Multivariate GARCH and Stochastic Volatility Models: estimation and model selection

Antonello Loddo

University of Perugia

October 14, 2004
Preface

VAR, GARCH and Stochastic Volatility Models (SVM) are important and powerful tools for today’s econometrics. Some issues still need to be challenged though, both on estimation and model selection. We derive an MCMC algorithm to find estimates for multivariate autoregressive models with stochastic volatilities. A Stochastic Search Model Selection algorithm allows efficient and flexible selection of the best subset of explanatory variables, which still represents a problem for multivariate models. Using a Modified Cholesky decomposition of the covariance matrix, and applying the Stochastic Search model selection, we are also able to identify restrictions on the covariance matrix itself.
Summary

• Background
  – VAR models
  – GARCH models
  – Stochastic Volatility models

• Our model: VAR with Stochastic Volatilities
  – the model
  – Stochastic Search Model Selection
  – estimation

• Conclusions
Background – VAR models

- historical importance of VAR models.
  - Vector Auto Regressive (VAR) modeling became widely used after the seminal work of Sims (1972, 1980), representing a huge breakthrough in the macroeconomic and econometric fields.
  - Sims criticized the "a priori" distinction between endogenous and exogenous variables typical of the preceding structural equation modeling (SE).
  - We can summarize his thought with a quote: "It should be feasible to estimate large scale macromodels as unrestricted reduced forms, treating all the variables as endogenous. Of course some restrictions, if only on lag-length, are essential..."
• How a VAR model look like:

$$y'_t = z'_t C + \sum_{j=1}^{L} y'_{t-j} A_j + \epsilon'_t,$$

for $t = 1, \cdots, T$, where $z_t$ is a $h$-dimensional vector of essential exogenous variable, $C$ is $h \times p$, $A_j$ is $p \times p$, $L$ is a known positive integer, $\epsilon_1, \cdots, \epsilon_T$ are independent identically distributed $N_p(0, \Sigma)$.

• It is often the case that $z_t$ is zero.
VAR estimation: frequentist approach

• In applications, OLS or MLE are often used to find estimates of $\Sigma$ and of $\Phi = (c', A'_1, \cdots, A'_L)'$.

• Those estimates are not free of problems though:
  – frequentist finite sample distributions of OLS or ML estimators for $\Sigma$ and of $\Phi$ are unavailable
  – on the other hand, a typical VAR involves a large number of parameters, and the sample size of data is often not large enough to justify the use of asymptotic theory
  – also, when nonlinear functions of the VAR coefficients (such as impulse response functions) are of interest, the asymptotic theory involves approximation of nonlinear functions
VAR estimation: Bayesian approach

- An alternative to asymptotic theory is represented by the Bayesian approach:
  - Bayesian approach combines a priori information and information from the data to form a finite sample posterior distribution
  - conjugate analysis allows to derive full conditionals for all the parameters of interest, making Gibbs sampler (Gelfand and Smith (1990)) a very feasible option
  - when a priori information is not available, or difficult to be parametrized, non-informative priors can be used
  - Dongchu Sun and Shawn Ni studied several alternative Bayesian procedures for estimating VAR models
ARCH and GARCH models: introduction

- In economical and financial data analysis, sometimes, the assumption of a constant variance is unrealistic: here the necessity of model the variance of a time series as well.

- Autoregressive Conditional Eteroschedasticity (ARCH) models have been first introduced with the seminal paper of Engle (1982), while GARCH models are a generalized version of ARCH models due to Bollersev (1986). For his contribution, Engle has been awarded of the Nobel prize for economics in 2003.

