# Solutions of the Sample Exam I
# Multivariate Econometrics
# VU Econometrics and Operations Research
# 2019 – 2020

*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 $\{x_t\}$ follows the following data generating process:

$$x_t = x_{t-1} + u_t,$$

where $x_0 = 0$ 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 $\sigma_\epsilon^2$, and finite fourth moment;

**(b)** $\sigma^2$ denotes the long run variance of $\{u_t\}$ and $\sigma_u^2$ denotes the contemporaneous variance of $\{u_t\}$.

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

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

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

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

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

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

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

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

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

# 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 of the concepts mentioned in the statement, **(ii)** your judgement about the statement about whether it is correct/wrong/unclear/incomplete, **(iii)** a correction of the statement. A formal answer using mathematics is possible, sometimes very useful but not always necessary.

**(a)** (5 points) Mixing and stationarity are important properties that a random sequence might have. About these properties the following is always true: If a process is mixing then it is stationary.

**(b)** (10 points) Spurious regression is a phenomena we encounter when we regress a random walk on a time trend. The resulting estimator from this regression converges to a random variable and the $t-$statistic of this regression is consistent.

**(c)** (10 points) Cross sectional dependence is encountered in panel data models. Suppose having the panel data variables $y_{i,t}$ and $x_{i,t}$ for $i = 1, \dots, N$ and $t = 1, \dots, T$ with the data generating process

$$y_{i,t} = \beta x_{i,t} + u_{i,t}.$$

The presence of cross sectional dependence implies correlation along time, such that

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

One way to eliminate the problems that might be induced by cross sectional dependence is to consider the regression of $y_{i,t}$ on $x_{i,t}$ and on the lags of $y_{i,t}$.

**Answer:**

Grading of these questions:

**(i)** short explanations of the concepts mentioned in the statement, (40% of total points for the question)

**(ii)** your judgement about the statement about whether it is correct/wrong/unclear/ incomplete and a discussion on your judgement, (40% of total points for the question)

**(iii)** a correction of the statement.(20% of total points for the question)

**(a)** (5 points) If for the Gaussian VAR(p) process $\{\mathbf{x}_t\}$, $\boldsymbol{\delta}_t = 0$, then

- $E(\mathbf{x}_t) = \mathbf{0}$;
- $Var(\mathbf{x}_t) = \boldsymbol{\Omega}_x$;
- $E(\mathbf{x}_t\mathbf{x}'_{t-j}) = \boldsymbol{\Lambda}^j\boldsymbol{\Omega}_x$.

We see that the Gaussian pair $\mathbf{x}_t$ and $\mathbf{x}_{t-j}$ are tending to independence as $j \to \infty$ because $E(\mathbf{x}_t\mathbf{x}'_{t-j}) = \boldsymbol{\Lambda}^j\boldsymbol{\Omega}_x \to \mathbf{0}$. This characteristic is called "restricted memory" or mixing and the process $\mathbf{x}_t$ is an example of a mixing process. So in general, the realization of the sequence at time $t$ is not informative about the realization at either $t-j$ or $t+j$, when $j$ is sufficiently large; the present is not informative about either the remote past or remote future.

A random sequence $\{\mathbf{x}_t\}$ is said to be *stationary in the wide sense* (covariance stationary), if the mean, the variance and the sequence of $j$-th order autocovariances for $j > 0$ are all independent of $t$. A random sequence $\{\mathbf{x}_t\}$ is said to be *stationary in the strict sense*, if for every $k > 0$, the joint distributions of all collections $(\mathbf{x}_t, \mathbf{x}_{t+1}, \ldots, \mathbf{x}_{t+k})$ do not depend in any way on $t$.

The statement is incorrect. A stationary process does not have to be mixing. A mixing process does not have to be stationary. Consider the following counter example. Let $x_t$ be an *i.i.d.* process with $E(x_t) = 0$, and $z$ is any random variable not depending on $t$, with $E(z) = 0$ and independent of $x_t$ for all $t$. Consider,

$$y_t = x_t + z.$$

We can show that $\{y_t\}$ is a stationary sequence. To check for the mixing property of $\{y_t\}$, we consider

$$Cov(y_t, y_{t-j}) \to Var(z) > 0,$$

which implies that $y_t$ is not mixing.

