Bayesian Estimation and Inference for the ECCC-GARCH Model in R

Tomasz Woźniak[a]

[a] *Department of Economics, University of Melbourne, 111 Barry Street, Carlton 3053, Australia*
email: tomasz.wozniak@unimelb.edu.au

Abstract

A random walk Metropolis-Hastings algorithm for the Bayesian estimation of the Vector Autoregressive models with the conditional volatility process being the Extended Constant Conditional Correlation GARCH(1,1) model is provided, as well as an appropriate estimator for the marginal data density. The codes are available under the GNU General Public License v3.0. To refer to the codes in publications, please, cite one of the following papers:

Woźniak, Tomasz (2015) Testing Causality Between Two Vectors in Multivariate GARCH Models, *International Journal of Forecasting*, 31(3), pp. 876–894.

Woźniak, Tomasz (2018) Granger-Causal Analysis of GARCH Models: a Bayesian Approach, *Econometric Reviews*, 37(4), pp. 325-346.

Keywords: R, Multivariate GARCH Models, Metropolis-Hastings Sampler, Marginal Data Density

1. The VAR-ECCC-GARCH Model

Model and likelihood function. The model under consideration is the Vector Autoregressive process of Sims (1980) for the conditional mean, and the Extended Constant Conditional Correlation Generalized Autoregressive Conditional Heteroskedasticity process of Jeantheau (1998) for conditional variances. The model presented in this manuscript was used in Woźniak (2015, in press). The exposition in this note is based on Woźniak (2015). The conditional mean process models linear relations between current and lagged observations of the considered variables:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \epsilon_t \tag{1a}$$

$$\epsilon_t = D_t r_t \tag{1b}$$

$$r_t \sim i.i.St^N\left(\mathbf{0}, \mathbf{C}, \nu\right), \tag{1c}$$

for all $t = 1, \ldots, T$, where y_t is a $N \times 1$ vector of data at time t, α_1 is a $N \times 1$ vector of constant terms, $\alpha_1, \ldots, \alpha_p$ are $N \times N$ matrices of autoregressive parameters, ϵ_t and r_t are $N \times 1$ vectors of residuals and standardized residuals respectively, $D_t = diag(\sqrt{h_{1t}}, \ldots, \sqrt{h_{Nt}})$ is a $N \times N$ diagonal matrix with conditional standard deviations on the diagonal. The standardized residuals follow a N-variate standardized Student t distribution with a vector of zeros as a location parameter, a matrix $\mathbf{C}$ as a scale matrix, and $\nu > 2$ degrees of freedom.

The conditional covariance matrix of the residual term ϵ_t is decomposed into:

$$H_t = D_t \mathbf{C} D_t \quad \forall t = 1, \ldots, T. \tag{2}$$

For the matrix H_t to be a positive definite covariance matrix, h_t must be positive for all t and $\mathbf{C}$ positive definite (see Bollerslev, 1990). A $N \times 1$ vector of current conditional variances is modeled with lagged squared residuals, $\epsilon^{(2)}_{t-1} = (\epsilon^2_{1t-1}, \ldots, \epsilon^2_{Nt-1})'$, and lagged conditional variances, h_{t-1}:

$$h_t = \omega + A\epsilon^{(2)}_{t-1} + Bh_{t-1}, \tag{3}$$

for all $t = 1, \ldots, T$, where $\omega > 0$ is a $N \times 1$ vector of positive constants, A and B are the coefficient matrices of ARCH and GARCH effects respectively. Matrices A and B are such that all their elements are non-negative, $A, B \geq 0$, and that the GARCH process is covariance stationary, i.e. the largest eigenvalue of $A + B$ is less than 1. The vector of conditional variances is then given by $E[\epsilon^{(2)}_t | I(t-1)] = \frac{\nu}{\nu-2} h_t$, where $I(t-1)$ is the information set available at time $t-1$.

The likelihood function has the following form:

$$p(\mathbf{y}|\theta) = \prod_{t=1}^{T} \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} ((\nu-2)\pi)^{-\frac{N}{2}} |H_t|^{-\frac{1}{2}} \left(1 + \frac{1}{\nu-2}\epsilon_t' H_t^{-1} \epsilon_t\right)^{-\frac{\nu+N}{2}}. \tag{4}$$

This model has its origins in the Constant Conditional Correlation GARCH (CCC-GARCH) model proposed by Bollerslev (1990). That model consisted of N univariate GARCH equations describing the vector of conditional variances, h_t. The CCC-GARCH model is equivalent to equation (3) with diagonal matrices $A(L)$ and $B(L)$. Its extended version, with non-diagonal matrices $A(L)$ and $B(L)$, was analyzed by Jeantheau (1998). He & Teräsvirta (2004) call this model the Extended CCC-GARCH (ECCC-GARCH).

