Exam in: ECON 4160: Econometrics: Modelling and Systems Estimation

Day of exam: 23 November 2020

Time of day: 09:00—14:00

This is a 5 hour home exam.

Guidelines:
In the grading, question A gets 40 %, B 40 % and C 20 %.

Answers notes are within «... ». Some of the questions are of the "explain in you own words ", the answers notes should then be interpreted as expressions of the main points of understanding under evaluation, not the exact wording.

Question A (40 %)

1. Consider the stochastic difference equation:

$$Y_t = 1 + 0.5Y_{t-1} + \epsilon_{yt}, \quad t = 1, 2, ..., T \tag{1}$$

where ϵ_{yt} is white noise. What is the characteristic root associated with this equation?

« Characteristic equation:

$$\lambda - 0.5 = 0.$$

and characteristic root: $\lambda = 0.5$. »

2. Explain why the time series generated by (1) is stationary.

« The requirement for stationarity is that the characteristic root is less than one in magnitude. Here it is 0.5, therefore the time series generated by (1) is stationary. »

3. What is the expectation of Y_t?.

«

$$E(Y_t) = \frac{1}{(1 - 0.5)} = 2$$

»

4. Calculate the variance of Y_t under the assumption that $Var(\epsilon_{yt}) = 1$.

«

$$Var(Y_t) = \frac{1}{(1 - 0.5^2)} = 1.33$$

»

5. Let ζ_j denote the autocorrelation function (ACF) of Y_t. What are the values of ζ_1, ζ_2 and ζ_{10}?

«

$$\begin{aligned} \zeta_1 &= 0.5 \\ \zeta_2 &= 0.25 \\ \zeta_{10} &= 0.0007656 \end{aligned}$$

»

6. Consider another equation:

$$X_t = 1 - 0.5X_{t-1} + \epsilon_{xt}, \quad t = 1, 2, ..., T \tag{2}$$

where ϵ_{xt} is white noise. Sketch a graph of the ACF of X_t given by (9) together with a graph of the ACF of Y_t given by (1). How will you characterize Y_t and X_t as negatively or positively autocorrelated?

« Y_t: Monotonous (non-linear) decline from 0.5. Positive autocorrelation (Some may use bars, others a line gaph)) X_t: Oscillates between negative and positive values (starting in -0.5), that decline in magnitude. Negative autocorrelation. »

7. Imagine that the model in question A1 is used to forecast Y_{T+1}, Y_{T+2},...,Y_{T+H}. For simplicity we abstract from parameter estimation uncertainty (i.e., we assume that the model is correctly specified within-sample). Under the assumption of a quadratic cost function and assuming $Y_T = 4$, what are the optimal point forecasts of Y_{T+1}, Y_{T+2} and Y_{T+10}?

« The optimal forecast is the conditional forecast of Y_{T+h} given Y_T because it minimizes the Mean Squared Forecast Error (MSFE). Hence, letting Y^f_{T+h} denote a forecast:

$$\begin{aligned} Y^f_{T+1} &= 1 + 0.5 * 4 = 3 \\ Y^f_{T+2} &= 1 + 0.5 * Y^f_{T+1} = 1 + 0.5 * (1 + 0.5 * 4) = 2.5 \\ Y^f_{T+10} &= E(Y_t) + 0.5^{10}(Y_T - E(Y_t)) = 2 + 0.5^{10}(4 - 2) \approx 2 \end{aligned}$$

Some may relevantly mention the glide-path interpretation (clearest in the last expression). »

8. Calculate 95% forecast confidence intervals for the forecasts of Y_{T+1} and Y_{T+2}, when it is assumed that $\varepsilon_{T+h} \sim IIN(0, 1)$, $h = 1, 2$.

« Forecast errors:

$$\begin{aligned} e_{T+1} &= Y_{T+1} - Y^f_{T+1} = \varepsilon_{T+1} \\ e_{T+2} &= Y_{T+2} - Y^f_{T+2} = \varepsilon_{T+2} + 0.5\varepsilon_{T+1} \end{aligned}$$

The conditional expectations (on T) are zero, the variances are:

$$\begin{aligned} Var(e_{T+1} \mid T) &= Var(\varepsilon_{T+1}) = 1 \\ Var(e_{T+2} \mid T) &= 1 + 0.5^2 = 1.25 \end{aligned}$$

since $Cov(e_{T+1}, e_{T+2} \mid T) = 0$ from the assumptions. The second variance can also be found from formula (12.14), as:

$$Var(e_{T+2} \mid T) = 1 \times \frac{1 - 0.5^4}{1 - 0.5^2}$$

Hence, the approximate forecast confidence interval are:

$$\begin{aligned} T + 1 : &= 3 \pm 2 \times 1 : [1, 5] \\ T + 1 : &= 2.5 \pm 2 \times \sqrt{1.25} : [0.26, 4.74] \end{aligned}$$

»

9. Imagine that when we get to evaluate the two first forecasts, we observe that Y_{T+1} and Y_{T+2} (the actuals) are well outside the forecast confidence intervals. How would you suggest that the forecasts of Y_{T+3},...Y_{T+10} can be revised in the light of the outcomes (actuals) for period $T+1$ and $T+2$?

