Alternative estimators of the covariance matrix in GARCH models

Calzolari, Giorgio and Fiorentini, Gabriele and Panattoni, Lorenzo

Universita' di Firenze, Italy., European University Institute, Firenze, Italy., IBM Semea, Livorno, Italy

1993

Giorgio Calzolari
Gabriele Fiorentini
Lorenzo Panattoni

ALTERNATIVE ESTIMATORS OF THE COVARIANCE MATRIX IN *GARCH* MODELS

Alternative Estimators of the Covariance Matrix in GARCH Models

Giorgio Calzolari Università di Firenze Firenze, Italy

Gabriele Fiorentini European University Institute, Firenze, Italy

Lorenzo Panattoni IBM Semea Livorno, Italy

Abstract

With most of the available software packages, estimates of the parameter covariance matrix in a GARCH model are usually obtained from the outer products of the first derivatives of the log-likelihoods (BHHH estimator). However, other estimators could be defined and used, analogous to the covariance matrix estimators in maximum likelihood studies described in the literature for other types of models (linear regression model, linear and nonlinear simultaneous equations, Probit and Tobit models). These alternative estimators can be derived from: (1) the Hessian (observed information), (2) the estimated information (expected Hessian), (3) a mixture of Hessian and outer products matrix (White's QML covariance matrix). Signifacant differences among these estimates can be interpreted as an indication of misspecification, or can be due to systematic inequalities between alternative estimators in small samples. Unlike other types of models, from our Monte Carlo study we do not encounter very large differences, presumably because GARCH estimation is usually applied when the sample size is rather large. However, analogously to other types of models we find in this Monte Carlo study that, even in absence of misspecification, the sign of the differences between some estimators is almost systematic. This suggests that, as for other types of models, the choice of the covariance estimator is not neutral, but the results of hypotheses testing are not strongly affected by such a choice. [1]

[1] For presentation at the *European Meeting of the Econometric Society*, Uppsala, 22-26 August 1993. The authors are indebted to Lucia Buzzigoli, Giampiero M. Gallo and Barbara Pacini for discussions and comments. Financial support from MURST 60% and CNR funds is gratefully acknowledged.

1 *Introduction*

The number of theoretical and applied works on Autoregressive Conditional Heteroskedasticity (ARCH) has rapidly grown after Engle (1982).

The most popular among all processes derived from ARCH seems to be the Generalized Autoregressive Conditional Heteroskedastic process (GARCH) proposed by Bollerslev (1986) and discussed in Engle and Bollerslev (1986). It was introduced mainly to allow for a more flexible lag structure with respect to the ARCH specification. As Bollerslev (1986, p.308) points out, the "extension of the ARCH process to the GARCH process bears much resemblance to the extension of the standard time series AR process to the general ARMA process and (....) permits a more parsimonious description".

Engle (1982) proposed an efficient two-step procedure for estimation of ARCH models. However, GARCH estimation is usually performed with maximum likelihood, assuming a conditional normal or a conditional Student-t distribution of the error terms. In this context, maximum likelihood estimation is usually performed by iterating to convergence a Newton-like algorithm where the Hessian matrix is replaced by the (asymptotically equivalent) matrix of outer products of the first order derivatives of the log-likelihoods. The method became popular in the econometric literature after it was proposed for systems of simultaneous equations in the well known paper by Berndt, Hall, Hall and Hausman (1974) and is usually referred to as the BHHH method. Application of this procedure is widely exemplified in the recent literature on conditional heteroskedasticity. Examples are the papers by Thomas (1991, p.10), Lamoureaux and Lastrapes (1990, p.227), Baillie and Myers (1991, p.116), Bollerslev, Engle and Wooldridge (1988), Engle, Lilien and Robins (1987), Nelson (1991), Baillie and De Gennaro (1990, p.208), Bollerslev (1987, p.544), Chou (1988, p.291), Baillie and Bollerslev (1989, p.300), De Santis and Sbordone (1990, p.6), Buzzigoli (1992); see also the papers in Engle and Rothschild (1992).

Many of the above and other applications have been performed using the excellent software developed by Tim Bollerslev and Ken Kroner, based on the BHHH algorithm with numerical computation of the first order derivatives. The main reason for using an algorithm like BHHH is that it does not require computation of derivatives beyond the first order, and the main reason for calculating derivatives numerically is the complexity of analytical derivatives in the GARCH context (see Engle and Bollerslev, 1986, p.25).

A natural consequence is that in hypotheses testing the inverse of the outer products matrix is used to estimate variances and covariances of the equation and GARCH errors parameters.

Calzolari and Fiorentini (1992) calculate analytical second order derivatives and investigate the computational benefits of the analytical Hessian in the maximization procedure. They also derive formulae for the estimated information matrix (expected Hessian). Their formulae will be used in this paper to estimate parameter variances and covariances in several different but asymptotically equivalent ways. Equivalence rests upon the property that, under correct specification of the model and suitable regularity conditions, all these covariance estimators asymptotically give the inverse of Fisher's information matrix.

The inverse of the matrix of second order derivatives of the log-likelihood (observed Hessian), with minus sign, is an estimator of the asymptotic covariance matrix of all parameters of the model (equation coefficients as well as α's and β's parameters of the GARCH error process). It would be natural to use this covariance estimator if some Newton-like maximization method were employed to calculate maximum likelihood estimates.

The first derivatives of the log-likelihood can be used to build a matrix of outer products, (OP matrix, as in Berndt, Hall, Hall and Hausman, BHHH, 1974), and its inverse can be used to estimate variances and covariances. In most applications of the literature it is customary to take advantage of the block-diagonal structure of the information matrix, and therefore the off-diagonal blocks of the matrix are set to zero. Of course we may also calculate the full matrix of outer products, including the two off-diagonal blocks.

Both matrices will be computed and used in this Monte Carlo study.

Calzolari and Fiorentini (1992) also perform maximum likelihood estimation using the method of scoring, thus employing the inverse of the estimated information matrix (expected Hessian). Also this matrix (which is block-diagonal) can be used to estimate variances and covariances of the parameters of the model.

Not associated with a particular maximization algorithm is the quasi maximum likelihood (QML) covariance matrix estimator, the use of which has become more and more popular in the last few years, after White (1982,1983) and Gourieroux et al. (1984). It gives the covariance matrix of the parameters when the error distribution process is not correctly specified (misspecification consistent). Its computation requires both matrices of second order derivatives and of outer products of first derivatives. As the block-diagonality of the information matrix depends on some particular features of the error process (like, for instance, zero third order moment), it seemed more sensible to use Hessian and OP matrices in full form, to construct this QML estimator. Of course, under correct specification, also this estimator is equivalent to the others, as it gives asymptotically the inverse of the Fisher's information matrix (thus, block-diagonal).

Although perfectly aware that these covariance estimators are equivalent only for large samples, we would probably expect that even for a small sample all groups of results had to be sufficiently close to one another, specially because all the matrices would in any case be computed at the same parameters values. In large samples, significant differences could be interpreted as an indication of misspecification (e.g. White 1982). For small samples, this is not necessarily true, as several Monte Carlo studies have shown for other types of models: the Probit model (Griffiths et al. 1987), linear and nonlinear simultaneous equations (Calzolari and Panattoni 1988a, 1988b), the linear regression model (Parks and Savin, 1990), the Tobit model (Calzolari and Fiorentini, 1993).

