Benchmarks and the accuracy of GARCH model estimation

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

www.reading.ac.uk/centaur

CentAUR

Benchmarks and the Accuracy of GARCH Model Estimation

Chris Brooks, ISMA Centre, University of Reading, UK[1]

Simon P. Burke, Department of Economics, University of Reading, UK

Gita Persand, ISMA Centre, University of Reading, UK

1. Introduction

The class of Generalised Autoregressive Conditionally Heteroscedastic (GARCH) models (See Bollerslev *et al*., 1992 for a comprehensive review) has become enormously popular for application to economic and in particular financial data over the past fifteen years. Such models are considered extremely useful for modelling and forecasting movements in asset return volatilities over time - for example, for use in pricing financial options, or in the context of risk management.

Along with development in these models has come a phenomenal improvement in computational speed and power, so that problems which were previously infeasible even with a supercomputer, can now be overcome with a desktop PC. The popularity of GARCH has also encouraged econometric software developers to incorporate into their programs routines which permit the estimation of these models. At the time of writing, almost all of the major software packages permit the estimation of GARCH-

type models, at least in some form[2]. Indeed, even many packages which are not primarily intended for time series analysis, such as LIMDEP[3] now, or will soon offer the ability to estimate GARCH(1,1) models.

Given the widespread estimation of conditional variance models, and the proliferation of packages that are available, it is perhaps surprising that, at the time of writing, no reviews have been conducted with the aim of determining the accuracy of the resulting coefficient and standard error estimates. Two obvious and related questions to ask are, do all packages give the same answers, and are those answers accurate? Since the outputs from GARCH estimations are often used as inputs to financial risk management systems for example, any software inconsistencies which go undocumented could have potentially disastrous consequences.

The purpose of the present paper is to review a number of the most widely used software packages, with particular reference to their estimation accuracy when judged against a respected benchmark. This work builds upon an earlier paper by McCullough and Renfro (1999), while our innovation is to consider the numerical consistency of GARCH and EGARCH estimation and forecasting using a large number of different econometric software packages. It is also worth stating at this stage the issues which we will not cover - including detailed descriptions of the packages, their user-friendliness, flexibility, speed etc. - for these have been amply reviewed elsewhere (see, for example, Brooks, 1997).

---

[2] One notable exception here being PC-Give.

[3] This review was undertaken on a test copy of a forthcoming version of LIMDEP, which at the time of writing is not commercially available, and which is likely to be called v8.0.

2. Software Accuracy: The Importance of Benchmarks

Financial econometrics often involves the use of large, relatively complex models which require nonlinear estimation techniques. Such models make the production of a benchmark more difficult, but at the same time, all the more important, for the potential for discrepancies between codes and errors in estimation are heightened. A potential candidate as a benchmark for the GARCH-(1,1) model is due to Fiorentini, Calzolari, and Panattoni (1996, hereafter FCP). The FCP benchmark concerns estimation of a GARCH(1,1) model of the form

$$y_t = \mu + \varepsilon_t \quad , \quad \varepsilon_t | \Omega_{t-1} \sim N(0, h_t) \tag{1}$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{2}$$

The conditional likelihood function for this model can be expressed as

$$L(\theta) = \sum_{t=1}^{T} -\frac{1}{2}\ln h_t - \frac{1}{2}\frac{\varepsilon_t^2}{h_t} \tag{3}$$

where θ is the vector of parameters. As it stands, the model is not fully specified, and thus cannot be estimated using the usual conditional maximum likelihood procedure; two further specifications are necessary - starting values for the parameters (μ, α_0, α_1, β_1) and initialisation of the two series ε_t^2 and h_t must also be supplied. Typically, these are done using a set of default initialisations in the packages, so that the user may be unaware that any assumptions concerning plausible values have been made. The usual default solution to parameter initialisations is to set parameter values in the conditional mean equal to those estimated using a "first pass" ordinary least squares regression, and those in the conditional variance equation to zero[4]. A popular initialisation for the conditional variance vector (h_t) is to set each element to the

4

average mean-adjusted sum of squares of the data, i.e. the residual sum of squares from a linear regression of the dependent variable on a constant, divided by the number of observations:

$$h_t = \varepsilon_t^2 = \frac{1}{T}\sum_{j=1}^{T} e_j^2 \qquad\qquad t \leq 0 \qquad\qquad\qquad (4)$$

Many published papers do not even discuss the initalisation of pre-sample values, and hence effectively these studies report the results from estimating an unknown model, which makes replication of their results a difficult task.

