# Written Examination in Time Series Analysis (B3)

## Spring 2016

## 2016-05-25 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. 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. 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)
6. All solutions must be on separate sheets. No solutions on the questionnaire! (If so, they will be disregarded.)
7. 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 $\beta_1$ the examinator will take that as a $\beta_1$ even though you may claim that it is given from the context it should be clear that you meant something else, like $\beta_3$. Thus, given this example, writing $\beta_1$, 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...

8. Please keep the questionaire.

9. Do well!

# Task 1

(20 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) (6p) Given the process

$$Y_t = \alpha_0 + \alpha_1 t + \phi_1 Y_{t-1} + e_t,$$

for what values for the parameters $\alpha_0$, $\alpha_1$ and $\phi_1$ is this process *trend-stationary?* Given those values (that make the process trend stationary - suggest a *transformation* that will make the process covariance stationary.

D) (6p) Given the process

$$Y_t = \alpha_0 + \alpha_1 t + \phi_1 Y_{t-1} + e_t,$$

for what values for the parameters $\alpha_0$, $\alpha_1$ and $\phi_1$ is this process *difference-stationary?* Given those values (that make the process trend stationary - suggest a *transformation* that will make the process covariance stationary.

# Task 2

(16 points in total)

A) (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.

Now, consider the following stochastic process

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

where

$$\phi(B) = (1 - \phi B)$$

and

$$\theta(B) = 1.$$

$e_t \sim N(0, \sigma^2)$ and $\phi \neq 0$.

B) (4p) Derive the expected value of the process.

C) (4p) Derive the variance of the process in $(1)$. Be explicit in what assumptions, if any, you make to be able to derive this result.

# Task 3

(30 points in total) Consider the following model

$$Y_t = Y_{t-1} + e_t. \tag{2}$$

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

Assume that the process started at $Y_1$ and that $Y_0 = 0$.

A) (4p) Derive the *expected value* of the process (2).

B) (4p) Derive the *variance* of the process (2).

C) (4p) Let us say that you have a realization of the process in (2) with 5000 (five thousand) observations, sketch how the SACF for that realization is likely to look.

D) (4p) Now, take second difference of the process above, that is

$$\nabla^2 Y_t$$

where $\nabla = (1 - B)$. Derive the resulting process when the the original process above is differenced twice. Given the respresentation ARMA(p,q) what are the values of p and q for this resulting model?

Is is stationay? Is it invertible?

E) (4p) Derive the expected value of the process $\nabla^2 Y_t$.

F) (4p) Derive the variance of the process $\nabla^2 Y_t$.

G) (4p) Derive the first autocorrelation (yes, auto*correlation*) of the process $\nabla^2 Y_t$.

# Task 4

(34 points in total)

In this task you will analyze a classic data set: The Nile flow data, yearly data from 1871 to 1970.

Please note that some or the output presented in this task may or may not be redundant for the purpose of solving all the subtasks. It is part of the task to know what output to use for the respective subtask.

A) (6p) Perform a unit root test, testing if the Nile flow data has a unit root, use significance level 5%. Document the test procedure as outlined in the test-template.

B) (6p) You try to estimate an AR(1) model to the Nile flow data. Perform a test to test if the first eight autocorrelations of the residuals from this model are simultaneously zero, against the alternative that at least one of them is not zero. Use significance level 5%. Document the test procedure as outlined in the test-template.

Given the result of this test, and this test *alone*, what would be your conclusion in terms of whether this model is a appropriate model or not for this data?

C) (6p) Now, a collegue of yours claims that it is well known since ancient times (and well documented in the literuature) that the Nile has specific flow for seven years, and then the flow changes (increase) the eight year.

Perform a test to test if the eight and only the eight autocorrelation of the residuals from the AR(1) model is significantly different from zero. Use significance level 5%. Document the test procedure as outlined in the test-template.