**(b)** (10 points) Let us again consider two *unrelated* variables $x_t$ and $y_t$ such that

$$x_t = x_{t-1} + u_t$$
$$y_t = y_{t-1} + v_t$$

where $u_t \sim$ i.i.d$(0, \sigma_x^2)$, $v_t \sim$ i.i.d$(0, \sigma_y^2)$, $x_0 = 0$ and $y_0 = 0$. Let's now run

$$y_t \;=\; \beta x_t + e_t.$$

Theory show that this regression will be a spurious regression. In a spurious regression of this setup we expect

- $\widehat{\beta} \xrightarrow{p} 0$
- $R^2 \xrightarrow{p} 0$
- $t_{\widehat{\beta}} \xrightarrow{p} t - \text{dist}$

But instead we have

- $\widehat{\beta} \xrightarrow{d} f(B)$
- $R^2 \xrightarrow{d} g(B)$
- $T^{-1/2} t_{\widehat{\beta}} \xrightarrow{d} h(B)$,

where $f(B), g(B)$ and $h(B)$ are functions of Brownian motions. So spurious regression occurs when we regress a random walk on another random walk. In this case, the resulting estimator indeed converges to a random walk. But the t-statistic is inconsistent. Because the t-statistic itself diverges under the null hypothesis, we can see it from the result $T^{-1/2} t_{\widehat{\beta}} \xrightarrow{d} h(B)$.

**(c)** (10 points) First of all the statement in the question states that, cross-sectional dependence implies

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

This is wrong. Above equation is about being correlated over time. Cross-sectional dependence is about correlation across cross-section units. The solution proposed to cross-sectional dependence is also wrong. Regression of $y_{i,t}$ on $x_{i,t}$ and on the lags of $y_{i,t}$ might perhaps help with the serial correlation in the errors. But it will not help to alleviate the problems of cross-sectional dependence.

If the error term of the model is cross-sectionally correlated we will have

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

So the error term that belongs to the model of $j^{th}$ cross section unit is correlated with the error term that belongs to the model of $i^{th}$ cross section unit. A potential solution to this problem is proposed by Pesaran (2006). He suggests augmenting the regression model with the cross-sectional averages of the variables. He shows in his paper that this would overcome the problems caused by cross-sectional dependence, under certain assumptions.

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

Suppose that we have the following bivariate data generating process for $\mathbf{w}_t = (y_t, x_t)'$.

$$\mathbf{w}_t = \boldsymbol{\Gamma}_1 \mathbf{w}_{t-1} + \boldsymbol{\Gamma}_2 \mathbf{w}_{t-2} + \boldsymbol{\varepsilon}_t,$$

where

$$\boldsymbol{\Gamma}_1 = \begin{pmatrix} \gamma_{1yy} & \gamma_{1yx} \\ \gamma_{1xy} & \gamma_{1xx} \end{pmatrix}, \quad \boldsymbol{\Gamma}_2 = \begin{pmatrix} \gamma_{2yy} & \gamma_{2yx} \\ \gamma_{2xy} & \gamma_{2xx} \end{pmatrix}$$

and

$$\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim IN \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right],$$

for $t = 1, \ldots, T$. This model can be written in a vector error correction model (VECM) form as

$$\Delta \mathbf{w}_t = \mathbf{A} \mathbf{w}_{t-1} + \mathbf{B} \Delta \mathbf{w}_{t-1} + \boldsymbol{\varepsilon}_t.$$

Answer the following questions.

**(a)** (10 points) Starting from the VAR(2) model derive the VECM and write $\mathbf{A}$ and $\mathbf{B}$ in terms of the parameters of the VAR(2).

**(b)** (10 points) Now consider the vector error correction model. Suppose that $\mathbf{A}$ can be written as

$$\mathbf{A} = \boldsymbol{\alpha} \boldsymbol{\beta}',$$

where $\boldsymbol{\alpha}$ is $2 \times 1$ and $\boldsymbol{\beta}$ is $2 \times 1$ and has the form

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} 1 \\ -\beta_1 \end{pmatrix}.$$

Furthermore, let

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ 0 & b_{22} \end{pmatrix}.$$

Derive the conditional error correction model of $y_t$ given $x_t$ and the past. How would you test for no-cointegration in this CECM. Under what condition(s) is $x_t$ weakly exogeneous for the parameters of interest $\phi = \{\alpha_1, \beta_1\}$.

**(c)** (5 points) Now let $\gamma_{1yx} = 0$ and $\gamma_{2yx} = 0$ and consider the model for $y_t$ only. Suppose that $\gamma_{1yy} = 0.7$ and $\gamma_{2yy} = -0.1$. Is $y_t$ stationary under these restrictions? Show your calculations.

