On Comparing Estimation Methods for VAR-ARCH Models

S. Liu, W. Polasek

Institute of Statistics and Econometrics,
University of Basel, CH-4051 Basel, Switzerland

Abstract: VAR-VARCH models are becoming increasingly important for applications in several fields of economics and finance. This paper focuses on the method of scoring for maximum likelihood estimation (MLE). We first estimate a bivariate VAR-ARCH model with a simulated data set and with an exchange rate time series and then compare the method with two others (one classical and one Bayesian) that are available in statistical and econometric program packages.

1 Introduction

Autoregressive conditional heteroskedastic (ARCH) models were first introduced by Engle (1982). Applications can be found in several fields of economics and finance and a recent survey can be found in e.g., Gourieroux (1997). The numerical techniques based on the BHHH method of Berndt et al. (1974) for maximum likelihood estimation (MLE) are used for the VAR-ARCH models in most current estimation programes. The method of scoring (see, e.g. Fomby et al. (1984)) for MLE using the exact information matrix is expected to obtain better estimates in small samples. The MCMC method is implemented in the Bayesian framework. We will estimate a bivariate VAR-ARCH model with simulated and real data and will compare the method of scoring with the other two.

2 MLE for heteroskedastic time series

Suppose a multivariate time series y_t, $t = 1, ..., T$, of dimension M is independently and identically normally distributed and specified by the conditional moment structure: the $M \times 1$ conditional mean vector is $E(y_t \mid \psi_{t-1}) = \mu_t$ and the $M \times M$ conditional variance matrix $Var(y_t \mid \psi_{t-1}) = V_t$, where ψ_{t-1} indicates the information set available at time $t-1$. Assume that $\mu_t = \mu_t(\theta)$ and $V_t = V_t(\theta)$ are functions of the $p \times 1$ vector of parameters θ. This multivariate conditional heteroskedastic model can be written as

$$y_t = \mu_t + u_t, \quad t = 1, ..., T, \tag{1}$$

where u_t is a $M \times 1$ disturbance vector, which has a normal distribution where the mean vector is $E(u_t \mid \psi_{t-1}) = 0$ and the conditional variance matrix is $E(u_t u_t' \mid \psi_{t-1}) = V_t$ with V_t being positive definite.

The process (1) is stationary and the stability conditions are discussed in Gourieroux (1997). The conditional log-likelihood function of $y = (y_1, ..., y_T)'$ in (1) is given by

$$\begin{aligned} L(y,\theta) &= \sum_{t=1}^{T} L_t(y,\theta) \\ &= -\sum_{t=1}^{T} \frac{1}{2} \log \mid V_t \mid - \sum_{t=1}^{T} \frac{1}{2} u_t' V_t^{-1} u_t, \quad t = 1, ..., T. \end{aligned} \tag{2}$$

To apply the method of scoring for MLE, we define the $p \times 1$ gradient vector $g = g(\theta)$ and the $p \times p$ Hessian matrix $H = H(\theta)$:

$$g = \sum_{t=1}^{T} \frac{\partial L_t(y,\theta)}{\partial \theta}, \tag{3}$$

$$H = \sum_{t=1}^{T} \frac{\partial^2 L_t(y,\theta)}{\partial \theta \partial \theta'}. \tag{4}$$

If E indicates the mathematical expectation with respect to y, then the $p \times p$ information matrix $F = F(\theta)$ is

$$F = -E(H). \tag{5}$$

In the following two lemmas, we use the matrix differential techniques advocated by Magnus and Neudecker (1999) to derive the gradient vector and the information matrix of the multivariate time series model (1).

Lemma 1: The gradient vector of model (1) is

$$g(\theta) = \frac{1}{2} \sum_{t=1}^{T} \left(\frac{\partial \mathrm{vech} V_t}{\partial \theta'} \right)' D' D \mathrm{vech} P_t + \sum_{t=1}^{T} \left(\frac{\partial \mu_t}{\partial \theta'} \right)' V_t^{-1} u_t, \tag{6}$$

where

$$P_t = V_t^{-1} u_t u_t' V_t^{-1} - V_t^{-1}, \tag{7}$$

vech denotes the vectorization operator which eliminates all supradiagonal elements of the matrix, $\mathrm{vech} V_t$ is a $N \times 1$ vector and D is the $M^2 \times N$ duplication matrix ($N = M(M+1)/2$).