Given the result of this test, and this test *alone*, what would be your conclusion in terms of whether this model is a appropriate model or not for this data?

D) (8p) There are outputs and correlograms from three additional models in the figures. Do a diagnostic evaluation of these three models using the diagnostic tools that you have at your disposal given the output. There is no need to do any formal tests in this subtask. What/which of these models would be appropriate for capturing the systematic variation of the Nile flow? If more than one model would be considered appropriate, which one would be the *best* in your opinion, and why is that?

E) (4p) Given the general form of a model written as

$$\Phi(B)\,\phi(B)\,Y_t = \Theta(B)\,\theta(B)\,e_t.$$

For the model you find best, write out the lag polynomials (all of them) for your model of choice.

F) (4p) Given the model that you specified the lag polynomials for in the previous subtask, do the algebra such that you in the end have the process on the form: $Y_t$ alone on the left hand side, and $Y_t$ *not* occuring on the right hand side. Also, the back shift operator $B$ must *not* be anywhere in the expression.

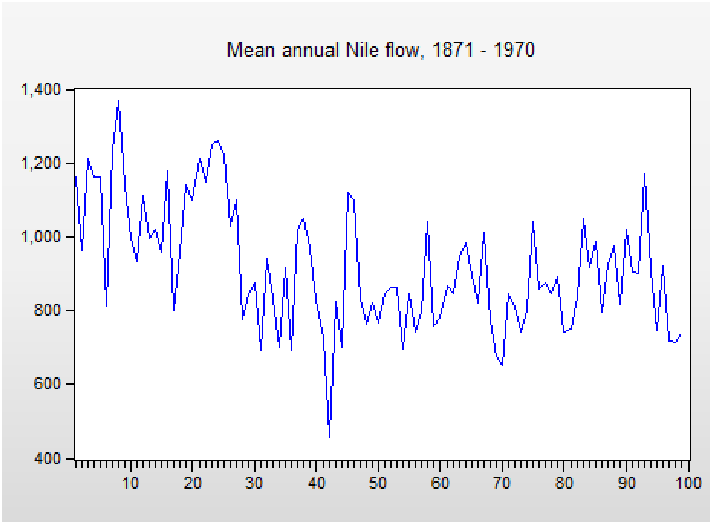



Figure 5.1

Date: 05/22/16 Time: 13:56
Sample: 1 100
Included observations: 99

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.488 | 0.488 | 24.317 | 0.000 |
| | | 2 | 0.387 | 0.195 | 39.753 | 0.000 |
| | | 3 | 0.312 | 0.088 | 49.867 | 0.000 |
| | | 4 | 0.225 | 0.004 | 55.199 | 0.000 |
| | | 5 | 0.214 | 0.061 | 60.075 | 0.000 |
| | | 6 | 0.238 | 0.106 | 66.163 | 0.000 |
| | | 7 | 0.202 | 0.023 | 70.610 | 0.000 |
| | | 8 | 0.271 | 0.138 | 78.677 | 0.000 |
| | | 9 | 0.127 | -0.131 | 80.463 | 0.000 |
| | | 10 | 0.084 | -0.051 | 81.263 | 0.000 |
| | | 11 | 0.216 | 0.204 | 86.577 | 0.000 |
| | | 12 | 0.201 | 0.064 | 91.217 | 0.000 |
| | | 13 | 0.234 | 0.068 | 97.562 | 0.000 |
| | | 14 | 0.189 | -0.058 | 101.75 | 0.000 |
| | | 15 | 0.152 | -0.004 | 104.51 | 0.000 |
| | | 16 | 0.177 | 0.065 | 108.29 | 0.000 |
| | | 17 | 0.145 | 0.022 | 110.87 | 0.000 |
| | | 18 | 0.188 | 0.097 | 115.23 | 0.000 |

Figure 5.2 Correlogram of Nile Flow (orignal data)