   «

   The most common reason for such failures are that the DGP has changed after the forecast was made, a post-forecast structural break or regime shift. When we learn that we have consistently under-predicted, the likely interpretation is that the constant term in the new regime is larger than it was before the break (smaller if the forecast over-predicted)

   Hence, one suggestion is to update the forecast of Y_{T+3},...Y_{T+10} based on what we have learned about the structural break.

$$Y_{T+3} = 1 + \text{add} + 0.5Y_{t-1} + \epsilon_{yt}, \ \ t = 1, 2, ..., T$$

   where "add" represents an add-factor. It can be based on the observed forecast error for period for $T+2$, or a weighted average of $T+2$ and $T+2$ forecast errors. This is topic is covered by Ch 12.7 and it was mentioned the lecture about forecasting. But not specifically for the exact AR(1) model we have here. »

Question B (40 %)

1. Assume that a time series Y_t is generated by:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_{yt}, \ \ t = 1, 2, ..., T \tag{3}$$

   where $\epsilon_{yt} \sim IIN(0, \sigma_y^2)$ for all t. (3) can be re-written as:

$$\Delta Y_t = \phi_0 - \pi Y_{t-1} - \phi_2 \Delta Y_{t-1} + \epsilon_{yt}, \ \ t = 1, 2, ..., T \tag{4}$$

   where

$$\pi \ = \ 1 - \phi_1 - \phi_2. \tag{5}$$

   Imagine that you estimate (4) by OLS and that the t-value of $-\hat{\pi}$ is -2.1. Can you reject the null hypothesis of $Y_t \sim I(1)$ at the 5 % level?

   You can take as given that the number of observations is large enough to validate the use of asymptotic critical values.

   « It is assumed that the estimated equation correspond to the DGP. The test situation is $H_0 : -\pi = 0$ against $H_1 : -\pi < 0$ (stationary with a stable causal solution). Since a process with a root larger than one in magnitude is stationary as well (but does not have a stable causal solution), some students may specify the alternative as $H_1 : -\pi \neq 0$. Since we have not always been precise about the 1- or 2-sidedness of the alternative, this is OK.

   The easiest procedure is to use Table 3 in the Ericsson and MacKinnon paper, since there is an intercept in the estimated equation. The correct asymptotic critical values under H_0 are given in the $k = 1$ part of the table. $H_0 : -\pi = 0$ is not rejected at any of the three significance levels. »

2. Table 1 at the back of the exam set shows unit-root tests for $TEMP$, a quarterly time series of western-hemisphere temperatures measured as deviations from 1950-1980 means. The table also contains results for $DTEMP$ (the difference of $TEMP$). Explain how you reach a conclusion about $TEMP$ being I(0) or I(1) based on the tests.

« For testing $TEMP \sim I(1)$ against $TEMP \sim I(0)$, the t-adf in the row labelled "D-lag 3" is the correct to use since the fourth t-DY_lag is seen to be insignificant. Since $t - adf = -0.8975$ the hypothesis $TEMP \sim I(1)$ is not rejected at any customary levels of significance. It is not a big error to use the t-adf in the row "D-lag 4", since it is just a matter of losing some efficiency (the level of the test is not compromised by this choice).

For testing $DTEMP \sim I(1)$ against $DTEMP \sim I(0)$ the t-adf in the row "D-lag 2" is correct to use. It rejects the null, hence we conclude that $TEMP \sim I(1)$ since the first difference is $I(0)$.

»

3. We want to investigate empirically whether there is a relationship between $TEMP$ and another time series variable, $CO2$, which measures CO_2 in the atmosphere. You can take as given that $CO2$ is an I(1) variable.

   Imagine that a friend, who studies at a business school, suggests that you can regress $TEMP$ on $CO2$ and a constant term, and compare the t-value of the regression coefficient to a 5 % critical value from the t-distribution.

   How would you explain to him that his method would put you in acute danger of falling into the spurious regression trap?

