Draft of Solutions of Exam 2020
Multivariate Econometrics
VU Econometrics and Operations Research

This page is intentionally left blank.

Instructions

(i) All questions should be answered to get full points.

(ii) Each question is worth 25 points.

(iii) Read the instructions in the questions carefully.

(iv) Answer the questions as detailed as possible. Use mathematical expressions when necessary. You can use words when you cannot provide a formal mathematical answer to the questions.

(v) If a question is not clear to you, make your own assumptions to clarify the meaning of the question and then answer the question based on your assumptions.

(vi) See the back of this page for some standard results that you may make use of while answering the questions.

Some standard results

Suppose that the scalar process $\{z_t\}$ follows the following data generating process:

$$z_t = z_{t-1} + u_t,$$

where $z_0 = 0$ and u_t has the following properties:

(a) $u_t = \psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ where $\sum_{j=0}^{\infty} j \cdot |\psi_j| < \infty$ and $\{\epsilon_t\}$ is an $i.i.d$ sequence with mean zero and variance σ_ϵ^2, and finite fourth moment;

(b) σ^2 denotes the long run variance of $\{u_t\}$ and σ_u^2 denotes the contemporaneous variance of $\{u_t\}$.

Note that under these assumptions z_t can be written as a partial sum as

$$z_t = \sum_{s=1}^{t} u_s.$$

Let $W(r)$ be a standard Brownian motion process associated with u_t. Then the following results hold:

(1) $T^{-1/2} \sum_{t=1}^{T} u_t \xrightarrow{d} \sigma W(1)$;

(2) $T^{-1} \sum_{t=1}^{T} u_t^2 \xrightarrow{p} \sigma_u^2$;

(3) $T^{-1} \sum_{t=1}^{T} z_{t-1} u_t \xrightarrow{d} \frac{1}{2}\sigma^2 \left[W(1)^2 - \frac{\sigma_u^2}{\sigma^2} \right]$;

(4) $T^{-3/2} \sum_{t=1}^{T} t u_{t-j} \xrightarrow{d} \sigma \left\{ W(1) - \int_0^1 W(r)dr \right\}$ for $j = 0, 1, \ldots$;

(5) $T^{-3/2} \sum_{t=1}^{T} z_{t-1} \xrightarrow{d} \sigma \int_0^1 W(r)dr$;

(6) $T^{-2} \sum_{t=1}^{T} z_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 W(r)^2 dr$;

(7) $T^{-5/2} \sum_{t=1}^{T} t z_{t-1} \xrightarrow{d} \sigma \int_0^1 r W(r)dr$;

(8) $T^{-3} \sum_{t=1}^{T} t z_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 r W(r)^2 dr$;

(9) $T^{-(v+1)} \sum_{t=1}^{T} t^v \rightarrow 1/(v+1)$ for $v = 0, 1, \ldots$;

(10) Suppose that the DGP of another time series process y_t follows the model

$$y_t = y_{t-1} + e_t,$$

where $y_t = 0$ and e_t satisfy the same assumptions as **(a)** and has a long run variance σ_e^2 and $W_e(r)$ is a standard Brownian motion processs associated with e_t, then

$$T^{-2} \sum_{t=1}^{T} z_t y_t \xrightarrow{d} \sigma\sigma_e \int_0^1 W(r) W_e(r) dr.$$

Question 1: Conceptual Questions (25 points out of 100 points)

Below you will find **3** statements. All these are related to the concepts/techqniques that have been discussed during the lectures. Some of these statements are correct, some are wrong, some need further clarification. You need to provide a brief, to the point answer that would contain **(i) short explanations/definitions of the concepts mentioned in the statement, (ii) your judgement about the statement about whether it is correct/wrong/unclear/incomplete, and an explanation of your judgement (iii) a correction of the statement.** The concepts that you need to explain and define are written in *italics*. A formal answer using mathematics is possible, sometimes very useful but not always necessary.

(a) (5 points) *Reduced form* representation of a VAR model can be uniquely converted to a *structural form* VAR model. It is always preferable to consider reduced from VARs.

(b) (5 points) Suppose that y_t and x_t are two $I(1)$ *processes* and these two processes are *cointegrated* with the cointegrating vector $(1, -\beta)$. Here, β can be estimated consistently by estimating the regression model

$$y_t = \beta x_t + u_t.$$

Furthermore the hypothesis of $\mathcal{H}_0 : \beta = 1$ can be tested by using normal distribution.

(c) (15 points) For a panel data variable $y_{i,t}$, a dynamic panel data model can be written as

$$y_{i,t} = \rho y_{i,t-1} + u_{i,t}.$$

Let N represent the number of cross-section units and T represent the number of time series periods. Suppose that $u_{i,t}$ follows

$$u_{i,t} = \mu_i + \epsilon_{i,t},$$

where $\mu_i \sim (0, \sigma_\mu^2)$ and $\epsilon_{i,t} \sim (0, \sigma_\epsilon^2)$. This is called the *interactive fixed effects* assumption for $u_{i,t}$. When T is fixed as $N \to \infty$, the *fixed effects estimator* of ρ is unbiased and consistent.

Answer:

(a) (5 points) Reduced form:

$$\mathbf{x}_t = \boldsymbol{\delta}_t + \boldsymbol{\Lambda}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t,$$

is the reduced form representation, the system in which all right hand-side variables are predetermined at time t: No variable has a direct contemporaneous effect on the other variables in the model. In most of the cases the reduced form does not directly represent the behavioural and technical relationships specified by economic theory: Economic theory considers contemporaneous relations.

