Written Examination in Time Series Analysis (B3)

Spring 2015

2015-04-29 08.00-12.00

Bergsbrunnagatan 15, room2.

Lars Forsberg, Department of Statistics, Uppsala University

Allowed means of assistance:

1. Pen or **pencil** (recommended) and eraser
2. **Calculators**,
   (a) 'programmable' calculator, e.g. calculator with graphing functions is OK.
   (b) Calculators with blue-tooth are not allowed.
   (c) Calculators with access to internet are not allowed.
   (d) Calcuators with which it is possible to send and recieve messages of any kind are not allowed.
3. **Physical (paper) dictionary** (no electronic dictionary allowed).
   (a) Dictionary must contain *no notes* of any kind.
   (b) Each student must have his/her own dictionary. It is not allowed for students to pass a dictionary between them.
4. **Ruler.**
5. Collection of formulae and Statistical Tables named '*Collection of Formulae and Statistical Tables for the B2-Econometrics and B3-Time Series Analysis courses and exams*', that the student brings to the exam location.
6. Please note that a collection of critical values for the Student´s t, Normal, Chi-square and F-distributions is given in the Appendix of the '*Collection of Formulae and Statistical Tables for the B2-Econometrics and B3-Time Series Analysis courses and exams*'.
7. About degrees of freedom in tests: If, by any chance, the degree of freedom number that you need for a critical value is not in the table, say that you need 125, but there is only 120 and 130 in the table, then choose the lower number of degrees of freedom, that is, in this case 120.

8. Also note that the '*Test template*', that should be used when performing tests, is given in the '*Collection of Formulae and Statistical Tables for the B2-Econometrics and B3-Time Series Analysis courses and exams'.*

That is:

1. NO BOOK (except paper-dictionary) is allowed.
2. NO (student-written) notes are allowed.
3. NO other document than the one 'Collection of Formulae and Statistical Tables for Time Series Exam' is allowed.

Instructions: **Please note the following:**

1. Start with reading through the instructions!
2. Make sure you **follow** the instructions!
3. Start with reading through the exam.
4. You may write your solutions in Swedish or English.
5. If you find something unclear or if you suspect a typo/mistake in any of the tasks - please do not hesitate to contact the staff at the exam-location for them to get in touch with the responsible teacher.
6. Total score is **100** points
   (a) If you want the ECTS grades, please indicate that on the cover page!
   (b) For each task the maximum number of points is given within parenthesis, e.g. (16p in total).
   (c) For each subtask the number of points is given within parenthesis, e.g. (2p)
7. All solutions must be on separate sheets. No solutions on the questionnaire! (If so, they will be disregarded.)
8. Make sure your solutions are: easy to read and easy to understand, that is:
   (a) For each task that you solve, please start with a new sheet: after Task 1, start with a blank sheet for Task 2, etc.

(b) Write the *task number* at the top of each page, in the

................................**MIDDLE OF THE PAGE!!!**..

Like:

....**TASK 1**..

- if you write it in the upper left corner, the staple will cover it, and there is no for way for the examinator to know if the text of that sheet belongs to the previous sub-task or what it is. The Examinators will not make any 'qualified guesses' of what is being displayed on any given page. It is the responsibility of the student to make sure that every task and sub-task is easily identifiable.

(c) If you continue a sub-task on the next sheet of paper - indicate that at the top of the page - IN THE MIDDLE OF THE PAGE, like, for example:

....'Task 1B (cont.)'...

(d) Please separate each subtask A, B etc with a horizontal line across the sheet

____

if they are on the same sheet of paper - that way it will be easy for the examinator to actually see where one subtask ends and next begins.

(e) For examinator readability, it is highly recommended that you use a pencil, (and not a pen), which will allow you to erase and rewrite if you make a mistake. Crossed-over text and corrections using 'tipp-ex' will just cause blurriness and confusion to the examinator.