   « His method rests crucially on the assumption that t-value has a t-distribution under the null hypothesis of no relationship. However, when both regressor and regressand are I(1) variables this does not hold. Instead the t-value has a different distribution called Dickey-Fuller distribution under null of no relationship. Using the t-distribution (or the normal), leads to massive false over-rejection of the null hypothesis of no relationship. Type-I error probability is much closer to 100 % than a nominal significance level of 5 %, for example. And it gets worse with larger samples. This is the spurious regression trap. Using conventional robust standard errors to correct the t-value before it is compared with a the 5 % critical value of the t-distribution does not solve the problem. »

4. Explain why a correct test of the null hypothesis of no relationship between $TEMP$ and $CO2$ can be based on the model estimated in Table 2. What is the result of the test?

   « A correct testing procedure is to regress $DTEMP$ on its lags and on current and lagged $DCO2$. The lag lengths should be long enough to "map up" the autocorrelation in the data (with the aim to make the model not mis-specified). Finally include the lagged level of $TEMP$ and $CO2$. Table 2 is an example.

   The test of the null hypothesis of no relationship between the two (levels) variables is made by comparing the t-value of the lagged $TEMP$ to the relevant critical value form the so called ECM-test in Ericsson and MacKinnon's paper (it has a Dickey-Fuller type distribution). The t-value of -5.89 is smaller than all three critical values in Table 1, section with $k = 2$, in the paper. Hence the null of no relationship is rejected statistically, and the alternative of cointegration between $TEMP$ and $CO2$ is accepted.

   If properly explained, a good answer can also mention (or use) the EG-test of no cointegration.

   A relevant remark could be that the ECM test is inferior to the Trace-test (Johansen method) if $CO2$ is not weakly exogenous with respect to the cointegration parameter. But not required to get full score. »

5. Table 3 shows estimation results after the lags of $TEMP$ and $CO2$ have been replaced by the variable TEMP-1.4*CO2_1. Show how the coefficient 1.4 has been calculated, and explain how it should be interpreted.

« When cointegration is accepted we find the estimated relationship in the usual way by assuming that $DTEMP = 0$ and $DCO2 = 0$ (more generally constant changes other than zero, but that gives more cluttered algebra). Hence the estimated static long-run relationship becomes (abstracting from the seasonals):

$$TEMP_t = \frac{-254.928}{-0.561265} Constant + \frac{0.784512}{-0.561265} CO2_t + u_t$$

where $u_t \sim I(0)$.

$$\frac{0.784512}{-0.561265} \approx 1.4$$

The number 1.4 shall therefore be interpreted as the (ECM) estimate of the true cointegration parameter. Not a mistake here to interpret it as the estimated long-run effect on $TEMP$ of a permanent change in $CO2$ (although, strictly speaking, to test that requires more evidence about exogeneity of CO2). »

6. Figure 1 shows a selection of graphs that can be used to evaluate the degree of parameter constancy of the model of $DTEMP$ in Table 3.

   noindent It can be shown that there are significant structural breaks in the marginal equation of $DCO2$, namely in 1991(3), 2010(1), 2015(4), 2016(2), 2016(3), 2016(4) and 2018(4). When the set of indicator variables (i.e., dummies) for these seven quarters are added to the conditional equation of $DTEMP$, the residual sum of squares becomes: $RSS = 30697.8882$ (the log-likelihood becomes $log - likelihood = -717.918$).

   What can be concluded about the relative invariance of the model estimated in Table 3?
   HINT: In the $F(7, 161)$ distribution, the 5 % critical value is 2.06. In the $\chi^2(7)$ it is 14.

   « Figure 1 shows that two or three of the 1-step Chow-test are significant at the 1 % level, but that can happen also under the null of stability (Type-I error). The plotted recursive 1-step residuals gives a similar impression, while the graph of the sequence of multi-period (" break-point") Chow-tests "Ndn" shows no significant breaks at all. Most importantly perhaps, the recursive estimate of the ECM-coefficient is stable (the sample size is very small at the start). In sum, the graphs in Figure 1 support a relatively high degree of parameter constancy of the conditional model of $DTEMP$ in Table 3.

   For the parameters of the conditional model to be characterized as invariant with respect structural breaks in the marginal model, the set of break indicators should be insignificant when added to the conditional model. Using the F-test of joint significance of a sub-set (i.e. based on the difference between RSS_U and RSS_R), we get

$$F(7, 161) = \frac{32001.0957 - 30697.8882}{30697.8882} \times \frac{(180 - 19)}{(7)} = 0.97641$$

   Using the hint, we see that this is nowhere near statistical significance. Hence we do not reject the hypothesis of invariance of the conditional model. The likelihood rate test becomes:

$$\chi^2(7) = -2 \times (-721.66 - (-717.918)) = 7.48$$

   »

Question C (20 %)

1. Explain what is meant by simultaneity bias of OLS when used to estimate the coefficients of an identified structural equation in a SEM.