**Answer:**

**(a)** (10 points) Starting from the VAR(2) model derive the VECM and write $\mathbf{A}$ and $\mathbf{B}$ in terms of the parameters of the VAR(2).

$$\begin{aligned}\mathbf{w}_t &= \mathbf{\Gamma}_1\mathbf{w}_{t-1} + \mathbf{\Gamma}_2\mathbf{w}_{t-2} + \boldsymbol{\varepsilon}_t \\ &= \mathbf{\Gamma}_1\mathbf{w}_{t-1} + \mathbf{\Gamma}_2\mathbf{w}_{t-1} - \mathbf{\Gamma}_2\mathbf{w}_{t-1} + \mathbf{\Gamma}_2\mathbf{w}_{t-2} + \boldsymbol{\varepsilon}_t \\ &= \mathbf{\Gamma}_1\mathbf{w}_{t-1} + \mathbf{\Gamma}_2\mathbf{w}_{t-1} - \mathbf{\Gamma}_2\Delta\mathbf{w}_{t-1} + \boldsymbol{\varepsilon}_t.\end{aligned}$$

Now we subtract $\mathbf{w}_{t-1}$ from both sides, this yields

$$\Delta\mathbf{w}_t = -\mathbf{w}_{t-1} + \mathbf{\Gamma}_1\mathbf{w}_{t-1} + \mathbf{\Gamma}_2\mathbf{w}_{t-1} - \mathbf{\Gamma}_2\Delta\mathbf{w}_{t-1} + \boldsymbol{\varepsilon}_t.$$

Rearranging yields

$$\Delta\mathbf{w}_t = (\mathbf{\Gamma}_1 + \mathbf{\Gamma}_2 - \mathbf{I}_2)\mathbf{w}_{t-1} - \mathbf{\Gamma}_2\Delta\mathbf{w}_{t-1} + \boldsymbol{\varepsilon}_t.$$

This is the VECM form. So we have

$$\begin{aligned}\mathbf{A} &= \mathbf{\Gamma}_1 + \mathbf{\Gamma}_2 - \mathbf{I}_2, \\ \mathbf{B} &= -\mathbf{\Gamma}_2\Delta\mathbf{w}_{t-1}.\end{aligned}$$

**(b)** (10 points) We can rewrite the error correction models for $y_t$ and $x_t$ by imposing the restrictions given in the question.

$$\Delta y_t = \alpha_1(y_{t-1} - \beta_1 x_{t-1}) + b_{11}\Delta y_{t-1} + b_{12}\Delta x_{t-1} + \varepsilon_{1,t},$$
$$\Delta x_t = \alpha_2(y_{t-1} - \beta_1 x_{t-1}) + b_{22}\Delta x_{t-1} + \varepsilon_{2,t},$$

Conditional error correction model of $y_t$ given $x_t$ can be obtained by including contemporaneous value of $\Delta x_t$ in the model of $\Delta y_t$. However, this will affect the coefficients of the model. In order to see how it effects the coefficients of the model, we need to acknowledge the fact that the error term of the conditional model should be independent of the error term of the marginal error correction model for $x_t$. Let $\varepsilon_{y.x,t}$ be the error term of the conditional error correction model. It should satisfy the following

$$E[\varepsilon_{y.x,t}\varepsilon_{2,t}] = 0$$

We can write $\varepsilon_{y.x,t} = \varepsilon_{1,t} - a\varepsilon_{2,t}$. Now, let's plug this in the above equation

$$\begin{aligned} E[(\varepsilon_{1,t} - a\varepsilon_{2,t})\varepsilon_{2,t}] &= E[\varepsilon_{1,t}\varepsilon_{2,t}] - aE[\varepsilon_{2,t}^2] \\ &= \sigma_{12} - a\sigma_{22} = 0. \end{aligned}$$

This means that

$$a = \frac{\sigma_{12}}{\sigma_{22}}.$$

The error term of the conditional model then should be

$$\varepsilon_{y.x,t} = \varepsilon_{1,t} - \frac{\sigma_{12}}{\sigma_{22}}\varepsilon_{2,t}.$$

In order to have this error term, we need to multiply the marginal ECM of $x_t$ with $\frac{\sigma_{12}}{\sigma_{22}}$ and subtract it from the ECM of $y_t$. This will give us the CECM of $y_t$ given $x_t$.

$$\frac{\sigma_{12}}{\sigma_{22}}\Delta x_t = \frac{\sigma_{12}}{\sigma_{22}}\alpha_2(y_{t-1} - \beta_1 x_{t-1}) + \frac{\sigma_{12}}{\sigma_{22}}b_{22}\Delta x_{t-1} + \frac{\sigma_{12}}{\sigma_{22}}\varepsilon_{2,t},$$

we have

$$\begin{aligned} \Delta y_t = &\left(\alpha_1 - \frac{\sigma_{12}}{\sigma_{22}}\alpha_2\right)(y_{t-1} - \beta_1 x_{t-1}) + \frac{\sigma_{12}}{\sigma_{22}}\Delta x_t \\ &+ b_{11}\Delta y_{t-1} + \left(b_{12} - \frac{\sigma_{12}}{\sigma_{22}}b_{22}\right)\Delta x_{t-1} + \varepsilon_{y.x,t}, \end{aligned}$$

This is the conditional error correction model.

