

Take-Home Final Exam

There are four questions. They have equal weight, though they are probably not of equal difficulty. The exam is meant to allow an A performance in four hours of work. Long answers are less welcome than shorter, correct answers. The exam is available at 10AM Wednesday, January 23 and is due at 10AM Thursday, January 24. Submission of answers by email is OK, and I will verify that I have received such submissions. Submission of paper answers should be at my office, 104 Fisher or to my assistant Judith MacLaury, two doors down the hall.

(1) Suppose

$$y_t = Ay_{t-1} + c + \varepsilon_t,$$

2×1

with ε_t the innovation in y_t and stationary with finite variance. If the true values of A and c are given by

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \quad c = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

and we estimate A and c by OLS, which linear combinations of elements of the resulting estimates \hat{A} and \hat{c} have non-Gaussian classical asymptotic distributions? Justify your answer. *A direct approach is possible here: It is fairly obvious that if we let $z_t = y_{1t} - y_{2t}$, we have a stationary random variable equal to $1 + \varepsilon_{1t} - \varepsilon_{2t}$, while y_{1t} itself is an $I(1)$ process with non-zero constant term, and thus is dominated by linear trend. Since the whole system is can then be expressed in terms of a constant term, a stationary stochastic term, and a term dominated by linear trend, **all** of the coefficients estimated by application of OLS will have the usual standard Gaussian asymptotics.*

To apply the formal analysis of the notes, one needs the Jordan decomposition of the complete system matrix one obtains by including the constant term in the system. That is, we write

$$\begin{bmatrix} y_t \\ k_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ k_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}.$$

Here of course the initial k_0 , and hence all later k_t 's, are 1. Then we observe that the Jordan decomposition we need is

$$\begin{bmatrix} 1 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & 1 \\ 1 & 0 & -2 \\ 0 & 0 & 2 \end{bmatrix}.$$

This gives us the components of the system from the three rows of the matrix on the right above: $y_{1t} - y_{2t} - 1$, $y_{1t} - 2$ and $.$ The first is from a block corresponding to a zero (hence stationary) eigenvalue, while the latter two are from a block dominated by deterministic polynomials. Then we conclude as before that all components of the system can be generated without any components dominated by stochastic trend, so all coefficients and linear combinations of them have standard Gaussian asymptotic distributions.

(2) In the structural VAR model

$$\Gamma_0 y_t = \Gamma_1 y_{t-1} + \varepsilon_t,$$

with as usual ε_t spanning the space of innovations in y_t and having $\text{Var}(\varepsilon_t) = I$, much applied work imposes for identification zero-restrictions on the Γ_0 matrix.

- (a) Explain why, if y_t is 3×1 and zero restrictions on Γ_0 are the only identifying restrictions, 3 restrictions are needed, generally, for exact identification. *The system and the restrictions on it are invariant under pre-multiplication by any orthonormal matrix. An orthonormal matrix of order n has n^2 elements and is subject to the $(n^2 + n)/2$ orthonormality restrictions, leaving $(n^2 - n)/2$ degrees of freedom, which in our $n = 3$ case is 3.*
- (b) Since in this model neither reordering equations nor reordering the elements of y changes the model, there are really only a few distinct patterns in which three zero restrictions can be imposed. Explain why the following two patterns do not in fact deliver identification:

$$\text{i) } \begin{bmatrix} x & x & 0 \\ 0 & x & x \\ 0 & x & x \end{bmatrix} ; \quad \text{ii) } \begin{bmatrix} x & x & 0 \\ x & x & 0 \\ x & x & 0 \end{bmatrix} .$$

In the first the latter two rows can still be pre-multiplied by a 2×2 orthonormal matrix without violating the restrictions, while in the second the whole matrix can be so transformed, even though both matrices have the minimum necessary 3 zero restrictions.

