Problem Set 8 - Solutions

Applied Statistics and Econometrics II
Spring 2018, NYU
Ercan Karadas

(Sent by 3pm, April 25)

[1] Consider the following data generating process

$$\begin{aligned} x_t + \beta y_t &= u_{1t}, \qquad \text{where} \quad u_{1t} = \theta u_{1,t-1} + \epsilon_{1t} \\ x_t + \alpha y_t &= u_{2t}, \qquad \text{where} \quad u_{2t} = \rho u_{2,t-1} + \epsilon_{2t} \end{aligned}$$

where $|\rho| < 1$ and

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \sim D\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \gamma \\ \gamma & \sigma_1^2 \end{pmatrix}\right]$$

where D denotes a generic distribution.

a) If $\theta = 1$ then $x_t + \beta y_t \sim I(1)$ but $x_t + y_t \sim I(0)$. Hence $\alpha \neq \beta$ and
   i. if $\alpha \neq 0$, $x_t \sim I(0)$ and $y_t \sim I(1)$ but $(1\,\alpha)$ is a cointegrating vector.
   ii. if $\alpha = 0$, $x_t \sim I(0)$ and $y_t \sim I(1)$

   If $|\theta| < 1$ then both x and y are stationary.

b) $\theta = 1, \alpha \neq 0, \beta \neq 0$; $(1\,\alpha)$

c) i.

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 1 & \beta \\ 1 & \alpha \end{pmatrix}^{-1} \begin{pmatrix} (1-L)^{-1} & 0 \\ 0 & (1-\rho L)^{-1} \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

Hence,

$$\begin{aligned} x_t &= \frac{1}{\alpha - \beta}\left(-\beta \sum \epsilon_{1t} + \alpha \sum \rho^t \epsilon_{2t}\right) \\ y_t &= \frac{1}{\alpha - \beta}\left(\sum \epsilon_{1t} - \sum \rho^t \epsilon_{2t}\right) \end{aligned}$$

ii.

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \frac{1}{\alpha - \beta} \begin{pmatrix} -\beta & \alpha \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \frac{1}{\alpha - \beta} \begin{pmatrix} -\beta & \alpha \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

$$\begin{aligned} x_t &= \frac{1}{\alpha - \beta}\left(-\beta x_{t-1} + \alpha\beta y_{t-1} - \beta\epsilon_{1t} + \alpha\epsilon_{2t}\right) \\ y_t &= \frac{1}{\alpha - \beta}\left(x_{t-1} - \rho\beta y_{t-1} + \epsilon_{1t} - \epsilon_{2t}\right) \end{aligned}$$

iii.

$$\Delta x_t = \frac{\beta(1-\rho)}{\alpha-\beta} z_{t-1} + \eta_{1t}$$
$$\Delta y_t = -\frac{(1-\rho)}{\alpha-\beta} z_{t-1} + \eta_{2t}$$

d) VAR in levels is guaranteed to be consistent but not as efficient as when we impose the cointegrating restrictions. However, imposing incorrect cointegrating restrictions can lead to misspecification. Additionally, it is even more complicated to obtain standard errors for impulse responses from cointegrated VARs.

[2] Consider the following data generating process

$$\begin{aligned} x_t + y_t &= v_t, \qquad & v_t(1-\rho_1 L) &= \epsilon_{1t} \\ 2x_t + y_t &= u_t, \qquad & u_t(1-\rho_2 L) &= \epsilon_{2t} \end{aligned}$$

where $|\rho| < 1$ and

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}\right]$$

where D denotes a generic distribution.

a) i. Stationary
   ii. Cointegrated
   iii. Nonstationary

b)

$$(1-L)x_t + (1-L)y_t = \epsilon_{1t}$$
$$2(1-\rho_2 L)x_t + (1-\rho_2 L)y_t = \epsilon_{2t}$$

rearrange to obtain

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2\rho_2 & \rho_2 \end{pmatrix}\begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

Finally, the reduced form is obtained by inverting the matrix of contemporaneous correlations

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 2\rho_2 - 1 & \rho_2 - 1 \\ 2(1-\rho_2) & 2-\rho_2 \end{pmatrix}\begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} -\epsilon_{1t} + \epsilon_{2t} \\ 2\epsilon_{1t} - \epsilon_{2t} \end{pmatrix}$$

c)