In this paper we first summarize the analytical formulae for the first and second order derivatives of the GARCH log-likelihood with respect to all parameters, presented in Calzolari and Fiorentini (1992) (first derivatives are also given in Bollerslev, 1986). We then use these derivatives to build the Hessian matrix, the outer products matrices, the estimated information matrix and the QML matrix. Finally, with a wide set of Monte Carlo experiments on three models, we compare these different estimators of the covariance matrix.

2 The GARCH (p,q) model, log-likelihood, first and second derivatives

In this section we summarize the formulae derived in Bollerslev (1986) for the first order, and in Calzolari and Fiorentini (1992) for the second order derivatives of the log-likelihood. We represent the GARCH (p,q) model as

[1] $$y_t = x_t' b + \epsilon_t$$

[2] $$\epsilon_t \mid \mathcal{I}_{t-1} \sim N(0; h_t)$$

[3] $$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} = z_t' \omega$$

where

y_t is the endogenous variable;
x_t is a $k \times 1$ vector of weakly exogenous variables in the sense of Engle, Hendry and Richard (1983);
b is a $k \times 1$ vector of unknown coefficients;
ϵ_t is a conditionally normal disturbance;
$\mathcal{I}_t$ is the information set;
z_t' $= (1, \epsilon_{t-1}^2, ..., \epsilon_{t-q}^2, h_{t-1}, ..., h_{t-p})$;
ω $= (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)'$ is the $(1+q+p) \times 1$ vector of unknown variance parameters.
θ $=(b', \omega')'$ is the $(k+1+q+p) \times 1$ vector of all unknown parameters.

Applying Schweppe's (1965) prediction error decomposition form of the likelihood function, the log-likelihood is, apart from some constant

[4] $$L_T(\theta) = \sum_{t=1}^{T} l_t(\theta)$$

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

For the pre-sample values of h_t and ϵ_t^2 we take an estimate of the unconditional expectation, so if $t \leq 0$

[6] $$h_t = \epsilon_t^2 = \frac{1}{T} \sum_{s=1}^{T} \hat{\epsilon}_s^2$$

where $\hat{\epsilon}_s^2$ are consistently estimated residuals.

Differentiating with respect to the variance parameters ω we get

[7] $$\frac{\partial l_t}{\partial \omega} = -\frac{1}{2} \frac{1}{h_t} \frac{\partial h_t}{\partial \omega} + \frac{1}{2} \frac{\partial h_t}{\partial \omega} \frac{\epsilon_t^2}{h_t^2} = \frac{1}{2} \frac{1}{h_t} \frac{\partial h_t}{\partial \omega} \left[\frac{\epsilon_t^2}{h_t} - 1 \right]$$

where

[8] $$\frac{\partial h_t}{\partial \omega} = z_t + \sum_{i=1}^{p} \beta_i \frac{\partial h_{t-i}}{\partial \omega}$$

moreover

[9] $$\frac{\partial}{\partial \omega} \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t^2 \right) = 0$$

therefore if $t \leq 0$

[10] $$\frac{\partial h_t}{\partial \omega} = 0$$

and this allows to calculate recursively derivatives of equation (8). Differentiating with respect to coefficients we get

[11] $$\frac{\partial l_t}{\partial b} = -\frac{1}{2} \frac{1}{h_t} \frac{\partial h_t}{\partial b} + \frac{\epsilon_t x_t}{h_t} + \frac{1}{2} \frac{\partial h_t}{\partial b} \frac{\epsilon_t^2}{h_t^2} = \frac{\epsilon_t x_t}{h_t} + \frac{1}{2} \frac{1}{h_t} \frac{\partial h_t}{\partial b} \left[\frac{\epsilon_t^2}{h_t} - 1 \right]$$

recalling that for $t \leq 0$

$$[12] \quad \frac{\partial \epsilon_t^2}{\partial b} = \frac{\partial h_t}{\partial b} = \frac{\partial}{\partial \omega}\left(\frac{1}{T}\sum_{s=1}^{T}\hat{\epsilon}_s^2\right) = \frac{1}{T}\sum_{s=1}^{T} -2\epsilon_s x_s = -\frac{2}{T}\sum_{s=1}^{T}\epsilon_s x_s$$

while for $t > 0$

$$[13] \quad \frac{\partial \epsilon_t^2}{\partial b} = -2\epsilon_t x_t$$

Apart from missprints, the equations above are the same as in Bollerslev (1986). Note that the expression given in equation (12) vanishes only asymptotically in a GARCH estimation (while it would be identically zero in OLS), so we have included its value in our computations.

$$[14] \quad \frac{\partial h_t}{\partial b} = -2\sum_{i=1}^{q}\alpha_i (x_{t-i}\epsilon_{t-i})^{I_{t-i}}\left[\frac{1}{T}\sum_{s=1}^{T}\epsilon_s x_s\right]^{(1-I_{t-i})}$$

$$+\sum_{j=1}^{p}\beta_j \frac{\partial h_{t-i}}{\partial b} \qquad \left[\begin{array}{ll} I_{t-i}=1 & \text{if } t-i>0 \\ I_{t-i}=0 & \text{if } t-i\leq 0\end{array}\right.$$

which is slightly different from Bollerslev (1986, p.316, eq.24), the difference being confined to the first q time periods. This makes our computations exactly comparable with those based on numerical first order differentiation, which automatically accounts for a non-zero value in equation (12). Equation (2) allows us to compute recursively derivatives with respect to coefficients recalling that pre-sample values of h_t, ϵ_t^2 and their derivatives are given in equations (6) and (12). Further differentiation gives the terms to build the Hessian matrix. Note that, although the information matrix is known to be block diagonal, the efficient implementation of Newton or Newton-like algorithms requires the full Hessian matrix, which is not block-diagonal in small samples. Differentiating with respect to the variance parameters we get

$$[15] \quad \frac{\partial^2 l_t}{\partial\omega\partial\omega'} = \left[\frac{\epsilon_t^2}{h_t}-1\right]\frac{\partial}{\partial\omega'}\left[\frac{1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial\omega}\right] - \frac{1}{2}\frac{1}{h_t^2}\frac{\partial h_t}{\partial\omega}\frac{\partial h_t}{\partial\omega'}\frac{\epsilon_t^2}{h_t}$$

$$= \left[\frac{\epsilon_t^2}{h_t}-1\right]\left[\frac{1}{2}\frac{1}{h_t}\frac{\partial^2 h_t}{\partial\omega\partial\omega'} - \frac{1}{2}\frac{1}{h_t^2}\frac{\partial h_t}{\partial\omega}\frac{\partial h_t}{\partial\omega'}\right] - \frac{1}{2}\frac{1}{h_t^2}\frac{\partial h_t}{\partial\omega}\frac{\partial h_t}{\partial\omega'}\frac{\epsilon_t^2}{h_t}$$