The variable $x_t$ are weakly exogenous for the parameters of interest $\phi$ in the model $\Delta y_t$ if and only if

- $\phi$ is a function of the parameters of the conditional model only.
- $\phi$ are variation free of the parameters of the marginal ECM for $x_t$.

This is ensures when $\alpha_2 = 0$. In this case the CECM model boils down to

$$\begin{aligned} \Delta y_t = \alpha_1(y_{t-1} - \beta_1 x_{t-1}) + \frac{\sigma_{12}}{\sigma_{22}}\Delta x_t + b_{11}\Delta y_{t-1} \\ + \left(b_{12} - \frac{\sigma_{12}}{\sigma_{22}}b_{22}\right)\Delta x_{t-1} + \varepsilon_{y.x,t}. \end{aligned}$$

**(c)** (5 points) Now we go back to the VAR(2) model and consider the model for $y_t$ only, which can be written as

$$y_t = \gamma_{1yy} y_{t-1} + \gamma_{1yx} x_{t-1} + \gamma_{2yy} y_{t-2} + \gamma_{2yx} x_{t-2} + \varepsilon_{1,t}.$$

We can now impose the restrictions of the question, which are $\gamma_{1yx} = 0$, $\gamma_{2yx} = 0$, $\gamma_{1yy} = 0.7$ and $\gamma_{2yy} = -0.1$. These restrictions yield

$$y_t = 0.7 y_{t-1} - 0.1 y_{t-2} + \varepsilon_{1,t}.$$

In order to check if $y_t$ is stationary we can write the AR(2) model in the lag polynomial form. This will yield

$$(1 - 0.7L + 0.1L^2) y_t = \varepsilon_{1,t}.$$

For stability we need the roots of the polynomial $|(1 - 0.7z + 0.1z^2)|$ to be outside the unit circle. The roots are $z_1 = 5$ and $z_2 = 2$. They are both outside the unit circle, which means that under these restrictions $y_t$ is stable. In addition to stability, the model for $y_t$ does not have any mean shifts and the error term is identically distributions along $t$. All these makes $y_t$ a wide-sense stationary process.

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

Suppose that we have the following data generating process for $\{y_t\}$ and $\{x_t\}$:

$$y_t = \mu + y_{t-1} + \varepsilon_{1,t}$$
$$x_t = \delta t + \varepsilon_{2,t},$$

for $t = 1, \ldots, T$. We assume that $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ satisfy the same assumptions as the error process $u_t$ that are indicated in **(a)** in Page 2. Additionally, we assume:

**(i)** $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are uncorrelated with each other;

**(ii)** Long run variances of $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are $\sigma_1^2$ and $\sigma_2^2$, respectively;

**(iii)** $y_0 = 0$ and $x_0 = 0$;

**(iv)** $\mu$ and $\delta$ are non-zero constants.

We consider the least squares estimation of the following model:

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

using a sample of $T$ observation pairs. Answer the questions below.

**(a)** (10 points) Derive and discuss the orders of probability and limiting distributions of

$$\sum_{t=1}^{T} x_t^2, \quad \sum_{t=1}^{T} x_t y_t.$$

**(b)** (5 points) Derive the limiting distribution of the OLS estimator $\hat{\beta}$. Interpret your results.

**(c)** (10 points) The $t-$statistic to test for the significance of $\beta$ can be written as

$$t_{\beta=0} = \left( \sum_{t=1}^{T} x_t^2 \right)^{1/2} \hat{\beta}\, \hat{\sigma}_e^{-1},$$

where $\hat{\sigma}_e^2 = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{\beta} x_t)^2$. Derive the limiting distribution of this $t-$statistic. Interpret your result. (Note: use words if you cannot provide a formal mathematical answer.)

**Answer:**

**(a)** (10 points) We need to derive and discuss the orders of probability and limiting distributions of

$$\sum_{t=1}^{T} x_t^2, \quad \sum_{t=1}^{T} x_t y_t.$$

We start with $\sum_{t=1}^{T} x_t^2$. We have $x_t = \delta t + \varepsilon_{2,t}$, then we can write

$$\sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T} (\delta t + \varepsilon_{2,t})^2 = \sum_{t=1}^{T} [\delta^2 t^2 + 2\delta t \varepsilon_{2,t} + \varepsilon_{2,t}^2]$$
$$= \delta^2 \sum_{t=1}^{T} t^2 + 2\delta \sum_{t=1}^{T} t\varepsilon_{2,t} + \sum_{t=1}^{T} \varepsilon_{2,t}^2.$$