Reduced Form	Structural
$\Psi_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\Psi_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}$
$\Psi_1 = \begin{pmatrix} 0 & -0.5 \\ 1 & 1.5 \end{pmatrix}$	$\Psi_1 = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 0 & -0.5 \\ 1 & 1.5 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0.5 \end{pmatrix}$
$\Psi_2 = \begin{pmatrix} -0.5 & -0.75 \\ 1.5 & 1.75 \end{pmatrix}$	$\Psi_2 = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} -0.5 & -0.75 \\ 1.7 & 1.75 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0.5 & 0.25 \end{pmatrix}$
$\Psi_s \to \infty,\ s \to \infty$	$\Psi_s \to \infty,\ s \to \infty$ but only for X_t

d)

$$x_t = -\frac{\epsilon_{1t}}{(1-L)} + \frac{\epsilon_{2t}}{(1-\rho_2 L)}$$
$$y_t = \frac{2\epsilon_{1t}}{(1-L)} - \frac{\epsilon_{2t}}{(1-\rho_2 L)}$$

hence the obvious cointegrating vector is (2 1)'.

e) In large samples the regression of of y_t on x_t will deliver an asymptotically consistent estimate of the coefficient λ in the regression

$$y_t = \lambda x_t + u_t$$

however, notice that the error term u_t is $u_t = -\frac{\epsilon_{2t}}{(1-\rho_2 L)}$. Typically this is of unknown form so a recommended strategy to correct for small sample bias is to use the Saikkonen, Phillips and Loretan, or Stock and Watson approach of including lags and leads of Δxt. Here, because the source of the correlation in the residuals is known, we have that an AR(1) correction would solve the problem since,

$$(1-\rho_2 L)y_t = -2x_t(1-\rho_2 L) - \epsilon_{2t}$$

[3] Consider the following bivariate VAR

$$y_{1t} = 0.3y_{1t-1} + 0.8y_{2t-1} + \epsilon_{1t}$$
$$y_{2t} = 0.9y_{1t-1} + 0.4y_{2t-1} + \epsilon_{2t}$$

where $E(\epsilon_{1t}\epsilon_{1\tau}) = 1$ for $t = \tau$ and 0 otherwise; $E(\epsilon_{2t}\epsilon_{2\tau}) = 2$ for $t = \tau$ and 0 otherwise; and $E(\epsilon_{1t}\epsilon_{2\tau}) = 0$ for all t, τ.

Answer the following questions.

a) To answer this question, verify the roots of the polynomial

$$\left| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.3 & 0.8 \\ 0.9 & 0.4 \end{pmatrix} z \right| = (1-0.3z)(1-0.4z) - (0.8z)(0.9z) = 1 - 0.7z - 0.6z^2$$

The roots are 0.83 and 2, hence the system is not stationary.

b)

$$\Psi_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \Psi_1 = \begin{pmatrix} 0.3 & 0.8 \\ 0.9 & 0.4 \end{pmatrix}, \ \Psi_2 = \begin{pmatrix} 0.81 & 0.56 \\ 0.63 & 0.88 \end{pmatrix}$$

Clearly, since

$$\Psi_s = \begin{pmatrix} 0.3 & 0.8 \\ 0.9 & 0.4 \end{pmatrix}^s$$

and the process is not stationary, then $\Psi_s \to \infty$.

c) Calculate the fraction of the MSE of the two period-ahead forecast error for variable 1, $E[y_{1,t+2} - \hat{E}(y_{1,t+2}|y_t, y_{t-1}, \ldots)]^2$, that is due to $\epsilon_{1,t+1}$ and $\epsilon_{1,t+2}$.

$$\begin{aligned} E[y_{1,t+2} - \hat{E}(y_{1,t+2}|y_t, y_{t-1}, \ldots)]^2 &= E[\epsilon_{1,t+2} + 0.3\epsilon_{1,t+1} + 0.8\epsilon_{2,t+1}]^2 \\ &= 1 + 0.3^2 + 0.8^2 \times 2 = 2.37 \end{aligned}$$

The fraction due to ϵ_1 is (1+0.32)/2.37 = 0.46 or 46%.