Differentiating with respect to the coefficients we get

$$[16] \quad \frac{\partial^2 l_t}{\partial b\partial b'} = -\frac{x_t x_t'}{h_t} - \frac{1}{h_t^2}\epsilon_t x_t\frac{\partial h_t}{\partial b'} + \left[\frac{\epsilon_t^2}{h_t}-1\right]\frac{\partial}{\partial b'}\left[\frac{1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial b}\right]$$

$$-\frac{1}{h_t^2}\epsilon_t\frac{\partial h_t}{\partial b}x_t' - \frac{1}{2}\frac{1}{h_t^2}\frac{\partial h_t}{\partial b}\frac{\partial h_t}{\partial b'}\frac{\epsilon_t^2}{h_t}$$

where

$$[17] \quad \frac{\partial}{\partial b'}\left[\frac{1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial b}\right] = \frac{1}{2}\frac{1}{h_t}\frac{\partial^2 h_t}{\partial b\partial b'} - \frac{1}{2}\frac{1}{h_t^2}\frac{\partial h_t}{\partial b}\frac{\partial h_t}{\partial b'}$$

The second order mixed derivatives are

$$[18] \quad \frac{\partial^2 l_t}{\partial b\partial\omega'} = -x_t\frac{\partial h_t}{\partial\omega'}\epsilon_t\frac{1}{h_t^2} - \frac{1}{2}\frac{\partial h_t}{\partial b}\frac{\partial h_t}{\partial\omega'}\frac{1}{h_t^2}\left[\frac{\epsilon_t^2}{h_t}-1\right]$$

$$+\frac{1}{2}\frac{1}{h_t}\frac{\partial^2 h_t}{\partial b\partial\omega'}\left[\frac{\epsilon_t^2}{h_t}-1\right] - \frac{1}{2}\frac{\epsilon_t^2}{h_t}\frac{\partial h_t}{\partial b}\frac{\partial h_t}{\partial\omega'}\frac{1}{h_t^2}$$

All the expressions involved in the above equations have been given before, with the exception of the second order derivatives of h_t, which are now derived. The blocks of the Hessian matrix are given in equations (2-2). The second order derivatives of h_t to be inserted into such equations have the following expressions

$$[19] \quad \frac{\partial h_t}{\partial\omega}\frac{\partial h_t}{\partial\omega'} = \frac{\partial z_t}{\partial\omega'} + \sum_{i=1}^{p}\frac{\partial h_{t-i}}{\partial\omega}\frac{\partial\beta_i}{\partial\omega'}\sum_{i=1}^{p}\beta_i\frac{\partial^2 h_{t-i}}{\partial\omega\partial\omega'}$$

$$= A + \sum_{i=1}^{p}B_i + \sum_{i=1}^{p}C_i$$

A is the $(1+q+p)\times(1+q+p)$ matrix

$$[20] \quad A = \frac{\partial\left[1,\epsilon_{t-1}^2,\ldots,h_{t-1},\ldots,h_{t-p}\right]'}{\partial\left[\alpha_0,\alpha_1,\ldots,\beta_1,\ldots,\beta_p\right]}$$

$$= \begin{bmatrix} 0 & 0 & \vdots & 0 & 0 & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \vdots & 0 & 0 & \vdots & 0 \\ \frac{\partial h_{t-1}}{\partial \alpha_0} & \frac{\partial h_{t-1}}{\partial \alpha_1} & \vdots & \frac{\partial h_{t-1}}{\partial \alpha_q} & \frac{\partial h_{t-1}}{\partial \beta_1} & \vdots & \frac{\partial h_{t-1}}{\partial \beta_p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial h_{t-p}}{\partial \alpha_0} & \frac{\partial h_{t-p}}{\partial \alpha_1} & \vdots & \frac{\partial h_{t-p}}{\partial \alpha_q} & \frac{\partial h_{t-p}}{\partial \beta_1} & \vdots & \frac{\partial h_{t-p}}{\partial \beta_p} \end{bmatrix}$$

the number of zero rows being $1 + q$.

B_i is the $(1 + q + p) \times (1 + q + p)$ matrix

[21] $$B_i = \frac{\partial h_{t-i}}{\partial \omega}\frac{\partial \beta i}{\partial \omega'} \qquad \frac{\partial \beta i}{\partial \omega'} = (0, 0, 0, \ldots, \underset{\substack{\uparrow \\ (1+q+i)th\ element}}{1}, \ldots, 0)$$

[22] $$B_i' = \begin{bmatrix} 0 & 0 & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \vdots & 0 \\ \frac{\partial h_{t-i}}{\partial \alpha_0} & \frac{\partial h_{t-i}}{\partial \alpha_1} & \vdots & \frac{\partial h_{t-i}}{\partial \beta_p} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \vdots & 0 \end{bmatrix}$$

so that

[23] $$\sum_{i=1}^{p} B_i' = A$$

The matrices C_i can simply be calculated recursively from the same equation (2) recalling that for the pre-sample values $\partial^2 h_t / \partial\omega\partial\omega' = 0$ for $t \leq 0$. Also recalling how the pre-sample values for ϵ_t^2 and h_t were computed for $t \leq 0$, we have

[24] $$\frac{\partial^2 h_t}{\partial b \partial b'} = \frac{2}{T}\sum_{s=1}^{T} x_s x_s' \qquad [for\ t \leq 0]$$

so we can calculate recursively

[25] $$\frac{\partial^2 h_t}{\partial b \partial b'} = 2\sum_{i=1}^{q} \alpha_i (x_{t-i}x_{t-i}')^{I_{t-i}} \left[\frac{1}{T}\sum_{s=1}^{T} x_s x_s'\right]^{(1-I_{t-i})} + \sum_{j=1}^{p} \beta_j \frac{\partial^2 h_{t-j}}{\partial b \partial b'} \quad \left[\begin{array}{ll} I_{t-i} = 1 & \text{if } t - i > 0 \\ I_{t-i} = 0 & \text{if } t - i \leq 0 \end{array}\right.$$

and

[26] $$\frac{\partial^2 h_t}{\partial \omega \partial b'} = \frac{\partial z_t}{\partial b'} + \sum_{i=1}^{p} \beta_i \frac{\partial^2 h_{t-i}}{\partial \omega \partial b'}$$

where

[27] $$\frac{\partial z_t}{\partial b'} = \begin{bmatrix} 0 \\ -\epsilon_{t-1}x_{t-1}' \\ \vdots \\ -\epsilon_{t-q}x_{t-q}' \\ \frac{\partial h_{t-1}}{\partial b'} \\ \vdots \\ \frac{\partial h_{t-p}}{\partial b'} \end{bmatrix}$$

These last formulae can be inserted into equations (2), (2) and (2) to produce diagonal and off-diagonal blocks of the Hessian matrix.

The information matrix is block diagonal (see Bollerslev, 1986, p.316), and the two diagonal blocks are estimated by the sample analogues of

[28] $$E\left[\frac{\partial^2 l_t}{\partial \omega \partial \omega'} \mid \mathcal{I}_{t-1}\right] = -\frac{1}{2}\frac{1}{h_t^2}\frac{\partial h_t}{\partial \omega}\frac{\partial h_t}{\partial \omega'}$$

[29] $$E\left[\frac{\partial^2 l_t}{\partial b \partial b'} \mid \mathcal{I}_{t-1}\right] = -\frac{x_t x_t'}{h_t} - \frac{1}{2}\frac{1}{h_t^2}\frac{\partial h_t}{\partial b}\frac{\partial h_t}{\partial b'}.$$

The matrix of outer products of the first derivatives used in the BHHH algorithm on GARCH models is usually taken block-diagonal, the first block being $(k \times k)$

$$\sum_{t=1}^{T} \frac{\partial l_t}{\partial b} \frac{\partial l_t}{\partial b'} \tag{30}$$

while the last $(1+q+p) \times (1+q+p)$ block is

$$\sum_{t=1}^{T} \frac{\partial l_t}{\partial \omega} \frac{\partial l_t}{\partial \omega'} \tag{31}$$

where derivatives are given in equations (7) and (11) Of course, the same derivatives can be used to build a full matrix, including the off-diagonal blocks that would vanish only asymptotically. Even if this matrix is usually taken in block-diagonal form in the literature, experiments in this paper have been performed with both types of matrices.