Equations (1) to (4) applied to the daily German mark / British pound exchange rate data (comprising 1974 observations) of Bollerslev and Ghysels (1996)[5] constitutes FCP's Model I. The benchmark coefficient estimates, to 6 significant figures, are found to be -0.00619041, 0.0107613, 0.153134, and 0.805974 for $\mu, \alpha_0, \alpha_1, \beta_1$ respectively.

FCP have assisted substantially in the development of a benchmark for GARCH(1,1) models by deriving closed-form analytical expressions for the second-derivatives of the conditional log-likelihood function in (3). This is preferable since using a finite-difference approximation to the Hessian, for example, can induce numerical error in the estimated standard error of the coefficients. The packages surveyed are considered in the light of this benchmark. A possible alternative to the benchmark using actual data, would be to use simulated data, whose distributional and other properties are

---

[4] In fact, assigning a plausible positive value is preferable, since zeros are usually local optima for the parameter vector, and given starting values of zero makes it more difficult for the optimisation routine to move towards the correct values.

[5] The data, also employed in the present study, are available at the *Journal of Business and Economic Statistics* data archive ftp site - see

known. However, even if data is constructed according to some known data generating process, it is not clear that the model under the DGP will be the optimal estimated model for any given Monte Carlo replication.

A caveat worth mentioning here relates to the distinction between conditional and unconditional maximum likelihood. All of the packages surveyed in this review use conditional maximum likelihood to estimate the model − that is, the estimation is conducted conditional upon the pre-sample initialisations of ε and h. On the other hand, unconditional maximum likelihood would estimate these parameters along with the others. Given that conditional maximum likelihood is employed, any benchmark that is developed would be conditional upon the chosen initialisations. Consequently, one may argue that since the choice of initialisation is to an extent made arbitrarily, the benchmark will also be arbitrary. Whilst there is no reason to suppose that the initialisation used by most packages and presented above is necessarily optimal, and others are available[6], the initialisation used by FCP is also the one used by Bollerslev (1986) in his seminal paper. This might suggest that any package offering GARCH(1,1) would use Bollserslev and FCP's initialisation, but in fact this is not the case.

3. Default Model Estimation with the Econometric Software Packages

In this paper, nine packages are employed and their results compared when GARCH models are estimated using the pre-set default settings. The packages are summarised in Table 1. It is desirable that both the parameter, and the squared error / conditional

variance vector intialisations can be specified by the user. In the case of the conditional variance parameters, it is often useful to drag the initialisations away from their typical default values of zero, and estimation of the FCP Model I requires that the conditional variance vector is started at the average mean-adjusted sum of squares of the data. For some packages, this is not possible (*e.g.*, E-VIEWS or GAUSS) and thus the benchmark model cannot be estimated.

FCP provide Monte Carlo evidence that the use of analytic derivatives and analytic Hessian leads to more accurate estimation than procedures based upon numerical techniques of computation. Thus it is also desirable for packages to permit the user to use additional information such as analytical derivatives where available, and moreover that this is done as default. The best packages in this regard are SAS, SHAZAM and TSP.

Even worse than using a sub-optimal technique, is for a package not to provide any details at all on how the models are estimated, so that the user has no idea where the results have come from, let alone how accurate they are. The best package in terms of supplying the user with extensive documentation on the estimation procedures, is GAUSS. Further, this help should be available on-line as well as in hard copy form. In the case of SHAZAM, additional examples and documentation to supplement the user manual are available on their web page[7]. The manual that will be shipped with the forthcoming version of LIMDEP is likely to contain extensive documentation.