- Several multivariate extensions of ARCH-GARCH models appeared in the 1990s.
ARCH and GARCH models

• Conditional to the information available at time t, call it $\mathcal{F}_t$, an univariate GARCH(g,q) can be written as follows:

$$\varepsilon_t \mid \mathcal{F}_t \sim N(0, \sigma^2_t),$$

where the conditional variance is

$$\sigma^2_t = \alpha_0 + \beta_1 \sigma^2_{t-1} + \ldots, + \beta_g \sigma^2_{t-g} + \alpha_1 \varepsilon^2_{t-1} + \ldots + \alpha_q \varepsilon^2_{t-q}.$$  

– ARCH models would have all the beta terms set to zero, modeling the conditional variance as a function only of the errors.

– an advantage of the GARCH representation is that often a GARCH(1,1) can estimate the conditional variance as well as ARCH models with a much larger number of lags.
ARCH and GARCH models: estimation

- ML is the most common way to estimate GARCH models:
  - if the mean of the process is constant, ML estimation of univariate GARCH models is feasible
  - asymptotic theory for GARCH models suffers the same problems we saw for VAR, and it is generally hard to derive
  - Unconstrained multivariate GARCH are not estimable, so most of the frequentist literature focuses on finding realistic constraints that allows the model to be estimated via QMLE or via MLE using EM algorithm or bootstrap techniques

- In the 1990s several Bayesian techniques have been introduced for the estimation of GARCH models. Most of them rely on the Metropolis-Hastings algorithm
Stochastic Volatility Models

- Stochastic Volatility Models can be viewed as an alternative to the ARCH-GARCH representation:
  - they look like stochastic autoregressive process for the volatilities, so they include an error term in the volatility equation
  - they generally are more parsimonious and perform better than GARCH models in presence of high Kurtosis processes (common in financial time series)
  - they present some more estimation difficulties than GARCH models
Stochastic Volatility Models

- We can write a Stochastic Volatility model as follows:

\[ \varepsilon_t \mid \mathcal{F}_t \sim \mathcal{N}(0, \sigma_t^2), \]

where the conditional variance is

\[ \log \sigma_t = \alpha + \beta \log \sigma_{t-1} + \cdots + \log \sigma_{t-q} = h \upsilon_t, \quad \upsilon_t \sim \mathcal{N}(0, 1). \]

- Jacquier, Polson and Rossi (1994) proposed a Metropolis-Hastings algorithm to find Bayesian estimates of SVM.

- Uhlig (1992, 1997) proposes a Gibbs sampling algorithm to find Bayesian estimates of SVM. This algorithm requires though a numerical integration step.
Our model: VAR with Stochastic Volatilities

• Consider the following VAR model:

\[ y_t' = z_t' C + \sum_{j=1}^{L} y_{t-j}' A_j + \epsilon_t', \]

for \( t = 1, \ldots, T \). \( z_t \) is a vector of exogenous regressors, \( \epsilon_t, t = 1, \ldots, T \) are \( N_p(0, \Sigma_t) \) errors, and \( \Sigma_t \) is an unknown \( p \times p \) positive definite matrix that can be decomposed as

\[ \Sigma_t = \Gamma A_t \Gamma', \]

where \( A_t \) is diagonal and \( \Gamma \) is lower triangular with a diagonal of ones. This decomposition is known as Modified Cholesky Decomposition (see, for example, Pourahmadi (1999)).
Each element of $\mathbf{A}_t$, $\lambda_{tj}, j = 1, \ldots, p$ is modeled as:

$$\log \lambda_{tj} = \alpha_j + \beta_j \log \lambda_{t-1,j} + \sigma_j^2 \upsilon_{tj}$$

where

$$\upsilon_{tj} \sim \mathcal{N}(0, 1)$$

Here the logarithm of the conditional volatility follows an autoregressive times series model; unlike the ARCH and GARCH models, both mean and log-volatility equations have separate error terms.
Model selection