The last matrix used to estimate the variances and covariances of coefficients and GARCH parameters is the Whyte-type or QML matrix (quasi maximum likelihood, White, 1982; see also Gourieroux, Monfort and Trognon, 1984)

$$\hat{\Psi} = (H^{-1})(OP)(H^{-1}) \tag{32}$$

For the experiments of this paper the two matrices needed to build $\hat{\Psi}$ have been used in their full form, including the off-diagonal blocks.

3 Models, simulation experiments, results

Here we describe the settings of the simulation experiments and the models employed. We then report and comment our findings.

3.1 Design of the Monte Carlo experiments

A wide set of Monte Carlo experiments have been performed on several models. For each model we start from a given vector of *true* parameters held fixed over all replications, and fix a sample period lenght. Explanatory exogenous variables have been kept fixed

at their historical values and for longer samples have been repeated consecutively. Only model 3 has exogenous variables (apart from the constant term). We have also performed experiments with randomly generated exogenous variables, in particular with large kurtosis. In similar studies for other types of models (e.g. Calzolari and Panattoni 1988b, Parks and Savin 1990) a strong leptokurtic design of the regressors has in fact been identified as a *critical* condition and it has been found to seriously affect the small sample performances of the alternative covariances estimators. Analogous critical condition has been identified by Chesher (1989) in the unbalanced design of the regression matrix.

Independently of the explanatory exogenous variables, we then generate the normal (0,1) random deviates that, using the *true* GARCH α's and β's parameters, provide the disturbance terms, ϵ_t, over the sample period. Values of the endogenous variable are finally computed with simulation. The generation of the random error terms with the assumed GARCH structure gets rid of misspecification problems. With the generated data we now estimate the coefficients of the equation and GARCH parameters by maximum likelihood. For the pre-sample values of ϵ_t^2 and h_t we take their unconditional mean computed using in-sample residuals. Finally we compute, at the values that maximize the likelihood, the five estimates of the covariance matrix: estimated information, Hessian, full outer products matrix, block-diagonal outer products matrix, QML matrix.

3.2 The Models

The first model experimented with is the simple random-walk with drift model for the exchange rate Deutsche mark - U.S. dollar with weekly data discussed in Baillie and Bollerslev (1989, table 5).

MODEL 1

$$100\Delta \log s_t = b_1 + \epsilon_t \tag{33}$$

where an ARCH(1) and a GARCH (1,1) specification have been adopted for the error process. In tables 1-4 results are related to experiments with sample periods varying from 100 to 400 observations.

Also the second model has a very simple specification (Buzzigoli, 1992)

MODEL 2

$$[34] \qquad 100\Delta \log y_t = b_1 100\Delta \log y_{t-1} + \epsilon_t$$

Data are daily observations of prices for the Olivetti equities at the Milan Stock Exchange. In tables 5 and 6 results are related to experiments with samples of 100 and 400 observations respectively, and an ARCH(1) error process. In tables 7 and 8 the sample period lengths are 150 and 500, and a GARCH(1,1) process is adopted.

The third model experimented with is the monthly model of long term interest rates used in Bianchi, Calzolari and Sterbenz (1991). In this model the U.S. interest rate depends upon the money supply, the inflation rate as measured by the consumer price index, and the unemployment rate. The specification is given as

MODEL 3

$$[35] \qquad R_t = b_1 + b_2 M_t + b_3 I_t + b_4 U_t + \epsilon_t$$

where R_t is the long term interest rate, M_t is the real money supply, I_t is the inflation rate, U_t is the unemployment rate, and ϵ_t is the error term. Also for this model we first adopt an ARCH(1) specification of the error process, then a GARCH(1,1). Results are summarized in tables 9-12.

3.3 Results on alternative variance estimators

In each case we perform 1000 replications of the Monte Carlo process. Each table displays the *true* value of the parameter, the mean estimates, the Monte Carlo variance and the mean variance estimated with the five methods. The last two columns display the number of times that we found an outer product estimated variance (full OP or

block-diagonal OP matrix) greater than the corresponding Hessian estimate. These are the inequalities that occur more often in our experiments, and are also displayed in graphical form in the final figures of this paper. Figures 3-4, in fact, display for the parameters of the second model the two forms of the outer product estimated variances vis-a-vis the corresponding Hessian estimates.

The results we find are quite surprising since we conducted the experiment bearing in mind the evidence of former studies on other types of models. The five alternative estimators give quite similar standard errors estimates On the contrary, similar studies on Probit, Tobit, simultaneous equations and linear regression models showed that especially the estimators based on the variance of the score (OP) diverged substantially from the other alternative, yet asymptotically equivalent, estimators. This seems not to be the case for GARCH models.

The estimates computed with the observed Hessian and the estimated information are quite the same. The first is on average a little bigger than the second. Griffiths et al. (1987) for the Probit models case and Calzolari and Fiorentini (1993) for the Tobit found that these two estimators were practically undistinguishable. On the contrary, Calzolari and Panattoni (1988a) found significant differences, presumably due to the shortness of the sample periods typically used for macro-systems of simultaneous equations.

The estimates based on the outer products of the first derivatives of the log-likelihoods are on average bigger than the others and this agrees with the previous literature. However, differences are usually not very big especially for what it concerns the block-diagonal version of the outer products.

The standard errors of the coefficients based on the block-diagonal form of the outer products are usually greater than the corresponding Hessian estimates half of the times. This is the percentage one would expect having in mind the standard OLS case. This percentage becomes much bigger for those models where the fourth moment of the errors process does not exist. See tables 5, 6, 9 and 10.

Parks and Savin (1990) and Calzolari and Fiorentini (1993) found that the White-type QML standard errors were on average the smallest. We find that in the GARCH case these estimators take on average an intermediate value.

The Monte Carlo MSE is also quite similar to the mean of the five different variance estimators. In a few cases it is smaller than the Monte Carlo average of all the other variance estimators. Calzolari and Fiorentini (1993) found the Monte Carlo MSE to be regularly in between the average of the Hessian based and the outer products based estimators.