- (c) The following pattern of zeros produces exact identification in part of the 6-dimensional space of $\text{Var}(\varepsilon_t)$ matrices (6d because the covariance matrix is symmetric):

$$\begin{bmatrix} 0 & x & x \\ x & 0 & x \\ x & x & 0 \end{bmatrix} .$$

Prove that under these restrictions there is an open subset of \mathbb{R}^6 corresponding to $\text{Var}(\varepsilon_t)$ matrices that do not satisfy

$$\Gamma_0^{-1} \left(\Gamma_0^{-1} \right)' = \text{Var}(\varepsilon_t)$$

for any Γ_0 satisfying the restrictions. Here in order for the problem to make any sense the $\text{Var}(\varepsilon_t)$ matrix being discussed has to be interpreted as the reduced form residual covariance matrix, not the $\text{Var}(\varepsilon_t)$ matrix that was already assumed equal to I at the outset. Of course if $\text{Var}(\varepsilon_t) = I$, there is certainly no open subset of the type you are asked to show exists. A student inquired during the exam about whether it was still true in this part that “ $\text{Var}(\varepsilon_t) = I$ ”, and I answered too hastily, not realizing that I had used $\text{Var}(\varepsilon_t)$ inconsistently where I had meant to write something like $\text{Var}(u_t)$, with u_t being the reduced form residuals. Some students did recognize that the question must be aimed at the mapping between $\Gamma_0^{-1}(\Gamma_0^{-1})'$ and Γ_0 and managed to construct useful answers, but the question was more of a puzzle than it should have been.

Proceeding with the question as it was meant to be, as opposed to as it was, note that if

$$\Gamma_0 = \begin{bmatrix} 0 & c & e \\ a & 0 & f \\ b & d & 0 \end{bmatrix},$$

then

$$\Gamma_0' \Gamma_0 = \begin{bmatrix} a^2 + b^2 & bd & af \\ bd & c^2 + d^2 & ce \\ af & bd & e^2 + f^2 \end{bmatrix}.$$

The 3×3 ones matrix certainly cannot be written in this form. If either b or d exceeds one in absolute value, for example, then one of the first two diagonal elements would exceed one. So with $bd = 1$, $b = d = 1$, which in turn, to make the diagonal elements 1, implies $a = c = 0$, which would force the first two elements of the third column to zero. So this matrix, and any matrix near it, is not of the form $\Gamma_0' \Gamma_0$. There are positive definite matrices near this one (since it is itself positive semi-definite), and thus an open set of inverses of such matrices that are not of the form of $\Gamma_0^{-1}(\Gamma_0^{-1})'$, for Γ_0 satisfying the restrictions.

- (d) Does this result from 2c imply that we should avoid using this pattern of restrictions in applied work? Why or why not? If substantive considerations led us to this form as a plausible restriction, there is no reason not to use it. It would be important to recognize, though, that despite the apparent exact identification there are restrictions being imposed. In computation, this could mean that when we try to solve for Γ_0 from the reduced form covariance matrix, there is no solution. In testing, it would mean that maximized likelihood for this “exactly identified” model could not necessarily be taken to be equivalent to the unrestricted likelihood maximum.

- (e) List the other distinct possible patterns of zero restrictions, and for each comment on whether it implies identification. *The distinct patterns of 3 zeros, besides those displayed above, are:*

$$\begin{bmatrix} x & 0 & 0 \\ x & x & 0 \\ x & x & x \end{bmatrix} \quad \begin{bmatrix} x & 0 & 0 \\ 0 & x & x \\ x & x & x \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ x & x & x \\ x & x & x \end{bmatrix}$$

The first is a standard recursive normalization scheme, which we know gives exact identification without imposing any restrictions. The last makes the second two rows indistinguishable, and thus does not allow identification, and at the same time makes the system incomplete, since it implies Γ_0 is singular. The middle pattern provides local exact identification, and probably imposes no global restrictions, though I don't have a proof of that.

One could also consider more or fewer than 3 zeros. (One student did, but no one tried to list them all.) Fewer is not interesting, since it always results in lack of identification. The patterns with four zeros consist of:

- Three that are a lower triangular matrix (from the first 3 restrictions) plus one zero on the diagonal. These yield singular Γ_0 and lack of identification.*
- Three that are a lower triangular matrix plus one additional off-diagonal zero. This yields identification.*
- $\begin{bmatrix} x & 0 & 0 \\ 0 & x & x \\ 0 & x & x \end{bmatrix}$, which does not allow identification.*
- Various patterns that make an entire row or column zero.*

The patterns with five zeros are, besides those that make an entire row or column zero,

$$\begin{bmatrix} x & 0 & 0 \\ 0 & x & 0 \\ x & 0 & x \end{bmatrix}, \quad \begin{bmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & x & x \end{bmatrix}.$$

With six zeros, the only pattern yielding identification is a diagonal matrix.

