0.0503	0.0514
GMM3												
g = 5	1.5149	0.3496	0.3499	0.3550	0.1346	0.0239	0.0284	0.0189	0.7069	0.0502	0.0506	0.0502
g = 10	1.5145	0.3471	0.3474	0.3266	0.1338	0.0235	0.0286	0.0175	0.7059	0.0504	0.0508	0.0462
g = 20	1.5164	0.3428	0.3432	0.2901	0.1360	0.0237	0.0275	0.0159	0.7019	0.0505	0.0505	0.0410

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 3.2

Monte Carlo simulation result

True distribution: standardized t distribution with 5 degrees of freedom

T = 2,000, R = 500

$\omega = 1.5, \alpha = 0.15, \beta = 0.7$

	ω				α				β			
	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e	mean	s.e.	RMSE	avg.asy.s.e.
NQMLE	1.6408	0.5798	0.5967	0.4967	0.1498	0.0420	0.0420	0.0379	0.6843	0.0824	0.0839	0.0714
MDE												
g = 5					0.1265	0.0445	0.0503	0.0414	0.6830	0.1223	0.1234	0.1148
g = 10					0.1304	0.0377	0.0425	0.0236	0.6844	0.0924	0.0937	0.0581
g = 20					0.1356	0.0360	0.0387	0.0189	0.6851	0.0787	0.0801	0.0424
GMM1												
g = 5	1.8319	0.7489	0.8192	0.4625	0.1351	0.0342	0.0373	0.0237	0.6626	0.0914	0.0987	0.0644
g =10	1.7059	0.5906	0.6255	0.3500	0.1344	0.0342	0.0376	0.0199	0.6743	0.0831	0.0870	0.0486
g = 20	1.6199	0.5546	0.5674	0.2794	0.1362	0.0351	0.0377	0.0166	0.6809	0.0809	0.0831	0.0381
GMM2	1.5542	0.4042	0.4078	0.4689	0.1481	0.0403	0.0404	0.0321	0.6938	0.0610	0.0613	0.0678
GMM3												
g = 5	1.5185	0.3970	0.3974	0.4088	0.1291	0.0333	0.0394	0.0215	0.7033	0.0605	0.0606	0.0593
g = 10	1.5166	0.4032	0.4036	0.3299	0.1300	0.0352	0.0405	0.0190	0.7021	0.0618	0.0618	0.0477
g = 20	1.5129	0.4082	0.4084	0.2760	0.1330	0.0366	0.0403	0.0164	0.6991	0.0637	0.0637	0.0391
TMLE	1.5673	0.4042	0.4097	0.3942	0.1497	0.0311	0.0311	0.0308	0.6934	0.0594	0.0598	0.0578

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 3.3

Monte Carlo simulation result

True distribution: standardized chi-square distribution with 2 degrees of freedom

T = 2,000, R = 500

$\omega = 1.5, \alpha = 0.15, \beta = 0.7$

	ω				α				β			
	mean	s.e	RMSE	avg.asy.s.e.	mean	s.e	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.
NQMLE	1.5814	0.5680	0.5738	0.5105	0.1553	0.0514	0.0517	0.0448	0.6874	0.0840	0.0850	0.0772
MDE												
g = 5					0.1237	0.0471	0.0540	0.0360	0.6943	0.1155	0.1156	0.1051
g = 10					0.1320	0.0442	0.0477	0.0204	0.6910	0.0886	0.0891	0.0490
g = 20					0.1402	0.0443	0.0453	0.0158	0.6900	0.0800	0.0806	0.0332
GMM1												
g = 5	1.7455	0.6765	0.7197	0.4534	0.1346	0.0429	0.0456	0.0233	0.6712	0.0900	0.0945	0.0640
g =10	1.6397	0.6187	0.6343	0.3295	0.1350	0.0437	0.0462	0.0186	0.6812	0.0861	0.0881	0.0436
g = 20	1.5413	0.5479	0.5494	0.2536	0.1383	0.0440	0.0455	0.0152	0.6873	0.0830	0.0840	0.0323
GMM2	1.4939	0.3504	0.3504	0.4919	0.1497	0.0416	0.0416	0.0334	0.7010	0.0536	0.0536	0.0725
GMM3												
g = 5	1.4521	0.3467	0.3500	0.4064	0.1295	0.0392	0.0443	0.0217	0.7099	0.0541	0.0550	0.0601
g = 10	1.4553	0.3740	0.3767	0.3030	0.1311	0.0427	0.0467	0.0178	0.7082	0.0602	0.0607	0.0434
g = 20	1.4648	0.3897	0.3913	0.2438	0.1352	0.0440	0.0465	0.0148	0.7043	0.0635	0.0637	0.0338

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 3.4

Monte Carlo simulation result

True distribution: standardized gamma distribution with 2 degrees of freedom

T = 2,000, R = 500

$\omega = 1.5, \alpha = 0.15, \beta = 0.7$

	ω				α				β			
	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.
NQMLE	1.6175	0.5172	0.5304	0.4722	0.1536	0.0392	0.0394	0.0372	0.6840	0.0737	0.0754	0.0697
MDE												
g = 5					0.1256	0.0432	0.0497	0.0361	0.6986	0.1174	0.1175	0.1063
g = 10					0.1329	0.0356	0.0395	0.0213	0.6854	0.0808	0.0821	0.0531
g = 20					0.1400	0.0364	0.0377	0.0173	0.6818	0.0749	0.0771	0.0397
GMM1												
g = 5	1.7905	0.6343	0.6976	0.4411	0.1368	0.0346	0.0371	0.0224	0.6680	0.0840	0.0899	0.0621
g =10	1.7078	0.5640	0.6010	0.3369	0.1370	0.0349	0.0372	0.0186	0.6747	0.0783	0.0823	0.0461
g = 20	1.6290	0.5049	0.5212	0.2688	0.1399	0.0353	0.0367	0.0154	0.6885	0.0755	0.0764	0.0360
GMM2	1.5487	0.4185	0.4213	0.4539	0.1512	0.0333	0.0334	0.0304	0.6935	0.0584	0.0588	0.0656
GMM3												
g = 5	1.5098	0.3906	0.3907	0.3909	0.1318	0.0320	0.0368	0.0205	0.7035	0.0573	0.0574	0.0569
g = 10	1.5106	0.3951	0.3952	0.3088	0.1324	0.0336	0.0379	0.0177	0.7022	0.0592	0.0592	0.0444
g = 20	1.5101	0.4066	0.4067	0.2562	0.1358	0.0346	0.0374	0.0152	0.6985	0.0612	0.0613	0.0360

