Draft of Solutions of the Sample Exam II
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 $\{z_t\}$ 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 $v = 0, 1, \ldots$

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) (10 points) *Granger causality* and *strong exogeneity* imply *super exogeneity.*

(b) (5 points) Consider the following model for $\{y_t\}$:

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

for $t = 1, \ldots, T$. We can *test for a unit root in* y_t, by estimating the model by OLS and testing the hypothesis, $\mathcal{H}_0 : \rho = 1$ against the alternative $\mathcal{H}_1 : |\rho| < 1$. The *test statistic obtained from this regression* will have a $t-$distribution in finite samples regardless of the serial correlation structure of ϵ_t.

(c) (10 points) *Nickell bias* in *dynamic panel data models* is caused by the fixed effects transformation to eliminate *cross-sectional dependence.* A way to eliminate the Nickell bias is to estimate directly by OLS the first differenced model.

Answer:

(a) (10 points) For a causal relation to be unambiguous a measurable time interval must separate the cause and effect. Granger (1969) uses this idea. $\mathbf{y}_t$ is said to not Granger cause $\mathbf{z}_t$ if knowledge of $\mathbf{y}_t$ does not aid the prediction of $\mathbf{z}_{t+j}$ for some $j > 0$:

$$D_z(\mathbf{z}_t|\mathcal{Y}_{t-1}, \mathcal{Z}_{t-1}) = D_z(\mathbf{z}_t|\mathcal{Z}_{t-1}).$$

- $\mathbf{y}_t$ is said to cause $\mathbf{z}_t$ in Granger's sense if knowledge of $\mathbf{y}_t$ aids the prediction of $\mathbf{z}_{t+j}$ for some $j > 0$.

- Granger causality can arise for example if there is a common factor that is effecting $\mathbf{y}_t$ with a lag of one period and $\mathbf{z}_t$ with a lag of two periods. (No need for a "real" causal relationship, all that matters is the observable predictive power.)
- Granger causality is not related to conditioning but to marginalization with respect to $\mathcal{Y}_{t-1}$.
- Note that weak exogeneity is defined at a point in time and not with respect to the full sample, but Granger causality is defined w.r.t the entire history.
- Granger causality is a relationship between the variables.
- $\mathbf{y}_t$ may Granger-cause $\mathbf{z}_t$ without violating the weak exogeneity of $\mathbf{z}_t$.

$\mathbf{z}_t$ is strongly exogenous for the parameters of interest $\boldsymbol{\theta}$ in the model $D_{y|z}$ if

- $\mathbf{z}_t$ is weakly exogenous for $\boldsymbol{\theta}$;
- $\mathbf{y}_t$ does not Granger-cause $\mathbf{z}_t$.

- "Weakness" of weak exogeneity comes from the fact that is does not rule out feedback from the endogenous variables to the exogenous variables, with a lag.
- The existence of this kind of feedback does not effect the efficiency of the inference procedures of $\boldsymbol{\theta}$.
- Strongly exogenous variables can be treated as fixed from a statistical point of view.

Super exogeneity combines weak exogeneity and the invariance of conditional parameters to interventions changing marginal parameters. In order to define it formally let us assume that the policy maker can conduct interventions/actions $\mathcal{A}$ and these belong to a certain class/set of actions $\mathcal{A} \in \mathcal{C}$.

$\mathbf{z}_t$ is super exogeneous for $\boldsymbol{\theta} = f(\boldsymbol{\psi}_1)$ (parameters of the conditional model $\mathbf{y}_t|\mathbf{z}_t$) if

- $\mathbf{z}_t$ is weakly exogenous for $\boldsymbol{\theta}$;
- $\boldsymbol{\psi}_1$ is invariant to the changes in $\boldsymbol{\psi}_2$ induced by interventions belonging to $\mathcal{C}$.

The statement is false, because the implication relation does not hold here. Super exogeneity involves a policy parameter whereas Granger causality and strong exogeneity do not.

