EViews: Introductory User Guide

Basic Estimation: Time Series Models

Learning support material for the courses:

- ✓ **NMST537 Time Series Analysis**
- ✓ **NEKN432 Econometrics**

Based on official EViews Tutorials & EViews Illustrated.

EViews: Introductory User Guide

TIME SERIES ESTIMATION: DATA

Data and Workfile Documentation

- ***Part10.wf1*** contains monthly data from January 1960 - December 2011.
  - ✓ ***M1*** – money supply, billions of USD
    (source: *Board of Governors of the Federal Reserve*)
  - ✓ ***IP*** –industrial production, index levels
    (source: *Board of Governors of the Federal Reserve*)
  - ✓ ***Tbill*** – 3-month US Treasury rate
    (source: *Board of Governors of the Federal Reserve*)
  - ✓ ***CPI*** – Consumer Price Index, level
    (source: *Bureau of Labor Statistics*)

EViews: Introductory User Guide

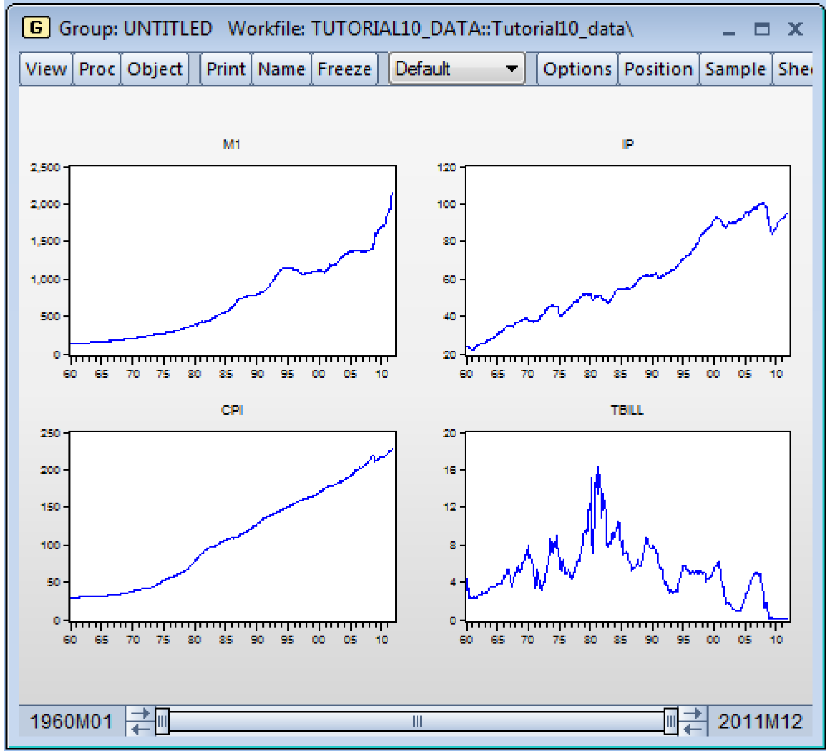
TIME SERIES ESTIMATION: BASICS

Time Series Estimation

- EViews has a built-in toolkit that allows you to estimate time series models ranging from the simplest to the most complex types.
- This tutorial demonstrates how to perform basic single equation ***time series*** regression techniques using EViews. For more details, see *User Guide*.
- <u>The main topics include</u>:
  - ✓ Specifying and Estimating Time Series Regressions
    - ✓ Static and dynamic models
    - ✓ Date functions
    - ✓ Trends and seasonality
  - ✓ Serial Correlation
    - ✓ Testing for Serial Correlation
    - ✓ Correcting for Serial Correlation: ARMA models
  - ✓ Heteroskedasticity and Autocorrelation
    - ✓ Testing for Heteroskedasticity and ARCH terms
    - ✓ HAC Standard Errors
    - ✓ Weighted Least Squares

Simple Time Series Regressions: Example 1 (Part I)

- Suppose you wish to estimate a model that captures movements in ***M1*** (money supply) based on other variables: ***IP*** (industrial production) ***CPI***, and ***Tbill*** rate.