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 3.5

Monte Carlo simulation result

True distribution: standardized t distribution with 5 degrees of freedom

T = 500, R = 500

$\omega = 1.5, \alpha = 0.15, \beta = 0.7$

	ω				α				β			
	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.
NQMLE	1.5750	1.2817	1.2839	1.0160	0.1379	0.0850	0.0859	0.0804	0.7046	0.1805	0.1806	0.1500
MDE												
g = 5					0.1062	0.0707	0.0832	0.2737	0.6846	0.2360	0.2365	0.5859
g = 10					0.1237	0.0595	0.0651	0.0479	0.6699	0.1784	0.1810	0.1201
g = 20					0.1379	0.0595	0.0607	0.0262	0.6656	0.1614	0.1651	0.0701
GMM1												
g = 5	2.2827	1.5057	1.6970	2.5058	0.1311	0.0608	0.0636	0.0376	0.6041	0.1711	0.1961	0.2991
g =10	1.9922	1.3264	1.4148	0.7433	0.1331	0.0626	0.0648	0.0292	0.6357	0.1667	0.1787	0.0967
g = 20	1.8485	1.3094	1.3550	0.5480	0.1390	0.0657	0.0666	0.0229	0.6458	0.1601	0.1690	0.0700
GMM2	1.8022	1.2425	1.2787	0.9691	0.1698	0.1016	0.1035	0.0678	0.6502	0.1715	0.1786	0.1364
GMM3												
g = 5	1.7538	1.2473	1.2729	0.7918	0.1299	0.0693	0.0722	0.0338	0.6746	0.1657	0.1677	0.1100
g = 10	1.7642	1.2673	1.2946	0.6043	0.1343	0.0699	0.0717	0.0271	0.6697	0.1695	0.1722	0.0810
g = 20	1.7707	1.2596	1.2884	0.4715	0.1425	0.0709	0.0713	0.0220	0.6613	0.1682	0.1726	0.0621
TMLE	1.6856	0.8697	0.8893	0.9661	0.1468	0.0612	0.0613	0.0618	0.6836	0.1240	0.1251	0.1378

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 3.6

Monte Carlo simulation result

True distribution: standardized t distribution with 5 degrees of freedom

T = 1,000, R = 500

$\omega = 1.5, \alpha = 0.15, \beta = 0.7$

	ω				α				β			
	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.
NQMLE	1.7067	0.8505	0.8753	0.7065	0.1484	0.0575	0.0575	0.0500	0.6780	0.1176	0.1196	0.0998
MDE												
g = 5					0.1038	0.0668	0.0812	0.2479	0.7291	0.1909	0.1931	0.4744
g = 10					0.1271	0.0545	0.0591	0.0358	0.6781	0.1426	0.1443	0.0913
g = 20					0.1366	0.0504	0.0521	0.0224	0.6753	0.1199	0.1224	0.0566
GMM1												
g = 5	1.9939	1.0232	1.1362	0.8291	0.1326	0.0477	0.0508	0.0301	0.6422	0.1335	0.1454	0.1110
g =10	1.7724	0.8525	0.8950	0.4942	0.1313	0.0478	0.0513	0.0241	0.6661	0.1193	0.1240	0.0671
g = 20	1.6646	0.8029	0.8196	0.3894	0.1352	0.0496	0.0517	0.0195	0.6735	0.1175	0.1204	0.0513
GMM2	1.5675	0.6227	0.6263	0.6574	0.1471	0.0643	0.0644	0.0441	0.6937	0.0940	0.0942	0.0946
GMM3												
g = 5	1.5096	0.5819	0.5820	0.5749	0.1228	0.0455	0.0530	0.0278	0.7091	0.0858	0.0862	0.0826
g = 10	1.5262	0.6048	0.6054	0.4581	0.1265	0.0506	0.0558	0.0233	0.7042	0.0912	0.0913	0.0641
g = 20	1.5307	0.6322	0.6329	0.3682	0.1320	0.0507	0.0538	0.0193	0.6980	0.0958	0.0958	0.0502
TMLE	1.6346	0.5121	0.5295	0.2996	0.1497	0.0414	0.0414	0.0218	0.6860	0.0750	0.0763	0.0439

