Vector Autoregressive Models

by

Helmut Lütkepohl

Department of Economics, European University Institute, Via della Piazzuola 43, I-50133 Firenze, ITALY, email: helmut.luetkepohl@eui.eu

Vector autoregressive (VAR) processes are popular in economics and other sciences because they are flexible and simple models for multivariate time series data. In econometrics they became standard tools when (Sims 1980) questioned the way classical simultaneous equations models were specified and identified and advocated VAR models as alternatives. A textbook treatment of these models with details on the issues mentioned in the following introductory exposition is available in (Lütkepohl 2005).

The model setup

The basic form of a VAR process is

$$y_t = Dd_t + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

where $y_t = (y_{1t}, \ldots, y_{Kt})'$ (the prime denotes the transpose) is a vector of K observed time series variables, d_t is a vector of deterministic terms such as a constant, a linear trend and/or seasonal dummy variables, D is the associated parameter matrix, the A_i's are $(K \times K)$ parameter matrices attached to the lagged values of y_t, p is the lag order or VAR order and u_t is an error process which is assumed to be white noise with zero mean, that is, $E(u_t) = 0$, the covariance matrix, $E(u_t u_t') = \Sigma_u$, is time invariant and the u_t's are serially uncorrelated or independent.

VAR models are useful tools for forecasting. If the u_t's are independent white noise, the minimum mean squared error (MSE) h-step forecast of y_{t+h} at time t is the conditional expectation given y_s, $s \leq t$,

$$y_{t+h|t} = E(y_{t+h}|y_t, y_{t-1}, \ldots) = Dd_{t+h} + A_1 y_{t+h-1|t} + \cdots + A_p y_{t+h-p|t},$$

where $y_{t+j|t} = y_{t+j}$ for $j \leq 0$. Using this formula, the forecasts can be computed recursively for $h = 1, 2, \ldots$. The forecasts are unbiased, that is, the forecast error $y_{t+h} - y_{t+h|t}$ has mean zero and the forecast error covariance is equal to the MSE matrix. The 1-step ahead forecast errors are the u_t's.

VAR models can also be used for analyzing the relation between the variables involved. For example, (Granger 1969) defined a concept of causality which specifies that a variable y_{1t} is causal for a variable y_{2t} if the information in y_{1t} is helpful for improving the forecasts of y_{2t}. If the two variables are jointly

generated by a VAR process, it turns out that y_{1t} is not Granger-causal for y_{2t} if a simple set of zero restrictions for the coefficients of the VAR process are satisfied. Hence, Granger-causality is easy to check in VAR processes.

Impulse responses offer another possibility for analyzing the relation between the variables of a VAR process by tracing the responses of the variables to impulses hitting the system. If the VAR process is stable and stationary, it has a moving average representation of the form

$$y_t = D^* d_t + \sum_{j=0}^{\infty} \Phi_j u_{t-j},$$

where the Φ_j's are $(K \times K)$ coefficient matrices which can be computed from the VAR coefficient matrices A_i with $\Phi_0 = I_K$, the $(K \times K)$ identity matrix. This representation can be used for tracing the effect of a specific forecast error through the system. For example, if $u_t = (1, 0, \ldots, 0)'$, the coefficients of the first columns of the Φ_j matrices represent the marginal reactions of the y_t's. Unfortunately, these so-called *forecast error impulse responses* are often not of interest for economists because they may not reflect properly what actually happens in a system of variables. Given that the components of u_t are typically instantaneously correlated, such shocks or impulses are not likely to appear in isolation. Impulses or shocks of interest for economists are usually instantaneously uncorrelated. They are obtained from the forecast errors, the u_t's, by some transformation, for example, $\varepsilon_t = B u_t$ may be a vector of shocks of interest if the $(K \times K)$ matrix B is such that $\varepsilon_t \sim (0, \Sigma_\varepsilon)$ has a diagonal covariance matrix Σ_ε. The corresponding moving average representation in terms of the ε_t's becomes

$$y_t = D^* d_t + \sum_{j=0}^{\infty} \Theta_j \varepsilon_{t-j},$$

where $\Theta_j = \Phi_j B^{-1}$.

There are many B matrices with the property that $B u_t$ is a random vector with diagonal covariance matrix. Hence, there are many shocks ε_t of potential interest. Finding those which are interesting from an economic point of view is the subject of *structural VAR analysis*.