3.4 *Wald statistic: a summary figure*

The results displayed in the tables and discussed in the previous section concerned only the estimators of the variances or standard errors of the parameters. We may now wish to consider also the behaviour of the alternative estimators of the covariances. This can be done in a synthetic way combining some or all the parameters errors into a single random variable, like the Wald statistic. Let γ_0 be a vector containing some of the *true* parameters of the model (for example, the equation coefficients only, or α's and β's parameters of the GARCH process), or even containing all the *true* parameters of the model, θ_0. Under the null hypothesis $H_0: \gamma = \gamma_0$ the Wald test statistic $(\hat{\gamma}-\gamma_0)'(\hat{\Psi}_\gamma)^{-1}(\hat{\gamma}-\gamma_0)$ is asymptotically distributed as a χ^2 with a number of degrees of freedom equal to the number of elements in $\hat{\gamma}$; for example, k if $\hat{\gamma}$ is the vector of estimated coefficients $\hat{b}$, $1+q+p$ if $\hat{\gamma}$ contains the estimates of the GARCH variance parameters $\hat{\omega}$, $k+1+q+p$ if $\hat{\gamma}$ contains the estimates of all parameters of the model $\hat{\theta}$.

It is worth to point out here that with our null hypothesis the parameters are well in the interior of the feasible parameter space so that the usual regularity conditions are satisfied. On the other hand we recall that when testing for zero restrictions on the GARCH parameters the one-sided nature of the test should be taken into account as discussed in Demos and Sentana (1991).

In each Monte Carlo replication $(\hat{\gamma}-\gamma_0)$ is the same and what change are only the different estimates of Ψ. Since it is the inverse

of the estimated Ψ that enters the Wald statistic, we should expect a value of the outer products based Wald systematically smaller than the corresponding value computed with the Hessian. Therefore, if we display the c.d.f. of these statistics, the curve related to the outer product matrices should be left-shifted with respect to the Hessian.

As far as the distribution of the QML Wald statistic (misspecification consistent) is concerned, we must recall how the covariance estimator $\hat{\Psi}$ is computed in this case

$$[36] \qquad \hat{\Psi} = (H^{-1})(OP)(H^{-1})$$

If the Hessian estimated covariance matrix (H^{-1}) is smaller than the corresponding outer product estimate (OP^{-1}), the product of matrices resulting from (36) should be even smaller (and therefore its inverse should be larger, and the distribution of the Wald test rightmost shifted). This consideration certainly holds for the diagonal terms of the matrix estimators, as shown in the tables, but does not hold for the whole matrices, as Parks and Savin (1990) showed on a simple linear regression model. Also for linear regression model Calzolari and Panattoni (1988b) showed that the QML estimator of the covariance matrix is in any case downward biased, bias becoming larger and larger as the fourth order moment increases.

These effects are evidenced in figures 1 and 2. These figures are related to model 3, and the Wald test statistics are displayed first for all the model's parameters, then for the coefficients and for the α's and β's parameters separately. In the first part of figures 1 and 2 exogenous regressors are kept at their historical value (as in tables 9 and 10). In the second part of the figures exogenous regressors have been randomly generated with same means and variances as the historical values, but with a very large fourth moment. Of course, the distance among the curves would become completely negligible when the sample period becomes very long. In fact, consistency of the estimators for the asymptotic covariance matrices ensures that, increasing the sample period lenght, the cumulative distribution of the four Wald statistics collapse over the χ^2 c.d.f. (For the OP case we use the full matrix and we display only the curve related to it.)

However, in this movement toward the χ^2 curve, the four sampling distributions maintain their relative positions.

The relative positions of the curves are exactly the same as those found in the previous similar studies. In our case they are much closer than they were found to be for example in Calzolari and Panattoni (1988a, 1988b). Especially the behaviour of the Wald statistic computed with the White-type (QML) matrix is remarkably good (for the historical regressors case) compared with what was found for other models when, in small samples, this curve was much more downward and rightward shifted.

The curves related to the Hessian and estimated information are very close to one-another. The Hessian curve is very slightly left-shifted with respect to the other, and therefore slightly closer to the theoretical χ^2 curve. It is clear, however, that the inequalities with respect to the outer product and White-type (QML) estimators are more relevant in practical applications.

4 Conclusion

We find that on average the Hessian and the estimated information matrix give rather similar results.

On the other hand variances estimated from the full outer product matrix are in most cases slightly larger than variances computed with the Hessian or with the estimated information. When the block-diagonal form of the outer products matrix is used, this inequality still remains in most cases for the α's and β's parameters, but is less evident or even absent for the equation coefficients. The inequality becomes much more evident when exogenous regressors are generated with a large fourth order moment.

The systematic inequalities clearly observed in the experiments have the same sign as for other types of models already analyzed in the literature, showing that the choice of the covariance estimator is not neutral and that hypotheses testing may be affected by such a choice for a great variety of models used in econometric applications. In the GARCH case however the choice of the variance estimator

seems not to have a great impact on hypothesis testing. This results are quite reassuring, mainly because most of the empirical studies with ARCH and GARCH were conducted, at an early stage, using the BHHH estimators for the parameters variance. This estimator had shown to behave quite wildly in many cases for other types of models. It seems not to be so for the type of models we have analyzed in this paper.

Table 1: **Model 1:** mean estim. param, mean estim. var.×100. ARCH(1) T=100.

par.	*True*	Est.	Var.	Inf.	Hes.	OP	b-d.OP	QML	OP>H	b-d.OP>H
b_1	-.294	-.297	.330	.356	.367	.401	.371	.365	77.4%	54.6%
α_0	.286	.293	.436	.468	.514	.621	.596	.485	70.0%	64.0%
α_1	.600	.572	4.11	4.28	4.77	5.63	5.44	4.60	75.6%	70.0%

Table 2: **Model 1:** mean estim. param, mean estim. var.×100. ARCH(1) T=300.

par.	*True*	Est.	Var.	Inf.	Hes.	OP	b-d.OP	QML	OP>H	b-d.OP>H
b_1	-.294	-.296	.116	.118	.119	.125	.121	.118	69.5%	54.8%
α_0	.286	.289	.151	.154	.157	.170	.167	.153	65.3%	60.8%
α_1	.650	.599	1.56	1.57	1.62	1.74	1.71	1.59	75.6%	64.3%

Table 3: **Model 1:** mean estim. param, mean estim. var.×100. GARCH(1,1) T=150.

par.	*True*	Est.	Var.	Inf.	Hes.	OP	b-d.OP	QML	OP>H	b-d.OP>H
b_1	-.294	-.300	.779	.734	.754	.807	.748	.761	74.8%	49.6%
α_0	.286	.353	3.56	5.32	5.04	7.74	7.58	5.15	84.0%	83.4%
α_1	.350	.351	1.79	1.85	1.99	2.51	2.46	1.96	80.4%	78.2%
β_1	.500	.459	2.81	4.25	3.68	6.23	6.11	3.62	82.0%	81.4%

Table 4: **Model 1**: mean estim. param, mean estim. var.×100. GARCH(1,1) T=400.

par.	*True*	Est.	Var.	Inf.	Hes.	OP	b-d.OP	QML	OP>H	b-d.OP>H
b_1	-.294	-.295	.259	.282	.286	.294	.285	.285	72.3%	48.3%
α_0	.286	.317	1.28	1.31	1.38	1.55	1.53	1.54	71.7%	70.7%
α_1	.350	.353	.700	.639	.656	.730	.722	.670	69.7%	69.1%
β_1	.500	.481	1.11	1.10	1.13	1.31	1.30	1.26	71.3%	70.8%

Table 5: **Model 2**: mean estim. param, mean estim. var.×100. ARCH(1) T=100.