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 3.7

Asymptotic standard error and Monte Carlo simulation result

True distribution: standardized t distribution with 5 degrees of freedom

T = 2,000, R = 500

$\omega = 0.15, \alpha = 0.15, \beta = 0.7$

	Asymptotic standard error			ω				α				β			
	ω	α	β	mean	s.e.	RMSE	avg.asy.s.e	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.
NQMLE	0.0536	0.0428	0.0762	0.1641	0.0580	0.0597	0.0497	0.1498	0.0420	0.0420	0.0379	0.6843	0.0824	0.0839	0.0714
MDE															
g = 5		0.0623	0.1636					0.1265	0.0445	0.0503	0.0414	0.6830	0.1223	0.1234	0.1148
g = 10		0.0479	0.0998					0.1304	0.0377	0.0425	0.0236	0.6844	0.0924	0.0937	0.0581
g = 20		0.0452	0.0854					0.1356	0.0360	0.0387	0.0189	0.6851	0.0787	0.0801	0.0424
GMM1															
g = 5	0.0526	0.0337	0.0749	0.1832	0.0749	0.0819	0.0462	0.1351	0.0342	0.0373	0.0237	0.6626	0.0914	0.0987	0.0644
g =10	0.0494	0.0335	0.0713	0.1706	0.0591	0.0625	0.0350	0.1344	0.0342	0.0376	0.0199	0.6743	0.0831	0.0870	0.0486
g = 20	0.0465	0.0333	0.0680	0.1620	0.0555	0.0567	0.0279	0.1362	0.0351	0.0377	0.0166	0.6809	0.0809	0.0831	0.0381
GMM2	0.0511	0.0338	0.0738	0.1552	0.0403	0.0406	0.0469	0.1482	0.0420	0.0421	0.0321	0.6940	0.0618	0.0621	0.0679
GMM3															
g = 5	0.0502	0.0308	0.0727	0.1520	0.0397	0.0397	0.0411	0.1288	0.0331	0.0393	0.0215	0.7035	0.0605	0.0606	0.0595
g = 10	0.0472	0.0306	0.0691	0.1515	0.0403	0.0404	0.0330	0.1299	0.0352	0.0405	0.0190	0.7025	0.0619	0.0619	0.0477
g = 20	0.0444	0.0304	0.0656	0.1514	0.0410	0.0411	0.0277	0.1330	0.0366	0.0404	0.0165	0.6991	0.0635	0.0636	0.0394
TMLE	0.0363	0.0300	0.0547	0.1567	0.0404	0.0410	0.0394	0.1497	0.0311	0.0311	0.0308	0.6934	0.0594	0.0598	0.0578

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 3.8

Monte Carlo simulaiton result

True distribution: standardized t distribution with 5 degrees of freedom

T = 2,000, R = 500

$\omega = 1.5, \alpha = 0.1, \beta = 0.75$

	ω				α				β			
	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.
NQMLE	1.7035	0.7972	0.8227	0.6458	0.1002	0.0351	0.0351	0.0313	0.7280	0.1014	0.1038	0.0829
MDE												
g = 5					0.0845	0.0387	0.0417	0.0657	0.7130	0.1590	0.1632	0.1709
g = 10					0.0885	0.0327	0.0346	0.0204	0.7167	0.1172	0.1219	0.0734
g = 20					0.0925	0.0307	0.0316	0.0161	0.7193	0.1013	0.1059	0.0528
GMM1												
g = 5	2.0810	1.0521	1.2019	0.6993	0.0909	0.0285	0.0300	0.0212	0.6847	0.1216	0.1380	0.0864
g =10	1.8596	0.8336	0.9079	0.5120	0.0899	0.0290	0.0307	0.0177	0.7074	0.1042	0.1126	0.0640
g = 20	1.7261	0.7509	0.7842	0.3963	0.0909	0.0298	0.0312	0.0148	0.7199	0.0989	0.1034	0.0494
GMM2	1.6033	0.5635	0.5729	0.6148	0.1030	0.0459	0.0460	0.0279	0.7360	0.0774	0.0786	0.0797
GMM3												
g = 5	1.5423	0.4907	0.4925	0.5558	0.0845	0.0267	0.0308	0.0187	0.7496	0.0679	0.0679	0.0719
g = 10	1.5340	0.4987	0.4998	0.4544	0.0852	0.0288	0.0324	0.0167	0.7496	0.0696	0.0696	0.0589
g = 20	1.5350	0.5073	0.5086	0.3828	0.0876	0.0303	0.0327	0.0145	0.7468	0.0714	0.0714	0.0490
TMLE	1.6031	0.5423	0.5520	0.5097	0.1001	0.0256	0.0256	0.0256	0.7395	0.0704	0.0712	0.0667

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 3.9

Monte Carlo simulation result

True distribution: standardized t distribution with 5 degrees of freedom

T = 2,000, R = 500

$\omega = 1.0, \alpha = 0.1, \beta = 0.8$

	ω				α				β			
	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.
NQMLE	1.0985	0.4984	0.5080	0.4210	0.0986	0.0309	0.0309	0.0286	0.7901	0.0706	0.0713	0.0601
MDE												
g = 5					0.0830	0.0391	0.0427	0.0804	0.7765	0.1389	0.1409	0.1728
g = 10					0.0858	0.0289	0.0323	0.0187	0.7920	0.0842	0.0846	0.0553
g = 20					0.0903	0.0272	0.0289	0.0137	0.7876	0.0698	0.0709	0.0349
GMM1												
g = 5	1.4021	0.8342	0.9260	0.4674	0.0913	0.0267	0.0281	0.0193	0.7542	0.1010	0.1109	0.0636
g =10	1.2047	0.5575	0.5939	0.3249	0.0892	0.0254	0.0276	0.0157	0.7762	0.0734	0.0772	0.0454
g = 20	1.1233	0.5077	0.5225	0.2422	0.0901	0.0266	0.0284	0.0126	0.7843	0.0701	0.0718	0.0329
GMM2	1.0481	0.3634	0.3666	0.3809	0.0999	0.0324	0.0324	0.0247	0.7936	0.0538	0.0542	0.0553
GMM3												
g = 5	1.0191	0.3705	0.3710	0.3488	0.0853	0.0244	0.0285	0.0169	0.8029	0.0541	0.0542	0.0506
g = 10	1.0210	0.3738	0.3744	0.2924	0.0855	0.0249	0.0289	0.0149	0.8021	0.0546	0.0547	0.0423
g = 20	1.0159	0.3842	0.3846	0.2298	0.0879	0.0266	0.0292	0.0126	0.8003	0.0569	0.0569	0.0329
TMLE	1.0528	0.3171	0.3214	0.3177	0.0993	0.0224	0.0225	0.0227	0.7950	0.0468	0.0470	0.0469