Structural form:

$$\mathbf{B}\mathbf{x}_t = \boldsymbol{\Gamma}\mathbf{d}_t + \mathbf{C}\mathbf{x}_{t-1} + \mathbf{u}_t,$$

where $\mathbf{B}$ full rank

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ b_{m1} & \dots & \dots & b_{mm} \end{pmatrix} \neq \mathbf{I}_m,$$

and $E(\mathbf{u}_t|\mathcal{X}_{t-1}) = 0$ and $E(\mathbf{u}_t\mathbf{u}_t'|\mathcal{X}_{t-1}) = \boldsymbol{\Sigma}$. Here For instance if $b_{12} \neq 0$, it means that the second element of $\mathbf{x}_t$ has a contemporaneous effect on the first element of $\mathbf{x}_t$.

$$\mathbf{B}\mathbf{x}_t = \boldsymbol{\Gamma}\mathbf{d}_t + \mathbf{C}\mathbf{x}_{t-1} + \mathbf{u}_t \qquad \text{vs.} \qquad \mathbf{x}_t = \boldsymbol{\delta}_t + \boldsymbol{\Lambda}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t.$$

$$\boldsymbol{\Lambda} = \mathbf{B}^{-1}\mathbf{C}, \;\; \boldsymbol{\delta}_t = \mathbf{B}^{-1}\boldsymbol{\Gamma}\mathbf{d}_t, \;\; \boldsymbol{\varepsilon}_t = \mathbf{B}^{-1}\mathbf{u}_t, \;\; \boldsymbol{\Omega} = \mathbf{B}^{-1}\boldsymbol{\Sigma}(\mathbf{B}')^{-1}$$

Note that many different structures corresponding to many different choices of $\mathbf{B}$ can correspond to the same reduced form VAR model. One can impose some restrictions on $\mathbf{B}$, $\boldsymbol{\Gamma}$, $\mathbf{C}$ and $\boldsymbol{\Sigma}$: i.e. omission of certain variables from certain equations.

So the statement is not correct because the conversion is not unique. And the second part of the statement is also not correct because: Economic theory considers contemporaneous relations.

(b) (5 points) An $I(1)$ *process* is a process that is integrated of order 1. This means that if we take the first difference of this process it becomes stationary and usual CLT and LLN's can be applied to the first difference. If a $m \times 1$ vector $\mathbf{z}_t \sim$ I(1), and there exists an $m \times 1$ vector $\boldsymbol{\gamma}$ such that $\boldsymbol{\gamma}'\mathbf{z}_t \sim$ I(0), then the elements of $\mathbf{z}_t$ are said to be cointegrated with cointegrating vector $\boldsymbol{\gamma}$. For the elements of $\mathbf{z}_t$ to be cointegrated, it is sufficient that the residual obtained after taking an appropriate linear combination of the variables is an I(0) process that can be modelled by ordinary time series techniques. A unique cointegrating vector (if $\boldsymbol{\beta}$ is unique) is consistently estimated by least squares regression. (Engle and Granger 1987): This holds true regardless of the simultaneity structure of the model generating $\mathbf{x}_t$ and the short-run dynamics. So the first part of the statement is true. If the cointegrating rank is 1 then we can estimate the cointegrating vector by using the static least squares estimator. However the second part is not true. Because simultaneity and autocorrelation creates problems in the inference when we want to do hypothesis testing.

(c) (15 points) In this statement, the model given for the error term of the model

$$u_{i,t} = \mu_i + \epsilon_{i,t},$$

is not the interactive fixed effects assumption. This is the one way error component model. We note from the lecture slides that

- $u_{i,t} = \mu_i + v_{i,t}$, where $\mu_i \sim (0, \sigma_\mu^2)$ and $v_{i,t} \sim (0, \sigma_v^2)$. This is called the one way error component model.
- $u_{i,t} = \mu_i + \lambda_t + v_{i,t}$, where $\mu_i \sim (0, \sigma_\mu^2)$, $\lambda_t \sim (0, \sigma_\lambda^2)$ and $v_{i,t} \sim (0, \sigma_v^2)$. This is called the two way error component model.
- $u_{i,t} = \lambda_i f_t + v_{i,t}$, where $\lambda_i \sim (0, \sigma_\lambda^2)$, $f_t \sim (0, \sigma_f^2)$ and $v_{i,t} \sim (0, \sigma_v^2)$. This is called the interactive fixed effects.

In order to eliminate μ_i from the model we apply the within transformation.Define $\overline{y}_i = T^{-1}\sum_{t=1}^{T} y_{i,t}$. When we consider our model and apply the within demeaning, we will end up with the model:

$$y_{i,t} - \overline{y}_i = \delta(y_{i,t-1} - \overline{y}_{i,-1}) + \boldsymbol{\beta}'(\mathbf{x}_{i,t} - \overline{\mathbf{x}}_i) + (v_{i,t} - \overline{v}_i).$$

This is the fixed effects model and a serious difficulty arises with the fixed effects model in this context, where we have large N, small T (micro panels).

Nickell (1981) shows that this arises because the demeaning process creates a correlation between the regressors and the regression error. Fixed effects estimator:

$$\hat{\rho}_{FE} \quad = \quad \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \overline{y}_{i,-1})(y_{i,t} - \overline{y}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \overline{y}_{i,-1})^2},$$

– $y_{i,t-1}$ is correlated with $\overline{v}_i$ by construction: the latter average contains $v_{i,t-1}$, which is correlated with $y_{i,t-1}$.

– $\overline{v}_i$ is correlated with $\overline{y}_{i,-1}$.

Nickell (1981), showed that for these reasons the within estimator (fixed effects estimator) is biased of order $O(1/T)$ and it is inconsistent for N large and T small. Note that this estimator is biased and inconsistent when N is large and T is small (fixed). Only when $T \to \infty$ the within estimator of δ and $\boldsymbol{\beta}$ will be consistent. Nickell demonstrates that if $\delta > 0$, then the bias is invariably negative so that the persistence of $y_{i,t}$ will be underestimated. Note that the bias is not caused by an autocorrelated error process.