par.	*True*	Est.	Var.	Inf.	Hes.	OP	b-d.OP	QML	OP>H	b-d.OP>H
b_1	.203	.196	.914	.749	.793	.952	.886	.811	74.0%	66.4%
α_0	.200	.207	.219	.232	.255	.296	.290	.253	62.6%	60.6%
α_1	.600	.564	4.33	4.23	4.73	5.64	5.64	4.73	74.0%	68.4%

Table 6: **Model 2**: mean estim. param, mean estim. var.×100. ARCH(1) T=400.

par.	*True*	Est.	Var.	Inf.	Hes.	OP	b-d.OP	QML	OP>H	b-d.OP>H
b_1	.203	.204	.233	.210	.215	.229	.223	.217	67.0%	62.3%
α_0	.200	.202	.050	.056	.057	.057	.059	.055	64.0%	61.5%
α_1	.600	.597	1.19	1.10	1.14	1.22	1.19	1.12	68.8%	63.5%

Table 7: **Model 2**: mean estim. param, mean estim. var.×100. GARCH(1,1) T=150.

par.	*True*	Est.	Var.	Inf.	Hes.	OP	b-d.OP	QML	OP>H	b-d.OP>H
b_1	.203	.203	.682	.696	.717	.774	.718	.734	72.2%	51.6%
α_0	.200	.247	1.77	2.96	2.50	4.13	4.01	2.33	82.8%	81.2%
α_1	.350	.344	1.79	1.83	1.97	2.51	2.43	1.95	81.0%	78.2%
β_1	.500	.464	2.72	4.67	3.75	6.53	6.36	3.47	80.8%	79.4%

Table 8: **Model 2**: mean estim. param, mean estim. var.×100. GARCH(1,1) T=500.

par.	*True*	Est.	Var.	Inf.	Hes.	OP	b-d.OP	QML	OP>H	b-d.OP>H
b_1	.203	.205	.226	.222	.224	.230	.224	.225	64.8%	50.1%
α_0	.200	.218	.451	.444	.453	.506	.501	.470	71.7%	69.6%
α_1	.350	.353	.556	.545	.557	.611	.603	.558	69.3%	66.0%
β_1	.500	.484	.931	.832	.838	.961	.951	.870	69.4%	68.2%

Table 9: **Model 3**: mean estim. param, mean estim. var.×100. ARCH(1) T=100.

par.	*True*	Est.	Var.	Inf.	Hes.	OP	b-d.OP	QML	OP>H	b-d.OP>H
b_1	3.19	3.24	141.	104.	110.	133.	126.	117.	77.5%	69.3%
b_2	-.204	-.199	3.80	2.81	2.79	3.56	3.37	3.10	77.6%	68.4%
b_3	.419	.404	3.29	2.66	2.85	3.47	3.28	3.01	77.6%	70.0%
b_4	.775	.771	.149	.120	.129	.154	.145	.143	70.8%	63.4%
α_0	.100	.098	.053	.050	.058	.073	.063	.063	78.6%	59.6%
α_1	.600	.584	4.23	4.25	5.15	5.97	5.27	6.19	74.5%	60.3%

Table 10: **Model 3:** mean estim. param, mean estim. var.×100. ARCH(1) T=400.

par.	*True*	Est.	Var.	Inf.	Hes.	OP	b-d.OP	QML	OP>H	b-d.OP>H
b_1	3.19	3.20	.871	.746	.763	.805	.790	.771	69.4%	61.6%
b_2	-.204	-.203	.039	.036	.037	.040	.039	.037	65.4%	61.7%
b_3	.419	.421	.391	.380	.385	.420	.412	.383	66.7%	62.8%
b_4	.775	.776	.015	.014	.015	.016	.015	.015	64.5%	60.5%
α_0	.100	.099	.012	.014	.014	.015	.015	.014	70.9%	57.1%
α_1	.650	.604	1.21	1.19	1.23	1.33	1.28	1.22	69.8%	61.7%

Table 11: **Model 3:** mean estim. param, mean estim. var.×100. GARCH(1,1) T=150.

par.	*True*	Est.	Var.	Inf.	Hes.	OP	b-d.OP	QML	OP>H	b-d.OP>H
b_1	3.19	3.18	182.	162.	168.	190.	179.	181.	72.5%	61.7%
b_2	-.204	-.201	3.06	2.74	2.86	3.20	3.00	3.07	71.1%	59.9%
b_3	.419	.437	6.22	5.97	6.12	7.21	6.77	6.29	74.5%	64.9%
b_4	.775	.779	.378	.319	.332	.375	.353	.365	67.2%	58.5%
α_0	.100	.120	.371	.554	.545	.847	.769	.658	85.6%	79.5%
α_1	.350	.358	2.04	1.88	2.11	2.64	2.43	2.26	80.7%	72.4%
β_1	.500	.454	2.90	3.88	3.63	6.02	5.53	4.32	83.3%	78.7%

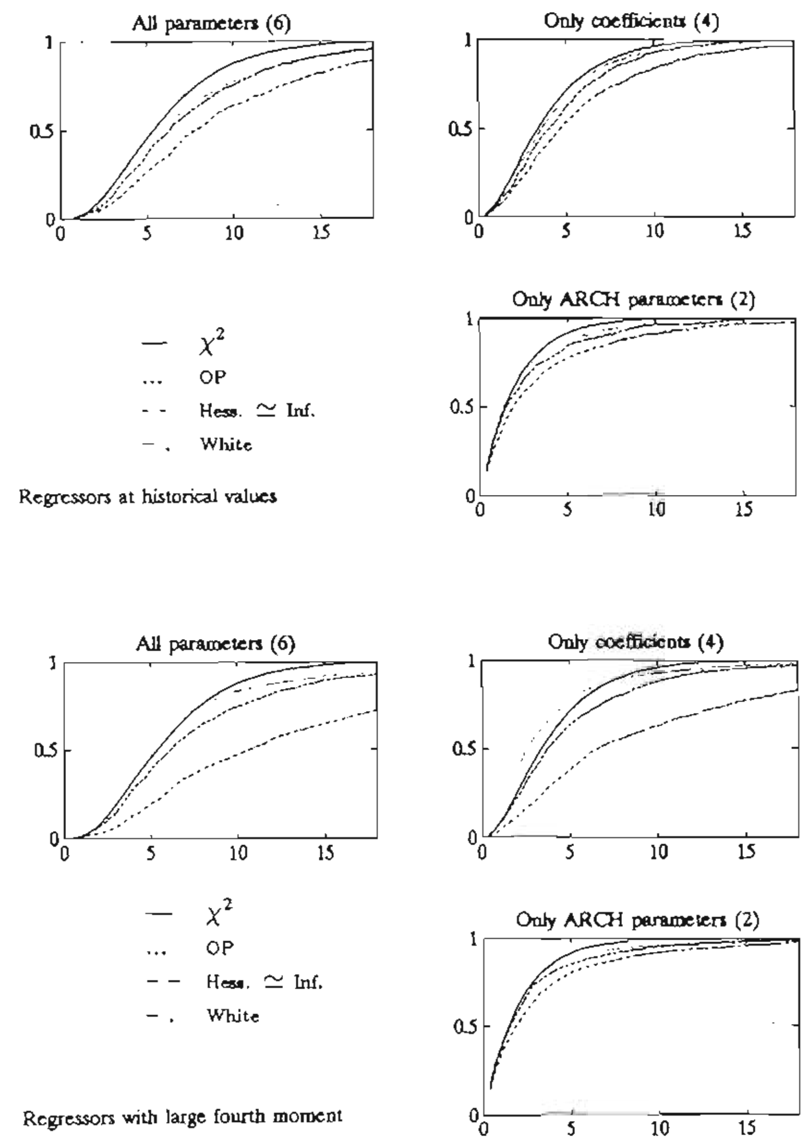

Figure 1: Model 3, ARCH(1), T=100. Cumulated distribution of the Wald Statistic.