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 3.10

Monte Carlo simulation result

True distribution: standardized t distribution with 5 degrees of freedom

T = 2,000, R = 500

$\omega = 2.0$, $\alpha = 0.2$, $\beta = 0.6$

	ω				α				β			
	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.	mean	s.e.	RMSE	avg.asy.s.e.
NQMLE	2.1671	0.6571	0.6780	0.5633	0.2007	0.0519	0.0519	0.0464	0.5803	0.0926	0.0947	0.0802
MDE												
g = 5					0.1673	0.0493	0.0592	0.0391	0.5854	0.1161	0.1170	0.1029
g = 10					0.1732	0.0461	0.0533	0.0293	0.5806	0.1018	0.1036	0.0648
g = 20					0.1800	0.0441	0.0485	0.0243	0.5856	0.0879	0.0891	0.0491
GMM1												
g = 5	2.2843	0.6843	0.7410	0.4703	0.1786	0.0412	0.0464	0.0275	0.5672	0.0922	0.0979	0.0662
g =10	2.2050	0.6461	0.6778	0.3793	0.1792	0.0419	0.0468	0.0237	0.5724	0.0918	0.0958	0.0520
g = 20	2.1022	0.6086	0.6172	0.3165	0.1816	0.0425	0.0463	0.0205	0.5786	0.0897	0.0922	0.0433
GMM2	2.0330	0.4707	0.4718	0.5295	0.1942	0.0420	0.0424	0.0385	0.5974	0.0700	0.0701	0.0763
GMM3												
g = 5	1.9805	0.4392	0.4396	0.4445	0.1725	0.0425	0.0506	0.0256	0.6073	0.0702	0.0706	0.0647
g = 10	1.9804	0.4406	0.4410	0.3665	0.1739	0.0449	0.0520	0.0227	0.6048	0.0725	0.0726	0.0524
g = 20	1.9774	0.4441	0.4447	0.3197	0.1778	0.0464	0.0514	0.0203	0.6000	0.0742	0.0742	0.0453
TMLE	2.0754	0.4595	0.4657	0.4461	0.1998	0.0383	0.0383	0.0370	0.5920	0.0684	0.0689	0.0660

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 3.11
Size of the 5% Wald test
T = 2,000, R = 500
$\omega = 1.5, \alpha = 0.15, \beta = 0.7$

	standard normal			standardized t(5)			standardized chisquare(2)			standardized gamma(2)		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β
NQMLE	0.034	0.044	0.026	0.086	0.096	0.086	0.074	0.106	0.078	0.086	0.096	0.086
MDE												
g = 5		0.096	0.024		0.204	0.046		0.318	0.086		0.214	0.074
g = 10		0.192	0.060		0.334	0.196		0.442	0.298		0.342	0.196
g = 20		0.210	0.088		0.368	0.298		0.512	0.434		0.402	0.282
GMM1												
g = 5	0.080	0.120	0.078	0.128	0.340	0.156	0.154	0.358	0.164	0.100	0.260	0.136
g =10	0.088	0.158	0.048	0.200	0.324	0.238	0.286	0.466	0.320	0.180	0.348	0.228
g = 20	0.106	0.206	0.120	0.290	0.378	0.316	0.380	0.528	0.450	0.284	0.416	0.344
GMM2	0.018	0.060	0.022	0.048	0.100	0.050	0.022	0.086	0.022	0.024	0.074	0.026
GMM3												
g = 5	0.030	0.302	0.184	0.068	0.316	0.088	0.540	0.392	0.064	0.052	0.312	0.074
g = 10	0.038	0.360	0.216	0.134	0.388	0.170	0.134	0.486	0.180	0.114	0.384	0.160
g = 20	0.056	0.436	0.248	0.220	0.442	0.252	0.244	0.518	0.308	0.202	0.456	0.262

GMM1--the score from Normal distribution & MDE
GMM2-- the scores from Normal distribution & rescaled t distribution
GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Chapter 4
AN EMPIRICAL STUDY

1 Introduction

In this chapter, we apply the augmented GMM estimators to a GARCH(1,1) model of the the hourly foreign exchange rate series of the West German deutschmark versus U.S. dollar (DM/$). The hourly data cover a six-month period in 1986, from 0:00 a.m. January 2, 1986 through 11:00 a.m. July 15, 1986. The data set contains a total of 3,190 trading hours. The exchange rate is taken from the average of the last five bid rates recorded at the end of each hour by the fifty largest banks in the foreign exchange market; for additional information, see Baillie and Bollerslev (1990).

This data set has been analyzed in a number of studies, including Baillie and Bollerslev (1990). They test the null hypothesis of a unit root in the logarithm of the exchange rate series, based on the methodology of Phillips (1987) and Phillips and Perron (1988), and find that it cannot be rejected. Therefore, our estimation is based on the first difference of the exchange rate,

$$\epsilon_t = 100 \cdot [ln(s_t) - ln(s_{t-1})], \quad t = 2, 3, ..., 3189.$$

where s_t is the exchange rate before the transformation.

We calculate the first four unconditional moments and the correlogram of the series, ϵ_t, as shown in Table 4.1. The mean is -0.004 and the variance is 0.035. The skewness and the kurtosis are, respectively, 0.171 and 10.027. The normal distribution would have skewness equal to zero and kurtosis equal to three. The series is clearly non-normal; it is an asymmetric distribution with thick tails. These are features of the unconditional distribution of the series, whereas the potential advantages of augmented GMM derive from non-normality of the conditional distribution. Nevertheless, the clear non-normality of the series should make the augmented GMM estimators potentially useful.

2. Estimation of the GARCH(1,1) Model

Before we begin estimation, we first ask what kind of model specification is suitable for this series. The correlogram in Table 4.1 shows that there is no linear correlation in the mean, but there is linear correlation in the second moments. Therefore an ARCH-type model would seem appropriate. Baillie and Bollerslev (1990) analyzed these hourly exchange rates using an MA(1)-GARCH(1,1) model. They found that the variables in the conditional variance of the GARCH(1,1) model were significant but the MA(1) coefficient was insignificant. This result, and the lack of correlation in the levels of the variable which we found in Table 4.1, suggest that the martingale-GARCH(1,1) model is appropriate. This is the specification later adopted by Baillie and Chung (1999). On the other hand, unlike Baillie and Bollerslev (1990), Baillie and Chung (1999) do not include hourly dummy variables and vacation day dummy variables in the model. We follow Baillie and Chung (1999) and adopt a martingale-GARCH(1,1) model without hourly or vacation dummy variables.