[4] Consider the following VAR

$$\begin{aligned} y_t &= (1+\beta)y_{t-1} - \beta\alpha x_{t-1} + \epsilon_{1t} \\ x_t &= \gamma y_{t-1} + (1-\gamma\alpha)x_{t-1} + \epsilon_{2t} \end{aligned}$$

Answer the following questions.

a) Stationarity requires that the values of z satisfying

$$\left| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1+\beta & -\beta\alpha \\ \gamma & (1-\gamma\alpha) \end{pmatrix} z \right| = 0$$

lie outside the unit circle. For $z = 1$, notice

$$\begin{vmatrix} -\beta & \beta\alpha \\ -\gamma & \gamma\alpha \end{vmatrix} = -\beta\gamma\alpha + \beta\gamma\alpha = 0$$

b)

$$\Phi(1) = \begin{pmatrix} -\beta & \beta\alpha \\ -\gamma & \gamma\alpha \end{pmatrix} = (-\beta \ -\gamma)'(1 \ -\alpha)$$

so that

$$\begin{aligned} \Delta y_t &= \beta(y_{t-1} - \alpha x_{t-1}) + \epsilon_{1t} \\ \Delta x_t &= \gamma(y_{t-1} - \alpha x_{t-1}) + \epsilon_{2t} \end{aligned}$$

c) Given the ECM in part (b), notice

$$
\begin{aligned}
-\gamma\Delta y_t + \beta\Delta x_t &= -\gamma\beta z_{t-1} - \gamma\epsilon_{1t} + \gamma\beta z_{t-1} + \beta\epsilon_{2t} \\
\Delta w_t = -\gamma\Delta y_t + \beta\Delta x_t &= -\gamma\epsilon_{1t} + \beta\epsilon_{2t}
\end{aligned}
$$

Next

$$
\begin{aligned}
y_t &= y_{t-1} + \beta(y_{t-1} - \alpha x_{t-1}) + \epsilon_{1t} \\
x_t &= x_{t-1} + \gamma(y_{t-1} - \alpha x_{t-1}) + \epsilon_{2t}
\end{aligned}
$$

$$
\begin{aligned}
y_t - \alpha x_t &= (y_{t-1} - \alpha x_{t-1}) + \beta z_{t-1} - \alpha\gamma z_{t-1} + \epsilon_{1t} - \alpha\epsilon_{2t} \\
z_t &= (1 + \beta - \alpha\gamma) z_{t-1} + \epsilon_{1t} - \alpha\epsilon
\end{aligned}
$$

$$
\begin{pmatrix} z_t \\ \Delta w_t \end{pmatrix} = \begin{pmatrix} 1 + \beta - \alpha\gamma & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ \Delta w_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & -\alpha \\ -\gamma & \beta \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}
$$

d) From part c)

$$
\begin{pmatrix} z_t \\ w_t \end{pmatrix} = \begin{pmatrix} 1 & -\alpha \\ -\gamma & \beta \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix}
$$

taking the inverse

$$
\frac{1}{\beta - \alpha\gamma} \begin{pmatrix} \beta & \alpha \\ \gamma & 1 \end{pmatrix} \begin{pmatrix} z_t \\ w_t \end{pmatrix} = \begin{pmatrix} x_t \\ y_t \end{pmatrix}
$$

and therefore

$$
\begin{aligned}
y_t &= \frac{\beta}{\beta - \alpha\gamma} z_t + \frac{\alpha}{\beta - \alpha\gamma} w_t \\
x_t &= \frac{\gamma}{\beta - \alpha\gamma} z_t + \frac{1}{\beta - \alpha\gamma} w_t
\end{aligned}
$$

w_t is $I(1)$ and z_t is $I(0)$, which is a version of the Beveridge-Nelson decomposition proposed by Gonzalo and Granger (1995).