Prior distribution. For the unrestricted VAR-GARCH model, the following prior specification is assumed. All of the parameters of the VAR process are *a priori* normally distributed with a vector of zeros as the mean and a diagonal covariance matrix with hyper-parameter λ_1 on the diagonal. A similar prior distribution is assumed for the constant terms of the GARCH process, with the difference that for ω the distribution is truncated to the constrained parameter space. The parameters modeling the dynamic part of the GARCH process, collected in matrices A and B follow a truncated normally-distributed prior with zero mean and diagonal covariance matrix with hyper-parameter λ_2 on the diagonal. The truncation of the distribution to the parameter space imposes the non-negativity and stationarity conditions. Each of the correlation parameters of the correlation matrix $\mathbf{C}$ follows a marginal uniform distribution on the interval $[-1, 1]$ (see Barnard, McCulloch & Meng, 2000, for the implications of assuming such a marginal prior distribution for the joint distribution of the correlation matrix). Finally the prior distribution proposed by Deschamps (2006) is assumed for the degrees of freedom parameter. To summarize, the prior specification for the considered model has a detailed form of:

$$p(\theta) = p(\alpha)p(\omega, A, B)p(\nu) \prod_{i=1}^{N(N-1)/2} p(\rho_i), \tag{5}$$

where each of the prior distributions is assumed:

$$\begin{aligned}
\alpha &\sim \mathcal{N}^{N+pN^2}\left(\mathbf{0}, \lambda_1 \cdot I_{N+pN^2}\right) \\
\omega &\sim \mathcal{N}^{N}\left(\mathbf{0}, \lambda_1 \cdot I_N\right) \mathcal{I}(\omega > 0) \\
(\mathrm{vec}(A)', \mathrm{vec}(B)')' &\sim \mathcal{N}^{N+N^2(q+r)}\left(\mathbf{0}, \lambda_2 \cdot I_{N+N^2(q+r)}\right) \mathcal{I}\left(A, B \geq 0 \wedge \text{ eigenvalue}(A+B) < 1\right) \\
\nu &\sim .04 \exp\left[-.04(\nu-2)\right] \mathcal{I}(\nu \geq 2) \\
\rho_i &\sim \mathcal{U}(-1, 1) \qquad \text{for } i = 1, \ldots, N(N-1)/2,
\end{aligned}$$

where $\alpha = (\alpha_0', \mathrm{vec}(\alpha_1)', \ldots, \mathrm{vec}(\alpha_p)')'$ stacks all the parameters of the VAR process in a vector of size $N + pN^2$. I_n is an identity matrix of order n. $\mathcal{I}(.)$ is an indicator function taking value equal to 1 if the condition in the brackets holds and 0 otherwise. ρ_i is the ith element of a vector stacking all the elements below the main diagonal of the correlation matrix, $\rho = (\mathrm{vecl}(\mathbf{C}))$. The values of hyper-parameters, (λ_1, λ_2), are to be specified by the investigator. Finally, the K-dimensional vector collecting all of the parameters of the model, where $K = N(2.5 + N(p + 2.5))$, is as follows:

$$\theta = (\alpha', \omega', \mathrm{vec}(A)', \mathrm{vec}(B)', \rho', \nu)' . \tag{6}$$

2. Bayesian estimation and inference

The posterior distribution of the parameters of the model is proportional to the product of the likelihood function (4) and the prior distribution of the parameters:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta). \tag{7}$$

Estimation of models. The form of the posterior distribution (7) for all of the parameters, θ, for the GARCH models, even with the prior distribution set to a proper distribution function, as in (5), is not in a form of any known distribution function. Moreover, none of the full conditional densities for any sub-group of the parameter vector has a form corresponding to a standard distribution. Still, the posterior distribution, although it is known only up to a normalizing constant, exists; this is ensured by the bounded likelihood function and the proper prior distribution. Therefore, the posterior distribution may be simulated with a Monte Carlo Markov Chain (MCMC) algorithm. Due to the above mentioned problems with the form of the posterior and full conditional densities, a proper algorithm to sample the posterior distribution (7) is, e.g. the Metropolis-Hastings algorithm (see Chib & Greenberg, 1995, and references therein). The algorithm was adapted for multivariate GARCH models by Vrontos, Dellaportas & Politis (2003).