The model to be estimated is

$$\epsilon_t|\Psi_{t-1} \sim D(0, h_t),$$

$$h_t = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}.$$

NQMLE, MDE, and three types of augmented GMM estimators are used. The augmented GMM estimators are GMM1, GMM2 and GMM3 as previously defined in chapters 2 and 3. The estimation was performed using the CML and CO procedures in GAUSS for optimization. The NEWTON (Newton-Raphson) method is the optimization algorithm. The weighting matrix has the specification suggested by Newey and West (1987), with the lag length selected with the automatic-lag selection criterion of Newey and West (1994). For NQMLE, we tried different initial values for the optimization, and obtained the same results. The initial values used in the estimation of the augmented GMM estimators are from the results of NQMLE.

Table 4.2 shows the estimation results. For NQMLE, the estimates of ω, α and β are significant and the sum of α and β is less than one. There is no evidence of integrability in variance. For MDE with at least 10 autocorrelations of ϵ_t^2, the estimates of β are close to those of NQMLE but with smaller standard errors. However, MDE gives a smaller estimate of α. For GMM1, the estimate of α is still less than those of NQMLE, while the estimates of ω and β are close to those of NQMLE. For all of the parameters, the GMM1 standard errors are smaller than those of NQMLE. For GMM2, all estimates are close to the results of NQMLE, but with smaller standard errors. For GMM3, the estimate of α is less than the NQMLE estimate, while the estimate of β is slightly larger than for NQMLE, and again the standard errors are smaller.

In conclusion, NQMLE and GMM2 are similar in terms of parameter estimates and standard errors. For MDE, GMM1, and GMM3, we get smaller estimates of α; there is no clear pattern on β. The various augmented procedures do give smaller standard errors. For example, the t-statistic of the estimated α is 4.74 for NQMLE but is 8.88 for GMM3 with g = 20.

3. Diagnostics

It is interesting to check the sample moments of the conditional distribution of the innovation $u_t = \epsilon_t h_t^{-1/2}$, which are reported in Table 4.3. The innovations are skewed and have severe excess kurtosis. Thus the conditions for possible efficiency gains from augmented GMM appear to exist.

Table 4.4 gives the sample autocorrelations of ϵ_t^2 and the theoretical (population) autocorrelations evaluated at the various estimates of α and β. The NQMLE and GMM2 values for the theoretical autocorrelations do not match the sample values very well. For MDE, GMM1 and GMM3, we should have closer agreement because the criterion minimized by the estimator includes the distance between the theoretical and sample autocorrelations. Still the discrepancies seem fairly large. This casts doubt on the validity of the model.

A formal test of the model can be carried out using the overidentification test of Hansen (1982). Since the augmented GMM estimators are overidentified (the number of moment conditions is larger than the number of parameters), the over-identifying restrictions can be tested to examine the validity of the moment restrictions. The test statistic is the minimized value of the GMM criterion function:

$$\bar{q}_T(\hat{\theta})'\hat{W}\bar{q}_T(\hat{\theta}), \tag{7}$$

where $\hat{\theta}$ and $\hat{W}$ are valued at the corresponding augmented GMM estimates. This statistic asymptotically has a χ^2 distribution with degrees of freedom equal to the number of moment conditions minus the number of parameters. Table 4.5 gives the values of the test statistic. For MDE, GMM2, GMM1, and GMM3 with g up to 10, we cannot reject the null hypotheses. For GMM1 and GMM3 with g = 20, the statistic exceeds the 5% critical value for the relevant χ^2 distribution. Thus, in these two cases, the null hypothesis that all moment conditions are satisfied is rejected.

We next apply the conditional moment test of Newey (1985). This test assumes the validity of one set of moment conditions, say q_{1t}, and tests the validity of a second set, say q_{2t}. In our case q_{1t} is the score function from the normal log likelihood and q_{2t} is the extra set of moment conditions for augmented GMM.

Suppose that the optimal GMM estimate $\tilde{\theta}$ is obtained using the first set of moment conditions, q_{1t}. Suppose also that $\hat{\theta}$ is a GMM estimate by adding the extra moments q_{2t}. The test statistic proposed by Newey (1985) is

$$m_T = H_T'Q^{-1}H_T,$$

where $H_T = \bar{q}_{2T}(\tilde{\theta}) - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}\bar{q}_{1T}(\tilde{\theta})$, and Q is the asymptotic covariance matrix of m_T, which can be consistently estimated by

$$\hat{\Omega}_{22} - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}\hat{\Omega}_{12} + (\hat{D}_2 - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}\hat{D}_1)(\hat{D}_1'\hat{\Omega}_{11}^{-1}\hat{D}_1)(\hat{D}_2 - \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}\hat{D}_1)'.$$

Here $\hat{\Omega}$ and $\hat{D}$ are as defined in chapter 2, and $\hat{D}_1$, $\hat{D}_2$,$\hat{\Omega}_{11}$,$\hat{\Omega}_{12}$ and $\hat{\Omega}_{22}$ are the appropriate submatrices. Asymptotically m_T is distributed as chi-squared with degrees of freedom equal to the number of moment conditions being tested. In our case, this is the same as the degree of overidentification (total number of moment conditions for augmented GMM, minus the number of parameters). Note that in our case the added conditions "q_{2t}" above can include the moment conditions called q_{2t} (GMM1), q_{3t} (GMM2) or both (GMM3) in chapter 2 and 3.

In our case, the reason why this is an interesting test to consider is that the extra moment conditions used by the augmented GMM estimators rely on stronger assumptions than the moment conditions used by the NQMLE. The validity of the NQMLE moment conditions requires only that the first two conditional moments be specified correctly (plus some regularity conditions). However, the validity of the extra moment conditions requires the representation (1) of chapter 1, which implies restrictions on the higher conditional moments. So, we are interested in testing these extra restrictions.

The test results are shown in Table 4.5. For GMM1 and GMM2, the values of m_T do not exceed the chi-squared critical value, and the hypothesis that $E[q_{2T}] = 0$ cannot be rejected. For GMM3, this hypothesis is rejected no matter how many moments based on autocorrelations of ϵ_t^2 are included.

