Multivariate Time Series Models

May 30, 2010

Junhui Qian

1 Introduction

In this chapter we consider vector-valued stochastic processes. We discuss VAR (Vector AutoRegression), Structural VAR, and multivariate conditional variance-covariance models.

2 VAR

We consider an r-dimensional vector autoregression (VAR) of the following form,

$$X_t = A_1 X_{t-1} + \cdots + A_p X_{t-p} + \varepsilon_t, \tag{1}$$

where (A_i) are VAR coefficient matrices and $\varepsilon_t \sim \mathrm{WN}(0, \Sigma)$. We call the above model p-th order VAR model or VAR(p) model.

We may represent the model in (1) as

$$A(L)X_t = \varepsilon_t,$$

where L is lag operator and $A(z) = I - A_1 z - \cdots - A_p z^p$ is a matrix of polynomials.

It is also useful to write the model in AR(1) form,

$$X_t^* = A X_{t-1}^* + \varepsilon_t^*, \tag{2}$$

where

$$X_t^* = \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{pmatrix}, \quad A = \begin{pmatrix} A_1 & \cdots & A_{p-1} & A_p \\ I & & & \\ & \ddots & & \\ & & I & \end{pmatrix}, \quad \varepsilon_t^* = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The eigenvalues of the matrix A satisfies

$$|\lambda^p I - \lambda^{p-1} A_1 - \cdots - A_p| = 0.$$

For the covariance stationarity of (X_t), all eigenvalues of A should be within the unit circle, so that any shock in ε_t eventually die out. This condition is equivalent to the one that requires all roots of $|A(z)| = 0$ lie outside the unit circle.

When the above condition holds, (X_t) has an MA(∞) representation, $X_t = \Phi(L)\varepsilon_t$, where $\Phi(z) = \sum_{i=0}^{\infty} \Phi_i z^i$ with Φ_i satisfying $\sum_{i=1}^{\infty} |\Phi_i| < \infty$. $|\cdot|$ here denotes any matrix norm. The MA coefficients can be obtained from the power series expansion of $A(z)^{-1}$, which exists on the unit disk in the complex plane since it is analytic.

2.1 Maximum Likelihood Estimation

Let Π be an $r \times (rp)$ matrix of parameters defined as

$$\Pi = [A_1 \ A_2 \ \cdots \ A_p].$$

If we define $Z_t = [X'_{t-1} X'_{t-2} \cdots X'_{t-p}]'$, we may write the original model in (1) as

$$X_t = \Pi Z_t + \varepsilon_t.$$

The conditional likelihood of X_t is

$$p(X_t, \theta|\mathcal{F}_{t-1}) = (2\pi)^{r/2} \left|\Omega^{-1}\right|^{1/2} \exp\left(-\frac{1}{2}(X_t - \Pi Z_t)'\Omega^{-1}(X_t - \Pi Z_t)\right),$$

where θ is the vector of parameters. The likelihood for the full sample conditional on $(X_0, X_{-1}, \dots, X_{1-p})$ is thus given by

$$p(X_T, \dots, X_t, \theta) = \prod_{t=1}^{T} p(X_t, \theta|\mathcal{F}_{t-1}).$$

The log conditional likelihood to be maximized is

$$\mathcal{L} = -\frac{Tr}{2}\log(2\pi) + \frac{T}{2}\log\left(|\Omega^{-1}|\right) + \frac{1}{2}\sum_{t=1}^{T}(X_t - \Pi Z_t)'\Omega^{-1}(X_t - \Pi Z_t). \tag{3}$$

We claim that the MLE of Π is the same as the OLS estimator:

$$\hat{\Pi} = \left(\sum_{t=1}^{T} X_t Z_t'\right)\left(\sum_{t=1}^{T} Z_t Z_t'\right)^{-1}.$$

To show this, first note that the MLE of Π shall minimize the sum in the last term in (3), which can be rewritten as

$$\begin{aligned}
& \sum_{t=1}^{T}(X_t - \Pi Z_t)'\Omega^{-1}(X_t - \Pi Z_t) \\
= & \sum_{t=1}^{T}(X_t - \hat{\Pi}Z_t + \hat{\Pi}Z_t - \Pi Z_t)'\Omega^{-1}(X_t - \hat{\Pi}Z_t + \hat{\Pi}Z_t - \Pi Z_t) \\
= & \sum_{t=1}^{T}(\hat{\varepsilon}_t + (\hat{\Pi} - \Pi)Z_t)'\Omega^{-1}(\hat{\varepsilon}_t + (\hat{\Pi} - \Pi)Z_t) \\
= & \sum_{t=1}^{T}\hat{\varepsilon}_t'\Omega^{-1}\hat{\varepsilon}_t + 2\sum_{t=1}^{T}\hat{\varepsilon}_t'\Omega^{-1}(\hat{\Pi} - \Pi)Z_t \\
& \quad + \sum_{t=1}^{T} Z_t'(\hat{\Pi} - \Pi)'\Omega^{-1}(\hat{\Pi} - \Pi)Z_t.
\end{aligned} \tag{4}$$