Null Hypothesis: NILE_FLOW has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=11)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -5.740625 | 0.0000 |
| Test critical values: | 1% level | -3.498439 | |
| | 5% level | -2.891234 | |
| | 10% level | -2.582678 | |

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

Figure 5.3 ADF test on Nile Flow (orignal) data

Dependent Variable: NILE_FLOW
Method: Least Squares
Date: 05/22/16 Time: 14:05
Sample (adjusted): 2 99
Included observations: 98 after adjustments
Convergence achieved after 3 iterations

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 910.6635 | 29.28521 | 31.09636 | 0.0000 |
| AR(1) | 0.493963 | 0.088150 | 5.603646 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.246473 | Mean dependent var | 914.8469 |
| Adjusted R-squared | 0.238624 | S.D. dependent var | 167.9162 |
| S.E. of regression | 146.5184 | Akaike info criterion | 12.83238 |
| Sum squared resid | 2060894. | Schwarz criterion | 12.88513 |
| Log likelihood | -626.7864 | Hannan-Quinn criter. | 12.85371 |
| F-statistic | 31.40085 | Durbin-Watson stat | 2.182517 |
| Prob(F-statistic) | 0.000000 | | |

| | |
|---|---|
| Inverted AR Roots | .49 |

Figure 5.4 Estimation output for an AR(1) estimated on the Nile Flow data.

Date: 05/22/16 Time: 14:08
Sample: 1 100
Included observations: 98
Q-statistic probabilities adjusted for 1 ARMA term

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | -0.094 | -0.094 | 0.8887 | |
| | | 2 | 0.089 | 0.081 | 1.6991 | 0.192 |
| | | 3 | 0.112 | 0.130 | 2.9989 | 0.223 |
| | | 4 | 0.009 | 0.024 | 3.0067 | 0.391 |
| | | 5 | 0.084 | 0.068 | 3.7583 | 0.440 |
| | | 6 | 0.085 | 0.085 | 4.5213 | 0.477 |
| | | 7 | -0.047 | -0.050 | 4.7622 | 0.575 |
| | | 8 | 0.236 | 0.202 | 10.816 | 0.147 |
| | | 9 | -0.019 | 0.010 | 10.856 | 0.210 |
| | | 10 | -0.093 | -0.139 | 11.819 | 0.224 |
| | | 11 | 0.156 | 0.090 | 14.561 | 0.149 |
| | | 12 | 0.036 | 0.078 | 14.707 | 0.196 |
| | | 13 | 0.121 | 0.116 | 16.389 | 0.174 |
| | | 14 | 0.066 | 0.031 | 16.896 | 0.204 |
| | | 15 | -0.022 | -0.019 | 16.952 | 0.259 |
| | | 16 | 0.130 | 0.056 | 18.971 | 0.215 |
| | | 17 | -0.020 | -0.042 | 19.021 | 0.268 |

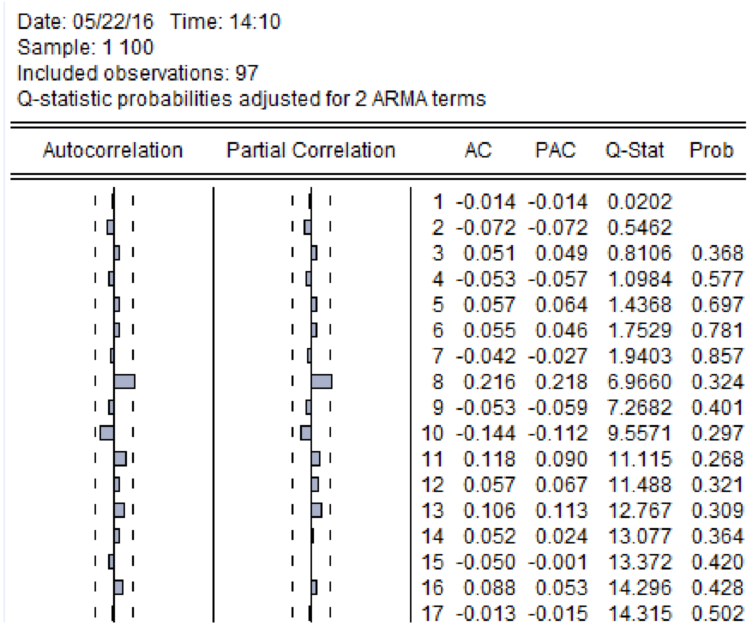
Figure 5.5 Correlogram of residuals from an AR(1) fitted on the Nile Flow data.