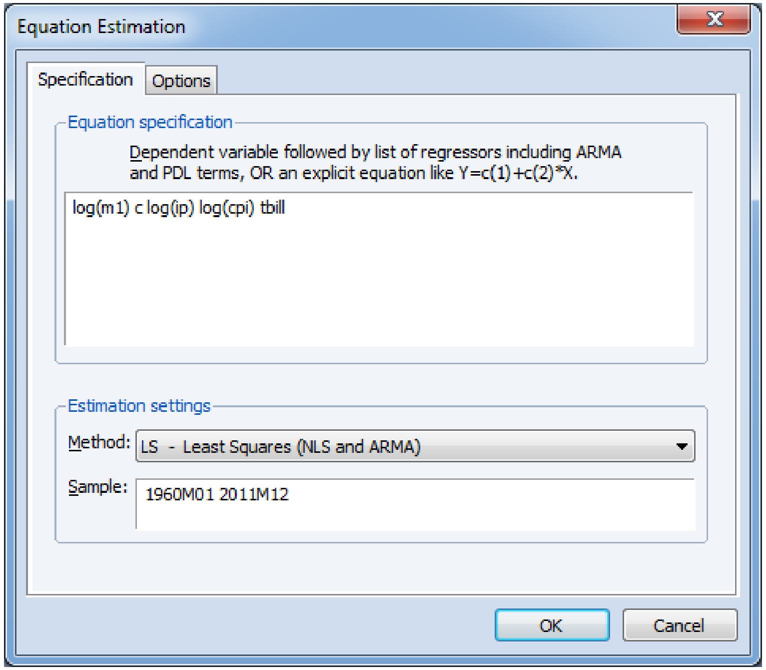
- As a first step, it may help to open these variables as a group, and plot the series in order to observe trends in the data.
- ***M1*** seems to grow over time, so adding a time trend may improve the fit of the model.
- ***CPI*** and ***IP*** seem to move together with ***M1*** and also grow over time.
- ***Tbill*** appears to have a different pattern from ***M1*** (and other series).

Simple Time Series Regressions: Example 1 (Part II)

- Specifying a time series equation in EViews is very easy and follows the same basic steps we introduced in ***Part 9***.

Estimation:

1. Open a workfile.
2. In the main menu, select ***Object*** → ***New Object*** → ***Equation*** and click **OK**.
3. The ***Equation Estimation*** box opens up. Specify here your variables:
   - ✓ ***log(m1)*** – dependent variable
   - ✓ ***c*** – constant
   - ✓ ***log(ip)*** – 1st independent variable
   - ✓ ***log(cpi)*** – 2nd independent variable
   - ✓ ***tbill*** – 3rd independent variable



The model is formulated as follows:

$$\log(M1_t) = \beta_0 + \beta_1 \log(IP_t) + \beta_2 \log(CPI_t) + \beta_3 TBILL_t + \varepsilon_t,$$
$$t = 1, \dots, T, \varepsilon_t \sim WN(0, \sigma^2).$$

Simple Time Series Regressions: Example 1 (Part III)

- The estimation output is displayed here.

- All variables appear to be highly statistically significant (based on *p-values/t-stats*).
- The *R-squared* value is very high: results imply that around 99.25% of the variation in ***log(m1)*** can be explained by the other variables in the model. Normally this would imply a very good fit for the model.
- We caution against these results: high *R-squared* does not necessarily imply that the model is a good or useful one.

Equation: EQ01 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(M1)
Method: Least Squares
Date: 03/25/13 Time: 02:57
Sample: 1960M01 2011M12
Included observations: 624

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.859324	0.042875	20.04261	0.0000
LOG(IP)	0.196359	0.025865	7.591598	0.0000
LOG(CPI)	1.055454	0.015406	68.51129	0.0000
TBILL	-0.021441	0.000986	-21.74826	0.0000

R-squared	0.992558	Mean dependent var	6.286356
Adjusted R-squared	0.992522	S.D. dependent var	0.825190
S.E. of regression	0.071357	Akaike info criterion	-2.435840
Sum squared resid	3.156971	Schwarz criterion	-2.407403
Log likelihood	763.9820	Hannan-Quinn criter.	-2.424789
F-statistic	27564.63	Durbin-Watson stat	0.027685
Prob(F-statistic)	0.000000		

