FIRST-HALF EXAM

Answer all 5 questions. You are to do the exam in two hours. You may make editorial corrections or transcribe your answers to more readable form outside the two hour limit. You may consult notes, reference books, static web pages. You may not discuss the exam with others or ask people or websites questions about the exam. The exam must be completed and handed in by 9AM Monday, March 15. The total of points on the exam is 110.

(1) **25 points** Suppose we have an i.i.d. sample $\{X_1, \dots, X_{10}\}$ of size 10 with the observed values

$$1, 2, 1, 4, 4, 1, 1, 1, 2, 1 \,.$$

We wish to characterize the uncertainty about a t-statistic for the null hypothesis that the mean of X is zero, so we construct a bootstrap sample and use it to generate the bootstrap distribution for $T = \bar{X}/s(X)$, where $\bar{X}$ is the sample mean and $s(X)$ is the sample standard deviation of X.

(a) With non-zero probability, this won't work because the bootstrap distribution will contain infinite values. Why?
Since it samples with replacement from the observed sample values, the bootstrap will with non-zero (though small) probability generate a bootstrap sample that consist of a single repeated value and which will therefore have a standard deviation of zero, making T infinite.

(b) What is the probability that a single bootstrap draw generated from this sample will give an infinite T?
All 1's, $.6^10$; all 2's: $.2^10$; all 4's: $.2^10$. Since these are mutually exclusive events, we sum up their probabilities to get 0.006046822.

(c) Would this problem go away if none of the observed values were identical?
No. The probability of a sample consisting of 10 equal values is smaller, but it is $.1^9 = 10^-9 > 0$.

(d) Would the Bayesian bootstrap run into a similar problem? Why or why not?
It would not run into this problem.
To connect the Bayesian bootstrap to the bootstrap, we need to think of the statistic T as a function mapping the distribution of the data to a real number. That is,

$$T = \frac{\sum_i p_i x_i}{\sqrt{\sum_i p_i x_i^2 - \left(\sum_i p_i x_i\right)^2}} \,.$$

(This assumes the "sample standard deviation" is the ML version rather than the unbiased version.)

Date: May 2, 2021.

The ordinary bootstrap keeps $p_i = 1/10$ and draws random x vectors from the sample distribution function; the Bayesian bootstrap keeps the x vector fixed and draws random p vectors from their posterior distribution, using a Dirichlet prior on the p vector. Since with probability one the p vector drawn from the Dirichlet posterior has all its elements positive, the denominator is always positive.
However, the Bayesian bootstrap of this sample shows a fat upper tail, reflecting the fact that very small values of the denominator are not unlikely.

(2) **25 points**One of the many uses of the Kalman filter is handling missing data. Suppose we have a bivariate VAR of the form

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = B \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix} .$$

We observe y_t only when t is an odd number and x_t only when t is even.

(a) Show how to set up the standard Kalman filter plant and observation equation in even number and odd number dates.
The state is $S_t = (y_t, x_t)$ and the plant equation is the VAR equation given above. The observation equation at odd dates, when we see y_t, is, with z_t representing what is observed,

$$z_t = y_t , \qquad \Leftrightarrow z_t = [1,0] S_t .$$

In our standard Kalman filter notation,

$$\begin{aligned} Plant: &\qquad S_t = BS_{t-1} + \eta_t \\ Observation: &\qquad z_t = HS_t + \nu_t , \end{aligned}$$

With $H = (1,0)$ at odd dates and $H = (0,1)$ at even dates, $\nu_t \equiv 0$, $\eta_t = (\varepsilon_{yt}, \varepsilon_{xt})'$.

(b) Does the fact that the Kalman filter equations differ from one date to the next create problems for generating a likelihood function for the sample observations as part of the Kalman filter output? Why or why not?
No problem. The filter generates a pdf for the next observation conditional on current and past data and evaluates that pdf at the observed value. The product of these conditional densities is the likelihood for the full sample, conditional on initial observations.

(c) What if there are some periods when we see neither x_t nor y_t?
Again no problem. The Kalman filter works even if $H = 0$. Another way to think of it is that when the next observation is at $t+2$ instead of at t, the plant equation becomes

$$S_{t+2} = B^2 S_t + \eta_t + B\eta_{t-1} .$$

(d) Suppose we need to estimate the unconditional variance of y_t and x_t. If we use the Kalman-filtered estimates of the time paths of y and x and simply compute their sample variances, would this give a consistent estimate? What about using the Kalman-smoothed estimates of the time paths this way? Is there a way to improve on these two estimators with Monte Carlo methods (be specific).