The middle term in (4) is zero, since

$$\begin{aligned}\sum_{t=1}^{T}\hat{\varepsilon}_t'\Omega^{-1}(\hat{\Pi}-\Pi)Z_t &= \operatorname{tr}\left(\sum_{t=1}^{T}\hat{\varepsilon}_t'\Omega^{-1}(\hat{\Pi}-\Pi)Z_t\right)\\ &= \operatorname{tr}\left(\sum_{t=1}^{T}\Omega^{-1}(\hat{\Pi}-\Pi)Z_t\hat{\varepsilon}_t'\right)\\ &= \operatorname{tr}\left(\Omega^{-1}(\hat{\Pi}-\Pi)\sum_{t=1}^{T}Z_t\hat{\varepsilon}_t'\right)\\ &= 0.\end{aligned}$$

where the last equality is due to the OLS first-order condition. The last term in (4), a non-negative quadratic term, is thus the only one that involves Π. It is now clear that $\hat{\Pi}$ is MLE of Π.

It can also be shown that the MLE of Ω is given by

$$\hat{\Omega}=\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t',$$

where

$$\hat{\varepsilon}_t = X_t-\hat{\Pi}Z_t.$$

2.2 Likelihood Ratio Test

The maximized log likelihood is given by

$$\begin{aligned}\mathcal{L} &= -\frac{Tr}{2}\log(2\pi)+\frac{T}{2}\log\left(|\Omega^{-1}|\right)-\frac{1}{2}\sum_{t=1}^{T}\hat{\varepsilon}_t'\hat{\Omega}^{-1}\hat{\varepsilon}_t\\ &= -\frac{Tr}{2}\log(2\pi)+\frac{T}{2}\log\left(|\Omega^{-1}|\right)-\frac{1}{2}\operatorname{tr}\left[\sum_{t=1}^{T}\hat{\varepsilon}_t'\hat{\Omega}^{-1}\hat{\varepsilon}_t\right]\\ &= -\frac{Tr}{2}\log(2\pi)+\frac{T}{2}\log\left(|\Omega^{-1}|\right)-\frac{1}{2}\operatorname{tr}\left[\sum_{t=1}^{T}\hat{\Omega}^{-1}\hat{\varepsilon}_t\hat{\varepsilon}_t'\right]\\ &= -\frac{Tr}{2}\log(2\pi)+\frac{T}{2}\log\left(|\Omega^{-1}|\right)-\frac{1}{2}\operatorname{tr}\left[\hat{\Omega}^{-1}(T\hat{\Omega})\right]\end{aligned}$$

$$= -\frac{Tr}{2}\log(2\pi) + \frac{T}{2}\log\left(|\Omega^{-1}|\right) - \frac{Tr}{2}.$$

Let the empirical covariance matrix under restriction be $\hat{\Omega}_0$ and that under no restriction be $\hat{\Omega}_1$. The likelihood ratio test statistic is given by

$$2(\mathcal{L}_1 - \mathcal{L}_0) = T\left(\log\left|\hat{\Omega}_0\right| - \log\left|\hat{\Omega}_1\right|\right).$$

Under the null hypothesis, this statistic has an asymptotic distribution of χ^2_m, where m is the number of restrictions.

2.3 Granger Causality Test

We say (X_t) does *not* Granger cause (Y_t) if and only if for all $m > 0$ the mean squared error of forecasting Y_{t+m} based on $(Y_t, Y_{t-1}, \ldots)$ does not exceed that based on $(X_t, X_{t-1}, \ldots, Y_t, Y_{t-1}, \ldots)$.