(b) (5 points) Unit root is implied by $\rho = 1$. For the OLS estimator of ρ we have the following asymptotic limit

$$T(\hat{\rho} - 1) = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}\epsilon_t}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2} \xrightarrow{d} \frac{[B(1)]^2 - \frac{\sigma_0^2}{\sigma_\epsilon^2}}{2\int_0^1 [B(r)]^2 dr}.$$

And the limit of the $t-$statistic is as follows

$$t_\lambda = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\{T^{-2}\sum_{t=1}^{T} y_{t-1}^2\}^{1/2} s_\epsilon} \xrightarrow{d} \frac{\frac{\sigma_\epsilon}{2\sigma_0}\left(B(1)^2 - \frac{\sigma_0^2}{\sigma_\epsilon^2}\right)}{\sqrt{\int_0^1 B(r)^2 dr}}.$$

The statement is false because the $t-$statistic does not have $t-$distribution. It has a nonstandard distribution. In general, this distribution depends on nuisance parameters which are the short-run and the long-run variances of the errors. In the case of serial uncorrelatedness, the long-run variance is equal to the short-run variance, which makes the nuisance parameters disappear from the regression. In this case, the asymptotic distribution is a Dickey Fuller distribution.

(c) (10 points) Consider the model

$$y_{i,t} = \delta y_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{i,t} + u_{i,t} \quad i = 1, \ldots, N;\ t = 1, \ldots, T$$

where δ is a scalar, $\mathbf{x}_{i,t}$ is $k \times 1$ and $\boldsymbol{\beta}$ is $k \times 1$. For now we assume

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

where $\mu_i \sim i.i.d.(0, \sigma_\mu^2)$ and $v_{i,t} \sim i.i.d.(0, \sigma_v^2)$.

We deal with unobserved heterogeneity, μ_i by applying the within (demeaning) transformation: this is the fixed effects model.

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:

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

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. In the model above

- $y_{i,t-1}$ is correlated with $\overline{v}_i$ by construction: the latter average contains $v_{i,t-1}$, which is correlated with $y_{i,t-1}$.
- $v_{i,t}$ 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. A major issue that arises in every panel data study that has potential implications on parameter estimation and inference is the possibility that the individual units are interdependent. In order to formally define cross-sectional dependence, we can consider the model

$$y_{i,t} = \boldsymbol{\beta}_i'\mathbf{x}_{i,t} + e_{i,t},$$

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.

The statement is wrong. Nickell bias in dynamic panel data models is caused by the fixed effects transformation to eliminate the unobserved heterogeneity. A way to eliminate the Nickell bias is to consider specific estimators such as Arrelano - Bond estimator or Anderson Hsiao estimator.

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

Suppose that we have the following bivariate error correction model for $\mathbf{w}_t = (y_t, x_t)'$.

$$\Delta \mathbf{w}_t = \mathbf{\Pi} \mathbf{w}_{t-1} + \mathbf{\Gamma} \Delta \mathbf{w}_{t-1} + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{\Pi}$ and $\mathbf{\Gamma}$ are 2×2 matrices. and

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

for $t = 1, \dots, T$. This model can be written in a vector moving average form (VMA) form as

$$\mathbf{w}_t = \mathbf{A}(L) \boldsymbol{\varepsilon}_t,$$

where $\mathbf{A}(L) = \sum_{i=0}^{\infty} \mathbf{A}_i L^i$. Answer the following questions.

(a) (10 points) Starting from the VECM model derive the VMA and write $\mathbf{A}_i$ for $i = 1, 2, \dots$ in terms of the parameters of the VECM.

(b) (5 points) Consider $\mathbf{\Pi}$ and answer the following questions.

- **(i)** Let rank$[\mathbf{\Pi}] = 0$. What does this imply for the elements of $\mathbf{w}_t$?
- **(ii)** Let rank$[\mathbf{\Pi}] = 1$. What does this imply for the elements of $\mathbf{w}_t$?
- **(iii)** Let rank$[\mathbf{\Pi}] = 2$. What does this imply for the elements of $\mathbf{w}_t$?