(f) For examinator readability: Write clearly, that is, letters, mathematical/statistical symbols and numbers should be easy recognizable!! Do not underestimate the correlation between readability and points scored, that is, when readability goes to zero, points scored also goes to zero, no matter your intentions or wheather *you* can read it or not.

(g) Also note that everything that you write will be taken at 'face value'. That is, for example, if you write β_1 the examinator will take that as a β_1 even though you may claim that it is given from the context it should be clear that you meant something else, like β_3. Thus, given this example, writing β_1, and that is not correct in that specific formula or statement, this will lead to subtraction of points, even if you will claim that it is just a typo, and that in another task or subtask, it is clear that you understand the issue.

(h) Please put the sheets in **order**, that is first Task 1, and then Task 2 etc...

9. Please keep the questionaire.

10. Do well!

Task 1

(24 points in total)

A) (2p) What is a stochastic process from a theoretical point of view? Explain using words, no formulae needed.

B) (6p) State the conditions for a stochastic process to be *covariance stationary*. For each condition, state that condition using formulae and also explain in words what it means.

To 'apply' the Box-Jenkins methodology, a nessecary condition is that the series in question is (at least) covariance stationary. If a process is *not* stationary, we need to transform it somehow to make it stationary before we can apply the Box-Jenkins methodology.

C) (4p) Give an example of a process that is *trend-stationary*. That is, write down the process using a formula, make sure you define all the 'parts' of the process. Explain in words why it is not stationary (if you by any chance find this sentence confusing - yes, it says and should say *not* stationary). Suggest a *transformation* that will make the process covariance stationary.

D) (4p) Give an example of a process that is *difference-stationary*. That is, write down the process using a formula, make sure you define all the 'parts' of the process. Explain in words why it is not stationary (if you by any chance find this sentence confusing - yes, it says and should say *not* stationary). Suggest a *transformation* that will make the process covariance stationary.

E) (8p) State the four stages of the Box-Jenkins methodology. For each stage, elaborate on the *purpose* of that specific stage, also give at least *one* example of a tool/method/statistical test that can be used in that specific stage.

Task 2

(14 points in total)

Consider the following process

$$\phi(B) Y_t = \theta(B) e_t \tag{1}$$

where $e_t \sim NID(0, \sigma^2)$.

Let

$$\phi(B) = (1 - \phi B) \tag{2}$$

and

$$\theta(B) = 1. \tag{3}$$

A) (2p) For the model defined by (1), (2) and (3), derive the *expected value* of the process. Be explicit in what assumptions, if any, you make to be able to derive this result. Also, state the nessecary condition (if one is needed) for the parameter, to ensure that the process is stationary.

B) (2p) For the model defined by (1), (2) and (3), derive the *variance* of the process. Be explicit in what assumptions, if any, you make to be able to derive this result. Also, state the nessecary condition (if one is needed) for the parameter, to ensure that the process is stationary.

C) (4p) For the model defined by (1), (2) and (3), derive the first *two* auto*covariances* for the process. Be explicit in what assumptions, if any, you make to be able to derive these results. You may write the autocovariances as functions of other autocovariances if you so see fit. Also, state the nessecary condition (if one is needed) for the parameter, to ensure that the process is stationary.

D) (2p) Given the variance and the autocovariance derived in previous subtasks, calculate the first two *autocorrelations* of the process. As your final representation of the autocorrelations, write the autocorrelations solely as functions of the parameter of the process, that is no autocorrelations on the right hand side or the expression(s).

E) (4p) For the model defined by (1), (2) and (3), sketch the correlogram (ACF and PACF) of Y_t. Note that you do *not* have to derive the ACF or the PACF in this subtask. For the scetch let $\phi = 0.9$. Use the number of lags you find appropriate.

Task 3

(18 points in total)

Consider (again) the following process

$$\phi(B) Y_t = \theta(B) e_t \tag{4}$$

where $e_t \sim NID(0, \sigma^2)$.