We need to find the limits of the three sums above. For the first term we can use the standard result (8). This tells us that the first term is $O(T^3)$ and if we divide it by $T^3$, it will be bounded. Now if we take the limit as $T \to \infty$ we get

$$\frac{1}{T^3} \delta^2 \sum_{t=1}^{T} t^2 \to \frac{\delta^2}{3}.$$

The second term is the sum of the multiplication of a time trend and a stationary random process. We can use the standard result (3) for this. The limit of this process is given by

$$T^{-3/2} 2\delta \sum_{t=1}^{T} t\varepsilon_{2,t} \xrightarrow{d} 2\delta\sigma_2 W(1) - 2\delta\sigma_2 \int_0^1 W(r)dr.$$

This means that $2\delta \sum_{t=1}^{T} t\varepsilon_{2,t} = O_p(T^{3/2})$. Now we analyze the third term, sum of squared errors. This is something that we are familiar with for many years. By law of large numbers we have:

$$T^{-1} \sum_{t=1}^{T} \varepsilon_{2,t}^2 \xrightarrow{p} \sigma_2^2.$$

This means that $\sum_{t=1}^{T} \varepsilon_{2,t}^2 = O_p(T)$. Combining the results for all three terms yields that the order of $\sum_{t=1}^{T} x_t^2$ is dominated by the first term, $\delta^2 \sum_{t=1}^{T} t^2$ which has an order of $O_p(T^3)$. This means that $\sum_{t=1}^{T} x_t^2$ has order $O_p(T^3)$. In order to find a bounded limit we need to divide it by $T^3$. Then we have

$$\frac{1}{T^3} \sum_{t=1}^{T} x_t^2 + o_p(1) \xrightarrow{p} \frac{\delta^2}{3},$$

because the second and third terms are $O_p(T^{3/2})$ and $O_p(T)$, respectively and asymptotically negligible when we divide them by $T^3$.

Now we need to analyse $\sum_{t=1}^{T} x_t y_t$. We can see that $y_t$ is not a random walk. But we can write it in terms of sum of a linear trend and a random walk (partial sum). We can write

$$\begin{aligned} y_t &= \mu + y_{t-1} + \varepsilon_{1,t} \\ &= \mu t + S_t, \end{aligned}$$

where

$$S_t = \sum_{s=1}^{t} \varepsilon_{1,s}.$$

Here the standard results given in the first page of the exam applies to $S_t$, not to $y_t$. Now we consider the sum we need to evaluate.

$$\begin{aligned} \sum_{t=1}^{T} x_t y_t &= \sum_{t=1}^{T} (\delta t + \varepsilon_{2,t})(\mu t + S_t) \\ &= \delta\mu \sum_{t=1}^{T} t^2 + \delta \sum_{t=1}^{T} S_t t + \mu \sum_{t=1}^{T} t\varepsilon_{2,t} + \sum_{t=1}^{T} S_t \varepsilon_{2,t}. \end{aligned}$$

So we need to find the orders and limits of these four terms by using the standard results. Let's start with the first term on the right hand side above. We can use standard result (8) for this. We have

$$\delta\mu \frac{1}{T^3} \sum_{t=1}^{T} t^2 \to \frac{\delta\mu}{3}.$$

So the first term has order $O_p(T^3)$. In order to evaluate the second term we can use the standard result (6). We have

$$\delta \frac{1}{T^{5/2}} \sum_{t=1}^{T} S_t t = \delta \frac{1}{T^{5/2}} \sum_{t=1}^{T} S_{t-1} t + \delta \frac{1}{T^{5/2}} \sum_{t=1}^{T} \varepsilon_{1,t} t \xrightarrow{d} \delta\sigma_1 \int_0^1 rW(r)dr,$$

In order to obtain this result in addition to standard result (8), we used standard result (3) which suggests that

$$\delta \frac{1}{T^{5/2}} \sum_{t=1}^{T} \varepsilon_{1,t} t = \delta \frac{1}{T} \left( \frac{1}{T^{3/2}} \sum_{t=1}^{T} \varepsilon_{1,t} t \right) = o_p(1).$$

So the second term has order $O_p(T^{5/2})$, which is smaller than the order of the first term. Now we analyze the third term $\mu \sum_{t=1}^{T} t\varepsilon_{2,t}$. We can use result number (3) directly, which implies

$$\mu \frac{1}{T^{3/2}} \sum_{t=1}^{T} t\varepsilon_{2,t} \xrightarrow{d} \sigma_2 \left\{ W(1) - \int_0^1 W(r)dr \right\},$$

which implies that the order of $\mu \sum_{t=1}^{T} t\varepsilon_{2,t} = O_p(T^{3/2})$. For the fourth term, we can use the standard result (2) which suggests that the sum is order $O_p(T)$. Combining all these will imply that the sum $\sum_{t=1}^{T} x_t y_t$ is dominated by the first term and the order is $T^3$. We can write

$$\frac{1}{T^3} \sum_{t=1}^{T} x_t y_t + o_p(1) \xrightarrow{p} \frac{\delta\mu}{3},$$

This answers the first part of the question.