Suppose the starting point of the Markov Chain is some value $\theta_0 \in \Theta$. Let $q(\theta^{(s)}, \theta'|\mathbf{y}, \mathcal{M}_i)$ denote the proposal density (candidate-generating density) for the transition from the current state of the Markov chain $\theta^{(s)}$ to a candidate draw θ'. The candidate density for model $\mathcal{M}_i$ depends on the data $\mathbf{y}$. In this study, a multivariate Student t distribution is used, with the location vector set to the current state of the Markov chain, $\theta^{(s)}$, the scale matrix $c\Omega_q$ and the degrees of freedom parameter set to five. The matrix, Ω_q, should be determined by preliminary runs of the MCMC algorithm, such that it is close to the covariance matrix of the posterior distribution, whereas c is a scalar that is used to achieve the requested acceptance rate in the Markov chain. Such a candidate-generating density should enable the algorithm to draw efficiently from the posterior density. A new candidate θ' is accepted with the probability:

$$\alpha(\theta^{(s)}, \theta'|\mathbf{y}, \mathcal{M}_i) = \min\left[1, \frac{p(\mathbf{y}|\theta', \mathcal{M}_i)p(\theta'|\mathcal{M}_i)}{p(\mathbf{y}|\theta^{(s)}, \mathcal{M}_i)p(\theta^{(s)}|\mathcal{M}_i)}\right], \tag{8}$$

and if it is rejected, then $\theta^{(s+1)} = \theta^{(s)}$. The sample drawn from the posterior distribution with the Metropolis-Hastings algorithm, $\{\theta^{(s)}\}_{s=1}^{S}$, should be diagnosed to ensure that it is a good sample from the stationary posterior distribution.

3. Estimation of marginal data densities

Having estimated the models, the marginal densities of the data (MDD) may be computed using one of the available methods. Since the estimation of the models is performed using the Metropolis-Hastings algorithm, a suitable estimator of the MDD is the one introduced by Chib & Jeliazkov (2001). This estimator of the MDD is computed using the so called marginal likelihood identity:

$$\log p\left(\mathbf{y}|\mathcal{M}_i\right) = \log p\left(\mathbf{y}|\theta^*, \mathcal{M}_i\right) + \log p\left(\theta^*|\mathcal{M}_i\right) + \log p\left(\theta^*|\mathbf{y}, \mathcal{M}_i\right), \tag{9}$$

where θ^* is a point in the parameter space for which the posterior density has high value (posterior means are used in the supplied algorithms), and on the right hand side of the formula above there are the ordinates of the likelihood function, the prior density function and the posterior density function respectively evaluated at θ^*. The ordinate of the posterior distribution evaluated at θ^* is estimated by:

$$\hat{p}\left(\theta^*|\mathbf{y},\mathcal{M}_i\right) = \frac{S^{-1}\sum_{s=1}^{S}\alpha(\theta^{(s)},\theta^*|\mathbf{y},\mathcal{M}_i)q(\theta^{(s)},\theta^*|\mathbf{y},\mathcal{M}_i)}{J^{-1}\sum_{j=1}^{J}\alpha(\theta^*,\theta^{(j)}|\mathbf{y},\mathcal{M}_i)}, \tag{10}$$

where $\{\theta^{(j)}\}_{j=1}^{J}$ is a sample drawn from $q(\theta^*,\theta'|\mathbf{y},\mathcal{M}_i)$.

The marginal data densities can be further used for hypotheses assessment (see e.g. Kass & Raftery, 1995). Woźniak (2015) presents their use for the assessment of the hypotheses of second-order Granger causality for the ECCC-GARCH models.

References

Barnard, J., McCulloch, R., & Meng, X.-l. (2000). Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Applicationto Shrinkage. *Statistica Sinica*, *10*, 1281–1311.

Bollerslev, T. (1990). Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model. *The Review of Economics and Statistics*, *72*, 498–505.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, *49*, 327.

Chib, S., & Jeliazkov, I. (2001). Marginal Likelihood from the Metropolis-Hastings Output. *Journal of the American Statistical Association*, *96*, 270– 281.

Deschamps, P. J. (2006). A Flexible Prior Distribution for Markov Switching Autoregressions with Student-t Errors. *Journal of Econometrics*, *133*, 153–190.

He, C., & Teräsvirta, T. (2004). An Extended Constant Conditional Correlation Garch Model and Its Fourth-Moment Structure. *Econometric Theory*, *20*, 904–926.

Jeantheau, T. (1998). Strong Consistency of Estimators for Multivariate ARCH Models. *Econometric Theory*, *14*, 70–86.

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*, 773–795.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, *48*, 1–48.

Vrontos, I. D., Dellaportas, P., & Politis, D. N. (2003). Inference for Some Multivariate ARCH and GARCH Models. *Journal of Forecasting*, *22*, 427–446.