But now let

$$\phi(B) = (1) \tag{5}$$

and

$$\theta(B) = \left(1 - \theta_1 B^1 - \theta_2 B^2\right). \tag{6}$$

A) (2p) For the model defined by (4), (5) and (6), derive the *expected value* of the process. Be explicit in what assumptions, if any, you make to be able to derive this result. Also, state any nessecary condition(s)for the parameter(s), to ensure that the process is stationary with respect to the mean.

B) (2p) For the model defined by (4), (5) and (6), derive the *variance* of the process. Be explicit in what assumptions, if any, you make to be able to derive this result. Also, state any nessecary condition(s) (if needed) for the parameters(s), to ensure that the process is stationary with respect to the variance.

C) (4p) For the model defined by (4), (5) and (6), derive the first *three* auto*covariances* for the process. Be explicit in what assumptions, if any, you make to be able to derive these results. You may write the autocovariances as functions of other autocovariances if you so see fit. Also, state any nessecary condition(s) (if needed) for the parameter, to ensure that the process is stationary with respect to the autocovariances.

D) (2p) Given the variance and the autocovariance derived in previous sub-tasks , derive the first *three autocorrelations* of the process. As your final representation of the autocorrelations, write the autocorrelations solely as functions of the parameter of the process, that is, no autocorrelations on the right hand side or the expression(s).

E) (4p) Sketch the correlogram (ACF and PACF) of Y_t. Note that you do *not* have to derive the ACF or the PACF in this subtask. For the scetch, choose parameter values for θ_1 and θ_2 that you see fit, as long as $\theta_1 \neq 0$ and $\theta_2 \neq 0$. Use the number of lags you find appropriate.

F) (4p) For the model defined by (4), (5) and (6) - now, set $\theta_2 = 0$ and derive the infinite AR-representation of the process, that is, derive $AR(\infty)$. State explicitely any condition(s) that has to be fulfilled for the infinite AR-representation to be a stationary representation of the process. The final result should have Y_t alone on the left hand side, and the backshift operator should *not* occur on the right hand side, write out at least five lags.

Task 4

(30 points in total)

Consider the dataset of GDP for United Kingdom[1], from hereon denoted 'GDP', albeit that it is denoted GDP_UK in the outputs.

List of figures:

1. Fig. 4.1: GDP - Time series plot of level
2. Fig. 4.2: Correlogram of GDP-Level
3. Fig. 4.3: ADF-test output for GDP-Level
4. Fig. 4.4: Time series plot of first difference of GDP
5. Fig. 4.5: Correlogram of first difference of GDP
6. Fig. 4.6: Estimation output - AR(1) on first difference of GDP
7. Fig. 4.7: Correlogram of residuals from AR(1) on first difference of GDP
8. Fig. 4.8: Estimation output - ARMA(1,1) on first difference of GDP
9. Fig. 4.9: Correlogram of residuals from ARMA(1,1) on first difference of GDP

The information in these figures are sufficient to solve the following subtasks.

A) (2p) For the original data, interpret the estimated second autocorrelation.

B) (2p) For the original data, interpret the estimated second *partial* autocorrelation.

C) (6p) By visual inspection of the time series plot in Figure 4.1, one might suspect that the series is not stationary. Perform an unit root test, testing if GDP has a unit root, use significance level 5%. Document the test procedure as outlined in the test-template.

D) (6p)) Perform a test to test the null hypothesis that the first three autocorrelations of the first difference of the GDP of UK are simultaneous zero, against the alternative the at least one is different from zero. Use significance level 1%. Document the test procedure as outlined in the test-template.

[1] Data is Gross domestic product at market prices - United Kingdom - Domestic (home or reference area), Total economy, Domestic currency (incl. conversion to current currency made using a fix parity), Chain linked volume (rebased), Non transformed data, Seasonally adjusted data, not calendar adjusted. Quarter 1 1955 to Quarter 4 2014.