---

[6] E-VIEWS, for example, uses an alternative method which they term "backcasting", based upon an exponential smoothing application to the squared residuals.

Default estimation results are presented in Table 2. At this stage, we do not intend to "pass judgement" on the packages that are good and those which are not, since defaults may differ between packages in terms of initialisations, starting values for parameters, convergence tolerances etc. Rather, we only present default results for comparison with one another. It can be seen that the coefficient estimates are broadly in agreement. However, when one considers the t-ratios associated with the coefficients, the discrepancies between methods used by each of the packages becomes far more apparent. For example, the t-ratio on the constant term in the conditional variance equation varies from 1.66 (TSP) to over 8 (EVIEWS and SAS)! Thus under TSP, we would conclude that this term was not significantly different from zero (which would also be the case for the benchmark with standard errors computed using analytic gradients and Hessian), while for all other packages, a zero null hypothesis for the coefficient would have been rejected at any standard confidence level. This is a result which researchers should find worrisome, for in all but the most clear-cut cases, statistical inferences from conditional variance-type models could be different, resulting from a standard error estimate that could be different by a factor of four or more compared with an estimate using the FCP approach with analytic gradient and Hessian. Interestingly, the rankings of the t-ratios across packages is similar for each coefficient – in other words, some packages seem to yield high t-ratios for all coefficients while others generate low values for all coefficients. The differences in t-ratios are probably due to differences in default calculation methods for the standard errors.

---

[7] http://shazam.econ.ubc.ca/intro/garch.htm

To summarise and conclude our findings under default estimation, we stress three points. First, software users should be told what the default estimation methods are, including such issues as convergence tolerance, method of initialisation of the series, how the standard errors are computed, and so on. Second, software users should always specify in their papers the package used in estimation, together with a description of which options were employed. Failure to do this could lead to results that are irreproducible. Third, t-ratios can differ from benchmark values because either the coefficients are wrong, or different methods for computing the standard errors have been employed. To this end, in the following section, we investigate the ability of the packages under investigation to hit the benchmark.

4. Hitting the Benchmark

Of the nine packages investigated, only one (TSP) hits the benchmark coefficients and Hessian-based standard errors using the default settings. Thus, it seems that TSP is the only package that uses Bollerslev's model as default. With regard to the remaining eight, we remove from this part of the analysis any packages which do not permit the estimation of FCP Model I because they do not have the necessary options - EVIEWS and GAUSS. We experiment with the options available for optimisation using the remaining seven packages. We follow McCullough and Renfro (1999) by presenting the accuracy of coefficient and standard error estimation using a measure known as the log relative error (LRE), given by

$$LRE = -\log_{10}[|x - c| / |c|] \tag{5}$$

where x is the estimate and c is the benchmark value. We also follow McCullough and Renfro in presenting the FCP GARCH benchmark to six digits, so that no package can obtain a score of more than six. The benchmark coefficient and standard error

estimates are presented to all six digits in McCullough and Renfro. The closer are the coefficient and benchmark values, the larger the LRE statistic will be. The integer part of the LRE shows the number of digits for which the estimated value and the benchmark agree[8].

LREs computed for the results from each package - both coefficients and standard errors - are presented in Table 3. Table 3 also shows which methods are available for standard error calculation under each package that permits estimation of the benchmark model. All of these remaining packages offer a reasonable degree of accuracy, although the closest to the benchmark seem to be TSP and LIMDEP.

LIMDEP, MICROFIT and MATLAB either hit, or come as close to hitting the benchmark as they can under default estimation. Several other packages, however, yielded results closer to the FCP benchmark values when non-default options were employed. Under TSP, the (non-default) optional commands "options double" before reading the data, and then "TOL=1e-x" (where x is any integer greater than 6) in the ARCH command, are required to obtain six digits of accuracy. In RATS, the "nlpar(subits=50,cvcrit=.0000001)" command should be used before the nonlinear estimation in order to increase the number of sub-iterations and the stringency of the convergence criterion relative to their default values. In RATS, results closest to the benchmark are also obtained using the "(method=bfgs,iters=500,robusterrors)" options with the "maximise" command. Under SAS, the "type=nonneg" and "optmethod=tr" options should be used after the model statement. Differences in

---

[8] For example, if the LRE is 3.7, the first three digits of the estimate are the same as those of the benchmark.

parameter estimates for estimation of the benchmark model could be attributable to a number of causes, including numerical error arising from the use of different algorithms (*e.g.*, BHHH instead of Newton) or numerical rather than analytic derivatives.