Woźniak, T. (2015). Testing Causality Between Two Vectors in Multivariate GARCH Models. *International Journal of Forecasting*, *31*, 876–894. DOI: http://dx.doi.org/10.1016/j.ijforecast.2015.01.005.

Woźniak, T. (in press). Granger-Causal Analysis of GARCH Models: a Bayesian Approach. *Econometric Reviews*, . DOI: http://dx.doi.org/10.1080/07474938.2015.1092839.

Description

A random-walk Metropolis-Hasting sampler for VAR(p) ECCC-GARCH(1,1) models.

Usage

```
MH(S, data, par0, sigma0, C=1, restrictions=NULL, lag=1,
           hyper.parameters = c(100,0.1), print.iterations=100)
```

Arguments

`S`	An integer specifying the number of iterations of the MH algorithm.
`data`	A $T \times N$ matrix of data.
`par0`	A K vector of admissible starting values for the parameters. The ordering of parameters in this vector is specified in equation (6).
`sigma0`	A $K \times K$ symmetric positive-definite matrix determining the scale matrix of the candidate density. This matrix corresponds to matrix Ω_q.
`C`	A positive scalar determining the scale matrix of the candidate density. This matrix corresponds to a scaling constant c.
`restrictions`	A K vector determining zero restrictions to be imposed on the parameters. Its elements equal to 1 set the zero restriction on the corresponding parameter in vector θ, whereas elements equal to 0 leave the corresponding parameters unrestricted. If equal to `NULL` then the unrestricted model is estimated.
`lag`	A positive integer specifying the lag order, p of the VAR model.
`hyper.parameters`	A vector with positive elements (λ_1, λ_2).
`print.iterations`	An integer specifying how often should the algorithm's iteration count be displayed.

Value

A list with the following arguments:

`KERNEL`	A S-vector containing the values of the kernel of the posterior distribution evaluated at each draw.
`LIKELI`	A S-vector containing the values of the likelihood function evaluated at each draw.
`PRIOR`	A S-vector containing the values of the prior distribution evaluated at each draw.
`rejections`	A S-vector containing the numbers of draws sampled from outside of the parameters space and rejected at each iteration.
`THETA`	A $S \times K$ matrix containing the MCMC draws.
`time`	The total time, in hours, of executing the simulation.

Author

Tomasz Woźniak

References

Woźniak, Tomasz (2015) Testing Causality Between Two Vectors in Multivariate GARCH Models, *International Journal of Forecasting*, **31**(3), pp. 876–894..

Woźniak, Tomasz (in press) Granger-Causal Analysis of GARCH Models: a Bayesian Approach, *Econometric Reviews*.

`ml.cj2001` *Marginal Data Density for ECCC-GARCH models*

Description

Estimates the marginal data density for VAR(p) ECCC-GARCH(1,1) models with the estimator of Chib & Jeliazkov (2001).

Usage

```
ml.cj2001(mcmc, kernel, rej=NULL, data, restrictions=NULL, sigma0, lag=1)
```

Arguments

`mcmc`	A $S \times K$ matrix containing the MCMC draws. It corresponds to object `THETA` from the output of function `MH`.
`kernel`	A S-vector containing the values of the kernel of the posterior distribution evaluated at each draw. It corresponds to object `KERNEL` from the output of function `MH`.
`rej`	A S-vector containing the numbers of draws sampled from outside of the parameters space and rejected at each iteration. It corresponds to object `rejections` from the output of function `MH`. If equal to `NULL` then the rejections are not taken into account in the estimation.
`data`	A $T \times N$ matrix of data.
`restrictions`	A K vector determining zero restrictions to be imposed on the parameters. Its elements equal to 1 set the zero restriction on the corresponding parameter in vector θ, whereas elements equal to 0 leave the corresponding parameters unrestricted. If equal to `NULL` then the MDD for the unrestricted model is estimated.
`sigma0`	A $K \times K$ symmetric positive-definite matrix determining the scale matrix of the candidate density. This matrix corresponds to matrix Ω_q.
`den`	A positive integer specifying the number of additional draws required for the computations. It corresponds to J from equation (10).
`lag`	A positive integer specifying the lag order, p of the VAR model.

Value

A scalar with the estimate of the natural logarithm of the marginal data density.

Author

Tomasz Woźniak

References

Chib, S., & Jeliazkov, I. (2001) Marginal Likelihood from the Metropolis-Hastings Output, *Journal of the American Statistical Association*, **96**(453), 270– 281.

Woźniak, T. (2015) Testing Causality Between Two Vectors in Multivariate GARCH Models, *International Journal of Forecasting*, **31**(3), pp. 876–894..

Woźniak, T. (in press) Granger-Causal Analysis of GARCH Models: a Bayesian Approach, *Econometric Reviews*.