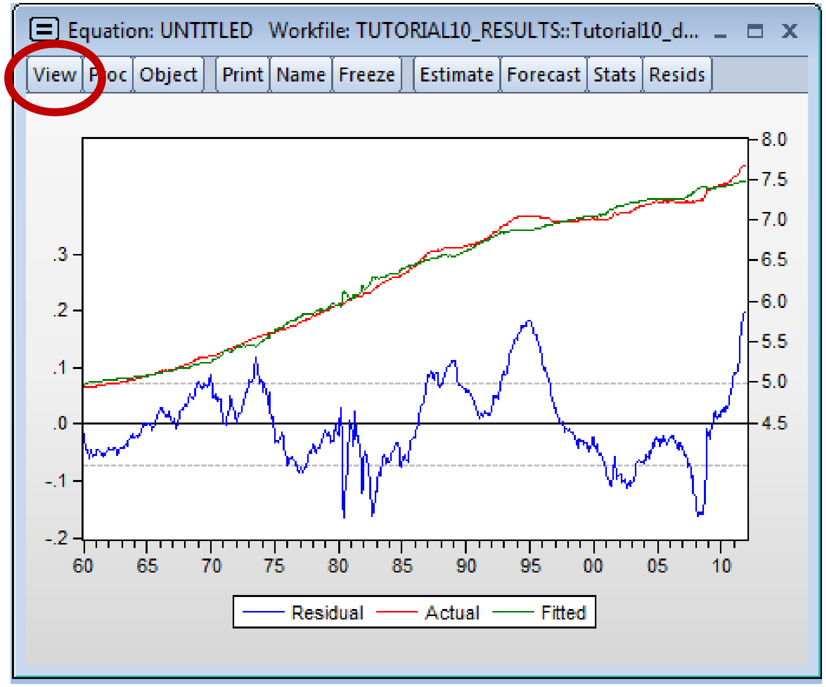
Simple Time Series Regressions: Example 1 (Part IV)

- Let's inspect the behavior of residuals in the ***Equation View***.

Graphical Examination of Residuals:

1. Click **View** on the ***Equation Object*** menu bar.
2. Select ***Actual, Fitted, Residual*** → ***Actual, Fitted, Residual Graph***.

Note that the residuals of this regression appear to have long periods of positive values followed by long periods of negative values, providing strong visual evidence of serial correlation.



Simple Time Series Regressions: Example 2 (Part I)

- *Distributed lag models* are easy to specify in EViews.
- In these models, one/more variables affects the dependent variable with a lag. Lags of dependent or independent variables can be specified directly in the equation box in EViews.

Estimation:

Suppose you want to examine how ***M1*** is affected by ***CPI*** and its first two lags in log form.

1. Open a workfile.
2. In the main menu, select ***Object*** → ***New Object*** → ***Equation*** and click **OK**.
3. The ***Equation Estimation*** box opens up. Specify here your variables:
   - ✓ ***log(m1)*** – dependent variable
   - ✓ ***c*** – constant
   - ✓ ***log(cpi)*** – 1st independent variable
   - ✓ ***log(cpi(-1))*** – 2nd independent variable
   - ✓ ***log(cpi(-2))*** – 3rd independent variable

   This model is formulated as follows:

$$\log(M1_t) = \beta_0 + \beta_1 \log(CPI_t) + \beta_2 \log(CPI_{t-1}) + \beta_3 \log(CPI_{t-2}) + \varepsilon_t,$$
$$t = 3, \dots, T, \varepsilon_t \sim WN(0, \sigma^2).$$

Simple Time Series Regressions: Example 2 (Part II)

- The estimation output is displayed here.

- ❑ Notice that we now have 622 instead of 624 observations because we are using two lags of ***CPI***.
- ❑ *R-squared* and *F-statistic* values are high, which is surprising especially since some coefficients do not appear to be very significant. For example, current ***CPI*** is statistically significant only at the 10% level, while the first lag of ***CPI*** is not significant.
- ❑ What causes these results? It turns out, there is substantial correlation between ***CPI***, ***CPI(-1)*** and ***CPI(-2)***.

Equation: EQ02 Workfile: TUTORIAL10_RESULTS::Tutorial10_data\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(M1)
Method: Least Squares
Date: 11/19/12 Time: 20:12
Sample (adjusted): 1960M03 2011M12
Included observations: 622 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.019323	0.026417	38.58627	0.0000
LOG(CPI)	-2.672317	1.531671	-1.744707	0.0815
LOG(CPI(-1))	-2.267378	2.754062	-0.823285	0.4107
LOG(CPI(-2))	6.117664	1.530592	3.996926	0.0001