5. Forecasting the Conditional Variance

Although for some researchers, the estimated coefficients from a GARCH model have an intrinsic beauty, many applied econometricians estimate such models only as a pre-cursor for forming out-of-sample forecasts of the conditional variance (see Brooks, 1998 or Brooks and Burke, 1999 for motivations, and descriptions of forecast production).

Clearly, if the estimated coefficients from a model are wrong, then the forecasts are also likely to be wrong. But even if the coefficients are right, or very close to their benchmark values, there is still scope for one's software to let one down again. One way to produce dynamic out-of-sample time-series forecasts, is to output the model coefficient estimates and the required end-of-sample fitted values, and to calculate the forecasts from the models using the relevant recursions. For a GARCH(1,1) model, this is not an algebraically difficult task. However, many computer packages have built-in forecasting routines which apparently save the user the trouble. In the case of RATS, the instructions for forecasting are not specific to a particular model or class of models, so that the user sets up the model as a system of equations, and the software then solves for the forecasts using an algorithm, such as the Gauss-Seidel. Thus the actual forecast recursions themselves, which would normally yield unique forecasts for given parameters and initial values, are not necessarily being used.

Again, many packages cannot produce out-of-sample conditional variance forecasts automatically, but those packages from our list which can are EVIEWS, GAUSS, MATLAB, MICROFIT, RATS, and SAS. The final value of the in-sample fitted conditional variance (observation 1974) is included in our table of results, followed by forecasts up to 8 steps ahead. It is fairly easy to compute the conditional variance forecasts from a GARCH(1,1) model by hand using a calculator or a spreadsheet. The one step ahead "correct" forecast for 1975 is given by

$$h_{1975}^f = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{\varepsilon}_{1974}^2 + \hat{\beta}_1 \hat{h}_{1974} \tag{6}$$

where the parameter estimates used are those of the FCP benchmark, and $\hat{\varepsilon}_{1974}^2$ and $\hat{h}_{1974}$ are the fitted values for the squared error and the conditional variance for observation 1974, and which take the values 0.2854 and 0.115 respectively[9]. Similarly, 2,3, ..., 8 step ahead forecasts can be generated by

$$h_{1975+i}^f = \hat{\alpha}_0 + (\hat{\alpha}_1 + \hat{\beta}_1)\hat{h}_{1974+i} \text{ for } i = 1,2,\ldots,7 \tag{7}$$

In order to produce conditional variance forecasts from MICROFIT and SAS, the user must extend the dataset, adding missing values for the 1974-1982 observation numbers.

The forecast values from the packages are presented in Table 4, together with the "correct" forecasts, given the starting values and coefficients. As can be seen, most of the packages do a surprisingly reasonable job of forecasting - GAUSS, MATLAB, MICROFIT, RATS and SAS have hit the correct values to within 1 figure at the third decimal place, while EVIEWS accuracy is well within most forecasters' acceptable

tolerance levels. The latter is particularly interesting since EVIEWS, unlike the other five packages in Table 4, cannot estimate FCP Model I and so bases its forecasts on different coefficients. The upshot is that the estimate on the lagged squared error is too low relative to that of FCP Model I, while that on the lagged conditional variance is too high, so that these errors approximately balance each other out when predictions are constructed using equation (7).

5. Benchmarks for other Models and Data

The class of non-linear models is infinite. Even within the family of autoregressive, conditionally heteroscedastic models, there is a large number of other, typically more complex, variants and extensions. The list of such models includes EGARCH (Nelson, 1991), GJR (Glosten, Jaganathan and Runkle, 1993), GQARCH (Sentana, 1995), and threshold GARCH (Rabemananjara and Zakoian, 1993). Although the development of a benchmark for every possible model is infeasible, one might expect (hope?) that benchmarks might be developed for all of the standard and most popular specifications.