Dependent Variable: NILE_FLOW
Method: Least Squares
Date: 05/22/16 Time: 14:03
Sample (adjusted): 3 99
Included observations: 97 after adjustments
Convergence achieved after 3 iterations

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 908.5152 | 37.16032 | 24.44853 | 0.0000 |
| AR(1) | 0.401059 | 0.101011 | 3.970455 | 0.0001 |
| AR(2) | 0.202085 | 0.100711 | 2.006572 | 0.0477 |

| | | | |
|---|---|---|---|
| R-squared | 0.278621 | Mean dependent var | 914.3505 |
| Adjusted R-squared | 0.263272 | S.D. dependent var | 168.7162 |
| S.E. of regression | 144.8139 | Akaike info criterion | 12.81922 |
| Sum squared resid | 1971280. | Schwarz criterion | 12.89885 |
| Log likelihood | -618.7319 | Hannan-Quinn criter. | 12.85141 |
| F-statistic | 18.15297 | Durbin-Watson stat | 2.000484 |
| Prob(F-statistic) | 0.000000 | | |
| Inverted AR Roots | .69 | -.29 | |

Figure 5.6 Estimation output from an AR(2) estimated on Nile Flow data



Date: 05/22/16 Time: 14:10
Sample: 1 100
Included observations: 97
Q-statistic probabilities adjusted for 2 ARMA terms

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | -0.014 | -0.014 | 0.0202 | |
| | | 2 | -0.072 | -0.072 | 0.5462 | |
| | | 3 | 0.051 | 0.049 | 0.8106 | 0.368 |
| | | 4 | -0.053 | -0.057 | 1.0984 | 0.577 |
| | | 5 | 0.057 | 0.064 | 1.4368 | 0.697 |
| | | 6 | 0.055 | 0.046 | 1.7529 | 0.781 |
| | | 7 | -0.042 | -0.027 | 1.9403 | 0.857 |
| | | 8 | 0.216 | 0.218 | 6.9660 | 0.324 |
| | | 9 | -0.053 | -0.059 | 7.2682 | 0.401 |
| | | 10 | -0.144 | -0.112 | 9.5571 | 0.297 |
| | | 11 | 0.118 | 0.090 | 11.115 | 0.268 |
| | | 12 | 0.057 | 0.067 | 11.488 | 0.321 |
| | | 13 | 0.106 | 0.113 | 12.767 | 0.309 |
| | | 14 | 0.052 | 0.024 | 13.077 | 0.364 |
| | | 15 | -0.050 | -0.001 | 13.372 | 0.420 |
| | | 16 | 0.088 | 0.053 | 14.296 | 0.428 |
| | | 17 | -0.013 | -0.015 | 14.315 | 0.502 |

Figure 5.7 Correlogram of residuals from an AR(2) fitted on Nile flow data.

