Dynamic factor models

Christopher A. Sims
Princeton University
sims@princeton.edu

April 30, 2015

The model

$$B(L)y_t = A(L)f_t + \varepsilon_t \qquad (*)$$

$$C(L)f_t = \nu_t \, .$$

y is $n \times 1$; f is $k \times 1$; B is diagonal; ε is i.i.d with identity covariance matrix; nu_t is the innovation in f_t, also with identity covariance matrix; ν_t and ε_s are independent of each other, all t and s. The appeal of the model is for the case $k << n$, which makes it much more tightly parameterized than an unconstrained VAR.

Normalization

Note that we can always, assuming we have an interpretation for $C^{-1}(L)$ (either because it converges or because we take the one-sided inverse and think of the model as starting at a specific date), write

$$B(L)y_t = A(L)C^{-1}(L)\nu_t + \varepsilon_t \,.$$

This has the same form as $(*)$, so we could normalize by insisting that f_t is i.i.d. $N(0, I)$. This would still leave the possibility of an orthonormal transformation of ν and $C(L)$, so we also need $(k^2 - k)/2$ additional constraints. For example people often choose to constrain A_0 to be triangular.

With finite lag lengths

Constraining f_t to be i.i.d. is just a normalization only if $A(L)$ and $C(L)$ are unconstrained and of infinite order. Just as an ARMA specification may be useful even though an arbitrary stationary process can be approximated arbitrarily well by either a pure finite order MA or a pure finite order VAR, it might be useful to separately parameterize C and A, if at least one is kept low-order to make root cancellation unlikely.

MCMC posterior simulation

This model allows a Gibbs sampling approach.

1. With B, A, C known and finite order, $(*)$ is the observation equation of a standard KF state-space system, so we can sample from its time path of states ($\{f_t\}$) and calculate its likelihood by standard methods.

2. With $\{f_t\}$ known, both $(*)$ and the AR equation for f_t are linear models — the two together form a VAR with a Granger causal ordering. So again, standard methods allow generating draws from the coefficients.

Initial conditions

- The initial conditions issue arises here as in any dynamic model. If the model is surely stationary, the best course is to use the unconditional distribution implied by the model as the distribution of initial conditions. This requires using Metropolis or M-H for each branch of the Gibbs scheme.

- If there may be some unit or explosive roots, but little is known about the forms of cointegrating vectors, using the unconditional distribution is not possible. Then a Minnesota prior is probably the best course.

Choosing the number of lags and the number of factors.

- This can be done with posterior odds ratios. For lag lengths, the Schwarz (i.e. BIC) criterion will work. For number of factors, it does not.

- What's wrong with BIC for factor count? If the number of factors is $k-1$ rather than k, the total parameter count declines by the number of elements in the last column of A plus the last column and last row of C. But if all the f_{kt}'s are zero, most of these parameters are unidentified. The Schwarz criterion assumes that parameters set to zero in the smaller model are identified, i.e., that if the smaller model were true, their coefficients would have small confidence bands around zero in large samples.

Shortcuts by making aggressive assumptions

- Suppose we multiply $(*)$ through by $B^{-1}(L)$. (Remember B is assumed diagonal.) Then suppose the resulting model has finite-order polynomials in the lag operator:

$$y_t = H(L)f_t + G(L)\varepsilon_t \qquad (\dagger)$$

- Stack all the lags of f that appear in $(\dagger)$ into f_t^*, and rewrite as

$$y_t = H^* f_t^* + G(L)\varepsilon_t \,.$$

- Then, assuming $\mathrm{Var}(f_t) = I$, $\mathrm{Var}(y_t) = HH' + \Lambda$, where Λ is diagonal.

Using principal components

- This form for $\mathrm{Var}(y_t)$ is exactly that of the standard factor analysis model — a less-than-full-rank positive definite matrix plus a diagonal matrix.

- Maximum likelihood factor analysis programs could therefore be applied to estimate the model, but further simplifications are possible if we are willing to assume that n is very large and much greater then k, and that the diagonal elements of Λ are all small relative to the eigenvalues of HH'. In fact we do not even need to require Λ diagonal if its elments remain small relative to HH'.

- Under these assumptions, we can approximately recover H by simply applying principal components to $\mathrm{Var}(y_t)$ — that is, we estimate the

space spanned by the columns of H by extracting the k eigenvectors of $\mathrm{Var}(y_t)$ that correspond to the k largest eigenvalues.

- While this approach has been widely applied because it is so easy with very large n, the asymptotic distribution theory invokes assumptions that seem unlikely to be even approximately correct. We mention it here because many applied papers in the literature that estimate "dynamic factor models" use this principal components approach, so you might be puzzled as to how that connects to the MCMC methods described above.