**(b)** (5 points) The distribution of the OLS estimator can be derived simply by using the results we obtained in the first part of the questions. OLS estimator can be written as

$$\hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2},$$

We know from part (a) that both the numerator and the denominator has the order $O_p(T^3)$ so if we divide both by $T^3$ they will converge to something bounded. We have

$$\hat{\beta} = \frac{\frac{1}{T^3}\sum_{t=1}^{T} x_t y_t}{\frac{1}{T^3}\sum_{t=1}^{T} x_t^2} \xrightarrow{p} \frac{\frac{\delta\mu}{3}}{\frac{\delta^2}{3}} = \frac{\mu}{\delta}.$$

This result implies that the OLS estimator of $\beta$ will converge in probability to $\frac{\mu}{\delta}$, a constant. It should have converge in probability to 0. As long as $\mu \neq 0$, the estimation of this regression will yield false results. We can consider this regression as a member of the family of spurious regressions.

**(c)** (10 points) The $t-$statistic to test for the significance of $\beta$ can be written as

$$t_{\beta=0} = \left( \sum_{t=1}^{T} x_t^2 \right)^{1/2} \hat{\beta} \; \hat{\sigma}_e^{-1},$$

where $\hat{\sigma}_e^2 = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{\beta} x_t)^2$. In order to derive the limiting distribution of this $t-$statistic, we need to use the limit results from parts (a) and (b). We have the limit result for $\hat{\beta}$, we have the limit result of $\sum_{t=1}^{T} x_t^2$. We need to obtain the limit of $\hat{\sigma}_e$. For this we write

$$\hat{\sigma}_e = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{\beta} x_t)^2 = \frac{1}{T} \sum_{t=1}^{T} y_t^2 + \hat{\beta}^2 \frac{1}{T} \sum_{t=1}^{T} x_t^2 - 2\hat{\beta} \frac{1}{T} \sum_{t=1}^{T} y_t x_t.$$

We need to find the limits of these terms. We know the limits of almost all of them, we are missing the limit of only the first term. We can obtain is as follows.

$$\sum_{t=1}^{T} y_t^2 = \sum_{t=1}^{T} (\mu t + S_t)^2 = \mu^2 \sum_{t=1}^{T} t^2 + 2\mu \sum_{t=1}^{T} S_t t + \sum_{t=1}^{T} S_t^2.$$

Using standard results (8), (6), (3) and (5) yields

$$\frac{1}{T^3} \sum_{t=1}^{T} y_t^2 = \mu^2 \frac{1}{T^3} \sum_{t=1}^{T} t^2 + o_p(1) \xrightarrow{p} \frac{\mu^2}{3}.$$

Now we go back to analysing $\hat{\sigma}_e$. If we combine all the limit results we have

$$\hat{\sigma}_e = T^2 \frac{1}{T^3} \sum_{t=1}^{T} y_t^2 + T^2 \hat{\beta}^2 \frac{1}{T^3} \sum_{t=1}^{T} x_t^2 - T^2 2\hat{\beta} \frac{1}{T^3} \sum_{t=1}^{T} y_t x_t.$$

Then we have

$$\begin{aligned} \frac{1}{T^2} \hat{\sigma}_e &\xrightarrow{p} \frac{\mu^2}{3} + \frac{\mu^2}{\delta^2} \frac{\delta^2}{3} - 2\frac{\mu}{\delta} \frac{\mu\delta}{3} \\ &= \frac{\mu^2}{3} + \frac{\mu^2}{3} - 2\frac{\mu^2}{3} \\ &= 0. \end{aligned}$$

We need to rewrite the test statistic now to have bounded limits for all the components. We have

$$\sqrt{T} t_{\beta=0} = \left( \frac{1}{T^3} \sum_{t=1}^{T} x_t^2 \right)^{1/2} \hat{\beta} \; \frac{1}{\frac{1}{T^2} \hat{\sigma}_e} \xrightarrow{p} \infty,$$

because $\frac{1}{T^2} \hat{\sigma}_e \xrightarrow{p} 0$. This means that the $t-$statistic is not consistent.

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

**(a)** An econometrics student from the VU wants to analyze money demand in the Netherlands using annual economic time series for the period 1917 − 2017. The variables she uses are the log of money stock (denoted by $m_t$), the log of the price index (denoted by $p_t$), the log of real GDP (denoted by $y_t$) and the central bank discount rate (denoted by $r_t$).