- Model selection plays a crucial role in VAR models:
  - For a VAR model with $p$ variables, $h$ exogenous variables (including a constant) and $L$ lags there are $2(Lp+h)p + \frac{p(p-1)}{2}$ competing models.
  - In 1990s Edward George and Robert McCulloch (1993) proposed a data driven Bayesian MCMC stochastic search algorithm for multiple regression that greatly reduced the amount of computation.
  - Often identification VARs economists are interested in identifying restrictions for $\Sigma$ and not much concerned about selection of regressors.
  - Stochastic Search Model Selection can be extended to be applied in our model, both for the regression coefficients and for the covariance matrices.
Model setup

- Define $x_t' = (z_t', y_{t-1}', \ldots, y_{t-L}')$. We then rewrite our model in the matrix form

$$Y = X\Phi + \epsilon,$$

where

$$Y = \begin{pmatrix} y_1' \\ \vdots \\ y_T' \end{pmatrix}, \quad X = \begin{pmatrix} x_1' \\ \vdots \\ x_T' \end{pmatrix}, \quad \Phi = \begin{pmatrix} c \\ A_1 \\ \vdots \\ A_L \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1' \\ \vdots \\ \epsilon_T' \end{pmatrix}.$$
• Also consider that

\[ \Sigma_t^{-1} = \Psi' \Lambda_t^{-1} \Psi, \]

where \( \Psi = \Gamma^{-1} \). \( \Psi \) has the same lower triangular form of \( \Gamma \). One of the great advantages of the Modified Cholesky decomposition is that \( \Psi \) is an unrestricted matrix, i.e. it does not have the constraint to be positive definite. This will allow to simplify enormously the inference on the covariance matrix.
How does the Stochastic Search work?

- Let \( m = (Lp + h)p \), the total number of unknown regression coefficients. Denote \( \phi = \text{vec}(\Phi) \). For given \( \gamma = (\gamma_1, \gamma_2, \cdots, \gamma_m) \), denote \( D = \text{diag}(h_1 \tau_1, \cdots, h_m \tau_m) \) and

\[
h_i = c_i^{\gamma_i} = \begin{cases} 
1, & \text{if } \gamma_i = 0, \\
c_i, & \text{if } \gamma_i = 1.
\end{cases}
\]

\( \tau_i \) are small and \( c_i \) are large constants. Here

\[
(\phi | \gamma) \sim N_m(0, DRD),
\]

where \( R \) is a known \( m \times m \) correlation matrix, that can be set to \( I \).
• If $\gamma_i = 0$, then the prior probability for the correspondent element $\phi_i$ will be concentrate around zero, while if $\gamma_i = 1$, then we will have the usual diffuse prior (normal with a very high variance) for the parameter.

• If $\gamma$ was set a priori, then we were just putting constraints to the model, but instead we are considering each $\gamma_i$ as a binary variable and we are computing its posterior distribution, making the shrinking towards zero of some elements of $\Phi$ completely data driven.

• An analogous technique it is been applied to the elements of $\Psi$. 
• So we can write a hierarchical model: for $t = 1, \cdots, T$

$$y_t' \mid \Lambda_t, \phi, \Psi \sim N(x_t' \Phi; \Gamma \Lambda_t \Gamma');$$

for each element of $\Lambda_t = \lambda_{t1}, \cdots, \lambda_{tp}$

$$\log \lambda_{tj} \mid \alpha_j, \beta_j, \sigma_j \sim N(\alpha_j + \beta_j \log \lambda_{t-1,j}; \sigma_j^2);$$

$$\phi \mid \gamma \sim N_m(0; DRD);$$

for $i = 1, \cdots, m$

$$\gamma_i \sim Bernoulli(p_i);$$
• Let $\eta_j$ be a vector containing the non redundant elements of
the $j^{th}$ column of $\Psi'$, i.e. $\eta_j = (\psi_{j1}, \cdots, \psi_{jj-1})'$. Then

$$(\eta_j \mid \omega_j) \overset{ind}{\sim} N_{j-1}(0, D_j R_j D_j), \text{ for } j = 2, \cdots, p,$$

where $R_j$ is $(j - 1) \times (j - 1)$ known. Here $\omega$ plays the same role for $\Psi$ that $\gamma$ does for $\Phi$;