To demonstrate the authors' view of the need for these benchmarks, we select one extension of the GARCH model, the EGARCH, and perform exactly the same exercise as above. There is no assumed "correct" answer in this case, since no explicit benchmark exists at present. But, given that the EGARCH model is a relatively simple extension of GARCH and that it has a number of appealing features (for example, it allows for asymmetry or "leverage", so that the conditional variance can be affected

---

[9] We do not force the packages to employ the correct parameter estimates from the FCP benchmark - those which hit the benchmark can be observed from table 2.

differently by positive and negative shocks of the same magnitude, and it does not require the imposition of any non-negativity constraints upon the parameters), it is a pity that most of the packages do not have sufficient flexibility to enable the estimation of the exponential model[10]. The packages which permit estimation are EVIEWS, GAUSS, MICROFIT, RATS, and SAS. The EGARCH model may be expressed in various different forms, but a popular specification is

$$\log(h_{,t}) = \omega + \beta \log(h_{t-1}) + \gamma \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \alpha \left[\frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} - \sqrt{\frac{2}{\pi}} \right] \tag{8}$$

with the conditional mean being specified as in equation (1). An additional wrinkle is that Nelson (1991) proposed that the model should be estimated assuming that errors are drawn from a Generalised Exponential Distribution (GED), rather than a normal distribution, which is usually assumed for most other members of the GARCH family of models. GAUSS does this compulsorily, and by default, while the other four packages cannot apply GED even as an option. SAS and GAUSS also use a quite different formulation of the model compared with (8); this in itself is a severe problem that makes interpretation and diagnostic checking of the model specification extremely difficult. Thus only EVIEWS, MICROFIT, RATS and SAS coefficient estimates are presented in Table 5. Since GAUSS uses GED, while all other packages use a Gaussian distribution, the coefficient estimates for the former are quite different from those of the latter, and hence are not reported. Under GED, there is also an additional parameter to estimate, which corresponds to the thickness of the tail.

---

[10] Note that we are specifically excluding instances where the user can, with difficulty, program up the routines in the pseudo-programming language available inside some packages. We thus consider only cases where the estimation of the EGARCH model is either fully automated, or that the instructions for estimation are widely available and distributed by the vendors rather than a third party.

Of the packages considered, EVIEWS estimates a slightly different model, with the term $\sqrt{\dfrac{2}{\pi}}$ removed from (8). Hence the intercept term in the conditional variance equation for this model will differ by a factor of $\alpha\sqrt{\dfrac{2}{\pi}}$ compared with that of RATS and SAS. Similarly, SAS uses a slightly different formulation than (8), which may be written

$$\log(h_{,t}) = \omega + \beta\log(h_{t-1}) + \gamma\left[\alpha\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \frac{\left|\varepsilon_{t-1}\right|}{\sqrt{h_{t-1}}} - \sqrt{\frac{2}{\pi}}\right] \tag{8'}$$

We allow for these differences by scaling the EVIEWS estimate by a factor of $\sqrt{2/\pi}$, so that the values of the coefficients are presented in the table in a consistent manner across the packages, although we do not modify the standard errors.

Results for default EGARCH estimations are presented in Table 4 for EVIEWS, MICROFIT, RATS and SAS. Although there is currently no benchmark, the mere fact that there are discrepancies between the estimates gives cause for concern. The MICROFIT, RATS and SAS coefficient values are "close" to one another, while those of EVIEWS are typically up to 10% away from the other two, although in the absence of a benchmark, this certainly does not constitute an indicator of inaccuracy.