**(i)** (5 points) She runs Dickey-Fuller tests (based on regression models with both a constant and a trend: $x_t = \alpha + \delta t + \rho x_{t-1} + u_t$) for the levels of $m_t$, $p_t$, $y_t$ and $r_t$ and also for the first differences of these series. The outcomes are as follows

| Variables | $t-$stat levels | $t-$stat first differences | 2.5 % $t-$distribution critical values | 2.5% Dickey-fuller critical values |
|---|---|---|---|---|
| $m_t$ | -2.04 | -4.30 | -1.98 | -3.95 |
| $p_t$ | -2.98 | -7.23 | -1.98 | -3.95 |
| $y_t$ | -1.45 | -4.00 | -1.98 | -3.95 |
| $r_t$ | -0.94 | -3.97 | -1.98 | -3.95 |

Explain why she is doing this analysis. What is being tested? What is the test regression? What are the null hypothesis and the alternative hypothesis? How can we draw conclusions from the results in the table? What are the conclusions?

**(ii)** (10 points) In her report she reports the following regression result as well:

$$\widehat{\Delta m_t} = \underset{(0.03)}{-0.12}\,(m_{t-1} - \underset{(0.56)}{3.12}y_{t-1} - \underset{(0.03)}{0.18}r_{t-1}) + \underset{(0.34)}{0.68}\Delta m_{t-1} + \underset{(0.20)}{0.10}\Delta p_{t-1},$$

$R^2 = 0.56$, where the numbers in the parentheses are the associated standard errors. Explain why she reports this result. What is she investigating? Interpret the results. Is it correct to estimate this single equation model? Do you have any suggestions for alternative models and estimators that would improve the accuracy of the conclusions?

**(b)** Another econometrics student from the VU is analyzing a panel data set on yearly housing expenses of 434 families over 5 years. Previous research shows that housing expenses data is dynamic in nature due to the presence of the lagged version of this variables in its model and it is highly suspected that an individual specific effect is present in the error term of the model that represents the heterogeneity among the families. The student realizes these and assumes the following model, where he denotes the variable by $c_{i,t}$:

$$c_{i,t} = \lambda c_{i,t-1} + u_{i,t},$$

where $u_{i,t} = \mu_i + \varepsilon_{i,t}$. Then she regresses $\widetilde{c}_{i,t} = c_{i,t} - T^{-1} \sum_{t=1}^{T} c_{i,t}$ on $\widetilde{c}_{i,t-1} = c_{i,t-1} - \sum_{t=1}^{T} c_{i,t-1}$ and obtains the pooled OLS estimator as

$$\widehat{\lambda}_{FE} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{c}_{i,t} \widetilde{c}_{i,t-1}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{c}_{i,t-1}^{2}}.$$

She finds that $\widehat{\lambda}_{FE} = 0.8$. She thinks that she can rely on this result to say something reliable about the dimension of the dynamic nature of household consumption.

**(i)** (5 points) Describe the situation faced by this student. Motivate the use of $\widehat{\lambda}_{FE}$ as an estimator. What is the problem with this estimator? When does it occur? Why does it occur? Explain in detail.

**(ii)** (5 points) Propose and discuss briefly an alternative estimator that is more proper/optimal to be used in this setup.

**Answer:**

**(a)** This is an empirical example of testing for unit roots and estimating the cointegrating relations.

**(i)** (5 points) This part is about Dickey-Fuller tests. The test regression is

$$x_t = \alpha + \delta t + \rho x_{t-1} + u_t.$$

So it is a Dickey-Fuller test regression with a constant and a drift. The econometrician want to test for unit roots in her variables. This is to test for the hypothesis

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

against the alternative

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

If $x_t$ represents the levels of these variables, then it means the econometrician wants to test for the unit root in levels. If $x_t$ represents the first differences of the variables then it means that the econometrician wants to test for presence of unit roots in the first differences. The test statistic obtained from running the regression above does not have a standard distribution. It has nonstandard distribution. And if the errors of the model are serially uncorrelated this distribution is nuisance parameter free. And the critical values of this distribution is tabulated by Dickey-Fuller. The econometrician needs to use the 2.5% critical values of the Dickey-Fuller distribution to draw conclusions about the hypothesis. In this case the 2.5% t-distribution critical values are useless.
By looking at the results in the table, we can conclude that the null hypothesis of a unit root cannot be rejected for the levels of any of the variables. When we look at the results of first differences we see that the for all variables we reject the null of unit root. So we can conclude that all the series involved are I(1) series.