Estimation and model specification

In practice the process which has generated the time series under investigation is usually unknown. In that case, if VAR models are regarded as suitable, the lag order has to be specified and the parameters have to be estimated. For a given VAR order p, estimation can be conveniently done by equationwise ordinary least squares (OLS). For a sample of size T, $y_1, \ldots, y_T$, and assuming that in addition presample values $y_{-p+1}, \ldots, y_0$ are also available, the OLS estimator of the parameters $B = [D, A_1, \ldots, A_p]$ can be written as

$$\hat{B} = \left(\sum_{t=1}^{T} y_t Z_{t-1}' \right) \left(\sum_{t=1}^{T} Z_{t-1} Z_{t-1}' \right)^{-1},$$

where $Z'_{t-1} = (d'_t, y'_{t-1}, \ldots, y'_{t-p})$. Under standard assumptions the estimator is consistent and asymptotically normally distributed. In fact, if the residuals and, hence, the y_t's are normally distributed, that is, $u_t \sim$ i.i.d.$n(0, \Sigma_u)$, the OLS estimator is equal to the maximum likelihood (ML) estimator with the usual asymptotic optimality properties. If the dimension K of the process is large, then the number of parameters is also large and estimation precision may be low if a sample of typical size in macroeconomic studies is available for estimation. In that case it may be useful to exclude redundant lags of some of the variables from some of the equations and fit so-called subset VAR models. In general, if zero or other restrictions are imposed on the parameter matrices, other estimation methods may be more efficient.

VAR order selection is usually done by sequential tests or model selection criteria. *Akaike's information criterion* (AIC) is, for instance, a popular model selection criterion ((Akaike 1973)). It has the form

$$\text{AIC}(m) = \log \det(\hat{\Sigma}_m) + 2mK^2/T,$$

where $\hat{\Sigma}_m = T^{-1} \sum_{t=1}^{T} \hat{u}_t \hat{u}'_t$ is the residual covariance matrix of a VAR(m) model estimated by OLS. The criterion consists of the determinant of the residual covariance matrix which tends to decline with increasing VAR order whereas the penalty term $2mK^2/T$, which involves the number of parameters, grows with m. The VAR order is chosen which optimally balances both terms. In other words, models of orders $m = 0, \ldots, p_{\max}$ are estimated and the order p is chosen such that it minimizes the value of AIC.

Once a model is estimated it should be checked that it represents the data features adequately. For this purpose a rich toolkit is available. For example, descriptive tools such as plotting the residuals and residual autocorrelations may help to detect model deficiencies. In addition, more formal methods such as tests for residual autocorrelation, conditional heteroskedasticity, nonnormality and structural stability or tests for parameter redundancy may be applied.

Extensions

If some of the time series variables to be modelled with a VAR have stochastic trends, that is, they behave similarly to a random walk, then another model setup may be more useful for analyzing especially the trending properties of the variables. Stochastic trends in some of the variables are generated by models with unit roots in the VAR operator, that is, $\det(I_K - A_1 z - \cdots - A_p z^p) = 0$ for $z = 1$. Variables with such trends are nonstationary and not stable. They are often called integrated. They can be made stationary by differencing. Moreover, they are called cointegrated if stationary linear combinations exist or, in other words, if some variables are driven by the same stochastic trend. Cointegration relations are often of particular interest in economic studies. In that case, reparameterizing the standard VAR model such that the cointegration relations appear directly may be useful. The so-called *vector error correction model* (VECM) of the form

$$\Delta y_t = Dd_t + \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t$$

is a simple example of such a reparametrization, where Δ denotes the differencing operator defined such that $\Delta y_t = y_t - y_{t-1}$, $\Pi = -(I_K - A_1 - \cdots - A_p)$ and $\Gamma_i = -(A_{i+1} + \cdots + A_p)$ for $i = 1, \ldots, p-1$. This parametrization is obtained by subtracting y_{t-1} from both sides of the standard VAR representation and rearranging terms. Its advantage is that Π can be decomposed such that the cointegration relations are directly present in the model. More precisely, if all variables are stationary after differencing once, and there are $K - r$ common trends, then the matrix Π has rank r and can be decomposed as $\Pi = \alpha\beta'$, where α and β are $(K \times r)$ matrices of rank r and β contains the cointegration relations. A detailed statistical analysis of this model is presented in (Johansen 1995) (see also Part II of (Lütkepohl 2005)).

There are also other extensions of the basic VAR model which are often useful and have been discussed extensively in the associated literature. For instance, in the standard model all observed variables are treated as endogenous, that is, they are jointly generated. This setup often leads to heavily parameterized models, imprecise estimates and poor forecasts. Depending on the context, it may be possible to classify some of the variables as exogenous and consider partial models which condition on some of the variables. The latter variables remain unmodelled.

One may also question the focus on finite order VAR models and allow for an infinite order. This can be done by either augmenting a finite order VAR by a finite order MA term or by accounting explicitly for the fact that the finite order VAR approximates some more general model. Details on these and other extensions are provided, e.g., by (Hannan & Deistler 1988) and (Lütkepohl 2005).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *in* B. Petrov & F. Csáki (eds), *2nd International Symposium on Information Theory*, Académiai Kiadó, Budapest, pp. 267–281.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* **37**: 424–438.

Hannan, E. J. & Deistler, M. (1988). *The Statistical Theory of Linear Systems*, Wiley, New York.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

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