Additionally the standard error estimates for EVIEWS yield *t*-statistics that are on occasions twice as large as those of RATS and SAS. The obvious implication of this is that the various packages use different default methods for computing the standard errors. To the extent that some standard errors may be better than others (see FCP, 1996), the potential for unreliable inference with this model is considerable. Clearly

there is a need for a benchmark and also some consensus as to what forms of the model are preferable, and the latter should be estimable by all packages. Such a benchmark should include a specification of the parameter starting values, the initialisations for the errors and conditional variances, with the method for calculating the standard errors being stated explicitly.

7. Conclusions

Econometric models have become increasingly sophisticated in the last two decades. While computing power and the availability of flexible of software packages have kept pace with this change, methods for determining the accuracy and robustness of the results from these models have not. This paper has shown that the results produced using a default application of several of the most popular econometrics packages differ considerably from one another, even using a relatively simple non-linear model on a dataset that it known to be "well-behaved". Only a handful of packages can estimate FCP's Model I, which is surprising, since it corresponds to Bollerslev's (1986) seminal formulation. We also find that forecasting results using the simple GARCH(1,1) model are not all the same, even for packages which hit the FCP benchmark coefficient estimates. Estimation for another popular model from the GARCH family, the EGARCH model, produced widely varying results.

Our results have a number of implications for published research and future software development. For example, many researchers do not present the initialisations that they employed, nor do they even state which package was used. We have shown that results are clearly software-dependent, which makes the results from two different papers, which probably used different packages not comparable. Authors should

therefore provide sufficient detail regarding model estimation to permit replication of their results, including the name (and version) of the package used, and explicit equations for estimation including starting values and initialisations.

Academic research effort should also, we believe, be invested in the development of benchmark estimation routines using given data, for all of the standard models, including GARCH-variants, threshold autoregressive and Markov switching models, for example.

Finally, our findings also have implications which software developers would do well to bear in mind, in particular the following points.

- Software developers should talk to each other! If packages produce different results for the same model including the initialisation, at best one set can be right, implying that the others are inaccurate, using inappropriate estimation techniques, or in some other way failing to correctly handle the data.

- The majority of software manufacturers do not supply adequate documentation to enable the user to determine how the results from estimating a given model were derived.

- Estimation methods that become known as best practice, such as the use of analytical gradients and Hessians when estimating GARCH(1,1) models[11], should apply as default, rather than being optional if the user has read a manual's appendices, or not available at all.

It is likely that the problems in estimation outlined above would be exacerbated by data which had additional regularities, such as large structural breaks, outliers, or other structure which is not parameterised in the model under consideration. In particular, the scope for numerical error is further heightened when the process is close to IGARCH, for the package will be required to invert a near-singular matrix. One wonders whether any two packages would produce the same results under these circumstances.

---

[11] There is still a role for numerical derivatives in estimating higher order GARCH models, or other members of the GARCH family (such as EGARCH), where analytic derivatives are simply too complex.

References

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

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992) ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence *Journal of Econometrics* 52(5), 5-59

Bollerslev, T. and Ghysels, E. (1996) Periodic Autoregressive Conditional Heteroscedasticity *Journal of Business and Economic Statistics* 14, 139-151

Brooks, C. (1998) Forecasting Stock Return Volatility: Does Volume Help? *Journal of Forecasting* 17, 59-80

Brooks, C. (1997) GARCH Modelling in Finance: A review of the Software Options *Economic Journal* 107(443) 1271-1276

Brooks, C. and Burke, S.P. (1999) "Forecasting Exchange Rate Volatility using Conditional Variance Models Selected by Information Criteria" *Economics Letters* 61, 273-278

Fiorentini, G., Calzolari, G., and Panattoni, L. (1996) Analytic Derivatives and the Computation of GARCH Estimates *Journal of Applied Econometrics* 11, 399-417

McCullough, B.D. and Renfro, C.G. (1999) Benchmarks and Software Standards: A Case Study of GARCH Procedures *Journal of Economic and Social Measurement* 25, 59-71

Nelson, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach *Econometrica* 59(2), 347-370

Rabemananjara, R and Zakoian, J.M. (1993) Threshold ARCH Models and Asymmetries in Volatility in Pesaran, M.H. and Potter, S.M. (eds) *Nonlinear Dynamics, Chaos and Econometrics* Chapter 11, 179-197, John Wiley and Sons, Chichester, England

