Part 2 and 3: Exercises

Solution proposal

10: Basic regression analysis with time series data

1. Explain the nature of time series data

   Time series data characterized by its (1) temporal ordering of data and where (2) realized data variables represent sequences of draws from a underlying stochastical process.The process can further be specified to be generated from a specific regression model.

2. We have the following finite distributed lagged model:

$$y_t = \alpha + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$$

3. Show that a) Temporary change by 1 in period t can be written as:

- t periods: δ_0
- t+1. periods: δ_1
- t+2. periods: δ_2
- t+3. periods: 0

$$y_t = \alpha + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$$
$$\text{Temporary change: } + 1 \text{ in period t, otherwise c}$$
$$y_{t-1} = \alpha + \delta_0 c + \delta_1 c + \delta_2 c + u_t$$
$$y_t = \alpha + \delta_0 (c+1) + \delta_1 c + \delta_2 c + u_t$$
$$y_{t+1} = \alpha + \delta_0 c + \delta_1 (c+1) + \delta_2 c + u_t$$
$$y_{t+2} = \alpha + \delta_0 c + \delta_1 c + \delta_2 (c+1) + u_t$$
$$y_{t+3} = \alpha + \delta_0 c + \delta_1 c + \delta_2 c + u_t$$

Which implies:

$$y_t - y_{t-1} = \delta_0$$
$$y_{t+1} - y_{t-1} = \delta_1$$
$$y_{t+2} - y_{t-1} = \delta_2$$
$$y_{t+3} - y_{t-1} = 0$$

b. Permanent change by 1 in period t

- t periods: δ_0
- t+1. periods: $\delta_0 + \delta_1$
- t+2. periods: $\delta_0 + \delta_1 + \delta_2$
- t+3. periods: $\delta_0 + \delta_1 + \delta_2$

$$\text{Permanent change: change} + 1 \text{ in period t and onwards}$$
$$y_{t-1} = \alpha + \delta_0 c + \delta_1 c + \delta_2 c + u_t$$
$$y_t = \alpha + \delta_0 (c+1) + \delta_1 c + \delta_2 c + u_t$$
$$y_{t+1} = \alpha + \delta_0 (c+1) + \delta_1 (c+1) + \delta_2 c + u_t$$
$$y_{t+2} = \alpha + \delta_0 (c+1) + \delta_1 (c+1) + \delta_2 (c+1) + u_t$$
$$y_{t+3} = \alpha + \delta_0 (c+1) + \delta_1 (c+1) + \delta_2 (c+1) + u_t$$

Which implies

$$y_t - y_{t-1} = \delta_0$$
$$y_{t+1} - y_{t-1} = \delta_0 + \delta_1$$
$$y_{t+2} - y_{t-1} = \delta_0 + \delta_1 + \delta_2$$
$$y_{t+3} - y_{t-1} = \delta_0 + \delta_1 + \delta_2 \text{ called the long-run multiplier}$$

4. Why do think the estimation of such coefficients could be of interest to policy makers?
   If correctly specified, the model lends itself to a causal interpretation (cerius paribus) which implies that the estimated coefficients can be used to evaluate the quantitative impact for both a temporary and permanent change to the policy variables.

5. What is meant by that OLS is BLUE?
   In the class of linear unbiased estimator (which in OLS is ensured by TS.1-TS.3), the method of OLS (i.e., find the coefficients that minimizes the residual sum of squartes) is the one that provide estimators with the least variance.

6. Is OLS of any use if (a) TS.3 is violated or (b) TS.5?
   If only TS.5 is violated, OLS will still be unbiased in small samples. As such valid point estimates can be obtained so that the model can be used for forecasting and R^2 provides valuable information. If TS.3 violates, coefficients unbiased and as such hold no usefull interpretation.

7. What is the problem with spurious regression?
   Spurious regression implies finding a strong relationship between two variables simply because each is growing over time, and such implies artificial high value for both R^2 and $R^{2,adj}$. This will violate TS.3/TS'3 assumption if not controlled for by adding a time trend.

8. How can the problem of spurious regression be detected and solved?
   Step 1.: Plot the time series involved. Step 2.: If the plots look suspicious, check whether an estimation of a trend variable provides a significant value.

9. How would you interpret a dummy variables?
   Binary state variable for whether a condition is active (=1) or inactive (=0)

10. Why do seasonality occur in time series data?
    If time series are observed at a monthly or quarterly intervals (or even weekly or daily), it may exhibit seasonality. For instance Christmas sale in December, viruses spread more easliy during some months of the year, or weather and temperature conditions affect harvesting.

11: Further issues in using OLS with time series data

We have that:

- a. MA(1) process: $y_t = \alpha e_{t-1} + e_t$
- b. AR(1) process: $y_t = \rho y_{t-1} + e_t$, where we assume that $|\rho| < 1$
- c. Random walk: $y_t = \rho y_{t-1} + e_t$
- d. Random walk with drift: $y_t = \alpha_0 + \rho y_{t-1} + e_t$ Where $e_t \sim i.\,i.\,d.\,N(0, \sigma_e^2)$

1. For (a), (b) find its (i) expected value and (ii) one period ahead forecast.
2. For (c) and (d) find its (i) expected value, (ii) variance, (iii) h-periods ahead forecast and (iv) correlation coefficient

1. MA(1): $y_t = \alpha e_{t-1} + e_t$

i.
$$E(y_t) = E(\alpha e_{t-1} + e_t) = \alpha E(e_{t-1}) + E(e_t)) = 0$$

ii.
$$y_{t+1} = \alpha e_t + e_{t+1} \Rightarrow E(y_{t+1}|y_t) = E(\alpha e_t + e_{t+1}) = \alpha e_t$$

