Error Bands for Impulse Responses, GCP

Christopher A. Sims
Princeton University
sims@princeton.edu

April 19, 2018

Impulse responses

- If $y_t = A(L)\varepsilon_t$ is a MAR of y_t, with $A(L)$ involving only non-negative powers of L (i.e., a one-sided, but not necessarily fundamental, MAR), the coefficients a_{ijs} in the i'th row, j'th column, of A_s, can be regarded as "impulse responses" when treated as a function of s for fixed i, j.

- They represent the time path of $y_{i,t+s}$ starting at time t, if ε_u=0, all u, except $\varepsilon_{jt} = 1$. I.e., the effect on the time path of y_i of a one-time unit disturbance in ε_j.

- This "impulse response" interpretation of the $a_{ij\cdot}$ is valid even if $A(L)$ is generated as the non-convergent one-sided inverse of the $B(L)$ in $B(L)y_t = \varepsilon_t$, when y is a non-stationary autoregressive process.

Orthogonalizing shocks

- In general innovations are correlated across variables. (Though sometimes, especially with fine-time-unit data, the correlations are small.)

- Responses to highly correlated shocks will be very similar, and this may make it hard to see the effects of some sources of disturbance.

- So it is common to transform the shocks to make them orthogonal:

$$y_t = A(L)WW'\varepsilon_t = C(L)\eta_t$$

$$\mathrm{Var}(\varepsilon_t) = (W')^{-1}W^{-1} \quad \therefore \mathrm{Var}(\eta_t) = I$$

Ways to orthogonalize

- Matrix square roots are not unique.

Error bands for impulse responses

- In one sense, this is straightforward: make draws from the posterior pdf of Σ and A (orβ) and for each draw calculate impulse responses $c_{ij}(t)$
- Plot the $c_{ij}t$ corresponding to the MLE (and/or the mean or median of the draws), together with the 5% upper and lower tails of the draws, for example. Could also plot HPD intervals.

These are not confidence intervals

- It's a Bayesian calculation. It gives intervals with a clear interpretation, by a straightforward procedure, but they are not confidence intervals (except asymptotically!).
- True confidence intervals for individual c_{ij}'s are not possible.
- This is a special case of a general point. If θ is the complete parameter vector for the distribution of the data X, then we can *always* (in principle) produce a 90% (say) confidence *region* for θ by constructiong a 90% significance level test for θ as H_0 for each θ in the parameter space. The set of θ's that are accepted in a given sample is an exact 90% confidence set.

- But if we try to construct a confidence set for an individual θ_i, the problem is that the distribution of any test statistic generally depends on all the parameters, not just θ_i.

- The normal linear regression model is a special case where there is a set of test statistics that depend on single θ_i's.

So what are frequentist CI's?

- There are "asymptotic" confidence intervals. These have coverage probabilities that converge to the correct ones for parameter values in some neighborhood of the true parameter value. They can be constructed by linearizing the mapping from B_{ij} (the AR coefficients) to C_{ij}, then transforming the normal asymptotic distribution for B to the corresponding approximate normal distribution for C.

- But these have no more frequentist asymptotic justification than do the Bayesian intervals.

- As the forecast horizon expands, the nonlinearities rapidly become more extreme, so the intervals based on linearization are always inaccurate at distant horizons.

- The Bayesian intervals are accurate whether or not A might have roots of one or larger in absolute value. The frequentist intervals are uninterpretable if that is true.

Pointwise bands, alternatives

- These bands are constructed from marginals for each c_{ij}.
- They cannot answer a question like "How likely is a hump-shaped response?"
- One could directly get a distribution for "up-life" from the MCMC sample.
- SZ alternative: display $1-\alpha$ probability bands for principal components of $\mathrm{Var}(c_{ij\cdot})$.
- Like displaying bands for principle components of the posterior of β in a regression, but with a more direct intuitive interpretation of the bounds.

Testing GCP

Suppose we have a system of the form

$$\begin{bmatrix} y_1(t) & y_2(t) \end{bmatrix} = [\mathrm{X}_1(t)\ \mathrm{X}_2(t)] \begin{bmatrix} \mathrm{B}_{11} & \mathrm{B}_{12} \\ \mathrm{B}_{21} & \mathrm{B}_{22} \end{bmatrix} + [\varepsilon_1(t)\ \varepsilon_2(t)]\,. \tag{1}$$

We can choose a γ to make $\nu(t) = \varepsilon_1(t) - \varepsilon_2(t)\gamma$ orthogonal to $\varepsilon_2(t)$ and define $C_1 = \mathrm{B}_{11} - \mathrm{B}_{12}\gamma$, $C_2 = \mathrm{B}_{21} - \mathrm{B}_{22}\gamma$, $\Omega = \mathrm{Var}([\nu(t)\ \varepsilon_2(t)])$ to allow us to rewrite the system as

$$\begin{bmatrix} y_1(t) - y_2(t)\gamma & y_2(t) \end{bmatrix} = [\mathrm{X}_1(t)\ \mathrm{X}_2(t)] \begin{bmatrix} \mathrm{C}_1 & \mathrm{B}_{12} \\ \mathrm{C}_2 & \mathrm{B}_{22} \end{bmatrix} + [\nu(t)\ \varepsilon_2(t)]\,. \tag{2}$$

Because the Jacobian of the transformation of parameters is the identity, and because the disturbances in the two blocks of equations in the transformed system are orthogonal, the likelihood factors into two pieces, one involving the parameters of the second equation only, the other involving the transformed first equation's parameters. Thus if our prior beliefs about the parameters of the transformed system make the parameters of the two blocks of equations independent, inference about the parameters of the second block of equations can be conducted by considering the likelihood for that block alone, as if there were no other equations in the system.

Of course in general, if we had started with prior beliefs expressed in terms of the original system's parameters, it would be a rare accident if our beliefs about the parameters of the two blocks in the transformed system were unrelated.

But testing for Granger causal priority (GCP) is a case where these results apply. GCP is the condition that, in a system in which $X(t)$ consists entirely of lagged values of y and we have grouped all lagged values of the first block y_1 of y into $X_1(t)$, $B_{12} = 0$. Thus a classical likelihood ratio test of the hypothesis that y_2 is GCP to y_1 is obtained by constructing twice the difference in log likelihoods and treating it as $\chi^2(df)$, where df, the degrees of freedom, is the number of elements in B_{12}. To be more specific, the classical LR test statistic is

$$T\left(\log(|S_{22}^R|) - \log(|S_{22}^U|)\right), \tag{3}$$

where S_{22}^R and S_{22}^U are the restricted and unrestricted cross-product matrices of residuals for the second block of equations.