(c) (10 points) Let the rank of $\mathbf{\Pi}$ be equal 1 and consider the decomposition

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$$

where $\boldsymbol{\alpha}$ is 2×1 and $\boldsymbol{\beta}$ is 2×1 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

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{yy} & \gamma_{yx} \\ \gamma_{xy} & \gamma_{xx} \end{pmatrix},$$

(i) Derive the conditional error correction model (CECM) of y_t given x_t and the past.

(ii) How would you test for no-cointegration in this CECM.

(iii) Under what condition(s) is x_t weakly exogeneous for the parameters of interest $\phi = \{\alpha_1, \beta_1\}$.

(iv) Propose a practical method to test for the weak exogeneity of x_t for the parameters of interest.

Answer:

(a) (10 points) Let's first obtain the VAR(2) representation. We have

$$\Delta \mathbf{w}_t = \mathbf{\Pi} \mathbf{w}_{t-1} + \mathbf{\Gamma} \Delta \mathbf{w}_{t-1} + \boldsymbol{\varepsilon}_t,$$

that we can rewrite as

$$\mathbf{w}_t = (\mathbf{I} + \mathbf{\Pi} + \mathbf{\Gamma}) \mathbf{w}_{t-1} - \mathbf{\Gamma} \mathbf{w}_{t-2} + \boldsymbol{\varepsilon}_t,$$

Let's defne $\mathbf{A}_1 = \mathbf{I} + \mathbf{\Pi} + \mathbf{\Gamma}$ and $\mathbf{A}_2 = -\mathbf{\Gamma}$, which allows us to rewrite the model as

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

this is the VAR(2) model. Now we can obtain the VMA model. $\mathbf{w}_t = \mathbf{B}(L)\boldsymbol{\varepsilon}_t$ by defining the lag polynomial $\mathbf{A}(L) = \mathbf{I} - \mathbf{A}_1 L - \mathbf{A}_2 L^2$. So we have

$$\begin{aligned}
\mathbf{B}(L) &= \mathbf{B}_0 + \mathbf{B}_1 L + \mathbf{B}_2 L^2 + \cdots \\
\mathbf{A}(L)\mathbf{B}(L) &= \mathbf{B}_0 + \mathbf{B}_1 L + \mathbf{B}_2 L^2 + \cdots - \mathbf{A}_1 L - \mathbf{A}_1 \mathbf{B}_1 L^2 - \mathbf{A}_1 \mathbf{B}_2 L^3 - \cdots \\
&\quad - \mathbf{A}_2 L^2 - \mathbf{A}_2 \mathbf{B}_1 L^3 - \mathbf{A}_2 \mathbf{B}_2 L^4 - \cdots \\
&= \mathbf{I}.
\end{aligned}$$

This gives

$$\begin{aligned}
\mathbf{B}_0 &= \mathbf{I} \\
(\mathbf{B}_1 - \mathbf{A}_1) L &= \mathbf{0} & &\Rightarrow \mathbf{B}_1 = \mathbf{A}_1 \\
(\mathbf{B}_2 - \mathbf{A}_1 \mathbf{B}_1 - \mathbf{A}_2 \mathbf{B}_0) L^2 &= \mathbf{0} & &\Rightarrow \mathbf{B}_2 = \mathbf{A}_1 \mathbf{B}_1 + \mathbf{A}_2 \mathbf{B}_0 \\
&\vdots & &\quad \vdots \\
(\mathbf{B}_k - \mathbf{A}_1 \mathbf{B}_{k-1} - \mathbf{A}_2 \mathbf{B}_{k-2}) L^k &= \mathbf{0} & &\Rightarrow \mathbf{B}_k = \mathbf{A}_1 \mathbf{B}_{k-1} + \mathbf{A}_2 \mathbf{B}_{k-2}.
\end{aligned}$$

If we know what $\mathbf{A}_1$ is, this will be equal to $\mathbf{B}_1$, by using $\mathbf{B}_0 = \mathbf{I}$, $\mathbf{A}_1$, $\mathbf{B}_1$ and $\mathbf{A}_2$; we can obtain $\mathbf{B}_2$. Then by recursive substitutions given the knowledge of all $\mathbf{A}_i$'s we can find the values for $\mathbf{B}_i$'s.