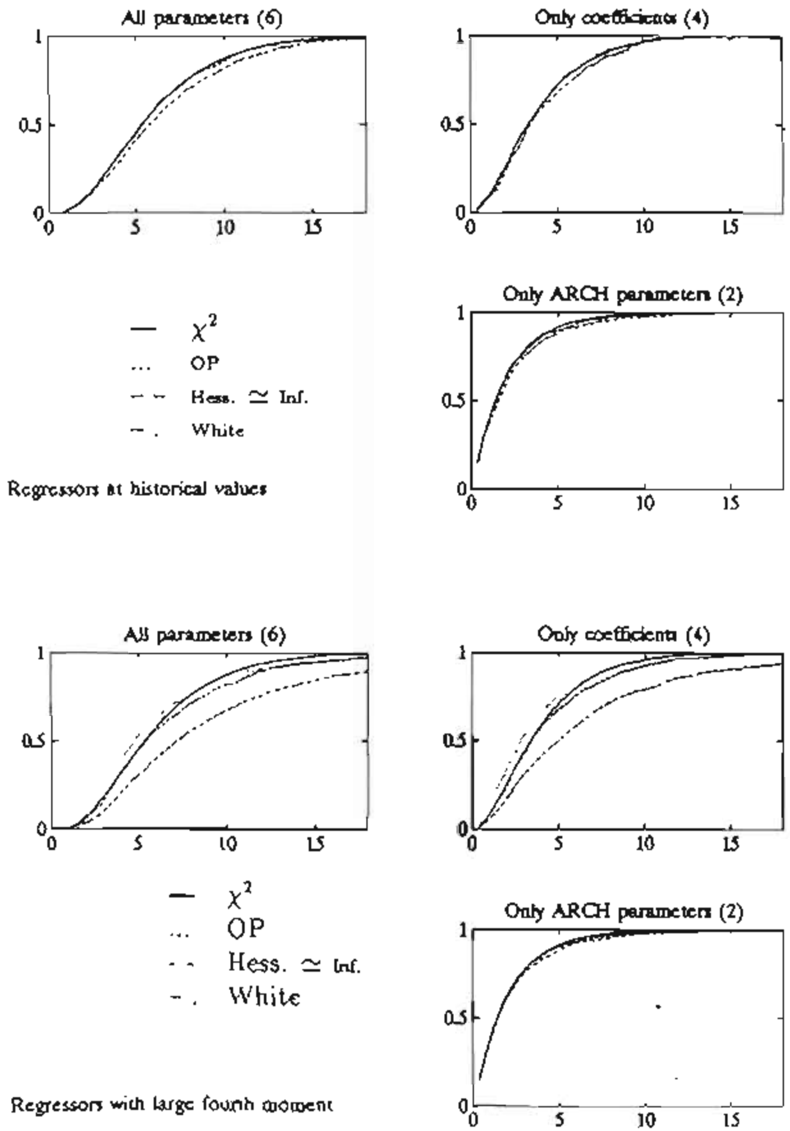

Figure 2: Model 3, ARCH(1), T=400. Cumulated distribution of the Wald Statistic.