Lemma 2: The information matrix F for model (1) is

$$F = E(R), \tag{8}$$

where $R = R(\theta) = E(H \mid \psi_{t-1})$ is the conditional expectation given by

$$\begin{aligned} R &= \frac{1}{2}\sum_{t=1}^{T}\left(\frac{\partial \mathrm{vech} V_t}{\partial \theta'}\right)' D'(V_t^{-1} \otimes V_t^{-1}) D \frac{\partial \mathrm{vech} V_t}{\partial \theta'} \\ &\quad + \sum_{t=1}^{T}\left(\frac{\partial \mu_t}{\partial \theta'}\right)' V_t^{-1} \frac{\partial \mu_t}{\partial \theta'}, \end{aligned} \tag{9}$$

where $\otimes$ indicates the Kronecker product.

The pseudo MLE approach for model (1) is considered by Gourieroux (1997) for the case where the underlying distribution is not known, but the conditional normal likelihood function is used; if the underlying distribution is known to be normal, the pseudo MLE approach is the same as the MLE approach. Lemma 1 (also valid for the pseudo MLE approach) and Lemma 2 for the univariate case can be found in Gourieroux (1997). Lemma 2 is given in Liu and Polasek (1999). We demonstrate the estimation procedure for a bivariate VAR(1)-ARCH(1) model with diagonal coefficient matrices using simulated and real data. The ML estimates are obtained by the method of scoring using the following iteration procedure (R replaces F):

$$\theta_{j+1} = \theta_j + \lambda_j R_j^{-1} g_j, \tag{10}$$

where $\lambda_j \leq 1$ and j is the iteration index for which the term is evaluated.

We compare this estimation method with the statistical and econometric packages of SPlus+GARCH of MathSoft (1996) and BASEL of Polasek (1999). To apply the BHHH algorithm for MLE in the SPlus+GARCH program one has to replace R_j in (10) by $g_j g_j'$.

The MCMC algorithm in the BASEL package uses the joint distribution of the VAR-ARCH model for the data $y = (y_1, ..., y_T)$ and the parameters θ

$$p(y, \theta) \propto \prod_{t=1}^{T}\prod_{m=1}^{M} N[y_t^m | \mu_t^m, V_t^m] p(\theta^m), \tag{11}$$

where $N[.]$ indicates the multivariate normal distribution, θ^m are the parameters of the m-th equation and $p(\theta) = \prod_{m=1}^{M} p(\theta^m)$ is the informative prior distribution for the M time series ($m = 1, ..., M$).

3 Diagonal VAR(1)-ARCH(1) model

Consider a special case of the bivariate VAR(1)-ARCH(1) model ($M = 2$), which is a "dvec" model of SPlus+GARCH:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, \quad t = 1, ..., T, \tag{12}$$

where $y_t = (y_{1t}, y_{2t})'$ is the t-th 2×1 vector of observations of time series $(y_1, \dots, y_T)$. The mean $\mu_t = (\mu_{1t}, \mu_{2t})'$ is specified by the "diagonal" VAR(1) equation

$$\begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & 0 \\ 0 & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix}, \tag{13}$$

where $\beta_{10}, \beta_{20}, \beta_{11}$ and β_{22} are scalar parameters. The conditional variance matrix of $u_t = (u_{1t}, u_{2t})'$ is the 2×2 positive definite matrix V_t:

$$V_t = \begin{pmatrix} v_{11t} & v_{12t} \\ v_{12t} & v_{22t} \end{pmatrix},$$

and vech$V_t = (v_{11t}, v_{12t}, v_{22t})'$ is parameterized by the "diagonal" model