Question 2: Modeling and stationarity (25 points out of 100 points)

(a) (15 points) Consider two processes y_t and x_t with the DGPs

$$y_t = A_1 y_{t-1} + B_1 y_{t-2} + A_2 x_{t-1} + u_{y,t},$$
$$x_t = A_3 y_{t-1} + A_4 x_{t-1} + u_{x,t}.$$

Suppose that

$$u_{y,t} = C_1 u_{y,t-1} + \varepsilon_{1,t},$$
$$u_{x,t} = \varepsilon_{1,t} + \varepsilon_{2,t},$$

where for $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t})'$.

$$\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim IN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right],$$

Now we define

$$\mathbf{w}_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix}.$$

(i) Write the two models for y_t and x_t given above in the form of a VAR(2) model for $\mathbf{w}_t$.

(ii) Show that for $\mathbf{w}_t$ we can obtain a model in the form

$$\Delta\mathbf{w}_t = \mathbf{E}\mathbf{w}_{t-1} + \mathbf{F}\Delta\mathbf{w}_{t-1} + \mathbf{G}\Delta\mathbf{w}_{t-2} + \mathbf{H}\boldsymbol{\varepsilon}_t,$$

where

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Write $\mathbf{E}$, $\mathbf{F}$ and $\mathbf{G}$ in terms of the model parameters A_1, A_2, A_3, A_4, B_1 and C_1.

(iii) Briefly discuss under what conditions at least one of the elements of $\mathbf{w}_t$ is I(2)?

(b) (10 points) Consider the bivariate system for y_t and x_t. Let $\mathbf{w}_t = (y_t, x_t)'$. The DGP for $\mathbf{w}_t$ can be written as

$$\Delta \mathbf{w}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{w}_{t-1} + \boldsymbol{\Gamma}\Delta\mathbf{w}_{t-1} + \boldsymbol{\epsilon}_t,$$

where

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ 0 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} 1 \\ -\beta_1 \end{pmatrix},$$

and

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0 & \gamma_1 \\ 0 & \gamma_2 \end{pmatrix}$$

and

$$\boldsymbol{\epsilon}_t = \begin{pmatrix} \epsilon_{y,t} \\ \epsilon_{x,t} \end{pmatrix} \sim IN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & 0 \\ 0 & \sigma_x^2 \end{pmatrix}\right].$$

Show that $\boldsymbol{\beta}'\mathbf{w}_t$ is weakly stationary if $|\gamma_2| < 1$, and $-2 < \alpha_1 < 0$. Write down the additional necessary conditions (if any). In addition to your derivations explain with your own words why do we need the conditions $|\gamma_2| < 1$, and $-2 < \alpha_1 < 0$ and what happens if these conditions are not satisfied.

(a) (15 points) Here (i) is 3 points, (ii) is 10 points, (iii) is 2 points.

(i) Writing the model in the form of a VAR(2), requires defining the matrices

$$\mathbf{A} = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}$$

and

$$\mathbf{B} = \begin{pmatrix} B_1 & 0 \\ 0 & 0 \end{pmatrix}$$

Then we can write

$$\mathbf{w}_t = \mathbf{A}\mathbf{w}_{t-1} + \mathbf{B}\mathbf{w}_{t-2} + \mathbf{u}_t,$$

where

$$\mathbf{u}_t = \begin{pmatrix} u_{y,t} \\ u_{x,t} \end{pmatrix}.$$

(ii) Now we need to obtain the VECM model that has $\boldsymbol{\varepsilon}_t$ as the error term. For this let us define

$$\mathbf{C} = \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix}$$

then we can write

$$\mathbf{u}_t = \mathbf{C}\mathbf{u}_{t-1} + \mathbf{H}\boldsymbol{\varepsilon}_t,$$

where $\mathbf{H}$ is defined in the question.If we want to eliminate $\mathbf{u}_t$ from our model we can obtain the model for $\mathbf{w}_t - \mathbf{C}\mathbf{w}_{t-1}$. We can do it by subtracting $\mathbf{C}\mathbf{w}_{t-1}$ from $\mathbf{w}_t$. We have

$$\begin{aligned} \mathbf{w}_t &= \mathbf{A}\mathbf{w}_{t-1} + \mathbf{B}\mathbf{w}_{t-2} + \mathbf{u}_t, \\ \mathbf{C}\mathbf{w}_{t-1} &= \mathbf{C}\mathbf{A}\mathbf{w}_{t-2} + \mathbf{C}\mathbf{B}\mathbf{w}_{t-3} + \mathbf{C}\mathbf{u}_{t-1}, \end{aligned}$$

Subtracting the lower equation from the upper one and rearranging gives

$$\mathbf{w}_t = (\mathbf{A} + \mathbf{C})\mathbf{w}_{t-1} + (\mathbf{B} - \mathbf{C}\mathbf{A})\mathbf{w}_{t-2} - \mathbf{C}\mathbf{B}\mathbf{w}_{t-3} + \mathbf{H}\boldsymbol{\varepsilon}_t$$