- (3) Suppose we have available time series on the log of an aggregate price index p_{it} for each of $i = 1, \dots, n$ countries over a period $t = 1, \dots, T$. We would like to construct an aggregate price index P_t for the whole collection of countries. One approach would be to set up this model:

$$\begin{aligned} p_{it} &= c_i + P_t + v_{it} \\ P_t &= \theta P_{t-1} + \xi_t \\ v_{it} &= \rho_i v_{i,t-1} + \varepsilon_{it}, \end{aligned}$$

where $\xi_t, \{\varepsilon_{it}, i = 1, \dots, n\}$ span the space of innovations in P_t and $\{v_{it}, i = 1, \dots, n\}$ and are assumed stationary, and Gaussian with a diagonal covariance matrix. Of course here P is unobservable and the idea is to estimate it.

- (a) Show how this setup fits in to a standard Kalman filtering framework, and explain how the Kalman filter would be used in constructing maximum posterior density (like MLE, but using a prior) estimates of the c vector, θ , and the ρ_i vector. Be explicit about what priors you would need to specify and discuss what would be reasonable and convenient choices of priors. *We take the state to be*

$$S_t = \begin{bmatrix} P_t \\ v_t \end{bmatrix},$$

where v_t is the vector of n country v_{it} 's. Then the state evolution equation is

$$S_t = AS_{t-1} + \zeta_t,$$

where ζ_t is ξ_t stacked over ε_t and A is a diagonal matrix with θ , then ρ_1, \dots, ρ_n down the diagonal. Then the observation equation is

$$p_t = HS_t + c, \quad \text{where } H = [\mathbf{1} \quad -I].$$

The KF would be used to evaluate the likelihood for given values of the parameter vector $\gamma = [\theta, \rho, \sigma]$, where σ is the vector of variances of the elements of the ζ vector. To evaluate the posterior we would need a prior over this parameter vector, of course, which we would multiply into the likelihood to get the posterior pdf. Also, to initiate the Kalman filter we would need an initial distribution for the state vector. If we are modeling the price level, it is likely that we will find non-stationarity somewhere in the system. It is not clear, though, whether we should expect it to be mainly in P or in v . If we were sure that, say, P were non-stationary and v were stationary, we might use a $N(c_i/(1 - \rho_i), \sigma_i^2/(1 - \rho_i^2))$ as the distribution for the initial v_{it} , and a very-large-variance normal for initial

P. However other choices could also make sense, and results should be checked for sensitivity to this aspect of the prior.

- (b) The Kalman smoother can give us estimates of the P sequence that use the full sample of data, but the associated covariance matrices do not take account of parameter uncertainty. Describe a convenient Markov chain Monte Carlo sampling scheme that would generate a posterior distribution for the P series that would take account of parameter uncertainty. *For a given set of parameter values, we can generate a sample draw from the $\{S_t\}$ sequence by a recursive algorithm very similar to the Kalman smoother. The difference is that at each point $t < T$ of the backward recursion one draws from the distribution of $S_t | \{\mathcal{I}_t, S_{t+1}\}$, where S_{t+1} is from the previous step of the backward recursion. Details on this are in the notes. This is not equivalent to drawing from the normal distribution produced by the smoother at each t .*

But this only gives us one draw. To get a distribution that reflects parameter uncertainty, we must use this draw, together with the data, to generate a draw of the parameters. Given the $\{S_t\}$ sequence, drawing from the posterior on the parameters is just an application of usual OLS estimates, at least if our prior is conjugate. If not, then the parameter draw might have to be made a Metropolis or Metropolis-Hastings draw. Then alternating between such parameter draws and state-sequence draws, one generates a complete artificial sample.

A few students suggested the sampling scheme Hamilton describes, which uses the Kalman smoother at each draw. This was not a bad answer, but not perfectly correct. For one thing, the method he describes does not take account of uncertainty in the covariance matrices that are taken as given by the Kalman filter, whereas the Gibbs sampling scheme above does. For another, what Hamilton proposes is not what is usually called a Markov chain Monte Carlo scheme (as the question requested). It is a Monte Carlo scheme, but generates iid draws. This is trivially a Markov process, but people knew how to do this before MCMC theory came on the scene.