Thus, for our empirical study, both types of specification test (overidentification test and conditional moment test) give mixed results. There is some doubt about the model but it is not decisively rejected. Perhaps adding the dummy variables used by

Baillie and Bollerslev (1990) would help to improve the model's conformance to the data.

4. Conclusions

We applied our augmented GMM estimators to a GARCH(1,1) model for the hourly DM/$ exchange rate. Using GMM2, the model passes our diagnostic tests, but both the estimates and their standard errors are very similar to NQMLE. The similarity is so strong that there is not very much point in using GMM2 instead of NQMLE. For GMM1 and GMM3, the results of the diagnostic tests are mixed. The estimates are only modestly different from the NQMLE estimates, but the standard errors are considerably smaller. For example, for GMM3 with g = 20, the asymptotic standard errors are about half as large as for NQMLE.

This raises the question of whether these efficiency gains are genuine, or just a reflection of an unreliably small asymptotic standard error. The simulations of chapter 3 give evidence on this point. In our empirical example we have $T = 3,190$, $\alpha = 0.18$, $\beta = 0.55$, and innovations that are slightly skewed and that have approximately the same degree of kurtosis as a t distribution with 5 degrees of freedom. The closest match in our simulations is in Table 3.10, where we have $T = 2,000$, $\alpha = 0.2$, $\beta = 0.6$, and the t_5 distribution. Here it was also true that GMM1 and GMM3 with large values of g had asymptotic standard errors that were about half as big as for NQMLE. The finite-sample (simulation) standard errors for GMM3 were also smaller than for NQMLE, but only $10 \sim 20\%$ smaller. Thus we might guess that, for our empirical example, the GMM1 or GMM3 estimates really are more precise than the NQMLE estimates, but not by as much as the asymptotic standard errors would indicate.

Table 4.1

The first four unconditional moments of the distribution of ε_t

mean	-0.004
variance	0.035
skewness	0.171
kurtosis	10.027

skewness -- $E\varepsilon_t^3/\sigma^3$

kurtosis -- $E\varepsilon_t^4/\sigma^4$

Correlogram

	Lag 1	2	3	4	5	6	7	8	9	10
ε_t	-0.022	0.020	0.024	-0.027	0.030	-0.024	-0.010	0.012	-0.007	0.027
ε_t^2	0.079*	0.075*	0.042*	0.055*	0.004	0.018	-0.007	-0.009	-0.019	0.017

* statistically significant at 1% level.

Table 4.2
Estimation result of DM/$ exchange rate

Model: $y_t = \varepsilon_t,\ \varepsilon_t|\Psi_{t-1} \sim D(0, h_t)$
$h_t = \omega + \alpha\, \varepsilon_{t-1}^2 + \beta\, h_{t-1}$

	ω		α		β	
	estimate	s.e.	estimate	s.e.	estimate	s.e.
NQMLE	0.01028	0.00262	0.17711	0.03736	0.54582	0.09157
MDE						
g=5	0.00990	n/a	0.08783	0.02476	0.62687	0.09514
g=10	0.01130	n/a	0.13056	0.02424	0.54372	0.05659
g=20	0.01112	n/a	0.13662	0.02270	0.54286	0.04909
GMM1						
g=5	0.01056	0.00167	0.13868	0.02322	0.54825	0.05507
g=10	0.01035	0.00137	0.15292	0.02208	0.53549	0.04498
g=20	0.01024	0.00125	0.15935	0.02073	0.53085	0.04119
GMM2	0.00997	0.00253	0.17852	0.03499	0.55343	0.08598
GMM3						
g=5	0.00937	0.00138	0.14727	0.01916	0.57000	0.04896
g=10	0.00943	0.00119	0.14417	0.01835	0.55819	0.04138
g=20	0.00942	0.00114	0.15514	0.01748	0.55852	0.03904
Semiparametric	0.00995		0.17147		0.55626	

GMM1--the score from Normal distribution & MDE
GMM2-- the scores from Normal distribution & rescaled t distribution
GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 4.3

The conditional distribution of $\varepsilon_t h_t^{-1/2}$

	mean	variance	skewness	kurtosis
NQMLE	-0.023	0.999	0.189	10.932
MDE				
g = 5	-0.022	0.989	0.183	10.416
g = 10	-0.023	1.008	0.177	10.652
g = 20	-0.023	1.012	0.178	10.692
GMM1				
g = 5	-0.023	1.036	0.181	10.728
g = 10	-0.024	1.058	0.183	10.830
g = 20	-0.024	1.066	0.183	10.877
GMM2	-0.023	1.001	0.192	10.952
GMM3				
g = 5	-0.024	1.060	0.191	10.824
g = 10	-0.024	1.092	0.188	10.821
g = 20	-0.024	1.071	0.191	10.877

skewness -- $E\varepsilon_t^3/\sigma^3$

kurtosis -- $E\varepsilon_t^4/\sigma^4$

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

Table 4.4
SACF and ACF

SACF of ε_t^2

Lag									
1	2	3	4	5	6	7	8	9	10
0.079	0.074	0.042	0.055	0.004	0.018	-0.007	-0.009	-0.019	0.017

ACF evaluated at α, β

	Lag									
	1	2	3	4	5	6	7	8	9	10
NQMLE*	0.160	0.091	0.052	0.029	0.017	0.010	0.005	0.003	0.002	0.001
NQMLE	0.211	0.152	0.110	0.080	0.058	0.042	0.030	0.022	0.016	0.011
MDE										
g = 5	0.098	0.070	0.050	0.036	0.025	0.018	0.013	0.009	0.007	0.005
g = 10	0.147	0.099	0.067	0.045	0.030	0.020	0.014	0.009	0.006	0.004
g = 20	0.155	0.105	0.071	0.049	0.033	0.022	0.015	0.010	0.007	0.005
GMM1										
g = 5	0.158	0.109	0.075	0.051	0.035	0.024	0.017	0.011	0.008	0.005
g = 10	0.176	0.121	0.083	0.057	0.039	0.027	0.019	0.013	0.009	0.006
g = 20	0.184	0.127	0.088	0.060	0.042	0.029	0.020	0.014	0.009	0.007
GMM2	0.214	0.157	0.115	0.084	0.061	0.045	0.033	0.024	0.018	0.013
GMM3										
g = 5	0.172	0.123	0.088	0.063	0.045	0.033	0.023	0.017	0.012	0.009
g = 10	0.166	0.117	0.082	0.058	0.040	0.028	0.020	0.014	0.010	0.007
g = 20	0.181	0.129	0.092	0.066	0.047	0.034	0.024	0.017	0.012	0.009