Now we can obtain the VECM model. We do it by adding and subtracting the necessary terms as follows

$$\begin{aligned} \mathbf{w}_t &= (\mathbf{A} + \mathbf{C})\mathbf{w}_{t-1} + (\mathbf{B} - \mathbf{C}\mathbf{A})\mathbf{w}_{t-2} - \mathbf{C}\mathbf{B}\mathbf{w}_{t-3} + \mathbf{H}\boldsymbol{\varepsilon}_t \\ &= (\mathbf{A} + \mathbf{C})\mathbf{w}_{t-1} + (\mathbf{B} - \mathbf{C}\mathbf{A})\mathbf{w}_{t-2} - \mathbf{C}\mathbf{B}\mathbf{w}_{t-2} + \mathbf{C}\mathbf{B}\mathbf{w}_{t-2} - \mathbf{C}\mathbf{B}\mathbf{w}_{t-3} + \mathbf{H}\boldsymbol{\varepsilon}_t \\ &= (\mathbf{A} + \mathbf{C})\mathbf{w}_{t-1} + (\mathbf{B} - \mathbf{C}\mathbf{A} - \mathbf{C}\mathbf{B})\mathbf{w}_{t-2} + \mathbf{C}\mathbf{B}\Delta\mathbf{w}_{t-2} + \mathbf{H}\boldsymbol{\varepsilon}_t \\ &= (\mathbf{A} + \mathbf{C} + \mathbf{B} - \mathbf{C}\mathbf{A} - \mathbf{C}\mathbf{B})\mathbf{w}_{t-1} - (\mathbf{B} - \mathbf{C}\mathbf{A} - \mathbf{C}\mathbf{B})\Delta\mathbf{w}_{t-1} + \mathbf{C}\mathbf{B}\Delta\mathbf{w}_{t-2} + \mathbf{H}\boldsymbol{\varepsilon}_t \end{aligned}$$

Final step is to subtract $\mathbf{w}_{t-1}$ from both sides. This will give the VECM model.

$$\Delta\mathbf{w}_t = (\mathbf{A}+\mathbf{C}+\mathbf{B}-\mathbf{CA}-\mathbf{CB}-\mathbf{I})\mathbf{w}_{t-1} - (\mathbf{B}-\mathbf{CA}-\mathbf{CB})\Delta\mathbf{w}_{t-1} + \mathbf{CB}\Delta\mathbf{w}_{t-2} + \mathbf{H}\boldsymbol{\varepsilon}_t.$$

This finding tell us that

$$\begin{aligned}\mathbf{E} &= \mathbf{A}+\mathbf{C}+\mathbf{B}-\mathbf{CA}-\mathbf{CB}-\mathbf{I},\\ \mathbf{F} &= -\mathbf{B}+\mathbf{CA}+\mathbf{CB},\\ \mathbf{G} &= \mathbf{CB}.\end{aligned}$$

This means

$$\begin{aligned}\mathbf{E} &= \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix} + \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} B_1 & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}\\ &- \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} B_1 & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\\ \mathbf{F} &= -\begin{pmatrix} B_1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix} + \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} B_1 & 0 \\ 0 & 0 \end{pmatrix},\\ \mathbf{G} &= \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} B_1 & 0 \\ 0 & 0 \end{pmatrix}.\end{aligned}$$

Some matrix algebra yields

$$\begin{aligned}\mathbf{E} &= \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix} + \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} B_1 & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} C_1A_1 & C_1A_2 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} C_1B_1 & 0 \\ 0 & 0 \end{pmatrix},\\ \mathbf{F} &= -\begin{pmatrix} B_1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} C_1A_1 & C_1A_2 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} C_1B_1 & 0 \\ 0 & 0 \end{pmatrix},\\ \mathbf{G} &= \begin{pmatrix} C_1B_1 & 0 \\ 0 & 0 \end{pmatrix}.\end{aligned}$$

Finally we get

$$\begin{aligned}\mathbf{E} &= \begin{pmatrix} A_1+C_1+B_1-C_1A_1-C_1B_1 & A_2-C_1A_2 \\ A_3 & A_4 \end{pmatrix},\\ \mathbf{F} &= \begin{pmatrix} -B_1+C_1A_1+C_1B_1 & C_1A_2 \\ 0 & 0 \end{pmatrix},\\ \mathbf{G} &= \begin{pmatrix} C_1B_1 & 0 \\ 0 & 0 \end{pmatrix}.\end{aligned}$$

(iii) In this question multiple approaches are acceptable. A short and concise answer that explains the cases where $\mathbf{w}_t$ is $I(2)$ is sufficient to get full points. If $\mathbf{B}+\mathbf{A}=\mathbf{I}$ and $\mathbf{C}=\mathbf{I}$, then we can show that $\mathbf{w}_t$ is $I(2)$. This is because we can rewrite

$$\begin{aligned}\mathbf{E} &= \mathbf{A} + \mathbf{C} + \mathbf{B} - \mathbf{CA} - \mathbf{CB} - \mathbf{I} \\ &= -(\mathbf{C} - \mathbf{I})(\mathbf{A} + \mathbf{B} - \mathbf{I}).\end{aligned}$$

And having $\mathbf{B} + \mathbf{A} = \mathbf{I}$ and $\mathbf{C} = \mathbf{I}$ gives two unit roots.