Dependent Variable: NILE_FLOW
Method: Least Squares
Date: 05/22/16 Time: 14:12
Sample (adjusted): 10 99
Included observations: 90 after adjustments
Convergence achieved after 4 iterations

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 882.8571 | 34.65853 | 25.47301 | 0.0000 |
| AR(1) | 0.446908 | 0.095723 | 4.668777 | 0.0000 |
| SAR(8) | 0.261188 | 0.096838 | 2.697164 | 0.0084 |

| | | | |
|---|---|---|---|
| R-squared | 0.275097 | Mean dependent var | 895.6556 |
| Adjusted R-squared | 0.258432 | S.D. dependent var | 154.5076 |
| S.E. of regression | 133.0532 | Akaike info criterion | 12.65214 |
| Sum squared resid | 1540174. | Schwarz criterion | 12.73547 |
| Log likelihood | -566.3463 | Hannan-Quinn criter. | 12.68574 |
| F-statistic | 16.50802 | Durbin-Watson stat | 2.153780 |
| Prob(F-statistic) | 0.000001 | | |

| | | | | |
|---|---|---|---|---|
| Inverted AR Roots | .85 | .60-.60i | .60+.60i | .45 |
| | .00+.85i | -.00-.85i | -.60-.60i | -.60-.60i |
| | -.85 | | | |

Figure 5.8 Estimation out put from a SARMA(1,0)*(1,0)$_8$ model estimated on Nile flow data.

Date: 05/22/16 Time: 14:17
Sample: 1 100
Included observations: 90
Q-statistic probabilities adjusted for 2 ARMA terms

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | -0.078 | -0.078 | 0.5658 | |
| | | 2 | 0.154 | 0.149 | 2.7942 | |
| | | 3 | 0.088 | 0.113 | 3.5379 | 0.060 |
| | | 4 | -0.106 | -0.119 | 4.6193 | 0.099 |
| | | 5 | -0.031 | -0.084 | 4.7149 | 0.194 |
| | | 6 | 0.085 | 0.110 | 5.4209 | 0.247 |
| | | 7 | -0.055 | 0.001 | 5.7212 | 0.334 |
| | | 8 | -0.020 | -0.068 | 5.7609 | 0.451 |
| | | 9 | -0.046 | -0.078 | 5.9741 | 0.543 |
| | | 10 | -0.134 | -0.107 | 7.8383 | 0.449 |
| | | 11 | 0.045 | 0.067 | 8.0540 | 0.529 |
| | | 12 | 0.021 | 0.070 | 8.1024 | 0.619 |
| | | 13 | 0.086 | 0.090 | 8.8976 | 0.631 |
| | | 14 | 0.020 | -0.027 | 8.9412 | 0.708 |
| | | 15 | -0.093 | -0.144 | 9.9013 | 0.702 |
| | | 16 | -0.006 | -0.007 | 9.9052 | 0.769 |
| | | 17 | -0.101 | -0.051 | 11.056 | 0.749 |
| | | 18 | 0.062 | 0.079 | 11.494 | 0.778 |
| | | 19 | -0.016 | -0.029 | 11.523 | 0.828 |
| | | 20 | -0.025 | -0.069 | 11.596 | 0.867 |

Figure 5.9 Correlogram of residuals from a SARMA(1,0)*$(1,0)_8$ model estimated on Nile flow data.

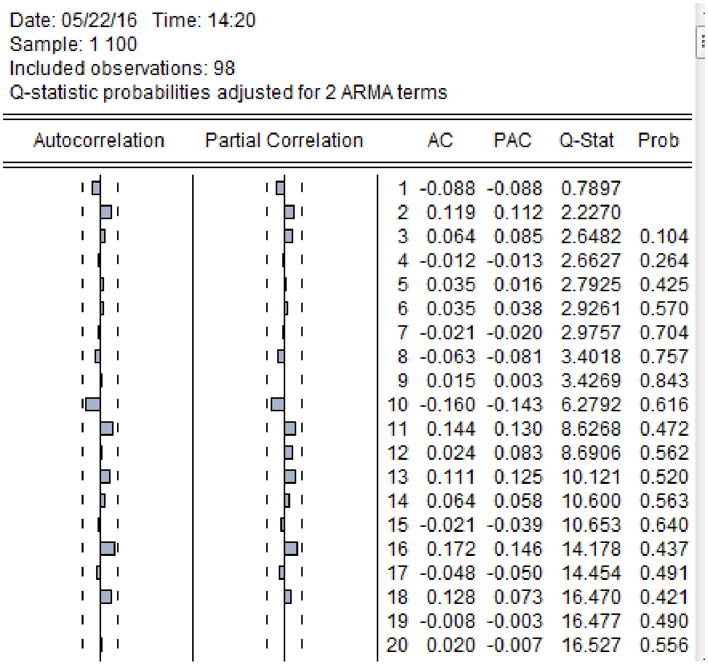
Dependent Variable: NILE_FLOW
Method: Least Squares
Date: 05/22/16 Time: 14:15
Sample (adjusted): 2 99
Included observations: 98 after adjustments
Convergence achieved after 6 iterations
MA Backcast: -6 1