$$\begin{pmatrix} v_{11t} \\ v_{12t} \\ v_{22t} \end{pmatrix} = \begin{pmatrix} \alpha_{10} \\ \alpha_{20} \\ \alpha_{30} \end{pmatrix} + \begin{pmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{pmatrix} \begin{pmatrix} u_{1t-1}^2 \\ u_{1t-1}u_{2t-1} \\ u_{2t-1}^2 \end{pmatrix} \tag{14}$$

with $\alpha_{10} > 0$, $\alpha_{20} > 0$, $\alpha_{30} > 0$, $\alpha_{11} \geq 0$, $\alpha_{22} \geq 0$ and $\alpha_{33} \geq 0$ such that $V_t > 0$ exists.

The parameter vector is partitioned as $\theta = (\beta', \alpha')'$, where $\beta = (\beta_{10}, \beta_{20}, \beta_{11}, \beta_{22})'$ contains all parameters only for the mean equation and $\alpha = (\alpha_{10}, \alpha_{20}, \alpha_{30}, \alpha_{11}, \alpha_{22}, \alpha_{33})'$ only for the variance equation. Based on Lemmas 1 and 2, we can get $g(\theta)$ and $R(\theta)$.

4 Numerical Example

We use S-PLUS for two examples of bivariate simulated and real data. We generate a time series of T=200 observations for the VAR(1)-ARCH(1) model introduced previously with the following parameter vector:

$$\begin{aligned} \theta &= (\beta', \alpha')' \\ &= (\beta_{10}, \beta_{20}, \beta_{11}, \beta_{22}, \alpha_{10}, \alpha_{20}, \alpha_{30}, \alpha_{11}, \alpha_{22}, \alpha_{33})', \end{aligned}$$

where

$$\begin{aligned} \beta_{10} &= 0.10, \quad \beta_{11} = -0.4, \\ \beta_{20} &= 0.30, \quad \beta_{22} = 0.05 \end{aligned}$$

are for the mean equation (12) and

$$\begin{aligned} \alpha_{10} &= 0.15, \quad \alpha_{11} = 0.01, \\ \alpha_{20} &= 0.05, \quad \alpha_{22} = 0.07, \\ \alpha_{30} &= 0.10, \quad \alpha_{33} = 0.05, \end{aligned}$$

for the variance equation (13).

Based on $g(\theta)$ and $R(\theta)$, we write a MLE program to find estimates by using the method of scoring. The iteration process of (10) is then conditionally based on the first two observations. The criterion used for checking convergence of the program is

$$\frac{\hat{\theta}_{j+1} - \hat{\theta}_j}{\hat{\theta}_j} \leq 0.01, \tag{15}$$

with $\lambda_j = 1/10j$ in (10). We compare the estimates after 100 iterations in two program packages, i.e. SPlus+GARCH of MathSoft (1996) and the BASEL package of Polasek (1999) with our own MLE procedure. All estimates of the coefficients converge at about 25 iterations. As an overall performance measure we compute $||\hat{\theta} - \theta||^2 = \sum(\hat{\theta}_i - \theta_i)^2$, the squared distances (or simple mean squared errors) of the estimates and known "real" values of the parameters. We denote the squared distances in the mean equation as "AR part", for the variance equation as "ARCH part" and for all parameters as "Total". The estimates are compared in Table 1.

Table 1: Comparisons of estimates from SPlus+GARCH, BASEL and MLE with Squared Distances

Parameters	True Values	SPlus+G Estimates	BASEL Estimates	MLE Estimates
β_{10}	0.10	0.1223657	0.1086234	0.1397224
β_{20}	0.30	0.2004163	0.2678540	0.2855772
β_{11}	-0.40	-0.5246969	-0.5083313	-0.5427349
β_{22}	0.05	0.0493810	0.0158123	0.0404794
α_{10}	0.15	1.0107307	0.0505836	0.2498541
α_{20}	0.05	-0.0001644	0.0519461	0.1158166
α_{30}	0.10	0.6928885	0.1056807	0.2946794
α_{11}	0.01	-0.1427515	0.0304871	0.1397502
α_{22}	0.07	-0.0180374	0.0378279	0.1243003
α_{33}	0.05	0.1011727	0.0279547	0.0774030
squared distance	MSE			
AR part		0.0259668	0.1400330	0.0222498
ARCH part		1.1435582	0.0127414	0.0708833
Total		1.1695250	0.1527744	0.0931331