(b) (5 points)

(i) If $r = 0$, then $\mathbf{w}_t$ is a vector of non-cointegrated I(1) series and the VECM boils down to a VAR in first differences.

(ii) If $r = 1$, then there is one cointegrating relation between the elements of $\mathbf{w}_t$.

(iii) If $r = 2$, $\mathbf{\Pi}$ is full rank, hence $\mathbf{w}_t$ is I(0).

(c) (10 points)

(i) (5 points)

$$\begin{aligned}
\Delta y_t &= \alpha_1(y_{t-1} - \beta_1 x_{t-1}) + \gamma_{yy}\Delta y_{t-1} + \gamma_{yx}\Delta x_{t-1} + \varepsilon_{y,t}; \\
\Delta x_t &= \alpha_2(y_{t-1} - \beta_1 z_{t-1}) + \gamma_{xy}\Delta y_{t-1} + \gamma_{xx}\Delta x_{t-1} + \varepsilon_{x,t},
\end{aligned}$$

In order to derive the conditional error correction model (CECM) of y_t given x_t and the past, we need to find $\varepsilon_{y.x,t}$ that is uncorrelated with $\varepsilon_{x,t}$, such that $E(\varepsilon_{y.x,t}\varepsilon_{x,t}) = 0$. Suppose that $\varepsilon_{y.x,t}$ follows

$$\varepsilon_{y.x,t} = \varepsilon_{y,t} - a\varepsilon_{x,t}.$$

In order to find a we write

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

this should be equal to zero, which gives

$$\sigma_{12} - a\sigma_{22} = 0 \longrightarrow a = \frac{\sigma_{12}}{\sigma_{22}}.$$

So then

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

This should be the error term of the conditional model. Then the way to obtain this conditional model is to first write the system in vector notation, which gives

$$\begin{aligned}
\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} &= \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} 1 & -\beta_1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} \\
&\quad + \begin{pmatrix} \gamma_{yy} & \gamma_{yx} \\ \gamma_{xy} & \gamma_{xx} \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix}
\end{aligned}$$

Now we premultiply both sides by

$$\begin{pmatrix} 1 \\ -\sigma_{12}/\sigma_{22} \end{pmatrix}.$$

This gives us

$$\begin{aligned} \begin{pmatrix} 1 & -\sigma_{12}/\sigma_{22} \end{pmatrix} \begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} &= \begin{pmatrix} 1 & -\sigma_{12}/\sigma_{22} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} 1 & -\beta_1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} \\ &+ \begin{pmatrix} 1 & -\sigma_{12}/\sigma_{22} \end{pmatrix} \begin{pmatrix} \gamma_{yy} & \gamma_{yx} \\ \gamma_{xy} & \gamma_{xx} \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \end{pmatrix} \\ &+ \begin{pmatrix} 1 & -\sigma_{12}/\sigma_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix} \end{aligned}$$

This can written as

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

This is the conditional error correction model.

(ii) (2 points) Testing for no cointegration can be done by testing whether the speed of adjustment coefficient is equal to zero or not.

$$\mathcal{H}_0 : \left(\alpha_1 - \frac{\sigma_{12}}{\sigma_{22}}\alpha_2\right) = 0.$$

(iii) (1 points) We do need additional assumptions to make the estimation more efficient. Because we see that the speed of adjustment parameter is a function of the model parameters of Δx_t. In this case if we ignore the marginal model for Δx_t and estimate the conditional model only, we will be ignoring useful information, that would in the end lead to loss of efficiency. The assumption we need here is the weak exogeneity of x_t from the parameters of the cointegrating vector, parameters of interest. This is ensured if α_2 is equal to zero. So if the marginal model for Δx_t is not error correcting, then x_t is weakly exogenous for the cointegration parameters.