GMM1–the score from Normal distribution & MDE
GMM2– the scores from Normal distribution & rescaled t distribution
GMM3– the scores from Normal distribution & rescaled t distribution and MDE
* – the estimates based on Baillie and Bollerslev (1990)

Table 4.5

Overidentification test and conditional moment test

	Overidentification test	conditional moment test	chi-square distribution	
	the minimized value of GMM criterion	m_T	degrees of freedom	the critical value*
MDE				
g = 5	3.48	n/a	2	5.99
g = 10	12.28	n/a	7	14.07
g = 20	23.25	n/a	17	27.59
GMM1				
g = 5	9.85	7.61	5	11.07
g = 10	16.97	13.86	10	18.31
g = 20	31.51	25.55	20	31.41
GMM2	0.03	0.76**	2	5.99
GMM3				
g = 5	12.34	19.14	7	14.07
g = 10	18.50	24.92	12	21.03
g = 20	32.43	38.91	22	33.93

GMM1--the score from Normal distribution & MDE

GMM2-- the scores from Normal distribution & rescaled t distribution

GMM3-- the scores from Normal distribution & rescaled t distribution and MDE

* -- with 0.95%

** -- there is no double sum in weighting matrix of GMM2 case

Chapter 5
CONCLUDING REMARKS

The normal quasi maximum likelihood estimator (NQMLE) is the MLE under the assumption of normality, but this assumption is likely to be violated in empirical data. Although in such cases the NQMLE is still consistent and asymptotically normal, it is inefficient. Therefore it is worthwhile to find more efficient estimators. We suggest augmented GMM estimators to improve the efficiency of the NQMLE under non-normality. We interpret the NQMLE as a GMM estimator, where the moment conditions represent the normal score function, and then we augment this set of moment conditions with other moment conditions that also do not depend on the validity of any particular distributional assumption.

We consider two sets of extra moments to augment the GMM (NQMLE) estimator. The first set of extra moments is from the autocorrelations of the squared innovations. The second set of extra moments is from the rescaled student's t distribution. The student's t distribution is of particular interests because its property of leptokurtosis (fat tail) is consistent with many empirical financial data sets. We consider three combinations of these extra moments, and the resulting augmented GMM estimators are called GMM1, GMM2, and GMM3.

We compare the performance of these different estimators by calculating and comparing the asymptotic standard errors. When the true density is non-Gaussian, our results show that the augmented GMM estimators have moderate improvements in efficiency compared with NQMLE. The efficiency gain is mostly in estimation of the parameter α; for ω and β there is very little improvement.

Monte Carlo results show that the augmented GMM estimators have an efficiency gain over the NQMLE when the true distribution is non-Gaussian, but this requires a rather large sample size, such as $T = 2,000$. For smaller sample sizes, adding moment conditions from the autocorrelations of ϵ_t^2 helps, especially for α, but extra moments from the rescaled t distribution do not seem to improve efficiency when the sample size is small (e.g. 500). The GMM3 estimator, which uses both extra sets of moment conditions, shows more of a gain in efficiency for all of sample sizes we consider.

Our simulation results show that the augmented GMM estimators that use a large number of moment conditions (GMM1 and GMM3) can be biased in finite samples, even when the sample size is rather large (e.g. $T = 2,000$), and inference based on the asymptotic distribution can be inaccurate. For example, tests suffer serious size distortions. GMM2 does not suffer from these problems. One possible reason for the poor finite-sample performance of GMM1 and GMM3 is that some of the moment conditions based on the autocorrelations of the squared data may be "irrelevant" (or nearly irrelevant), as discussed by Hall and Peixe (1999). They show that including irrelevant moment conditions may lead to bias and size distortions.

We give an empirical application to the DM/$ exchange rate to illustrate the use of the augmented GMM estimators and to compare the results with those of NQMLE. GMM2 is not very different from NQMLE, while for GMM1 and GMM3 the estimate of α is somewhat smaller and the standard errors are smaller.

A promising line of future research would be to investigate the use of moment conditions from rescaled asymmetric distributions. This is motivated by the fact that many financial data are asymmetric as well as fat-tailed.

BIBLIOGAPHY

1. Ahn, Seung C. and P. Schmidt (1995), "A Separability Result for GLS Prediction and Conditional Moment Tests," *Econometric Reviews*, 14, 19-34.

2. Andersen, T. G. and B.E. Sorensen (1996), "GMM Estimation of a Stochastic Volatility Model: a Monte Carlo Study," *Journal of Business and Economic Statistics*, 14, 328-352.

3. Attanasio, O. P. (1991), "Risk, Time-Varying Second Moments and Market Efficiency," *Review of Economic Studies*, 58, 479-494.

4. Baillie, R. T. and T. Bollerslev (1989), "The Message in Daily Exchange Rates: A Conditional-Variance Tale," *Journal of Business and Economic Statistics*, 7, 297-305.

5. Baillie, R. T. and T. Bollerslev (1990), "Intra-Day and Inter-Market Volatility in Foreign Exchange Rates," *Review of Economic Studies*, 58, 565-585.

6. Baillie, R. T. and H.M. Chung (1999), "Estimation of GARCH Models from the Autocorrelations of the Squares of a Process," Working Paper, Michigan State University.

7. Black, F. (1976), "Studies in Stock Price Volatility Changes," American Statistical Association, 177-181.

8. Bodurtha, J. N. Jr. and N. C. Mark (1991), "Testing the CAPM with Time-Varying Risks and Returns," *Journal of Finance*, 46, 1485-1505.

9. Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

10. Bollerslev, T. (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *The Review of Economics and Statistics*, 69, 542-547.