Neither the filtered nor the smoothed estimated time paths for the state will work. Both give, conditional on different data, expected values of the state, and have some error. The unconditional variance will be the variance of these expectations plus the variance of the errors, hence larger than the sample variances of either filtered or smoothed estimates.

However, the backward recursion we discussed in class generates realized values, drawn from the posterior distribution of the state time path conditional on all the observed data. Averaging the sample variances of these draws, over many draws, would give a consistent estimate of the unconditional variance and also of its variance.

The previous two paragraphs are what I expected as answers and what most people gave as answers, in whole or in part. However, other approaches are possible, as suggested in what follows.

The question seems to assume that the coefficient matrices and disturbance covariance matrices defining the filter are known. If so, we can calculate the unconditional variance analytically from those matrices. If the matrices are not known, but B and $\mathrm{Var}((\varepsilon_{yt}, \varepsilon_{xt}))$ are constant, we can estimate them by using the KF to evaluate likelihood and maximizing likelihood.

If all that is needed is a consistent estimate of the unconditional variance, the sample variance of the data would work.

All of this assumes that the data are from a stationary process, so that the unconditional variance is well defined.

(3) **20 points** Here's a possibly incorrect assertion: "The standard random effects model, assuming group constants are random and independent of right-hand-side variables, allows the econometrician, by varying her prior distribution over the variances of the group effects and the idiosyncratic disturbances, to arrive at arbitrary weighted averages of the between and within regression estimators. The fixed effects estimator, because it insists that the group constant parameters are non-random, avoids this source of bias." Discuss this. Is every assertion in it correct? None of it correct? Some of it correct?

The within regression is the fixed effects estimator. It is true that the random effects estimator can be expressed as a weighted average of the between and within estimators. It is also true that the weights are affected by the priors on the variances of the group effects and the individual-specific disturbances. However here as in most standard cases, any given prior becomes dominated by the likelihood in large samples. Usually, the weights are mainly determined by the likelihood, with only

small effects of the prior. To have a large effect on the weights, the researcher's prior, in a large sample, would have to be "dogmatic", making the analysis unconvincing to readers with less sharp prior beliefs. Perhaps more important is the fact that under the assumptions that justfy the random effects estimator, both between and within estimators are unbiased, so that manipulating the weights on them does not create any bias — a weighted average of two unbiased estimators is still unbiased.

The fixed effects estimator can be biased by a dogmatic prior also, if a prior on the variance of the group effects is imposed and there is correlation of the group effects with the X's. Usually, though, the fixed effects estimator is given a purely frequentist interpretation, and from this perspective, with a flat prior, its coefficient estimators are not biased.

(4) **25 points** Consider a first-order VAR system of the form $y_t = Ay_{t-1} + \varepsilon_t$. For the A matrices below determine i) which if any of the variables in the model are stationary, ii) whether cointegration is present, and iii) if cointegration is present, which linear combinations of the variables are stationary. You will want to use the eigen function in R or a similar command in another matrix programming language.

(a)

$$\begin{bmatrix} 2 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & .8 \end{bmatrix}$$

(b)

$$\begin{bmatrix} 1.25 & 0 & -.25 \\ .5 & 1 & -.5 \\ .25 & 0 & .75 \end{bmatrix}$$

(c)

$$\begin{bmatrix} .575 & .325 & -.225 \\ .65 & .35 & .65 \\ -.225 & .325 & .575 \end{bmatrix}$$

Recall that if $A = P\Lambda P^{-1}$ is the Jordan decomposition of A, then $P^{-1}y = z$ transforms y into a vector of z variables each of which is either a univariate autoregression or a group of variables corresponding to a repeated value on the diagonal of Λ. When Λ is diagonal, $P^{-1}y$ is the z vector, and the stationary and non-stationary linear combinations of y are just the rows of $P^{-1}y$ corresponding to eigenvalues less than one in absolute value, or larger, respectively. For (c), this leads to a simple answer:

```
> eigen(A3)
  eigen() decomposition
  $values
  [1]  1.0  0.8 -0.3
```

```
$vectors
     [,1]          [,2]       [,3]
[1,] 0.4082483 -7.071068e-01  0.4082483
[2,] 0.8164966  1.645736e-16 -0.8164966
[3,] 0.4082483  7.071068e-01  0.4082483
> solve(eigen(A3)$vectors)
          [,1]          [,2]      [,3]
[1,]  0.6123724  6.123724e-01 0.6123724
[2,] -0.7071068 -1.570092e-16 0.7071068
[3,]  0.6123724 -6.123724e-01 0.6123724
```

Because the first column of P (`eigen(A3)$vectors`) has no zero entries, we know that every element of y is non-stationary (puts weight on the component of z corresponding to the unit eigenvalue). Because there is just one unit eigenvalue and no repeating roots (non-singular P), we know that the coefficients of the stationary linear combinations are the last two rows of P^{-1}, or, rescaled, $(-1,0,1)$ and $(1,-1,1)$.