[5] Consider the following VECM

$$
\Delta y_t = c + \alpha\beta' y_{t-1} + \epsilon_t, \qquad \epsilon_{it} \sim iid\,(0, \sigma^2)
$$

where $\alpha = (\alpha_1, 0)'$ and $\beta = (1, -\beta_2)'$. Equation by equation, the system is given by

$$
\begin{aligned}
\Delta y_{1t} &= c_1 + \alpha_1(y_{1,t-1} - \beta_2 y_{2,t-1}) + \epsilon_{1t} \\
\Delta y_{2t} &= c_2 + \epsilon_{2t}
\end{aligned}
$$

Answer the following questions.

a)

$$\Pi = \begin{pmatrix} \alpha_1 & \alpha_2\beta_2 \\ 0 & 0 \end{pmatrix}, \quad A = \begin{pmatrix} \alpha_1 + 1 & -\alpha_1\beta_2 \\ 0 & 1 \end{pmatrix}$$

b)

$$\alpha_\perp = \begin{pmatrix} 0 \\ k \end{pmatrix}, \quad \beta_\perp = \begin{pmatrix} \beta_2 k \\ k \end{pmatrix}, \quad k \neq 0.$$

c) Using the hint $\Pi\Psi(1)' = 0$. It is easy to show that

$$\beta_\perp \left(\alpha'_\perp I_2 \beta'_\perp\right)^{-1} \alpha_\perp = \begin{pmatrix} 0 & \beta_2 \\ 0 & 1 \end{pmatrix}$$

and therefore

$$\begin{pmatrix} \alpha_1 & -\alpha_1\beta_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \beta_2 \\ 0 & 1 \end{pmatrix} = 0$$

d) All you need to remember is that from the B-N decomposition, the trends are the linear combinations captured in $\Psi(1)y_t$, which in this case turns out to be $\beta_2 y_{1t} + y_{2t}$. Notice that this combination is orthogonal to the cointegrating vector.

e) Let $z_t = y_{1t} - \beta_2 y_{2t}$ be the cointegrating vector. From the equations for y_1 and y_2 we have

$$z_t = c_1 + (\alpha_1 + 1)y_{1,t-1} - \alpha_1\beta_2 y_{2,t-1} + \epsilon_{1t} - \beta_2 c_2 - \beta_2 y_{2,t-1} - \beta_2 \epsilon_{2t}$$

combining terms

$$z_t = (c_1 - \beta_2 c_2) + (\alpha_1 + 1)z_{t-1} + v_t, \quad v_t = \epsilon_{1t} - \beta_2 \epsilon_{2t}$$

which is an AR(1) whose stationarity requires that $|\alpha_1 + 1| < 1$ or the equivalent condition $-2 < \alpha_1 < 0$. When $\alpha_1 = 0$, z_t is no longer stationary, so there is no cointegration for any value of β_2. y_1 and y_2 are in this case two independent random walks.

[6] a) One approach is to try to propagate the process forward for different values of β. It turns out that for this example, the process is stationary for any β.

b) Assuming normality

$$Q(\beta) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(1) - -\frac{\sum_{t=1}^{T}(y_t - \ln(\beta y_{t-1}))^2}{2}$$

First order conditions,

$$\frac{\partial Q}{\partial \beta} = \frac{1}{\beta}\sum_{t=1}^{T}(y_t - \ln(\beta y_{t-1}))^2$$

gives

$$\hat{\beta} = exp\left\{\overline{y} - \overline{\ln(y)}\right\}$$

c)

$$\frac{\partial^2 Q}{\partial \beta^2} = -\frac{T}{\beta^2}\sum_{t=1}^{T}(y_t - \ln(\beta y_{t-1})) - \frac{1}{\beta^2}\sum_{t=1}^{T}(y_t - \ln(\beta y_{t-1}))$$

By QMLE results and assuming the regularity conditions are met, we know that

$$p\lim\left(-\frac{1}{T}\frac{\partial^2 Q}{\partial \beta^2}\right) = \frac{1}{\beta^2}$$
$$p\lim\left(-\frac{1}{T}\frac{\partial Q}{\partial \beta}\frac{\partial Q}{\partial \beta'}\right) = \frac{1}{\beta^2} \qquad (\text{since } \sigma_\epsilon^2 = 1)$$

and therefore

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \longrightarrow N(0, \beta^2)$$

d) Notice that

$$\lambda(\beta) = \ln \beta$$
$$\lambda'(\beta) = \frac{1}{\beta}$$

Using the delta method

$$\sqrt{T}\left(\ln \hat{\beta} - \ln \beta\right) \longrightarrow N(0, \frac{1}{\beta}\beta^2\frac{1}{\beta}) = N(0,1)$$

Notice that this is the result is equivalent to the result we would obtain from the least squares regression

$$y_t - \ln y_{t-1} = \gamma + \epsilon_t$$

with $\hat{\gamma} = \ln \beta$ and since by assumption $\sigma_\epsilon^2 = 1$, the result immediately follows.

e) Basic steps:

i. Draw with replacement a sample of size T from the residuals $\hat{\epsilon}_t$, re-centered.
ii. Given $\hat{\beta}$ and the initial condition, reconstruct the bootstrap sample of size T of y_t^*
iii. Estimate the model and obtain $\hat{\beta}^b$.
iv. Replicate steps 1-3, B times.