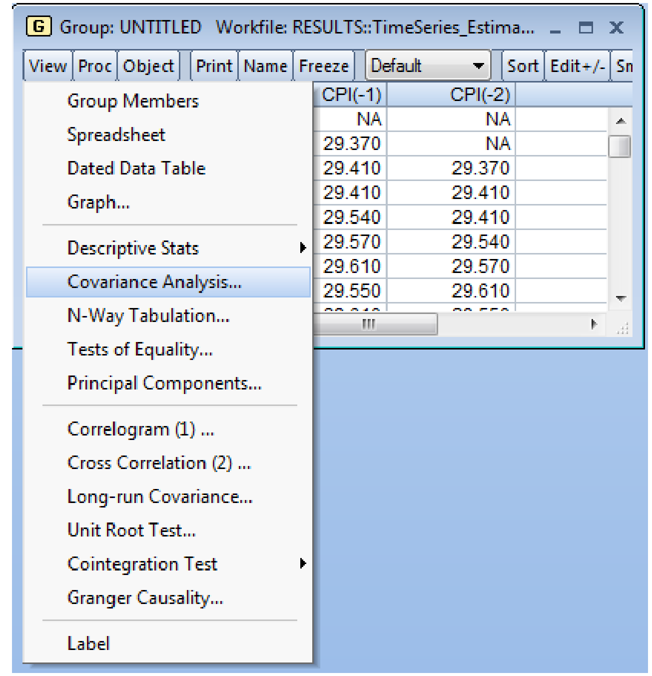
R-squared	0.986847	Mean dependent var	6.290681
Adjusted R-squared	0.986783	S.D. dependent var	0.822974
S.E. of regression	0.094614	Akaike info criterion	-1.871608
Sum squared resid	5.532245	Schwarz criterion	-1.843100
Log likelihood	586.0700	Hannan-Quinn criter.	-1.860528
F-statistic	15455.37	Durbin-Watson stat	0.031493
Prob(F-statistic)	0.000000		

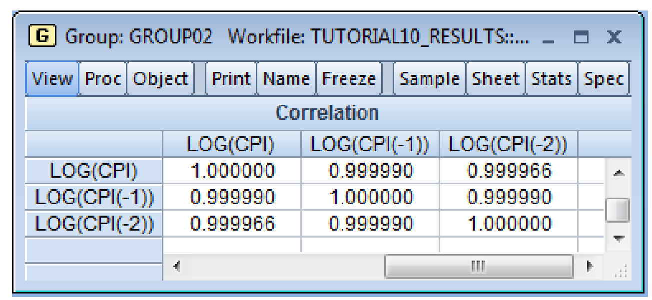
Simple Time Series Regressions: Example 2 (Part III)

- One can inspect the correlation matrix of ***CPI*** and its first two lags. To create the correlation matrix, follow these steps:
  1. Type in the command window:

     ```
     show cpi cpi(-1) cpi(-2)
     ```

     to open these series as a group.
  2. On the top menu of the group, click on ***View*** → ***Covariance Analysis***.
  3. The **Covariance Analysis** box opens up. Click the **Correlation** box under ***Statistics***.





- As you can see from the correlation matrix, the series are highly correlated (multicollinearity). This multicollinearity makes it difficult to estimate the effect at each lag.

Simple Time Series Regressions: Example 2 (Part IV)

- A more formal way to investigate collinearity among regressors is the **Coefficient Variance Decomposition** from the ***View*** menu of the Equation box. This test is particularly useful if there is a linear relationship between regressors which the simple correlation matrix may fail to detect.

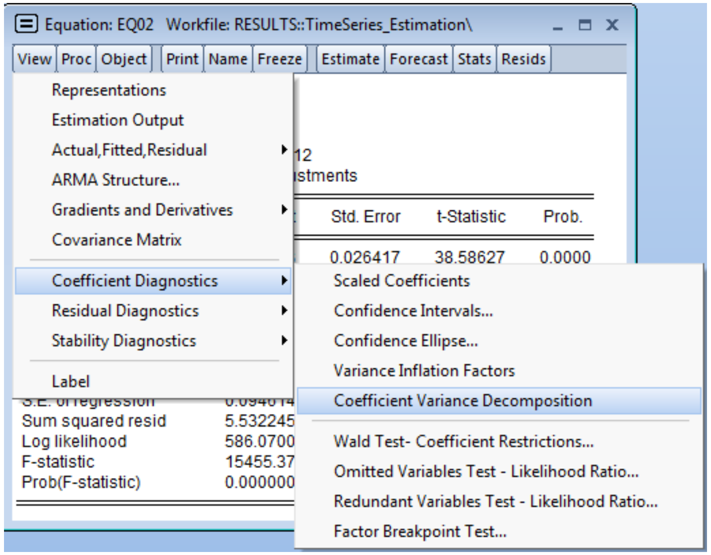
<u>**Coefficient Variance Decomposition:**</u>

1. Open an equation.
2. On the top menu of the equation box, click on ***View → Coefficient Diagnostics → Coefficient Variance Decomposition***.