For (a) and (b), the eigenvalue decomposition is less directly helpful, because the P matrix is singular, indicating repeated roots. In a first-order system created by stacking the lags of y in a model with more than one lag, the presence of repeated roots makes checking for cointegrating vectors complicated. In that case the principle that the number of cointegrating vectors is n (the dimension of the y vector) minus the number of unstable roots no longer holdls. But in a pure first order system, where all elements of the y vector are dated t, the number of stationary linear combinations is always n minus the number of unstable roots, and the cointegrating vectors are the left-eigenvectors corresponding to the stable roots.

For (b):

```
> eigen(A2)
eigen() decomposition
$values
[1] 1 1 1

$vectors
     [,1]      [,2]      [,3]
[1,]    0 0.6859943 0.3641975
[2,]    1 0.2425356 0.8571583
[3,]    0 0.6859943 0.3641975
```

Since there are three unit roots, there is no cointegration and no element of the y vector is stationary.

For (a),

```
> eigen(A1)
eigen() decomposition
```

```
$values
[1] 1.0 1.0 0.8

$vectors
[,1]      [,2] [,3]
[1,] 0.7071068 0.7071068    0
[2,] 0.7071068 0.7071068    0
[3,] 0.0000000 0.0000000    1
```

There is one stable root. We can't find the corresponding left eigenvector by inverting the matrix of right eigenvectors, because it's singular. We could apply **eigen()** to the transpose of $A1$, but that's unncessary because it's clear by inspection of $A1$ that the third element of y is stationary — it's a stationary univariate AR. Thus the one cointegrating vector is (0,0,1).

(5) **15 points** Proposition: "If the impulse responses of a VAR all go to zero as the time horizon grows, then none of the variables in the VAR can be non- stationary." Either prove this or give a counterexample.

If we stack the system into first order form $y_t = Ay_{t-1} + \varepsilon_t$, the impulse responses at t are the elements of the $n \times n$ upper left block of the matrix A^t, where n is the length of the original, unstacked y vector. If the Jordan decomposition of A is $P\Lambda P^{-1}$, $A^t = P\Lambda^t P^{-1}$. This can be expressed equivalently as $\sum_j p_j \lambda_j p^j$, where p_j is the columns of P corresponding to the j'th Jordan block, p^j is the corresponding rows of P^{-1}, and λ_j is the Jordan block. For a block with an unstable root on the diagonal, λ_j^t either converges to a constant as $t \to \infty$ (only when λ is 1 by 1 and equal to 1) or explodes in absolute value. All the components of the sum corresponding to stable roots converge to zero. Thus A^t can't go to zero as $t \to \infty$ if there are any roots equal to or larger than 1 in absolute value.

Since the irf's are in the upper left $n \times n$ submatrix of $P\Lambda^t P^{-1}$, couldn't that part of the matrix go to zero, while the rest did not? To see that that's not possible, note that the irf's of $z = P^{-1}y$ are Λ^t. For any Jordan block λ with diagonal elements greater than or equal to 1. Λ^t fails to go to zero. But since $z_t = P^{-1}y_t$, at least some elements of the irf's of y must not go to zero in order for the irf's of an element of z not to go to zero.

There are processes in which impulse responses go to zero but the process is non-stationary. For example, the process with the moving average representation

$$y_t = \sum_{s=0}^{\infty} \varepsilon_{t-s} / (1 + s^{.5}) .$$

For this process to exist, we must think of its shocks as shrinking in size as we go back in time, e.g. by assuming $\varepsilon_t = 0$ for $t < 0$. The variance of y grows toward infinity as t increases if the variances of the ε_t's is constant for $t > 0$, but the impulse response to its own shock goes to zero.

But a process like this cannot arise from a finite order VAR. So to answer this questionn properly required using the fact that we are discussing a VAR model.