Downloaded from: http://sdw.ecb.europa.eu/quickview.do?SERIES_KEY=
320.MNA.Q.S.GB.W2.S1.S1.B.B1GQ._Z._Z._Z.XDC.LR.N

E) (6p) The reasearcher fits an AR(1) model to the data. Perform a test to test the null hypothesis that the first autocorrelation of the residuals from this model is zero, against the altervative that it is *greater than* zero. Use significance level 1%. Document the test procedure as outlined in the test-template.

F) (4p) Now, without doing any formal tests, comment on the results of the AR(1) model. Is this is good model? Why or why not?

G) (4p) Consider the results from the ARMA(1,1) fitted to the first difference of the GDP series. Would you say this is a good model for the first difference of the GDP? Motivate without doing any formal tests. Which one of the AR(1) and ARMA(1) would you prefer?

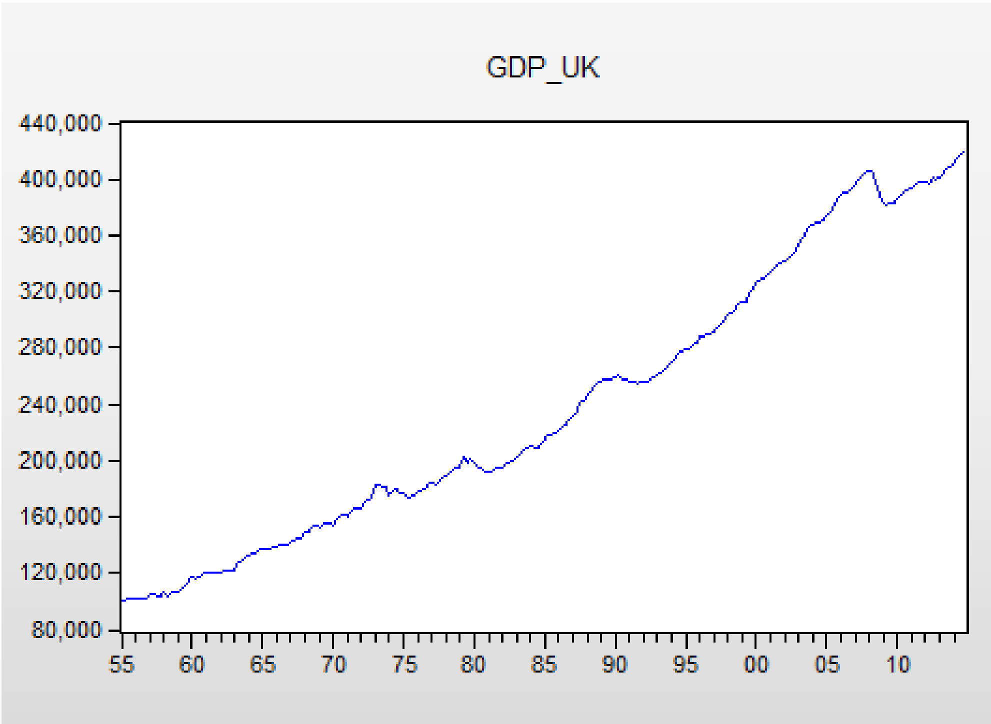
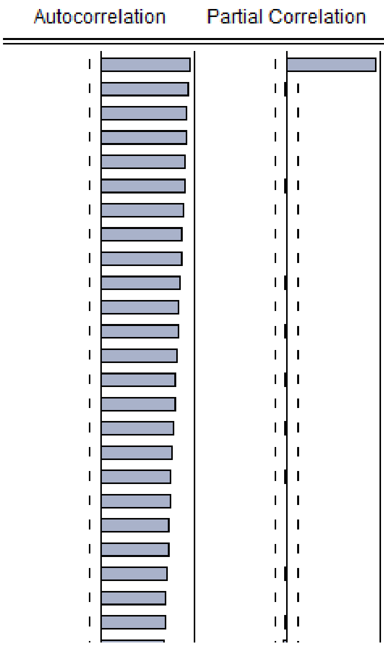