The question does not ask that you obtain an asymptotic refinement so you can construct the confidence interval directly from the B estimates $\hat{\beta}^b$ and choose the 5th and 95th percentiles. Alternatively, even though the variance of $\hat{\beta}$ is a function of β, the t-ratio is a pivotal statistic and hence a 95% confidence interval can be constructed with the percentile t-method by obtaining B estimates of the t-ratio, say t^* and constructing the 95% confidence interval as

$$\left[\hat{\beta} - t^*_{0.975}S_{\hat{\beta}}, \quad \hat{\beta} + t^*_{0.025}S_{\hat{\beta}}\right]$$

f) In this case, notice that

$$\sqrt{T}\left(\ln\hat{\beta} - \ln\beta\right) \longrightarrow N(0,1)$$

so that the percentile method is pivotal (since the variance is known and equal to one) and the bootstrap provides an asymptotic refinement.

g) BHHH step

$$\hat{\beta}_{j+1} = \hat{\beta}_j + \left(\frac{1}{T}\frac{\partial Q}{\partial\beta}\frac{\partial Q}{\partial\beta'}\right)^{-1}_{\hat{\beta}_j}\frac{\partial Q}{\partial\beta}|_{\hat{\beta}_j}$$

$$\hat{\beta}_2 = \hat{\beta}_1 + \left(\frac{1}{T\hat{\beta}_1^2}\sum_{t=1}^{T}\left(y_t - \ln(\hat{\beta}_1 y_{t-1})\right)^2\right)^{-1}\left(\frac{1}{\hat{\beta}_1^2}\sum_{t=1}^{T}\left(y_t - \ln(\hat{\beta}_1 y_{t-1})\right)\right)$$

[7] a) Assuming that returns are stationary, the dividend-price ratio is also stationary, because it is a linear combination of forecasts of stationary random variables. Since d_t has a unit root, it follows that p_t also has a unit root and that it is cointegrated with d_t. Hence, I would fit a VEC model for dividend growth and returns

$$\begin{bmatrix}\Delta d_t \\ r_t\end{bmatrix} = \begin{bmatrix}b_1 \\ b_2\end{bmatrix}(d_{t-1} - p_{t-1}) + \sum_{j=1}^{p}\Psi_j\begin{bmatrix}\Delta d_{t-j} \\ r_{t-j}\end{bmatrix} + \epsilon_t.$$

b) Because the restrictions are nonlinear, the restricted VECM would be difficult to estimate. Hence LM and LR tests would be computationally demanding. Thus I would choose a Wald test. This involves estimating an unrestricted VECM and then forming a nonlinear Wald statistic

$$W = R(\theta)'V_R^{-1}R(\theta),$$

where θ is a vector of VECM parameters, $V_R = \frac{\partial R}{\partial\theta}V_\theta\frac{\partial R'}{\partial\theta}$, and V_θ is the asymptotic variance for θ. The formula for V_R follows from the delta method. Alternatively, one could estimate V_R by Monte Carlo simulation, drawing θ from its asymptotic distribution and estimating $V_R = N^{-1}\sum_{i=1}^{N}(\theta_i - \overline{\theta})(\theta_i - \overline{\theta})'$, where θ_i is the ith Monte Carlo draw and $\overline{\theta}$ is simulation average, $\overline{\theta} = N^{-1}\sum_{i=1}^{N}\theta_i$.

c) W is asymptotically chi-square with degrees of freedom equal to the number of elements in $R(\theta)$. θ is asymptotically normal. By the delta method, we can approximate $R(\theta)$ as a vector of linear restrictions on θ. Higher order terms do not matter asymptotically. W can then be regarded as a quadratic form in normal random variables, hence a chi-square statistic. The number of degrees of freedom follows from the fact that θ is an unrestricted estimator, hence we don't need to adjust the degrees of freedom.