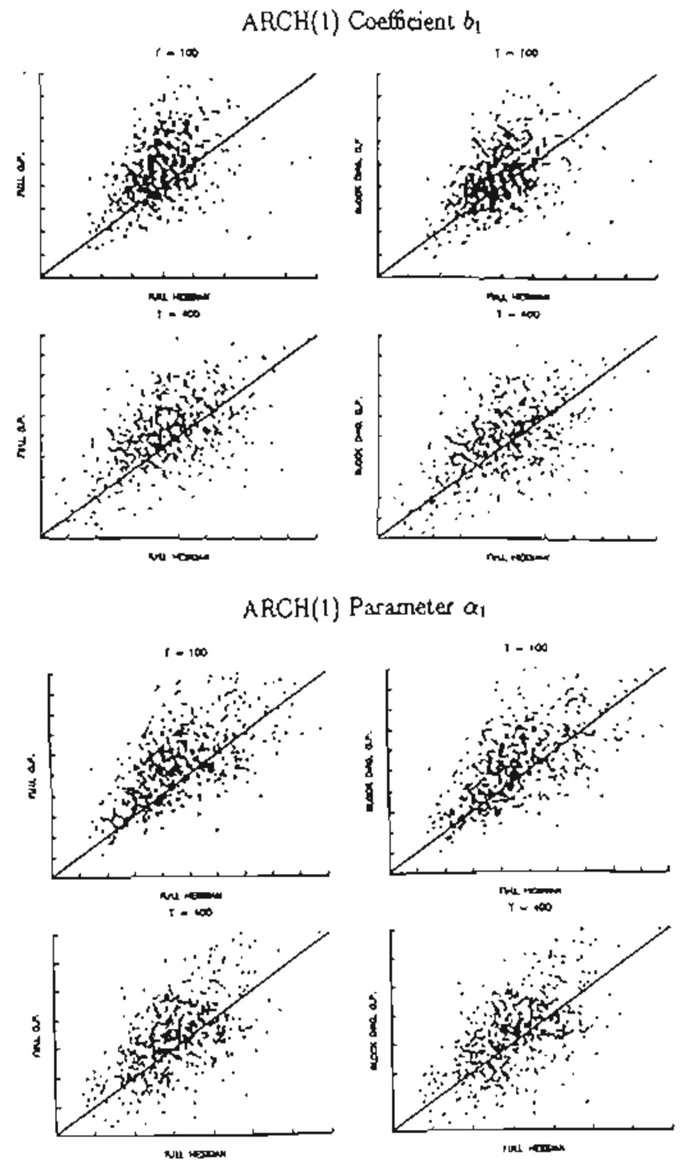

Figure 3: MODEL 2 $100\Delta\log y_t = b_1 100\Delta\log y_{t-1} + \epsilon_t$

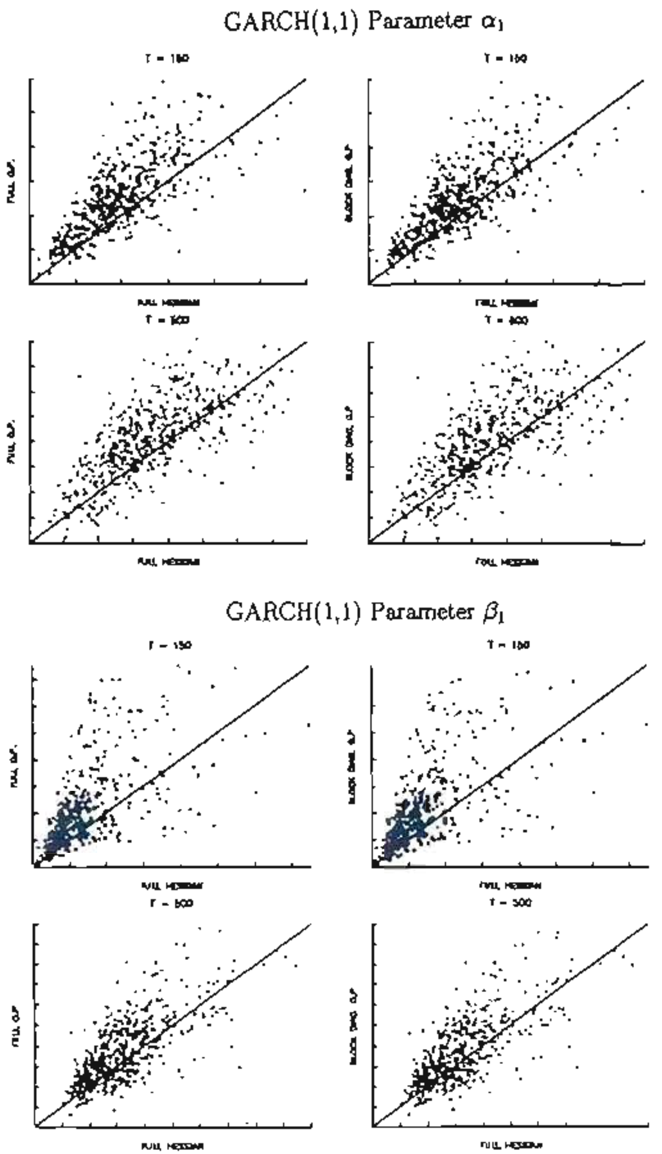

Figure 4: MODEL 2 $100\Delta \log y_t = b_1 100\Delta \log y_{t-1} + \epsilon_t$

Table 12: **Model 3:** mean estim. param, mean estim. var.×100. GARCH(1,1) T=500.

par.	*True*	Est.	Var.	Inf.	Hes.	OP	b-d.OP	QML	OP>H	b-d.OP>H
b_1	3.19	3.19	2.39	2.21	2.23	2.32	2.26	2.29	61.6%	54.6%
b_2	-.204	-.205	.099	.095	.096	.096	.098	.099	58.3%	51.6%
b_3	.419	.420	.766	.748	.757	.757	.775	.773	62.4%	56.4%
b_4	.775	.725	.035	.031	.033	.032	.032	.033	60.9%	53.1%
α_0	.100	.108	.100	.099	.103	.103	.112	.105	69.2%	64.2%
α_1	.350	.357	.598	.548	.570	.570	.605	.588	67.0%	61.0%
β_1	.500	.482	.864	.796	.819	.819	.905	.859	66.4%	62.7%

References

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

Baillie, R. T., and R. P. De Gennaro (1990): "Stock Returns and Volatility", *Journal of Financial and Quantitative Analysis* 25, 203-214.

Baillie, R. T., and R. J. Myers (1991): "Bivariate GARCH Estimation of the Optimal Commodity Futures Hedge", *Journal of Applied Econometrics* 6, 109-124.

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman (1974): "Estimation and Inference in Nonlinear Structural Models", *Annals of Economic and Social Measurement* 3, 653-665.

Bianchi, C., G. Calzolari, and F. P. Sterbenz (1991): "Simulation of Interest Rate Options Using ARCH". Universita' di Messina, discussion paper presented at the *European Meeting of the Econometric Society,* Cambridge, U.K., September 2-6.

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

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

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988): "A Capital Asset Pricing Model with Time-varying Covariances", *Journal of Political Economy* 96, 116-131.

Buzzigoli, L. (1992): "Modelli di Analisi delle Serie Temporali Finanziarie: Rassegna e Applicazioni di Modelli di Tipo ARCH". Universita' di Padova: Dipartimento di Scienze Statistiche, PhD dissertation.

Calzolari, G., and G. Fiorentini (1993, forthcoming): "Alternative Covariance Estimators of the Standard Tobit Model", *Economics Letters*.

Calzolari, G., and G. Fiorentini (1992): "Alternative Methods for GARCH Estimation", Universita' di Firenze: Dipartimento Statistico, Working Paper No. 44, presented at the *European Meeting of the Econometric Society,* Bruxelles, August 24-28.

Calzolari, G., and L. Panattoni (1988a): "Alternative Estimators of FIML Covariance Matrix: A Monte Carlo Study", *Econometrica* 56, 701-714.

Calzolari, G., and L. Panattoni (1988b): "Finite Sample Performance of the Robust Wald Test in Simultaneous Equation Systems", in *Advances in Econometrics,* Vol. 7, ed. by G. F. Rhodes, Jr., and T. B. Fomby. Greenwich, CO: JAI Press Inc., 163-191.

Chesher, A. (1989): "Hájek Inequalities, Measures of Leverage and the Size of Heteroskedasticity Robust Wald Tests", *Econometrica* 57, 971-977.

Chou, R. Y. (1988): "Volatility Persistence and Stock Valuations: Some Empirical Evidence using GARCH", *Journal of Applied Econometrics* 3, 279-294.

Demos, A. and E. Sentana, (1991): "Testing for GARCH effects: A One-sided Approach", discussion paper presented at the *European Meeting of the Econometric Society,* Cambridge, U.K., September 2-6.

De Santis, G., and A. M. Sbordone (1990): "A CAPM with a Multivariate Generalized ARCH Process: An Empirical Analysis of the Italian Financial Market". University of Chicago: Graduate School of Business, discussion paper.

Engle, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflations", *Econometrica* 50, 987-1007.

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

Engle, R. F., D. F. Hendry, and J. F. Richard (1983): "Exogeneity", *Econometrica* 51, 277-304.

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

Engle, R. F., and M. Rothschild, eds. (1992): "ARCH Models in Finance" *Journal of Econometrics* 52, 1-311.

Gourieroux, C., A. Monfort, and A. Trognon (1984): "Pseudo Maximum Likelihood Methods: Theory", *Econometrica* 52, 681-700.

Griffiths, W. E., R. C. Hill, and P. J. Pope (1987): "Small Samples Properties of Probit Model Estimators", *Journal of the American Statistical Association* 82, 929-937.

Lamoureaux, C. G., and W. D. Lastrapes (1990): "Persistence in Variance, Structural Change, and the GARCH Model", *Journal of Business & Economic Statistics* 8, 225-234.

MacKinnon, J. G., and H. White (1985): "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties", *Journal of Econometrics* 29, 305-325.

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

Parks, R. W. and N. E. Savin (1990): "The Choice of Coefficients Covariance Matrix Estimator: Outer Product versus Hessian". The University of Iowa, Department of Economics, working paper presented at the *6th World Congress of the Econometric Society,* Barcelona, August 22-28.

Schweppe, F. (1965): "Evaluation of Likelihood Functions for Gaussian Signals", *IEEE Transactions on Information* 11, 61-70.

Thomas, A. (1991): "Implied Non-linear ARCH Models in Computing Optimal Auction Bids". Toulouse: GREMAQ, discussion paper.

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

White, H. (1983): "Corrigendum", *Econometrica* 51, 513.

Contents