« Simultaneity bias refers to the case when the model equation in question includes at least two variables that are endogenous in the SEM. The equation will be normalized on one them, by the OLS estimator of the coefficient of other endogenous variable be inconsistent because it is correlated (also asymptotically) with the error-term of the structural equation. This follows logically from the reduced form of the model. »

2. Explain why estimation by IV or 2SLS gives consistent estimation of the coefficients of the equation.

   « Since the equation is identified, the order condition will imply that there is exactly as many exogenous or pre-determined variables excluded form the equation in question, as there are endogenous variables in the equation, minus one (the normalization on one endogenous). Hence, in principle, there is no problem of "finding a valid instrumental variable". It (or they) are defined by the multiple-equation model.

   Since the instrumental variables are asymptotically uncorrelated with the error-term of the structural equation, the IV estimator is consistent. In the case of exact identification, there is a single IV-estimator. In the case of over-identification, there is more than one IV estimator, and one interpretation of 2SLS is that it is an IV estimator (also called GIVE) with optimal instruments. In practice, that optimalization happens in the first stage, by the estimation of the reduced form to obtain fitted values for the endogenous variables that are included in the structural equation.

   Implicit in the above is that the SEM is well specified. If there is residual autocorrelation and the structural equation is dynamic, IV/2SLS gives inconsistent estimation. They solve one specific estimation problem. »

Table 1: Dickey-Fuller tests of unit-root in $TEMP$ and $DTEMP$.

```
The sample is: 1959(4) - 2019(4) (247 observations and 1 variables)

TEMP: ADF tests (T=241, Constant; 5%=-2.87 1%=-3.46)
 D-lag    t-adf      beta Y_1    sigma   t-DY_lag  t-prob
  4     -1.012       0.97326   0.1602     0.8884  0.3753
  3    -0.8975       0.97653   0.1601     -2.788  0.0057
  2     -1.340       0.96492   0.1624     -5.428  0.0000
  1     -2.311       0.93721   0.1718     -2.830  0.0051
  0     -2.981*      0.91992   0.1743

Unit-root tests

The sample is: 1959(4) - 2019(4) (246 observations and 1 variables)

DTEMP: ADF tests (T=241, Constant; 5%=-2.87 1%=-3.46)
D-lag     t-adf      beta Y_1     sigma   t-DY_lag  t-prob
  3     -9.806**    -0.85801    16.02    -0.7548  0.4511
  2     -13.94**    -0.95408    16.00      2.970  0.0033
  1     -17.35**    -0.64283    16.26      5.791  0.0000
  0     -19.20**    -0.21570    17.33
```

Table 2: Estimation results for a model of $DTEMP$ conditional on $DCO2$ (the first difference of $CO2$).

Modelling DTEMP by OLS
The estimation sample is: 1975(1) - 2019(4)

	Coefficient	Std.Error	t-value	t-prob
DTEMP_1	0.0474593	0.09568	0.496	0.6205
DTEMP_2	0.0250048	0.08436	0.296	0.7673
DTEMP_3	0.0734019	0.07503	0.978	0.3293
Constant	-254.928	53.03	-4.81	0.0000
DCO2	9.14585	3.421	2.67	0.0083
DCO2_1	5.42229	3.536	1.53	0.1271
DCO2_2	-2.47831	3.571	-0.694	0.4886
DCO2_3	8.04115	3.448	2.33	0.0209
TEMP_1	-0.561265	0.09522	-5.89	0.0000
CO2_1	0.784512	0.1438	5.45	0.0000
Seasonal	-46.6643	26.17	-1.78	0.0764
Seasonal_1	-21.3409	36.24	-0.589	0.5567
Seasonal_2	18.1724	25.68	0.708	0.4802

sigma	13.8428	RSS	32000.9952
R^2	0.500221	F(12,167) =	13.93 [0.000]**
Adj.R^2	0.464308	log-likelihood	-721.66
no. of observations	180	no. of parameters	13

AR 1-5 test:	F(5,162) =	0.25670 [0.9359]
ARCH 1-4 test:	F(4,172) =	2.4967 [0.0446]*
Normality test:	Chi^2(2) =	4.2469 [0.1196]
Hetero test:	F(21,158) =	1.8113 [0.0215]*
Hetero-X test:	F(57,122) =	1.2197 [0.1813]

Table 3: Estimation results for a model of $DTEMP$ conditional on $DCO2$, with $TEMP\_1$ and $CO2\_1$ replaced by $TEMP - 1.4 * CO2\_1$.