- (c) If this model is being estimated over a period in which the countries have differing long-term monetary policies, and thus different long term inflation rates, yet they are closely linked by trade and cross-country investments, the model is likely not to fit well. Explain why. *The model constrains all cross-country correlation to come via P , which enters with a unit*

coefficient in every country. Thus the common component of variation is constrained to grow at the same rate everywhere. Different trend inflation rates would have to enter via non-stationarity in v_i , and would then imply very weak relationships between country inflation rates.

- (d) Suppose we want to test the model specification by comparing it to an unrestricted reduced form VAR with given lag length k . How might we do this? *The models aren't nested, but they are both dynamic models of the same time series, so the extended Schwarz criterion applies. A tricky point here is that the index model naturally conditions on no initial conditions, while the VAR model is most easily handled if the likelihood conditions on initial conditions. To make the two likelihoods comparable, the VAR model would have to be used to generate a distribution over the whole sample, by introducing a prior over the unobserved pre-sample initial conditions.*

(4) Consider the model

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + c + \varepsilon_t,$$

$k \times 1$

where ε_t is the innovation in y_t and is stationary and Gaussian.

- (a) Describe as explicitly as possible the restrictions on A_1 and A_2 implied by the claim that the y process is cointegrated. (We are not restricting the number of cointegrating vectors here, except to claim that there is at least one.) The idea is to put the restriction in a form such that it would be easy to write a computer program to check whether it holds or not.

Form the matrix

$$B = \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix}.$$

If Matlab is available, form $[v, d] = \text{eig}(B)$. Check the number of (near) unit eigenvalues on the diagonal of d . If this is k or more, there is no cointegration. If it's 0, then also there is no cointegration. If it's between 0 and k , check that the columns of v corresponding to unit roots form a full column rank matrix. If not, there are repeated unit roots and therefore no cointegration (in the usual sense of the term). Otherwise, there is cointegration.

- (b) Suppose we wished to test the hypothesis that y is cointegrated and had a way to maximize likelihood with the restriction imposed. The usual asymptotic argument that says the Schwarz criterion gives the right decision on which model is correct does not hold here. Why not? *The usual asymptotic argument depends on the assumption that $T\Omega_T$ converges to a non-singular constant matrix as $T \rightarrow \infty$, where Ω_T is the inverse Hessian of the log likelihood. That is not true when unit roots are present.*
- (c) Suggest a modification of the Schwarz criterion that would work here. A generic suggestion will not earn full credit. Be as explicit as you can be about what should be done in this model. You get credit for your suggestion being convenient as well as for its being asymptotically justified. *The generic answer is: Use the extended version of the Schwarz criterion, that uses Ω_T directly instead of assuming asymptotic behavior for it. A shortcut is available, though, if we are willing to (which we probably shouldn't be) maintain the usual assumptions of cointegration analysis. That is, we could assume that roots are all either stationary or equal to 1 and that there are no repeated roots. Then, if there is no cointegration, the sum of squares of all component y series are $O(T^3)$, and $T^{(3/2)}\Omega_T$ converges to a constant. So the Schwarz correction to the log likelihood under that assumption would be*

$(3n/4 + (n + 1)/2) \log T$, instead of the $((2n + 1)/2) \log T$ of the standard Schwarz criterion. If cointegration is in fact present, then the normalizing factor is $(3(n - q)/4 + (n + q + 1)/2) \log T$, where $n - q$ is the number of unit roots present. Estimation subject to the cointegration constraint should in fact produce some number $n - q > 0$ of roots exactly equal to one. So we could save some effort by checking whether the cointegrated model is favored with a Schwarz correction of between $(n + \frac{1}{2}) \log T$ and $(5n + 2)/4 \log T$. If the answer is the same regardless of which version of the SC we choose, we can avoid calculating Ω_T , at least from the viewpoint of the asymptotic theory. Nonetheless, it is probably a good idea to do the extended SC calculation as a check. It is not uncommon for the extended SC to contradict the standard SC even in cases where there are no problems with the asymptotics of the standard SC.