for $i = 1, \cdots, p$, $j = 1, \cdots, p - 1,$

$$\omega_{ij} \sim Bernoulli(q_{ij}).$$

• Let $\delta_j' = (\alpha_j, \beta_j)$. For the volatility parameters we are using a noninformative prior distribution

$$[\delta_j, \sigma_j] \propto \sigma_j^{-2}. $$
Posterior

- We build an MCMC algorithm to sample from the single conditional posterior distributions. We derived closed forms for the conditional posteriors of $\Phi, \eta, \delta, \gamma, \omega$, so we can sample from them using the well known Gibbs sampler (Gelfand and Smith (1990)).

- The full conditional distributions of the lambdas do not have closed form: this is why most of the literature uses a Metropolis-Hasting algorithm to sample from lambda.
  - Since the logarithm of the posterior density of every single lambda, though, can be shown to be concave, we can use the more efficient adaptive rejection sampling algorithm given by Gilks and Wild (1992), called Gilks sampler.
**Algorithm**

- For cycle $k$:
  - draw $(\alpha_{(k)}, \beta_{(k)}) \mid \lambda_{11,(k-1)}, \cdots, \lambda_{Tp,(k-1)} \phi_{(k-1)}$;
  - draw $(\sigma_{1,(k)}, \cdots, \sigma_{p,(k)} \mid \lambda_{11,(k-1)}, \cdots, \lambda_{Tp,(k-1)} \phi_{(k-1)})$ separately for each $j$;
  - draw $(\lambda_{11,(k)}, \cdots, \lambda_{Tp,(k)} \mid \alpha_{k-1}, \beta_{k-1}, \phi_{(k-1)})$ using the Gilks sampler (we need to sample separately each lambda);
  - draw $(\eta_{j,(k)} \mid \lambda_{11,(k)}, \cdots, \lambda_{Tp,(k)}, \phi_{(k-1)}, \gamma_{(k-1)}, \omega_{(k-1)}, Y)$ separately for each $j$;
  - draw $(\omega_{j,(k)} \mid 
    \eta_{(k)}, \lambda_{11,(k)}, \cdots, \lambda_{Tp,(k)}, \phi_{(k-1)}, \gamma_{(k-1)}, \omega_{(k-1)}, Y)$
  - draw $(\phi_{(k)} \mid \gamma_{(k-1)}, \Sigma_{t,(k)}, \omega_{(k)}; Y)$;
  - draw $(\gamma_{(k)} \mid \phi_{(k)}, \lambda_{11,(k)}, \cdots, \lambda_{Tp,(k)}, \eta_{(k)}, \omega_{(k)}, Y)$. 

22
Conclusions

- We derive an MCMC algorithm to find bayesian estimates of multivariate autoregressive models with stochastic volatilities, and to perform a Stochastic Search Model Selection to efficiently identify restrictions on $\Phi$ and $\Sigma_t$:
  - even if Metropolis-Hastings is the most used algorithm to sample from the volatility distributions, it generally shows very low acceptance ratios in this kind of problem. Gilks sampler is much more efficient. This is crucial, since we have to sample separately from each lambda, so we need $T \times p$ separate samplers just for the volatilities
  - dimensionality of the model requires an efficient algorithm
- The stochastic search model selection is embedded in the MCMC algorithm, allowing us to identify nonzero elements of the matrix of regressors $\Phi$ and of the covariance matrices $\Sigma_t$.

- Our model selection is very flexible: while the most used techniques just find the minimum number of lags of the model, ours allows different lag patterns for each dependent variable.

- The model could be generalized using an unconstrained time variant covariance matrix, but in that case dimensionality of the hyperparameter can represent a big problem.

- We can generalize the Stochastic Volatility equation by introducing exogenous variables. This will allow to build a very general model, of which of both GARCH and Stochastic Volatility models are just special cases.