   Note that decomposition calculations follow Besley, Kuh and Welsch (2004). In general, there is a high degree of collinearity if:

   ✓ A condition number is smaller than 1/900 (0.001)

   ✓ There are two or more variables with higher covariance decomposition proportion than 0.5 (associated with a small condition number).



Simple Time Series Regressions: Example 2 (Part V)

- The Equation View displays a table showing *Eigenvalues*, *Condition Numbers*, *Variance Decomposition Proportions* and *Eigenvectors:*
  - ✓ The top portion shows the Eigenvalues sorted from the smallest to the largest.
  - ✓ In our case, the condition numbers are much smaller than 1/900=0.001, which indicates the presence of collinearity. (Note that the last condition number is always equal to 1).
  - ✓ The middle panel displays the variance decomposition proportions. The proportions associated with the smallest condition number. As you can see, 3 of the values are above 0.5, indicating that there is a high degree of collinearity between CPI, CPI(-1) and CPI(-2).

Equation: EQ02 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Coefficient Variance Decomposition
Date: 04/26/13 Time: 01:32
Sample: 1960M01 2011M12
Included observations: 622

Eigenvalues	11.37741	0.896239	0.000637	2.28E-07
Condition	2.00E-08	2.54E-07	0.000358	1.000000

Variance Decomposition Proportions

	Associated Eigenvalue			
Variable	1	2	3	4
C	6.30E-05	0.101787	0.898145	5.06E-06
LOG(CPI)	0.808877	0.191122	1.17E-06	3.19E-08
LOG(CPI(-1))	1.000000	2.78E-08	4.34E-07	9.86E-09
LOG(CPI(-2))	0.808856	0.191142	1.66E-06	3.19E-08

Eigenvectors

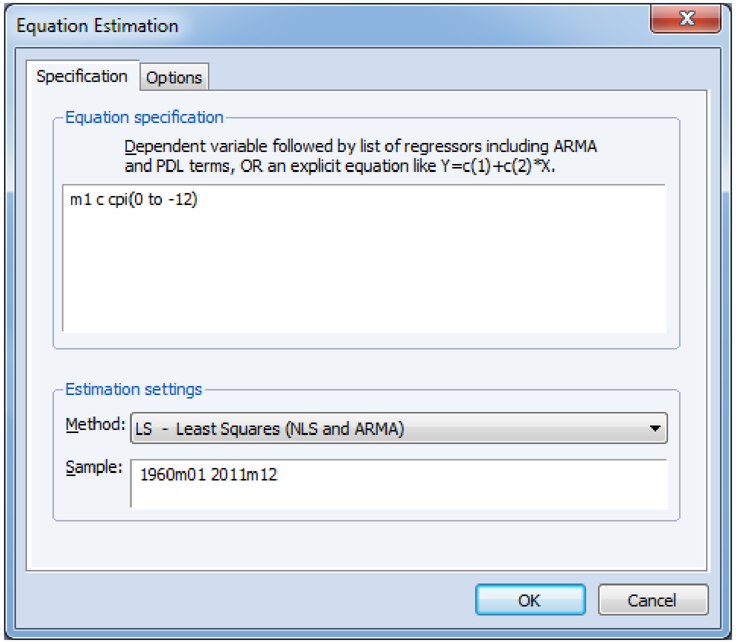
	Associated Eigenvalue			
Variable	1	2	3	4
C	-6.22E-05	0.008903	0.992187	-0.124440
LOG(CPI)	0.408399	-0.707309	-0.065526	-0.573263
LOG(CPI(-1))	-0.816492	-0.000485	-0.071895	-0.572863
LOG(CPI(-2))	0.408106	0.706848	-0.078115	-0.572462

Simple Time Series Regressions: Example 3 (Part I)

- You can just as easily specify a model containing a large number of lags without having to explicitly type out all the lags.

Estimation:

1. Suppose you would like to examine how ***M1*** is affected by current ***CPI*** values and its 12 lags. On the main menu, select ***Object*** → ***New Object*** → ***Equation*** and click **OK.**
2. The ***Equation Estimation*** box opens.
3. Specify here your variables:
   - ✓ ***m1*** – dependent variable
   - ✓ ***c*** – constant
   - ✓ ***CPI(0 to -12)*** – current and lagged values of CPI.