Figure 4.1: GDP for UK, level data

Correlogram of GDP_UK

Date: 04/23/15 Time: 15:39
Sample: 1955Q1 2014Q4
Included observations: 240



	AC	PAC	Q-Stat	Prob
1	0.988	0.988	237.41	0.000
2	0.977	-0.013	470.17	0.000
3	0.965	-0.004	698.33	0.000
4	0.953	-0.008	921.90	0.000
5	0.941	-0.003	1141.0	0.000
6	0.930	-0.009	1355.5	0.000
7	0.918	-0.005	1565.5	0.000
8	0.906	-0.003	1771.1	0.000
9	0.895	0.001	1972.4	0.000
10	0.883	-0.019	2169.2	0.000
11	0.871	-0.004	2361.7	0.000
12	0.859	-0.015	2549.7	0.000
13	0.847	-0.006	2733.4	0.000
14	0.835	-0.020	2912.5	0.000
15	0.823	0.002	3087.3	0.000
16	0.811	-0.013	3257.7	0.000
17	0.798	-0.006	3423.8	0.000
18	0.786	-0.011	3585.5	0.000
19	0.774	0.001	3743.0	0.000
20	0.762	0.005	3896.4	0.000
21	0.751	0.001	4045.8	0.000
22	0.739	-0.011	4191.2	0.000
23	0.727	-0.006	4332.7	0.000
24	0.715	-0.015	4470.2	0.000

Figure 4.2: Correlogram of GDP for UK.

Augmented Dickey-Fuller Unit Root Test on GDP_UK

Null Hypothesis: GDP_UK has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 2 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.081038	0.5533
Test critical values:	1% level	-3.997250	
	5% level	-3.428900	
	10% level	-3.137898	

*MacKinnon (1996) one-sided p-values.