To test the Granger causality (or more precisely, non-causality) in VAR framework, we write,

$$\begin{pmatrix} Y_{t+1} \\ X_{t+1} \end{pmatrix} = \begin{pmatrix} a_{1,11} & a_{1,12} \\ a_{1,21} & a_{1,22} \end{pmatrix} \begin{pmatrix} Y_t \\ X_t \end{pmatrix} + \cdots + \begin{pmatrix} a_{p,11} & a_{p,12} \\ a_{p,21} & a_{p,22} \end{pmatrix} \begin{pmatrix} Y_{t-p+1} \\ X_{t-p+1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \tag{5}$$

Let the null hypothesis be the assertion that X does not Granger cause Y. The null and the alternative hypotheses are stated as follows,

$$\begin{aligned} \mathrm{H}_0 &: a_{1,12} = \cdots = a_{p,12} = 0 \\ \mathrm{H}_1 &: a_{i,12} \neq 0 \text{ for some } i = 1, ..., p. \end{aligned}$$

We may employ the usual F test,

$$F = \frac{(SSR_0 - SSR_1)/p}{SSR_0/(T-2p-1)},$$

where SSR_0 and SSR_1 are, respectively, restricted and unrestricted sums of squared errors.

The Granger causality test is sensitive to the choice of p, the order of autoregression. It is thus necessary to select p in an objective way. For example, we may select p that minimizes some information criterion (e.g., AIC) before conducting the Granger causality test.

3 Structural VAR

The structural VAR explicitly allows contemporary relation between variables. We write the model as

$$BX_t = B_1 X_{t-1} + \cdots + B_p X_{t-p} + e_t, \tag{6}$$

where the covariance matrix of e_t, Λ, is a diagonal matrix. Compare the model with that in (1), where the covariance matrix of ε_t is generally non-diagonal. We usually call the model in (1) reduced-form VAR and that in (6) structural-form VAR. Correspondingly, the residual vector ε_t in (1) is called reduced-form error, and e_t in (6) the structural innovation. The reduced-form error and the structural innovation are obviously related by

$$B\varepsilon_t = e_t.$$

Without restriction, the SVAR in (6) is not identified. That is, two SVAR with different parameter values may reduce to the same reduced-form VAR, which implies the same data generating process for (X_t). Putting it another way, different SVAR's may be observationally equivalent. For example, consider the following two SVAR's,

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} X_t = \begin{pmatrix} .5 & -1 \\ .5 & -1.5 \end{pmatrix} X_{t-1} + e_t, \quad \Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & .5 \\ 0 & 1 \end{pmatrix} X_t = \begin{pmatrix} .5 & -1.25 \\ 0 & -0.5 \end{pmatrix} X_{t-1} + e_t, \quad \Lambda = \begin{pmatrix} .5 & 0 \\ 0 & 2 \end{pmatrix}.$$

Both SVAR's imply the following reduced-form VAR,

$$X_t = \begin{pmatrix} .5 & -1 \\ 0 & -0.5 \end{pmatrix} X_{t-1} + \varepsilon_t, \quad \Sigma = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

To achieve identification, we normally impose restrictions on B. To see how many restrictions are needed for identification, suppose a reduced-form VAR is given with error covariance matrix Σ. Since we have,

$$B\Sigma B' = \Lambda. \tag{7}$$

The fact that Λ is a diagonal matrix generates $r(r-1)/2$ restrictions on B, where r is the dimension of X_t. If there are $r(r-1)/2$ free parameters in the matrix B, then the model is identified. If there are less than $r(r-1)/2$ free parameters, then the model is over-identified. Take the above example, if we specify

$$B = \begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix},$$

then B has one free parameter β. Since there is one restriction, the model is identified as follows,

$$\begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \beta \\ 1 \end{pmatrix} = \beta - 1 = 0.$$

Different specifications of matrix B yields different SVAR.

If B is identified, we may obtain B by solving (7). We may also obtain Λ using (7) and calculate B_i by $B_i = BA_i$.

Impulse Response Function

We may write

$$X_t = \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t-i} = \sum_{i=0}^{\infty} \Pi_i e^*_{t-i},$$

where e^*_t is normalized from e^*_t to have unit variance and we have

$$\Pi_i = \Phi B^{-1} \Lambda^{1/2}.$$

Then the response in period i of the p-th variable to an impulse in the q-th structural innovation is given by (p, q)-th element in Π_i. Note that the unit shock in e^*_t is identical to one standard deviation shock to the corresponding e_t. When B is restricted to a lower triangular matrix with unit diagonals, we have an interesting equality,

$$B = \Lambda^{1/2} L^{-1},$$

where L is a lower triangular matrix from the Cholesky decomposition of Σ. To see this, note that $B\Sigma B' = (BL)(BL)' = \Lambda$ implies $BL = \Lambda^{1/2}$. Then we have