(b) (10 points) Under the restrictions: $\alpha_1 \neq 0$, $\alpha_2 = 0$, $|\gamma_2| < 1$; the model becomes

$$\begin{aligned}\Delta y_t &= \alpha_1(y_{t-1} - \beta_1 x_{t-1}) + \gamma_1 \Delta x_{t-1} + \epsilon_{1,t}; \\ \Delta x_t &= \gamma_2 \Delta x_{t-1} + \epsilon_{2,t}.\end{aligned}$$

The second equation allows us to write

$$(1 - \gamma_2 L)\Delta x_t = \epsilon_{2,t}.$$

This gives

$$\Delta x_t = (1 - \gamma_2 L)^{-1}\epsilon_{2,t}.$$

For this operation to yield a stable inversion $|\gamma_2| < 1$. Now we need to obtain the model for $\boldsymbol{\beta}'\mathbf{w}_t$. We can write

$$\begin{aligned}\boldsymbol{\beta}'\Delta\mathbf{w}_t = \Delta y_t - \beta_1 \Delta x_t &= \alpha_1(y_{t-1} - \beta_1 z_{t-1}) + \gamma_1(1 - \gamma_2 L)^{-1}\epsilon_{2,t} \\ &\quad + \epsilon_{1,t} - \beta_1(1 - \gamma_{22} L)^{-1}\epsilon_{2,t} \\ &= \alpha_1(y_{t-1} - \beta_1 z_{t-1}) + \epsilon_{1,t} + (\gamma_1 - \beta_1)(1 - \gamma_{22} L)^{-1}\epsilon_{2,t}\end{aligned}$$

which can be written as

$$\boldsymbol{\beta}'\Delta\mathbf{w}_t = \alpha_1 \boldsymbol{\beta}'\mathbf{w}_{t-1} + \epsilon_{1,t} + (\gamma_1 - \beta_1)(1 - \gamma_{22} L)^{-1}\epsilon_{2,t},$$

We can write this by using the lag polynomial notation as

$$[1 - (\alpha_1 + 1)L]\boldsymbol{\beta}'\mathbf{w}_t = \epsilon_{1,t} + (\gamma_1 - \beta_1)(1 - \gamma_{22} L)^{-1}\epsilon_{2,t},$$

So for the stationarity of $\boldsymbol{\beta}'\mathbf{w}_t$ we need the roots of the polynomial $[1 - (\alpha_1 + 1)z]$, lie outside the unit circle. This is ensured when $-2 < \alpha_1 < 0$.

Question 3: Asymptotic Derivations (25 points out of 100 points)

(a) (15 points) **A nonparametric test for unit roots:**

Suppose that we have the following data generating processes (DGP) for $\{y_t\}$

$$y_t = \rho y_{t-1} + u_t,$$

for $t = 1, \ldots, T$. We assume:

- $u_t = \psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ where $\sum_{j=0}^{\infty} j \cdot |\psi_j| < \infty$ and $\{\epsilon_t\}$ is an *i.i.d* sequence with mean zero and variance σ_ϵ^2, and finite fourth moment;
- σ^2 denotes the long run variance of $\{u_t\}$ and σ_u^2 denotes the contemporaneous variance of $\{u_t\}$;
- $y_0 = 0$.

We would like to test for unit roots. So we are interested in testing

$$\mathcal{H}_0 : \ \rho = 1,$$

against the alternative

$$\mathcal{H}_1 : \ |\rho| < 1.$$

For this reason, we consider the test statistic

$$VR = \frac{T\left(\sum_{t=1}^{T} \Delta y_t\right)^2}{\sum_{t=1}^{T} y_t^2}$$

(i) Derive carefully the order of probability and the limiting distribution of VR under the null hypothesis.

(ii) Obtain the order of probability of VR under the alternative hypothesis.

(iii) What do you conclude about the consistency of this test?

(b) (10 points) Suppose that we have the following data generating processes (DGP) for $\{y_t\}$ and $\{x_t\}$

$$y_t = \delta + y_{t-1} + u_{y,t}$$
$$x_t = \mu + u_{x,t}$$

for $t = 1, \ldots, T$. We assume:

- $u_{y,t} = \psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ where $\sum_{j=0}^{\infty} j \cdot |\psi_j| < \infty$ and $\{\epsilon_t\}$ is an $i.i.d$ sequence with mean zero and variance σ_ϵ^2, and finite fourth moment. σ_y^2 denotes the long run variance of $\{u_{y,t}\}$ and $\sigma_{u,y}^2$ denotes the contemporaneous variance of $\{u_{y,t}\}$;
- $u_{x,t} \sim i.i.d(0, \sigma_{u,x}^2)$;
- $y_0 = 0$;

We consider the estimation of the regression model

$$y_t = \beta x_t + error,$$

using a sample of T observations.

Consider the least squares estimator

$$\widehat{\beta} = \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2}.$$

(i) Derive and discuss the orders of probability and limiting distributions of the numerator and the denominator of $\widehat{\beta}$. Write down any necessary additional assumptions.

(ii) Use your findings for **(i)** to comment on the asymptotic behaviour of $\hat{\beta}$. What is wrong in this analysis?

Answer:

(a) (15 points) Ideally each item (i), (ii), (iii) is 5 points for correct answers but this might change for partially correct answers.

(i) We need to derive the order of probability and the limiting distribution of VR under the null hypothesis.

$$VR = \frac{T\left(\sum_{t=1}^{T}\Delta y_t\right)^2}{\sum_{t=1}^{T} y_t^2}$$

Under the null hypothesis we have

$$y_t = y_{t-1} + u_t,$$

which means we can use the standard results. Let's start with the numerator.

$$T\left(\sum_{t=1}^{T}\Delta y_t\right)^2 = T\left(\sum_{t=1}^{T} u_t\right)^2$$

Standard Result (1) is about this sum. It says $T^{-1/2}\sum_{t=1}^{T} u_t \xrightarrow{d} \sigma W(1)$. So we have $\sum_{t=1}^{T} u_t = O_p(T^{1/2})$. Then this implies for the denominator the following

$$T\left(\sum_{t=1}^{T}\Delta y_t\right)^2 = T\left(\sum_{t=1}^{T} u_t\right)^2 = O_p(T) \times [O_p(T^{1/2})]^2 = O_p(T^2).$$

And we can write

$$\frac{1}{T^2}T\left(\sum_{t=1}^{T}\Delta y_t\right)^2 \xrightarrow{d} \sigma^2[W(1)]^2$$

Now let's focus on the denominator.