Figure 4.3: ADF output of Unit Root test of level of GDP of UK

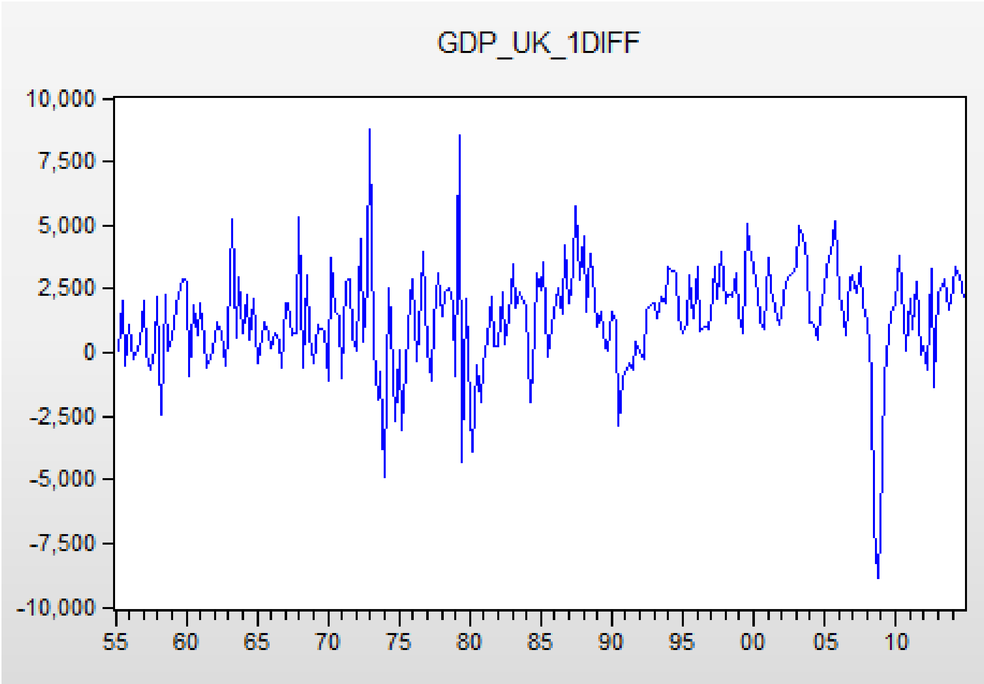
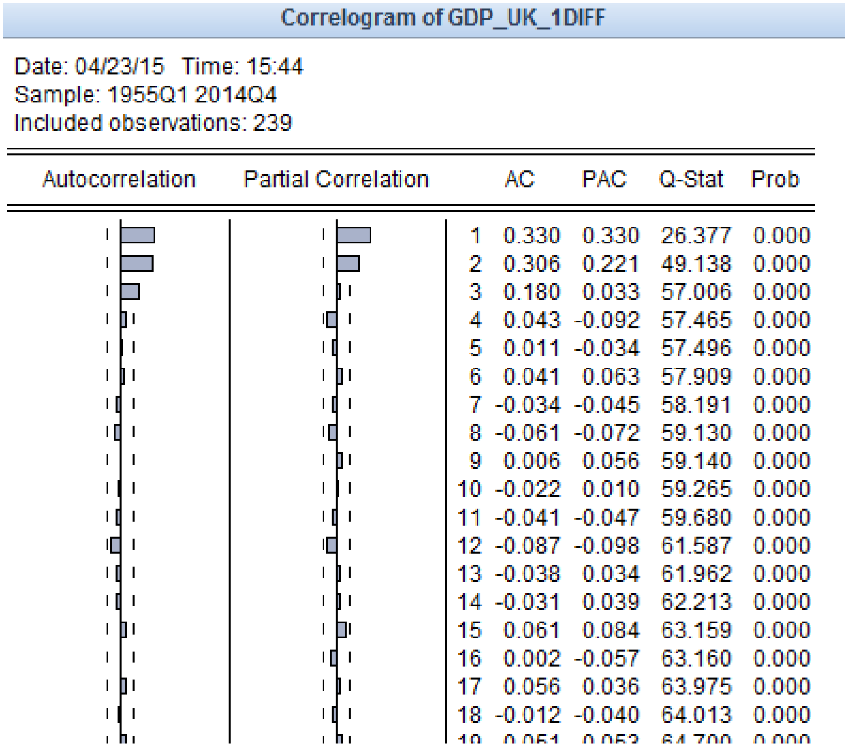


Figure 4.4: Time Series plot of first difference of GDP of UK.



Correlogram of GDP_UK_1DIFF

Date: 04/23/15 Time: 15:44
Sample: 1955Q1 2014Q4
Included observations: 239

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.330	0.330	26.377	0.000
		2	0.306	0.221	49.138	0.000
		3	0.180	0.033	57.006	0.000
		4	0.043	-0.092	57.465	0.000
		5	0.011	-0.034	57.496	0.000
		6	0.041	0.063	57.909	0.000
		7	-0.034	-0.045	58.191	0.000
		8	-0.061	-0.072	59.130	0.000
		9	0.006	0.056	59.140	0.000
		10	-0.022	0.010	59.265	0.000
		11	-0.041	-0.047	59.680	0.000
		12	-0.087	-0.098	61.587	0.000
		13	-0.038	0.034	61.962	0.000
		14	-0.031	0.039	62.213	0.000
		15	0.061	0.084	63.159	0.000
		16	0.002	-0.057	63.160	0.000
		17	0.056	0.036	63.975	0.000
		18	-0.012	-0.040	64.013	0.000
		19	0.051	0.053	64.700	0.000