Sentana, E. (1995) Quadratic ARCH Models *Review of Economic Studies* 62, 639-661

Table 1: Summary Statistics for Econometric Packages and GARCH Estimation

Package and Version used	Contact information	Can user specify ε_t^2 and h_t initialisations?		Analytic Gradient and Hessian available?	
				Gradient	Hessian
E-VIEWS 3.1	QMS Software, Suite 336, 4521 Campus Drive, Irvine, CA 92612, USA. 949 856 3368 http://www.eviews.com	Yes	No	No	No
GAUSS 3.2.39-FANPAC 1.1.11/2	Aptech Systems Inc, 23804 S.E. Kent, Langley Road, Maple Vallet, WA 98038, USA 425 432 7855 http://www.aptech.com	No	No	Yes	No
LIMDEP 8.0[1]	Econometric Software, 15 Gloria Place, Plainview, NY 11803 USA. 516 938 5254 http://www.limdep.com	No	No	Yes	Yes
MATLAB 11	The MathWorks Inc., 3 Applie Hill Drive, Natick, MA 01760 2098, USA 508 647 7000 http://www.mathworks.com	Yes	Yes	No	No
MICROFIT 4.0, Windows Version	Oxford Electronic Publishing, Oxford University Press, Walton Street, Oxford, UK. (+44) 1865 379 300 http://www.microfit.com	Yes	No	No	No
RATS 4.3	Estima, PO Box 1818, Evanston, IL 60204-1818, USA 847 864 8772 http://www.estima.com	Yes	Yes	No	No
SAS 6.12	SAS Institute, Campus Drive, Cary NC 27513 USA. 919 677 8000 http://www.sas.com	Yes	Yes	Yes	No
SHAZAM 8.0	Department of Economics, 997-1873 East Mall, Vancouver, B.C. Canada V6T 1Z1. 604 822 2876 http://shazam.econ.ubc.ca	Yes	No	Yes	No
TSP 4.5	TSP International, POB 61015 Station A, Palo Alto, CA 94306, USA. 650 326 1927 http://www.tspintl.com	Yes	Yes	Yes	Yes

Notes: Contact numbers given are for telephone; LIMDEP 7.0 is the currently available version of the program, which does not support GARCH model estimation. This review was undertaken on a test copy of a forthcoming version, which at the time of writing is not commercially available, and which is likely to be called v8.0. [1]This can be specified in version 7.0 or greater.

Table 2: Default Estimation of a GARCH(1,1) Model on the Bollerslev and Ghysels Data

							Parameters and *t*-ratios	
Package	μ	$t(\mu)$	α_0	$t(\alpha_0)$	α_1	$t(\alpha_1)$	β_1	$t(\beta_1)$
E-VIEWS	-0.00540	-0.64	0.0096	8.01	0.143	11.09	0.821	53.83
GAUSS-FANPAC	-0.00600	-0.75	0.0110	3.67	0.153	5.67	0.806	23.71
LIMDEP	-0.00619	-0.71	0.0108	3.44	0.153	5.61	0.806	26.73
MATLAB	-0.00619	-0.73	0.0108	8.13	0.153	10.96	0.806	48.67
MICROFIT	-0.00621	-0.73	0.0108	3.78	0.153	5.78	0.806	24.02
SAS	-0.00619	-0.74	0.0108	8.15	0.153	10.97	0.806	48.60
SHAZAM	-0.00613	-0.73	0.0107	5.58	0.154	7.91	0.806	36.96
RATS	-0.00625	-0.71	0.0108	3.76	0.153	5.79	0.806	23.93
TSP	-0.00619	-0.67	0.0108	1.66	0.153	2.86	0.806	11.11