(iv) (2 points) We can consider the ECM of z_t and test whether α_2 is equal to 0 or not.

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

(a) (10 points) Suppose that we have the following data generating process for $\{y_t\}$:

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

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 $\{u_t\}$;
- $y_0 = 0$;
- δ is a non-zero constant.

We consider the following regression model:

$$y_t = \mu t + e_t,$$

using a sample of T observation pairs. Consider the least squares estimator

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

Derive and discuss the orders of probability and limiting distributions of the numerator and the denominator of $\widehat{\mu}$. Derive the limiting distribution of the OLS estimator $\hat{\mu}$. Interpret your results. Does it make sense to use $\widehat{\mu}$ as an estimator for δ? Explain.

(b) (15 points) Suppose that we have the following data generating process 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 $\{u_t\}$;
- $y_0 = 0$;
- $\rho = 1$.

The t-statistic for

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

can be written as

$$t_{\rho=1} = \frac{\widehat{\rho} - 1}{\sqrt{\widehat{\sigma}_u^2 / \sum_{t=1}^{T} y_{t-1}^2}},$$

where $\widehat{\rho}$ is the OLS type estimator that has the form

$$\widehat{\rho} = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2},$$

where $\widehat{\sigma}_u^2$ is the residual variance estimator that satisfies

$$\widehat{\sigma}_u^2 \xrightarrow{p} \sigma_u^2.$$

Answer the following questions.

(i) Find the order of probability and the limiting distribution of $t_{\rho=1}$ under the null hypothesis.

(ii) Find the order of probability of $t_{\rho=1}$ under the alternative hypothesis.

(iii) Is this t-statistic consistent?

Answer:

(a) (10 points) First of all we need to rewrite y_t such that we can use the standard results from the literature. We have

$$y_t = \delta t + S_t,$$

where $S_t = \sum_{s=1}^{t} u_t$. We know that we can use the standard results for S_t. Now, let's start with the numerator

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

Standard result (9) with $v = 2$ and (7) can be used here. But in order to apply (7) we need a further step. We know that

$$S_t = S_{t-1} + u_t,$$

then we can write

$$\sum_{t=1}^{T} y_t t = \delta \sum_{t=1}^{T} t^2 + \sum_{t=1}^{T} S_{t-1} t + \sum_{t=1}^{T} u_t t,$$

Now results (9), (7) and (4) (with $j = 0$) gives

$$\sum_{t=1}^{T} y_t t = O_p(T^3) + O_p(T^{5/2}) + O_p(T^{3/2}),$$

So the dominant term is the first term and the order of the sum is determined by this first term. We can write

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

Now we analyze the denominator. We have

$$\sum_{t=1}^{T} t^2 \rightarrow \frac{1}{3},$$

by standard result (9). Combining these give

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

The limiting distribution of the OLS estimator of μ is equal to δ. Which means that if we would like to estmate δ we can use $\widehat{\mu}$ as an estimator for δ.