Figure 4.5: Correlogram of First difference of GDP of UK

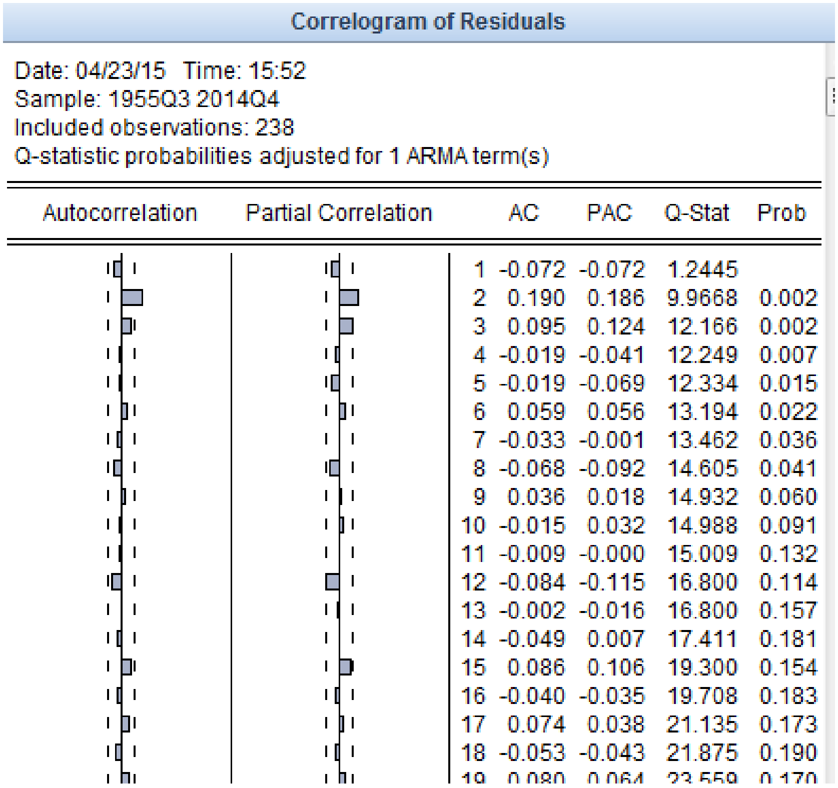
Dependent Variable: GDP_UK_1DIFF
Method: Least Squares
Date: 04/23/15 Time: 15:47
Sample (adjusted): 1955Q3 2014Q4
Included observations: 238 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1346.857	193.7844	6.950284	0.0000
AR(1)	0.330319	0.061411	5.378851	0.0000

R-squared	0.109206	Mean dependent var	1342.778
Adjusted R-squared	0.105431	S.D. dependent var	2116.709
S.E. of regression	2002.018	Akaike info criterion	18.05007
Sum squared resid	9.46E+08	Schwarz criterion	18.07925
Log likelihood	-2145.958	Hannan-Quinn criter.	18.06183
F-statistic	28.93204	Durbin-Watson stat	2.142373
Prob(F-statistic)	0.000000		

Inverted AR Roots	.33

Figure 4.6: Estimation output for an AR(1) estimated on the first difference of the GDP for UK



Correlogram of Residuals

Date: 04/23/15 Time: 15:52
Sample: 1955Q3 2014Q4
Included observations: 238
Q-statistic probabilities adjusted for 1 ARMA term(s)

Lag	AC	PAC	Q-Stat	Prob
1	-0.072	-0.072	1.2445	
2	0.190	0.186	9.9668	0.002
3	0.095	0.124	12.166	0.002
4	-0.019	-0.041	12.249	0.007
5	-0.019	-0.069	12.334	0.015
6	0.059	0.056	13.194	0.022
7	-0.033	-0.001	13.462	0.036
8	-0.068	-0.092	14.605	0.041
9	0.036	0.018	14.932	0.060
10	-0.015	0.032	14.988	0.091
11	-0.009	-0.000	15.009	0.132
12	-0.084	-0.115	16.800	0.114
13	-0.002	-0.016	16.800	0.157
14	-0.049	0.007	17.411	0.181
15	0.086	0.106	19.300	0.154
16	-0.040	-0.035	19.708	0.183
17	0.074	0.038	21.135	0.173
18	-0.053	-0.043	21.875	0.190
19	0.080	0.064	23.559	0.170

Figure 4.7: Correlogram of the residuals from the estimated AR(1) on the first difference of the GDP for UK.