$$\sum_{t=1}^{T} y_t^2$$

This is a standard sum, of which the limit is given in Standard Result (6). It implies $T^{-2}\sum_{t=1}^{T} y_t^2 \xrightarrow{d} \sigma^2 \int_0^1 W(r)^2 dr$. Then we have

$$\sum_{t=1}^{T} y_t^2 = O_p(T^2)$$

and

$$\frac{1}{T}\sum_{t=1}^{T} y_t^2 \xrightarrow{d} \sigma^2 \int_0^1 W(r)^2 dr.$$

Combining the results for numerator and the denominator, we get

$$VR = O_p(1)$$

and

$$VR = \frac{\frac{1}{T^2}T\left(\sum_{t=1}^{T}\Delta y_t\right)^2}{\frac{1}{T^2}\sum_{t=1}^{T} y_t^2} \xrightarrow{d} \frac{[W(1)]^2}{\int_0^1 W(r)^2 dr}.$$

Really nice nuisance parameter free distribution.

(ii) Under the alternative hypothesis, we have $|\rho| < 1$ which means we are in the stationary world. So the stationary asymptotics will apply. Also Δy_t is no longer equal to u_t. It is enough if you give a heuristic results that says

$$\sum_{t=1}^{T}\Delta y_t = O_p(T),$$

or ($\sum_{t=1}^{T}\Delta y_t = O_p(\sqrt{T})$) which gives

$$T\left(\sum_{t=1}^{T}\Delta y_t\right)^2 = O_p(T) \times [O_p(T)]^2 = O_p(T^3).$$

because Δy_t is stationary. For the denominator we have

$$\sum_{t=1}^{T} y_t^2 = O_p(T).$$

If we combine these results we get

$$VR = \frac{T\left(\sum_{t=1}^{T}\Delta y_t\right)^2}{\sum_{t=1}^{T} y_t^2} = \frac{O_p(T^3)}{O_p(T)} = O_p(T^2).$$

So under the alternative hypothesis the test statistic diverges to infinity.

(iii) Since under the null the test statistic is $O_p(1)$ and it is $O_p(T)$ under the alternative, we can conclude that the test is a consistent test for testing unit roots.

(b) (10 points)

(i) We need to derive and discuss the orders of probability and limiting distributions of the numerator and the denominator of $\widehat{\beta}$.

$$\widehat{\beta} = \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2}.$$

Here

$$y_t = \delta + y_{t-1} + u_{y,t} = \delta t + S_t$$
$$x_t = \mu + u_{x,t}$$

where $S_t = \sum_{s=1}^{t} u_{y,t}$ and S_t is a partial sum that obeys the standard results given in the first page of the exam.

Let's analyse the numerator. We have

$$\begin{aligned}\sum_{t=1}^{T} y_t x_t &= \sum_{t=1}^{T} (\delta t + S_t)(\mu + u_{x,t}) \\ &= \sum_{t=1}^{T} \delta\mu t + \sum_{t=1}^{T} \delta u_{x,t} + \sum_{t=1}^{T} S_t \mu + \sum_{t=1}^{T} S_t u_{x,t} \\ &= \delta\mu \sum_{t=1}^{T} t + \delta \sum_{t=1}^{T} u_{x,t} + \mu \sum_{t=1}^{T} S_t + \sum_{t=1}^{T} S_t u_{x,t}\end{aligned}$$

Standard results (10), (1), (5), (3), gives

$$\begin{aligned}\delta\mu \sum_{t=1}^{T} t &= O_p(T^2), \\ \delta \sum_{t=1}^{T} u_{x,t} &= O_p(T^{1/2}), \\ \mu \sum_{t=1}^{T} S_t &= O_p(T^{3/2}), \\ \sum_{t=1}^{T} S_t u_{x,t} &= O_p(T).\end{aligned}$$

Which means that the dominant term of the numerator is the first term. This implies

$$\sum_{t=1}^{T} y_t x_t = O_p(T^2)$$

and

$$\frac{1}{T^2}\sum_{t=1}^{T} y_t x_t = \frac{1}{T^2}\delta\mu\sum_{t=1}^{T} t + o_p(1) \xrightarrow{p} \frac{\delta\mu}{2}.$$

Now the denominator. We have

$$\begin{aligned}\sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T}(\mu + u_{x,t})^2 &= \sum_{t=1}^{T}\mu^2 + 2\mu\sum_{t=1}^{T} u_{x,t} + \sum_{t=1}^{T} u_{x,t}^2 \\ &= T\mu^2 + 2\mu\sum_{t=1}^{T} u_{x,t} + \sum_{t=1}^{T} u_{x,t}^2\end{aligned}$$

Standard results (1) and (2) gives

$$\sum_{t=1}^{T} u_{x,t} = O_p(T^{1/2}),$$

$$\sum_{t=1}^{T} u_{x,t}^2 = O_p(T),$$

and obviously

$$T\mu^2 = O_p(T).$$

Combining these gives

$$\sum_{t=1}^{T} x_t^2 = O_p(T)$$

and

$$\frac{1}{T}\sum_{t=1}^{T} x_t^2 = \mu^2 + \frac{1}{T}\sum_{t=1}^{T} u_{x,t}^2 + o_p(1) \xrightarrow{p} \mu^2 + \sigma_{u,x}^2$$

(ii) So we found that the numerator of $\hat{\beta}$ is $O_p(T^2)$ and the denominator is $O_p(T)$. This means that

$$\hat{\beta} = O_p(T).$$

It diverges to infinity. This means that are doing something wrong by considering this regression. We can also write

$$\frac{1}{T}\hat{\beta} \xrightarrow{p} \frac{\frac{\delta\mu}{2}}{\mu^2 + \sigma_{u,x}^2}.$$

In order to see what we are doing wrong we need to look at the DGPs of these processes. y_t is a random walk with a drift and x_t is driven by a constant and a stationary error. It doesn't make sense to consider a regression of y_t on x_t and this question we show that what happens when we consider this regression. The estimator of β diverges to infinity as T increases, whereas it should be zero.