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 912.2783 | 35.96081 | 25.36868 | 0.0000 |
| AR(1) | 0.490381 | 0.089382 | 5.486360 | 0.0000 |
| MA(8) | 0.294315 | 0.098428 | 2.990157 | 0.0036 |

| | | | |
|---|---|---|---|
| R-squared | 0.299466 | Mean dependent var | 914.8469 |
| Adjusted R-squared | 0.284718 | S.D. dependent var | 167.9162 |
| S.E. of regression | 142.0140 | Akaike info criterion | 12.77986 |
| Sum squared resid | 1915958. | Schwarz criterion | 12.85899 |
| Log likelihood | -623.2132 | Hannan-Quinn criter. | 12.81187 |
| F-statistic | 20.30544 | Durbin-Watson stat | 2.173234 |
| Prob(F-statistic) | 0.000000 | | |

| | | | | |
|---|---|---|---|---|
| Inverted AR Roots | .49 | | | |
| Inverted MA Roots | .79+.33i | .79-.33i | .33+.79i | .33-.79i |
| | -.33-.79i | -.33+.79i | -.79-.33i | -.79+.33i |

Figure 5.10 Estimation output from a SARMA(1,0)*(0,1)$_8$ model estimated on the Nile flow data.



Date: 05/22/16 Time: 14:20
Sample: 1 100
Included observations: 98
Q-statistic probabilities adjusted for 2 ARMA terms

| | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|
| 1 | -0.088 | -0.088 | 0.7897 | |
| 2 | 0.119 | 0.112 | 2.2270 | |
| 3 | 0.064 | 0.085 | 2.6482 | 0.104 |
| 4 | -0.012 | -0.013 | 2.6627 | 0.264 |
| 5 | 0.035 | 0.016 | 2.7925 | 0.425 |
| 6 | 0.035 | 0.038 | 2.9261 | 0.570 |
| 7 | -0.021 | -0.020 | 2.9757 | 0.704 |
| 8 | -0.063 | -0.081 | 3.4018 | 0.757 |
| 9 | 0.015 | 0.003 | 3.4269 | 0.843 |
| 10 | -0.160 | -0.143 | 6.2792 | 0.616 |
| 11 | 0.144 | 0.130 | 8.6268 | 0.472 |
| 12 | 0.024 | 0.083 | 8.6906 | 0.562 |
| 13 | 0.111 | 0.125 | 10.121 | 0.520 |
| 14 | 0.064 | 0.058 | 10.600 | 0.563 |
| 15 | -0.021 | -0.039 | 10.653 | 0.640 |
| 16 | 0.172 | 0.146 | 14.178 | 0.437 |
| 17 | -0.048 | -0.050 | 14.454 | 0.491 |
| 18 | 0.128 | 0.073 | 16.470 | 0.421 |
| 19 | -0.008 | -0.003 | 16.477 | 0.490 |
| 20 | 0.020 | -0.007 | 16.527 | 0.556 |

Figure 5.11 Correlogram from a SARMA(1,0)*$(0,1)_8$ model estimated on the Nile flow data.