(b) (15 points)

(i) (5 points) First we need to consider the following error-of-estimator:

$$\hat{\rho} - 1 = \frac{\sum_{t=1}^{T} y_{t-1} u_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

We can show that $\hat{\rho}$ is T consistent and the limit can be written as

$$T(\hat{\rho} - 1) = \frac{T^{-1} \sum_{t=1}^{T} y_{t-1} u_t}{T^{-2} \sum_{t=1}^{T} y_{t-1}^2} \xrightarrow{d} \frac{[W(1)]^2 - \frac{\sigma_u^2}{\sigma^2}}{2 \int_0^1 [W(r)]^2 dr}.$$

This implies that we need to write

$$t_{\rho=1} = \frac{T(\widehat{\rho} - 1)}{\sqrt{\widehat{\sigma}_u^2 / \frac{1}{T^2} \sum_{t=1}^{T} y_{t-1}^2}},$$

Now we need to analyze

$$\frac{1}{T^2} \sum_{t=1}^{T} y_{t-1}^2.$$

This is $O_p(1)$ according to standard result (6). Then, we have

$$t_{\rho=1} = \frac{T(\widehat{\rho} - 1)}{\sqrt{\widehat{\sigma}_u^2 / \frac{1}{T} \sum_{t=1}^{T} y_{t-1}^2}} \xrightarrow{d} \frac{O_p(1)}{\sqrt{O_p(1)/O_p(1)}} = O_p(1).$$

Order of probability under the null is $O_p(1)$ and the limiting ditribution is given by

$$t_{\rho=1} \xrightarrow{d} \frac{\frac{\sigma}{2\sigma_u} \left(B(1)^2 - \frac{\sigma_u^2}{\sigma^2} \right)}{\sqrt{\int_0^1 W(r)^2 dr}}.$$

(ii) (5 points) Under the alternative hypothesis $\mathcal{H}_1 : \ |\rho| < 1$, the standard asymptotics for stationary processes apply and $\hat{\rho}$ becomes $\sqrt{T}-$consistent. We have

$$\hat{\rho} - 1 = (\hat{\rho} - \rho) + (\rho - 1) = O_p(T^{-1/2}) + O_p(1),$$

If y_t is stationary we know by LLN that

$$\frac{1}{T} \sum_{t=1}^{T} y_{t-1}^2 = O_p(1),$$

So we have

$$t_{\rho=1} = \frac{(\widehat{\rho} - 1)}{\sqrt{\widehat{\sigma}_u^2 / \sum_{t=1}^{T} y_{t-1}^2}} \xrightarrow{d} \frac{O_p(1)}{\sqrt{O_p(1)/O_p(T)}} = O_p(\sqrt{T}).$$

$$|t_{\rho=1}| = O_p(\sqrt{T}),$$

(iii) (5 points) The test statistic is consistent because it is converging to a distribution under the null hypothesis and it is diverging to infinity under the alternative.

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

(a) (10 points) An econometrics student from the VU wants to analyze the evolution of CO2 level and its relation with economic activiy using long annual time series for the period 1918 - 2018. The retained variables are the log of CO2 emissions ($\ln c_t$), the log of domestically produced goods in the Netherlands ($\ln x_t$). He first tests for a unit root in the log of domestically produced goods and concludes that it has a unit root.

Then, his analysis continues by fitting some ARMA models to the series. He obtains the following results

$$(1 - \underset{(0.5)}{1.5}\ L - \underset{(0.20)}{0.5}\ L^2)\ \ln c_t = (1 + \underset{(0.02)}{0.80}\ L)\ \widehat{\varepsilon}_t.$$

Standard errors are given in parentheses.

Third step in his analysis is to estimate a potential long run relation between the logarithm of CO2 emissions and log of domestically produced goods by OLS. This yields the following result

$$\widehat{ln\ x_t} = \underset{(0.30)}{-4.32} \quad + \quad \underset{(0.21)}{0.63}\ ln\ c_t.$$

An ADF test on the residuals yields the test statistic -2.46. He concludes that there exists a long run relation between the CO2 emissions and domestically produced goods.

You are asked to interpret the results reported above and comment on the appropriateness of his analysis. In particular,

(i) Calculate the roots of the MA and of he AR polynomials. Given these, comment on the stability and stationarity and invertibility of the series.

(ii) Is there evidence in favor of the existence of a long run relation between CO2 emissions and domestically produced goods? Explain.