Question 4: Empirical Application (25 points out of 100 points)

(a) (10 points) Samwise is an econometrics student from Rohan University. He wants to analyze the evolution of average atmospheric temperature and gross domestic product over time and use his findings to investigate the relation between these variables. He collects annual data on these variables from two countries for the period 1920 - 2020. These countries are called Gondor and Isengard. The retained variables are

- log of average annual atmpospheric temperature of Gondor ($temp_{g,t}$),
- log of gross domestic product of Gondor ($gdp_{g,t}$)
- log of average annual atmpospheric temperature of Isengard ($temp_{i,t}$)
- log of gross domestic product of Isengard ($gdp_{i,t}$)

He first tests for unit roots by using the Dickey-Fuller test for the null hypothesis of unit root against alternative hypothesis of trend stationarity.

Variables	DF-statistic	5 % $t-$distribution critical values	5% Dickey-fuller critical values without trend	5% Dickey-fuller critical values with trend
$temp_{g,t}$	-1.89	-1.64	-2.89	-3.45
$gdp_{g,t}$	-1.67	-1.64	-2.89	-3.45
$temp_{i,t}$	-2.56	-1.64	-2.89	-3.45
$gdp_{i,t}$	-3.32	-1.64	-2.89	-3.45

Given this unit root test output he wants to estimate a vector error correction model for the 4 variables that he has data on.

(i) Comment on the results of the unit root tests. Discuss whether the choice of the unit root test is justifiable. Do you have any recommendations to Samwise that would improve his unit root analysis? If yes, what are they? Discuss in detail.

(ii) Given the unit root test results, what are your suggestions for the VECM model that Samwise wants to construct and analyze? If he wants to discover the potential cointegrating relations between the 4 variables, how should he proceed? Explain the steps he needs to follow in detail.

(b) (15 points) Another econometrics student from Rohan University named Frodo is analyzing a panel data set of daily wind speed and insurance claims for 150 towns over 360 days. Let $w_{i,t}$ denote the average wind speed in town i on day t and $c_{i,t}$ denote the total daily amount of insurance claims made from town i on day t. Frodo considers the model

$$c_{i,t} = \beta_i w_{i,t} + u_{i,t}.$$

He suspects that there might be correlation between the error terms of the models for wind speeds of different towns, such that

$$Cov(u_{i,t}u_{j,t}) \neq 0, \text{for } i \neq j.$$

He starts looking for an advice on how to analyze this panel data set.

(i) One of the ways to model this correlation is to assume a "spatial" model. Another way is to assume the existence of an "unobserved common factor". Explain to Frodo how these two approaches can be used model to the correlation between $u_{i,t}$ and $u_{j,t}$ for $i \neq j$.

(ii) Assume that there is an unobserved factor that affects $u_{i,t}$ and $w_{i,t}$ for all $i = 1, \ldots, N$. Explain to Frodo why it is not a good idea to ignore this factor structure even if he wants to estimate the individual specific coefficients, β_i by using OLS.

(iii) Now assume that β_i follows the random coefficient model, such that

$$\beta_i = \beta + \nu_i,$$

where $\nu_i \sim i.i.d(0, \omega_\nu)$, for $i = 1, \ldots, N$ and $|\beta| < K < \infty$, $0 < \omega_\nu < \infty$. In the presence of unobserved common factors, explain to Frodo, how he can use Pesaran's Common-Correlated-Effects (CCE) estimator to obtain a consistent estimator for β_i and for β.

Answer:

(a) (10 points) Each item is 5 points.

(i) We need to comment on the results of the unit root tests. We need to compare the test statistics with the Dickey-Fuller critical values with trend. Because the test Samwise is considering is the Dickey-Fuller test against the alternative hypothesis of trend stationarity. According to the test statistics for all series we do not reject the null hypothesis of unit root. So we can conclude that all 4 series have unit root. But Samwise, in his analysis does not take into account the possible serial correlation. What we saw in the lectures is that when there is serial correlation in the errors, we Dickey-Fuller test is no longer valid to use. We discussed some test such as the Augmented Dickey Fuller test or the Phillips Perron. A discussion of these tests is sufficient.

(ii) Now Samwise has 4 series with unit roots. In order to uncover the potential cointegrating relations between these 4 unit root series, Samwise can follow the systems approach of Johansen. In order to do it first he needs to determine the number of cointegrating vectors. This can be determined by Johansen's trace test or maximum eigenvalue test. And then Samwise should proceed by estimating by Maximum Likelihood the VECM model. Then the cointegrating relations will be revealed. You need to explain these methods in order to get points from this question.

(b) (15 points) Each item is 5 points.

(i) We need to comment on the results of the unit root tests. We need to compare the test statistics with the Dickey-Fuller critical values with trend. Because the test Samwise is considering is the Dickey-Fuller test against the alternative hypothesis of trend stationarity. According to the test statistics for all series we do not reject the null hypothesis of unit root. So we can conclude that all 4 series have unit root. But Samwise, in his analysis does not take into account the possible serial correlation. What we saw in the lectures is that when there is serial correlation in the errors, we Dickey-Fuller test is no longer valid to use. We discussed some test such as the Augmented Dickey Fuller test or the Phillips Perron. A discussion of these tests is sufficient. Factor structure approach assumes that there are global factors that are effecting each cross-section unit. This effect is not required to be the same. A global technology shock will affect all the countries. But the effect of this shock on each country will depend on the infrastructure regarding the production processes in this country. An earthquake will affect all the households in a neighbourhood, but the influence of this shock on each household will depend on the households' economic situation.