11. Bollerslev, T (1988), "On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process," *Journal of Time Series Analysis*, 9, 121-131.

12. Bollerslev, T. and R. Y. Chou, and K.F. Kroner (1992), "ARCH Modeling in Finance," *Journal of Econometrics*, 52, 5-59.

13. Bollerslev, T., R. F. Engle, and D.B. Nelson (1994), "ARCH Models," in : R.F. Engle and D.L. McFadden , eds., *Handbook of Econometrics*, Vol IV. Amsterdam: North-Holland, 2959-3038.

14. Bollerslev, T. and J. M. Wooldridge (1992), "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances," *Econometric Reviews* , 11, 143-172.

15. Bound, J., D. A. Jaeger, and R. M. Baker (1995), "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90, 443-450.

16. Breusch T., H. Qian, P. Schmidt, and D.J. Wyhowski (1999), "Redundancy of Moment Conditions," *Journal of Econometics*, 91, 89-111.

17. Chiang, C. L. (1956), "On Regular Best Asymptotically Normal Estimates," *Annals of Mathematical Statistics*, 27, 336-351.

18. Diebold, F. X. (1986), *Empirical Modeling of exchange Rate Dynamics*, Springer Verlag.

19. Ding, Z. and C. W. J. Granger (1996), "Modeling Volatility Persistence of Speculative Returns: A New Approach," *Journal of Econometrics*, 73, 185-215.

20. Engle, R. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987-1008.

21. Engle, R. F. and T. Bollerslev (1986), "Modelling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1-50.

22. Engle, R. F. and G. Gonzalez-Rivera (1991), "Semiparametric ARCH Models," *Journal of Business and Economic Statistics*, 9, 345-359.

23. Engle, R. F., P. M. Lilien and R. P. Robins (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55, 391-407.

24. Fama, E. F. (1965), "The Behavior of Stock Market Prices," *Journal of Business*, 38, 34-105.

25. Ferguson, T. S. (1958), "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities," *Annals of Mathematical Statistics*, 29, 1046-1062.

26. French, K. R., F. W. Schwert and R. F. Stambaugh (1987), "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3-30.

27. Frideman, B. M. and K. Kuttner (1988), "Time-Varying Risk Perceptions and the Pricing of Risky Assets," National Bureau of Economic Research Working Paper: 2694.

28. Geweke, J. (1986), "Modeling the Persistence of Conditional Variances: Comment," *Econometric Reviews*, 5, 57-61.

29. Glosten, L. R., R. Jagannathan and D. E. Runkle, (1993), "On the Relation between the Expected Value and the Volatility of the Nominal Excess Returns on Stocks," *Journal of Finance*, 48, 1779-1801.

30. Hall, A. R. and F. P. M. Peixe (1999), "A Consistent Method for the Selection of Relevant Instruments," *working paper.*

31. Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

32. Hansen, L. P., J. Heaton and a. Yaron (1996), "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262-280.

33. Hendry D. F. (1984), "Monte Carlo Experimentation in Econometrics," *Handbook of Econometrics*, Vol II. Amsterdam: North-Holland, 939-976.

34. Higgins, M. L., A. K. Bera (1992), "A Class of Nonlinear ARCH Models," *International Economic Review*, 33, 137-158.

35. Hong, C. H. (1987), "The Integrated Generalized Autoregressive Conditional Heteroskedastic Model: The Process Estimation and Some Monte Carlo Experiments," Unpublished manuscript, University of California, San Diego, Dept. of Economics.

36. Hsieh, D. A. (1988), "The Statistical Properties of Daily Foreign Exchange Rates," *Journal of International Economics*, 24, 129-145.

37. 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 & Economic Statistics* , 13, 1-10.

38. Lumsdaine, R. L. (1996), "Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models," *Econometrica*, 64, 575-596.

39. Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394-419.

40. Nelson, C. R. and R. Startz (1990), "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967-976.

41. Nelson, C. R. and R. Startz (1990), "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One," *Journal of Business*, 63, s125-s140.

42. Nelson, D. B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347-370.

43. Newey, W. K. (1985), "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229-256.

44. Newey, W. K. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in : R.F. Engle and D.L. McFadden , eds., *Handbook of Econometrics*, Vol IV. Amsterdam: North-Holland, 2113-2241.

45. Newey, W. K. and D. G. Steigerwald (1997), "Asymptotic Bias for Quasi-Maximum-Likelihood Estimators in Conditional Heteroskedasticity Models," *Econometrica*, 65, 587-599.

46. Newey, W. K. and K. D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

47. Newey, W. K. and K. D. West (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631-653.

48. Ogaki, M. (1993), "Generalized Method of Moments : Econometric Applications," *Handbook of Statistics*, 11, 455-482.

49. Pantula, S. G. (1986), "Modeling the Persistence of Conditional Variances: Comment," *Econometric Reviews*, 5, 71-74.

50. Patel, J. K., C. H. Kapadia and D. B. Owen (1976), *Handbook of Statistical Distributions*, Marcel Dekker.

51. Phillips, P. C. B. (1987), "Time Series Regression With a Unit Root," *Econometrica*, 55, 277-310.

52. Phillips, P. C. B. and Perron, P. (1985), "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335-346.

53. Schwert, G. W. (1989), "Heteroskedasticity is Stock Returns," it National Bureau of Economic Research Working Paper: 2956, 20.

54. Staiger, D. and J. H. Stock (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.

55. Taylor, S. J. (1986), *Modelling Financial Time Series*, Wiley.

56. Theil, H (1971), *Principles of Econometrics*, John Wiley & Sons.

57. Weiss, A. A. (1986), "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory*, 2, 107-131.

58. White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.

59. White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press.

60. Wooldridge, J.M. (1994), "Estimation and Inference for Dependent Processes," in : R.F. Engle and D.L. McFadden , eds., *Handbook of Econometrics*, Vol IV. Amsterdam: North-Holland, 2639-2738.

61. Zakoian, J. M. (1990), "Threshold heteroskedastic model," Unpublished Manuscript, INSEE.