$$\Pi_i = \Phi L.$$

Forecast Error Variance Decomposition

Write the k-step forecast error as

$$X_{t+k} - \mathbb{E}_t X_{t+k} = \sum_{i=0}^{k-1} \Pi_i e^*_{t+k-i}.$$

The forecast error variance of the p-th component of X_{t+k} is then given by

$$\sum_{i=0}^{k-1} \left(\sum_{j=1}^{r} \pi^2_{i,pj} \right),$$

where $\pi_{i,pj}$ is the (p, j)-th element of the matrix Π_i. The q-th component of e_t contributes to the above forecast error variance by

$$\sum_{i=0}^{k-1} \pi_{i,pq}^2.$$

This is called forecast error variance decomposition.

4 Multivariate Volatility Models

The multivariate volatility model improves on VAR by considering time-varying conditional covariance matrix. To be more specific, write

$$X_t = \mu_t + \omega_t,$$

where $\mu_t,\ \omega_t \in \mathbb{R}^d$, $\mu_t = \mathbb{E}(X_t|\mathcal{F}_{t-1})$, and $\Sigma_t = \text{var}(\omega_t|\mathcal{F}_{t-1})$ is a time-varying covariance matrix. As in univariate ARCH/GARCH models, we may represent ω_t as

$$\omega_t = \Sigma_t^{1/2}\varepsilon_t,$$

where ε_t is a vector of white noise with an identity covariance matrix. Multivariate volatility modeling is concerned with the time-varying structure of Σ_t.

4.1 Separable Multivariate GARCH

The simplest case would be the one where Σ_t is diagonal for all t, and each element on the diagonal satisfies one of various GARCH specifications. This treatment is equivalent to the separate modeling of each element in ω_t as an univariate GARCH.

A more realistic model would require that

$$\Sigma_t = D_t C D_t,$$

where C is a constant correlation matrix and D_t is a time-varying diagonal matrix. Each element on the diagonal of D_t may be given a GARCH structure. This model is called a CCC (Constant Conditional Correlation) model.

For example, consider a bivariate CCC model with,

$$C_t = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

where ρ is the constant conditional correlation coefficient.

Furthermore, we may specify that the correlation matrix be time-varying as well. A common specification is given by

$$\rho_t = \frac{\exp(z_t)}{1 + \exp(z_t)},$$

where z_t may depend on some exogenous variables as well as lagged observations of ρ_t, ω_t, and σ_t^2. For example, we may impose a GARCH(1,1) structure on z_t as follows,

$$z_t = c + a\frac{\omega_{1,t-1}\omega_{2,t-1}}{\sigma_{1t}\sigma_{2t}} + b\rho_{t-1},$$

where c, a, and b are constant coefficients.

4.2 General Multivariate GARCH

The diagonal vector error correction model (DVEC) specify Σ_t as

$$\Sigma_t = C + \sum_{i=1}^{p} A_i \odot (\omega_t \omega_t') + \sum_{i=1}^{q} B_i \odot \Sigma_{t-i},$$

where $\odot$ denotes Hadamard product, and C, (A_i) and (B_i) are all symmetric positive definite matrices. We may call the above model as DVEC(p, q) model.

To impose positive definitiveness on Σ_t, we may specify

$$\Sigma_t = CC' + \sum_{i=1}^{p} (A_i A_i') \odot (\omega_t \omega_t') + \sum_{i=1}^{q} (B_i B_i') \odot \Sigma_{t-i},$$

where C, (A_i), (B_i) are all lower triangular matrices. This model may be called "matrix-matrix" DVEC model.

We may define "vector-vector" DVEC by

$$\Sigma_t = CC' + \sum_{i=1}^{p} (a_i a_i') \odot (\omega_t \omega_t') + \sum_{i=1}^{q} (b_i b_i') \odot \Sigma_{t-i},$$

where (a_i) and (b_i) are nonzero vectors. Similarly we may define "scalar-scalar" DVEC and hybrids such as "matrix-vector", "scalar-vector", etc.. The simpler the model is, the more stringent restrictions are placed on the dynamics of the model.

To model richer dynamics in Σ_t, we may consider BEKK model, which is proposed in Engle and Kroner (1995),

$$\Sigma_t = CC' + \sum_{i=1}^{p} A_i (\omega_t \omega_t') A_i' + \sum_{i=1}^{q} B_i \Sigma_{t-i} B_i',$$

where C is lower triangular, and (A_i) and (B_i) are unrestricted square matrices.