Table 3: Accuracy of Estimates

Package	Parameter	Coefficient	Standard Error				
			Hessian	OPG	QMLE	IM	BW
LIMDEP	μ	6.0	-	-	-	-	6.0
	α_0	6.0	-	-	-	-	6.0
	α_1	6.0	-	-	-	-	6.0
	β_1	6.0	-	-	-	-	6.0
MATLAB	μ	4.6	-	4.7	-	-	-
	α_0	6.0	-	5.1	-	-	-
	α_1	4.9	-	5.1	-	-	-
	β_1	5.6	-	5.2	-	-	-
MICROFIT	μ	2.5	2.9	-	-	-	-
	α_0	4.2	3.5	-	-	-	-
	α_1	2.7	2.7	-	-	-	-
	β_1	3.8	4.0	-	-	-	-
RATS	μ	1.9	1.4	3.4	1.1	-	-
	α_0	4.1	2.5	2.3	2.8	-	-
	α_1	4.4	2.8	2.3	2.5	-	-
	β_1	3.8	2.4	2.6	2.4	-	-
SAS	μ	2.6	3.1	5.0	2.8	-	-
	α_0	4.4	4.5	4.7	4.8	-	-
	α_1	4.6	4.9	4.6	5.0	-	-
	β_1	5.2	4.8	4.9	5.1	-	-
SHAZAM	μ	3.2	-	3.1	-	4.3	5.0
	α_0	3.4	-	3.0	-	3.4	3.6
	α_1	4.1	-	3.5	-	3.9	4.2
	β_1	4.5	-	3.3	-	3.7	4.0
TSP	μ	6.0	6.0	6.0	6.0	-	-
	α_0	6.0	6.0	6.0	6.0	-	-
	α_1	6.0	6.0	6.0	6.0	-	-
	β_1	6.0	6.0	6.0	6.0	-	-

Notes: Cell entries are errors measured relative to the FCP benchmark values from FCP using the same method for estimating the first and second derivatives. 6.0 is the maximum possible score for any given package in this exercise since only six digits of accuracy are given in the FCP paper.

Table 4: Conditional Variance Forecasts

	Last in-sample	1 step ahead	2 steps ahead	3 steps ahead	4 steps ahead	5 steps ahead	6 steps ahead	7 steps ahead	8 steps ahead
Observation No.	1974	1975	1976	1977	1978	1979	1980	1981	1982
EVIEWS [1]	0.115	0.145	0.149	0.153	0.157	0.161	0.165	0.169	0.171
GAUSS	0.115	0.147	0.152	0.156	0.161	0.165	0.169	0.173	0.177
MATLAB	0.115	0.147	0.152	0.156	0.161	0.165	0.169	0.173	0.176
MICROFIT	0.115	0.147	0.152	0.156	0.161	0.165	0.169	0.173	0.177
RATS	0.115	0.147	0.152	0.156	0.161	0.165	0.169	0.173	0.177
SAS	0.115	0.147	0.151	0.156	0.160	0.164	0.168	0.172	0.176
"Correct Values"	0.115	0.147	0.152	0.156	0.161	0.165	0.169	0.173	0.177

Note: [1] EVIEWS did not obtain the FCP coefficient values. EVIEWS does obtain the "correct" forecast values, given the coefficient estimates that were actually obtained.

Table 5: Default Estimation for EGARCH(1,1) Model

	EVIEWS	MICROFIT	RATS	SAS
μ	-0.0106	-0.0116	-0.0117	-0.0116
$t(\mu)$	-1.32	-1.40	-1.32	-1.34
ω	-0.112 *	-0.127	-0.127	-0.127
$t(\omega)$	†	-4.65	-4.46	-4.38
β	0.916	0.912	0.912	0.912
$t(\beta)$	112	56.2	54.4	53.0
α	0.322	0.333	0.333	0.333
$t(\alpha)$	15.0	8.59	8.20	7.88
γ	-0.0370	-0.0385	-0.0385	-0.0385
$t(\gamma)$	-2.76	-2.10	-2.01	†

Notes: * this coefficient has been modified - see text for details; † denotes that a *t*-ratio cannot be given for this coefficient, since a slightly different version of the model is estimated by EVIEWS and SAS, and therefore one cannot infer what the value of the standard error would have been had equation (4) itself been estimated.