**(ii)** (10 points) The model she is estimating is the error correction model for $m_t$, the log of the money stock. In part (i) we found that all the variables are I(1). After this finding, it is natural to investigate if there is a long-run relation between these variables, whether they share a common stochastic trend. This can be done by using error correction models. So the econometrician is interested in the cointegrating relation between $m_{t-1}$, $y_{t-1}$ and $r_{t-1}$. Interestingly, she does not include the log of the price index in her analysis. The results suggest that the error correction term $(m_{t-1} - \beta_1 y_{t-1} - \beta_2 r_{t-1})$ is significantly in the model. This is implied by the significant coefficient estimate of $\alpha$ in the model

$$\Delta m_t = \alpha(m_{t-1} - \beta_1 y_{t-1} - \beta_2 r_{t-1}) + \gamma_1 \Delta m_{t-1} + \gamma_2 \Delta p_{t-1} + \epsilon_t.$$

The results suggest that all coefficients are significantly different than zero, except $\gamma_2$, which is the coefficient of the first difference of $\Delta p_{t-1}$. It is not correct to estimate a single equation error correction model before testing for certain assumptions. First of all, let's consider the case that there is only one (up to a normalization) cointegrating relation in the system, that is $(m_{t-1} - \beta_1 y_{t-1} - \beta_2 r_{t-1})$. Estimating the ECM for $\Delta m_t$ only causes loss of efficiency if the same error correction term is in the models for $y_t$ and $r_t$. In order to alleviate the loss of efficiency, the econometrician should estimate the conditional error correction model for $m_t$, conditional on $y_t$ and $r_t$. In this case for efficiency $m_t$ and $r_t$ has to be weakly exogenous for the parameters of the conditional error correction model. Now let's move away from the assumption of a single cointegrating relation. Suppose that there are more than one cointegrating vectors that are linearly independent from each other. Then, this single equation estimation will yield meaningless results. In this case we need to adapt a system estimation technique. This is provided by Johansen. First, we need to determine the number of linearly independent cointegrating vectors by using a Trace test or a maximum eigenvalue test. Then, after imposing the estimated number of cointegrating vectors, we need to use full information maximum likelihood estimation technique to estimate the cointegrating vectors and all other model parameters.

**(b)** This is an empirical exercise involving a panel data approach.

**(i)** (5 points) The student is facing a panel data problem that occurs when there is a dynamic panel data model with fixed effects. In the model she is analysing, the error term follows

$$u_{i,t} = \mu_i + \varepsilon_{i,t}.$$

Here, $\mu_i$ is the unobserved fixed effects. In short panels, where one has a large dimension for the cross-section units but a small dimension for the time-periods, this setup creates a problem. The dynamic nature of the panel creates a bias. Because $y_{i,t}$ is a function of $\mu_i$. $y_{i,t-1}$ is also a function of $\mu_i$. Hence the regressor $y_{i,t-1}$ is correlated with the error term $u_{t,i}$. The OLS estimator is biased and inconsistent even if the $v_{i,t}$ are serially uncorrelated. For this reason, she decides to use a fixed effects estimator. The fixed effects transformation of the model eliminates the $\mu_i$ from the models of individual cross-section units. Then one can use

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

as an estimator. This estimator is problematic too. It suffers from the well known Nickell bias problem. This is due to the following reason. Let

$$y_{i,t} - \overline{y}_i = \lambda(y_{i,t-1} - \overline{y}_{i,-1}) + (\varepsilon_{i,t} - \overline{\varepsilon}_i),$$

In the model above

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

**(ii)** (5 points) An alternative to the fixed effects estimator is the Anderson-Hsiao estimator. This estimator is unbiased but it is not the most optimal estimator. The idea behind this estimator is as follows. The first difference of the original model such that

$$y_{i,t} = \lambda y_{i,t-1} + \mu_i + \varepsilon_{i,t},$$

can be transformed into

$$\Delta y_{i,t} = \lambda \Delta y_{i,t-1} + \Delta \varepsilon_{i,t},$$

This transformation removes the individual effect but there is still correlation between the first differenced lagged dependent variables and the error term of the new model $\Delta v_{i,t}$. But a straightforward instrumental variables estimator is available: We may construct instruments for $\Delta y_{i,t-1}$ from the second and third lags of $y_{i,t}$: $y_{i,t-2}$ or $\Delta y_{i,t-2}$. Using $y_{i,t-2}$ or $\Delta y_{i,t-2}$ as instruments for $\Delta y_{i,t-1}$ solves the correlation problem if $v_{i,t}$ is $i.i.d.$: the instruments will be highly correlated with $\Delta y_{i,t-1}$ and uncorrelated with the error term $\Delta v_{i,t}$. This IV estimator leads to consistent estimates, but these estimators are not necessarily efficient, because it does not make use of all the available moment conditions and does not take into account the differenced structure on the residual disturbances.