Surprisingly, the other popular SPlus+GARCH program produces estimates which come out worst. E.g. instead of the true value $\alpha_{10} = 0.15$, SPlus+GARCH estimates $\hat{\alpha}_{10} = 1.01$. Note also that there is no positivity constraint in SPlus+GARCH and α_{11} and α_{22} are found to be negative while their true values are positive.
Furthermore, we estimate ARCH models for the exchange rates of the German Mark and the British Pound both against the US Dollar (weekly, 2 Jan. 1985 - 21 Jan. 1998).

Table 2: Estimates and Std.Errors from SPlus+GARCH and MLE

θ	SPlus+G Estimates	SPlus+G Std.Errors	BASEL Estimates	BASEL Std.Errors	MLE Estimates	MLE Std.Errors
β_{10}	-0.0020984	5.448e-04	-0.4024365	0.0938321	-0.0008335	1.869e-04
β_{20}	0.0030424	3.610e-04	-0.3230922	0.0692153	0.0005573	1.863e-04
β_{11}	-0.1706515	2.116e-02	-0.2024305	0.0837954	0.0053737	1.165e-02
β_{22}	-0.2610397	1.281e-02	0.0530685	0.0296580	0.0225780	1.198e-02
α_{10}	0.0001816	7.898e-06	0.0554886	0.0204951	0.0018142	9.934e-06
α_{20}	-0.0001057	4.271e-06	0.0367933	0.0282775	0.0002432	6.787e-06
α_{30}	0.0001066	2.761e-06	0.0663147	0.0259067	0.0017844	9.605e-06
α_{11}	0.5440509	6.776e-02	0.0400850	0.0261359	0.0443390	1.850e-02
α_{22}	0.5307577	5.132e-02	0.1091828	0.0821582	-0.0111523	1.414e-02
α_{33}	0.5126058	4.233e-02	0.0862747	0.0599910	0.1605119	1.774e-02

The estimation results for weekly returns are in Table 2. We see that the estimates are quite different, especially the AR parameters in the mean equation. The standard deviations of the parameter estimates are rather small for the MLE estimates.Overall, the BASEL estimates seem to produce better interpretable results.

5 Conclusions

We have presented a new ML estimation method based on the gradient vector and the information matrix of a large class of multivariate conditional heteroskedastic time series models. For the diagonal VAR(1)-ARCH(1) model, we have calculated two examples by using the method of scoring. In this special case, the method of scoring works rather well compared with other classical methods.

References

BERNDT, E., HALL, B., HALL, R. and HAUSMAN, J. (1974): Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement, 3, 653-665.*

ENGLE, R. F. (1982): *Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987-1006.*

FOMBY, T., HILL, R. C. and JOHNSON, S. R. (1984): Advanced Econometric Methods. Springer, New York

GOURIEROUX, C. (1997): ARCH Models and Financial Applications. Springer, New York

LIU, S. and POLASEK, W. (1999): *Maximum likelihood estimation for the VAR-VARCH model: A new approach. In Modelling and Decisions in Economics, Essays in Honor of Franz Ferschl, Leopold-Wildburger, U., Feichtinger, G. and Kistner, K.-P., eds, Physica-Verlag, Heidelberg, pp. 99-113.*

MAGNUS, J. R. and NEUDECKER, H. (1999): Matrix Differential Calculus with Applications in Statistics and Econometrics, revised edition. John Wiley and Sons, Chichester

MathSoft (1996): S+GARCH User's Manual, Version 1.0. Data Analysis Products Division, MathSoft, Seattle

POLASEK, W. (1999): The BASEL package: A Bayesian Sampling Environment Language. University of Basel, Basel

, , , ,