(ii) The factor structure approach gives

$$y_{i,t} = \boldsymbol{\beta}_i'\mathbf{x}_{i,t} + e_{i,t},$$
$$e_{i,t} = \boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{i,t}.$$

If $\mathbf{f}_t$ and $\boldsymbol{\gamma}_i$ are independently distributed of $\boldsymbol{\beta}_i$ and $\mathbf{x}_{i,t}$ for all i, t, then the estimators $\widehat{\boldsymbol{\beta}}_i$, $\widehat{\boldsymbol{\beta}}_P$ and $\widehat{\boldsymbol{\beta}}_{MG}$ are still consistent. This means if the global factors are affecting only the variable $y_{i,t}$ but not the variables contained in $\mathbf{x}_{i,t}$, then the consistency of the ordinary estimators is not affected by presence of cross-sectional dependence. However, it is not realistic to assume, for example, that a global oil price shock affects only the GDP $(y_{i,t})$ of the countries but not their exports, imports, investments $(\mathbf{x}_{i,t})$. So we need to allow the global factors to affect $\mathbf{x}_{i,t}$ as well. We do this by assuming that $\mathbf{x}_{i,t}$ has the following DGP

$$\mathbf{x}_{i,t} = \boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{i,t},$$

where $\mathbf{f}_t$ are the unobserved common factors that are affecting $y_{i,t}$ as well, $\boldsymbol{\Gamma}_i$ is a $m \times k$ factor loading matrix and $\mathbf{v}_{i,t}$ are the individual specific components. Now this is a problem because the presence of $\mathbf{f}_t$ in both y_t and $\mathbf{x}_t$ introduces the problem of enodgeneity.

(iii) Here you need to explain the method of Pesaran (2006). Pesaran writes $\mathbf{z}_{i,t} = (y_{i,t}, \mathbf{x}_{i,t}')'$:

$$\mathbf{z}_{i,t} = \mathbf{C}_i'\mathbf{f}_t + \mathbf{u}_{i,t},$$

where

$$\mathbf{u}_{i,t} = \begin{pmatrix} \varepsilon_{i,t} + \boldsymbol{\beta}'\mathbf{v}_{i,t} \\ \mathbf{v}_{i,t} \end{pmatrix},$$

and

$$\mathbf{C}_i = \begin{pmatrix} \boldsymbol{\gamma}_i + \boldsymbol{\Gamma}_i\boldsymbol{\beta}_i & \boldsymbol{\Gamma}_i \end{pmatrix},$$

here $\mathbf{C}_i$ is a $m \times (k+1)$ matrix. So we have

$$\mathbf{z}_{i,t} = \mathbf{C}_i'\mathbf{f}_t + \mathbf{u}_{i,t}.$$

We can use this to find suitable proxies for the unobserved factor space. In order to obtain suitable proxies, we will take the cross-sectional averages of the observed variables. This yields

$$\overline{\mathbf{z}}_t = \overline{\mathbf{C}}'\mathbf{f}_t + \overline{\mathbf{u}}_t,$$

where

$$\overline{\mathbf{z}}_t = \frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_{i,t},\ \overline{\mathbf{C}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{C}_i,\ \overline{\mathbf{u}}_t = \frac{1}{N}\sum_{i=1}^{N}\mathbf{u}_{i,t}.$$

We can rearrange

$$\overline{\mathbf{z}}_t = \overline{\mathbf{C}}'\mathbf{f}_t + \overline{\mathbf{u}}_t,$$

and solve it for $\mathbf{f}_t$. Here, under certain assumptions we can show that

$$\overline{\mathbf{u}}_t \xrightarrow{q.m.} \mathbf{0},$$

as $N \to \infty$ for each t. Here, "$q.m.$" signifies convergence in quadratic mean. And we can show that

$$\overline{\mathbf{z}}_t - \overline{\mathbf{C}}'\mathbf{f}_t \xrightarrow{p} 0,$$

as $N \to \infty$. Pesaran (2006) suggests using the cross-sectional averages as a proxy for the unobserved factors and estimate the following augmented regression model.

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \overline{\mathbf{Z}}\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i^*$$

Pesaran (2006) is not interested in estimating $\boldsymbol{\theta}_i$. For this reason we can use the orthogonal projection matrix while constructing our estimator. We define

$$\overline{\mathbf{M}} = \mathbf{I}_T - \overline{\mathbf{Z}}(\overline{\mathbf{Z}}'\overline{\mathbf{Z}})^{-1}\overline{\mathbf{Z}}'.$$

For more information on this $\overline{\mathbf{M}}$ matrix, search for "orthogonal projection matrix". Then the estimators can be written as

$$\widehat{\boldsymbol{\beta}}_{i,CCE} = (\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{X}_i)^{-1}\mathbf{X}_i'\overline{\mathbf{M}}\mathbf{y}_i,$$

$$\widehat{\boldsymbol{\beta}}_{P,CCE} = \left(\sum_{i=1}^{N} \mathbf{X}_i' \overline{\mathbf{M}} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \overline{\mathbf{M}} \mathbf{y}_i,$$

$$\widehat{\boldsymbol{\beta}}_{MG,CCE} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\beta}}_{i,CCE} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}_i' \overline{\mathbf{M}} \mathbf{X}_i)^{-1} \mathbf{X}_i' \overline{\mathbf{M}} \mathbf{y}_i$$

Pesaran shows that these estimators are consistent for $\boldsymbol{\beta}$ under certain assumptions and conditions placed on the growth rates of N and T.