(iii) Can you use this static least squares regression to test the null hypothesis of unit CO2 elasticity of domestically produced goods? If not, why not? If not, what would you propose?

(b) (15 points) Another econometrics student from the VU is analyzing a panel data set of real house prices and real income for 134 countries over 145 quarters. Let $p_{i,t}$ denote the real house price in country i at time period t and $y_{i,t}$ denote the real income in country i at time period t. The econometrician considers the model

$$p_{i,t} = \beta_i y_{i,t} + u_{i,t},$$

where $\beta_i = \beta + \eta_i$ with $\eta_i \sim i.i.d(0, \sigma_\eta^2)$.

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

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

He believes that this correlation is due the presence of an unobserved common shock that affects the house prices of all countries. He assumes

$$u_{i,t} = \lambda_i f_t + e_{i,t}, \tag{1}$$

where $e_{i,t}$ is independently and identically distributed across i and t. Furthermore he suspects that the same factor affects the real incomes of all the countries so he assumes

$$y_{i,t} = \gamma_i f_t + \epsilon_{i,t}, \tag{2}$$

where $\epsilon_{i,t}$ is independently and identically distributed across i and t.

He starts looking for an advice on what would be the effects of ignoring f_t and on how to proceed in this situation.

- **(i)** Inform the econometrician about the consequences of the presence of the unobserved f_t in (1)–(2).
- **(ii)** Pesaran (2006) proposes a method to estimate β_i and β in this set up. Discuss the method he proposes. Discuss the important assumptions of this method.

Answer:

(a) (10 points)

(i) (5 points) The roots of the polynomial can be calculated by writing

$$1 - 1.5z - 0.5z^2 = 0,$$

If we solve this to find z_1 and z_2 we obtain, $z_1 = -3.562$ and $z_2 = 0.56$. One root is inside the unit circle the other root is outside the unit circle. Hence the process is unstable, hence nonstatinary. The roos of the MA polynomial can be found by solving

$$1 + 0.8z = 0,$$

for z, which gives $z = -1.25$, which is outside the unit circle, that means the MA part is invertible.

(ii) (2 points) The evidence in favor of a long run relation between CO2 emissions and domestically produced goods can be seen from the OLS regression output. We need to look at the test statistic for the ADF test on te residuals. Because this is a way to test for cointegration. The statistic is given as -2.26. In order to make a conclusion we need to compare this number with the Mc Kinnon critical values. If the reject the null of non-stationarity then it means that the two variables are cointegrated.

(iii) (3 points) It can be used but it will be valid only when there is no serial correlation. Static least squares usually suffers from the serial correlation and endogeneity. The limiting distribution of the test statistics depend on nuisance parameters. So the tests are invalid while the estimators are consistent. One can use methods that take into account the serial correlation such at FMLS.

(b) (15 points)

(i) (5 points) This means if the global factors are affecting only the variable $p_{i,t}$ but not the variables contained in $\mathbf{y}_{i,t}$, then the consistency of the ordinary estimators is not affected by presence of cross-sectional dependence. So the usual OLS type estimators will suffer from endogeneity.

(ii) (10 points) Pesaran (2006) defines

$$\mathbf{z}_{i,t} = \begin{pmatrix} p_{i,t} \\ \mathbf{y}_{i,t} \end{pmatrix}$$

and proposes to use this to obtain approximations for the space spanned by the unobserved factors. By using the models given in the question we can write

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

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 $N \to \infty$.

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

as $N \to \infty$.

* This result suggests that we can use the cross-sectional averages of $\mathbf{z}_{i,t}$ as a proxy for the unobserved common factors.
* Note that we do not know the values of the elements in $\overline{\mathbf{C}}$, so we can not really consistently estimate $\mathbf{f}_t$ themselves, but we can estimate the space spanned by $\mathbf{f}_t$, which is enough for the consistent estimation of $\boldsymbol{\beta}_i$ and $\boldsymbol{\beta}$.

Assumptions are given in the slides of Part III.