2. AR(1): $y_t = \rho y_{t-1} + e_t$

i.

$$E(y_t) = E(\rho(\rho y_{t-2} + e_{t-1}) + e_t) = \ldots \text{(we skip this calcluations in this course)} \ldots = E\left(\sum_{j=0}^{\infty} \rho^j e_t\right) = 0$$

ii.
$$y_{t+1} = \rho y_t + e_{t+1} \Rightarrow E(y_{t+1}|y_t) = E(\rho y_t + e_{t+1}|y_t) = \rho y_t$$

3. Random walk:

$$y_t = y_{t-1} + e_t$$
$$(i)$$
$$E(y_t) = E(y_{t-2} + e_{t-1} + e_t) = E(y_0 + e_1 + e_2 + \ldots + e_t) = E(y_0) = y_0$$
$$(ii)$$
$$Var(y_t) = Var(y_0 + e_1 + e_2 \ldots + e_t) = t\sigma_e^2$$
$$(iii)$$
$$E(y_{t+h}|y_t) = E(y_{t+h} + e_{t+h}|y_t) = E(y_t + e_{t+h} + e_{t+h-1} + \ldots + e_t|y_t) = y_t$$
$$(iv)$$
$$Corr(y_{t+h}, y_t) = \frac{\sigma^2}{\sigma_{t+h}\sigma_t} = \frac{\sigma}{\sigma_{t+h}} = \sqrt{t/(t+h)}$$
this since $Cov(y_{t+h}, y_t) = Cov(y_t + e_{t+h-1} + e_{t+h-2} + \ldots + e_t, y_t) = \sigma_t^2 y_t = y_{t-1} + e_t$

4. Random walk with drift:

$$y_t = \alpha_0 + y_{t-1} + e_t$$
$$(i)$$
$$E(y_t) = E(\alpha_0 + \alpha_0 + y_{t-2} + e_{t-1} + e_t) = E(t\alpha_0 + y_0 + e_1 + e_2 \ldots + e_t) = E(t\alpha_0 + y_0) = t\alpha_0 + y_0$$
$$(ii)$$
$$Var(y_t) = Var(ty_0 + e_1 + e_2 \ldots + e_t) = t\sigma_e^2$$
$$(iii)$$
$$E(y_{t+h}|y_t) = E(\alpha_0 + y_{t+h} + e_{t+h}|y_t) = E(h\alpha_0 + y_t + e_{t+h} + e_{t+h-1} + \ldots + e_t|y_t) = h\alpha_0 + y_t$$
$$(iv)$$
$$Corr(y_{t+h}, y_t) = \frac{\sigma^2}{\sigma_{t+h}\sigma_t} = \frac{\sigma}{\sigma_{t+h}} = \sqrt{t/(t+h)}$$
this since $Cov(y_{t+h}, y_t) = Cov(y_t + e_{t+h-1} + e_{t+h-2} + \ldots + e_t, y_t) = \sigma_t^2 >$

2. What is the difference between weakly and strongly exogenous regressors?

Weakly exogenous regressors consist only of exogenous regressors that are included in the model specification related to the data generating process. Strongly exogenous regressors consist also of all the future and lagged variables of these variables.

3. What is meant that time series is (a) $I(0)$ and (b) $I(1)$?

Weakly dependent time series are integrated of order zero ($I(0)$). If a time serie variable has to be differentiated in order to obtain a weakly dependent time series, it is called integrated of order one ($I(1)$).

4. List some examples of time series that are (a) weakly dependent (b) highly persistent time series?

Examples of possible weakly dependent time series: unemployment, real interest rate \ Examples of highly persistent time series: oil prices, stock market (short term) and bit-coins

5. Why is the assumption of $u_t \sim N(0, \sigma_u^2)$ not needed in large sample for the estimated coefficient to be (asymptotically) normally distributed?

This is implicitly taken care of in TS'.1 (low of large numbers and central limit theorem) which implies that normality hold as an approximation for large samples

12: Serial correlation and heteroskedasticity in time series regression

1. What is the impact of serial correlation in a time series regression?
   OLS no longer efficient in small samples and statistical inference can not carried out.

2. Is R^2 measured correctly under the presence of serial correlation?
   Yes, given that TS'1.-TS'3 are satisfied.

3. What is meant by that OLS consistent but ~~not~~ biased?
   That OLS is consistent implies that as the sample size grows, the estimated coefficients will converge towards its true values. Unbiased implies that before the realization of the data sample, it is more likely that the estimated coefficients is either higher or lower than its true value.

4. Explain the difference between TS.1 and TS'1.
   The difference is that TS'.1 also assumes (1) Stationary (or covariance) stocastich process and (2) weekly dependent time series

5. Why use robust serial correlation (HAC)?
   Even with the presence of serial correlation, statistical inference (e.g., t-tests, F-testing) can be carried out in the form of rubust estimators.

16 Simultanous equations models (SEM)

Only for MSB104

1. Explain the three different forms of endogeneity problem?
   1: Ommited variables (cf. part I) 2: Measurement errors (cf. part I) 3: Simultaneity: The dependent variable in the regression model intereacts simultanously with at least one of its regressors.

2. What is meant by a structural equation?
   A structural equation is an equation founded in economic theory, as such provides the causal relationship between the variables included in the equation.

3. What is the problem of applying OLS to structural equation?
   If simultaneity occurs, TS'.3 will be violated (which implies that TS.3), so estimators will no longer be consisten or unibased.

4. Which other method should be used and what is the identification condition for this to work properly (2-variable system)?
   Instrumental variable approach in form of a two-stage least square procedure can be used. The rank condition is sufficient and necessary condition for this method to work properly: At least one variable from the equation we are estimating is included as a explanatory variable in the other equation.