Dependent Variable: GDP_UK_1DIFF
Method: Least Squares
Date: 04/23/15 Time: 16:03
Sample (adjusted): 1955Q3 2014Q4
Included observations: 238 after adjustments
Convergence achieved after 20 iterations
MA Backcast: 1955Q2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1358.055	253.1566	5.364484	0.0000
AR(1)	0.702643	0.113668	6.181536	0.0000
MA(1)	-0.412369	0.145382	-2.836459	0.0050

R-squared	0.142373	Mean dependent var	1342.778
Adjusted R-squared	0.135074	S.D. dependent var	2116.709
S.E. of regression	1968.568	Akaike info criterion	18.02053
Sum squared resid	9.11E+08	Schwarz criterion	18.06429
Log likelihood	-2141.443	Hannan-Quinn criter.	18.03816
F-statistic	19.50601	Durbin-Watson stat	2.070499
Prob(F-statistic)	0.000000		

Inverted AR Roots	.70
Inverted MA Roots	.41

Figure 4.8: Estimation output from an ARMA(1,1) fitted on the first difference of GDP.

Date: 04/23/15 Time: 16:06
Sample: 1955Q3 2014Q4
Included observations: 238
Q-statistic probabilities adjusted for 2 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.036	-0.036	0.3196	
		2	0.094	0.092	2.4436	
		3	0.030	0.036	2.6564	0.103
		4	-0.080	-0.087	4.2187	0.121
		5	-0.057	-0.070	5.0146	0.171
		6	0.041	0.053	5.4235	0.247
		7	-0.046	-0.025	5.9500	0.311
		8	-0.075	-0.093	7.3573	0.289
		9	0.039	0.028	7.7337	0.357
		10	0.002	0.030	7.7344	0.460
		11	-0.014	-0.016	7.7803	0.556
		12	-0.090	-0.123	9.8209	0.456
		13	-0.016	-0.022	9.8893	0.540
		14	-0.038	-0.000	10.253	0.594
		15	0.090	0.091	12.313	0.502
		16	-0.021	-0.040	12.430	0.572
		17	0.069	0.043	13.649	0.552

Figure 4.9: Correlogram of the residuals from the ARMA(1,1) estimated on the first difference of GDP.

Task 5

(14 points in total)

Consider the general model

$$\phi(B)\bigtriangledown^{d}\bigtriangledown_{s}^{D}Y_{t}=\Theta(B)\,\theta(B)\,e_{t}$$

where $e_t \sim NID\left(0,\sigma^2\right)$.

A) (3p) Specify d, D and s, and all the lag (backshift) polynomials, such that the resulting model is an ARMA(2,1). That is, first write out $D = ...$, $d = ...$, and $s = ...$, and then write out each and every polyonomial $\phi(B) = ..., \Theta(B) = ...$, etc, and that is sufficient for this subtask.

B) (3p) Now, using the values of d, D and s and the lag polynomials that you have specified in Task A, write out the process in such a way that you have Y_t alone on the left hand side, and 'everything else' on the right hand side, and the backshift operator should *not* be anywhere in the expression.

C) (4p) Specify d, D and s, and all the lag (backshift) polynomials, such that the resulting model is an SARIMA$(0,0,1)\times(0,0,1)_4$. That is, first write out $D = ...$, $d = ...$, and $s = ...$, and then write out each and every polyonomial $\phi(B) = ..., \Theta(B) = ...$, etc, and that is sufficient for this subtask.

D) (4p) Now, using the values of d, D and s and the lag polynomials that you have specified in Task C, write out the process in such a way that you have Y_t alone on the left hand side, and 'everything else' on the right hand side, and the backshift operator should *not* be anywhere in the expression.