Modelling DTEMP by OLS
The estimation sample is: 1975(1) - 2019(4)

	Coefficient	Std.Error	t-value	t-prob
DTEMP_1	0.0475378	0.09534	0.499	0.6187
DTEMP_2	0.0251967	0.08369	0.301	0.7637
DTEMP_3	0.0736135	0.07424	0.992	0.3228
Constant	-255.339	49.73	-5.13	0.0000
DCO2	9.12105	3.235	2.82	0.0054
DCO2_1	5.39361	3.297	1.64	0.1037
DCO2_2	-2.50755	3.325	-0.754	0.4518
DCO2_3	8.01630	3.263	2.46	0.0150
TEMP-1.4*CO2_1	-0.561248	0.09494	-5.91	0.0000
Seasonal	-46.6725	26.09	-1.79	0.0754
Seasonal_1	-21.3284	36.12	-0.590	0.5557
Seasonal_2	18.1921	25.59	0.711	0.4782

sigma	13.8015	RSS	32001.0957
R^2	0.500219	F(11,168) =	15.29 [0.000]**
Adj.R^2	0.467495	log-likelihood	-721.66
no. of observations	180	no. of parameters	12

AR 1-5 test:	F(5,163) =	0.25726 [0.9356]
ARCH 1-4 test:	F(4,172) =	2.4898 [0.0451]*
Normality test:	Chi^2(2) =	4.2626 [0.1187]
Hetero test:	F(19,160) =	2.0083 [0.0106]*
Hetero-X test:	F(47,132) =	1.3249 [0.1093]

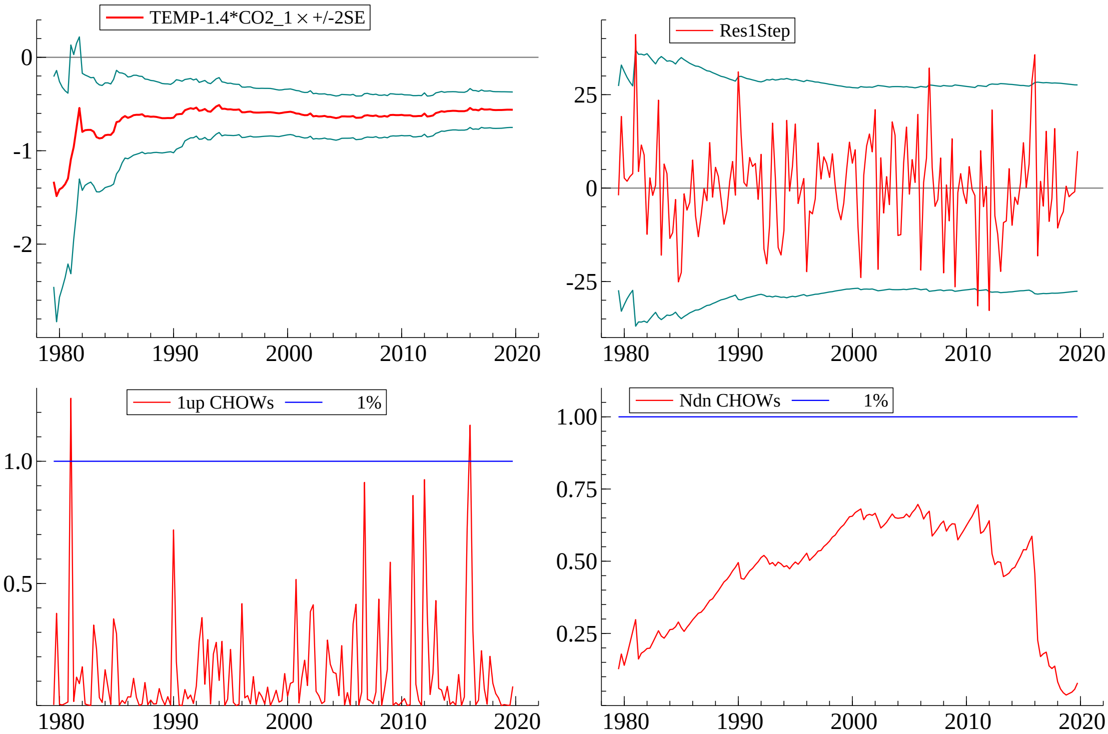

Figure 1: Some recursive graphs of the model estimated in Table 3.