This model is formulated as follows:

$$M1_t = \beta_0 + \sum_{i=0}^{12} \beta_i CPI_{t-i} + \varepsilon_t,$$

$$t = 13, \ldots, T, \varepsilon_t \sim WN(0, \sigma^2).$$

Simple Time Series Regressions: Example 3 (Part II)

- The estimation output is displayed here.

- ❑ Again, taken individually, none of the coefficients (except lag 12) is statistically significant.
- ❑ However, the *R-squared* value and the *F-statistic* are very high.
- ❑ Similar to the previous example, the issue here is that there is a high degree of collinearity among all regressors (since these are lags of the same variable).
- ❑ A way to deal with high collinearity is to fit a ***polynomial distributed lag model*** (not discussed in this tutorial).

Equation: EQ02A Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: M1
Method: Least Squares
Date: 12/10/12 Time: 00:06
Sample (adjusted): 1961M01 2011M12
Included observations: 612 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-104.7026	8.063640	-12.98453	0.0000
CPI	-4.664840	10.94955	-0.426030	0.6702
CPI(-1)	-1.266860	20.40829	-0.062076	0.9505
CPI(-2)	-5.402067	22.00400	-0.245504	0.8062
CPI(-3)	6.963252	22.13892	0.314525	0.7532
CPI(-4)	-0.714523	22.17279	-0.032225	0.9743
CPI(-5)	-2.113961	22.32003	-0.094711	0.9246
CPI(-6)	-2.018993	22.43649	-0.089987	0.9283
CPI(-7)	4.547749	22.34103	0.203560	0.8388
CPI(-8)	-4.054178	22.21367	-0.182508	0.8552
CPI(-9)	-1.175856	22.21645	-0.052927	0.9578
CPI(-10)	3.759427	22.10938	0.170038	0.8650
CPI(-11)	-23.07787	20.53068	-1.124067	0.2614
CPI(-12)	37.26513	11.04690	3.373355	0.0008

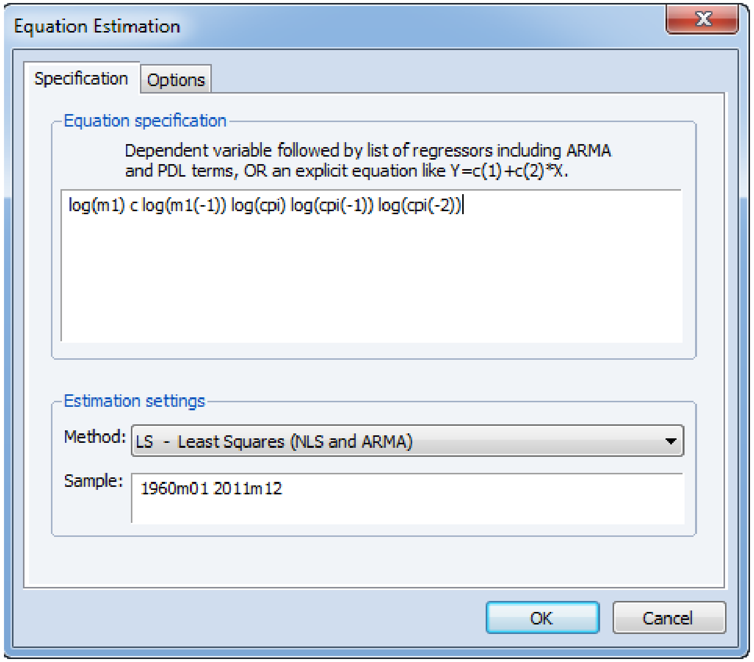
R-squared	0.968625	Mean dependent var	736.1742
Adjusted R-squared	0.967943	S.D. dependent var	501.4292
S.E. of regression	[illegible]	Akaike info criterion	11.85516
Sum squared resid	4819906.	Schwarz criterion	11.95620
Log likelihood	-3613.679	Hannan-Quinn criter.	11.89446
F-statistic	1420.155	Durbin-Watson stat	0.027281
Prob(F-statistic)	0.000000		

Simple Time Series Regressions: Example 4 (Part I)

- You can just as easily specify a dynamic model containing lags of dependent and independent variables.

Estimation:

1. On the main menu, select ***Object*** → ***New Object*** → ***Equation*** and click **OK.**
2. The ***Equation Estimation*** box opens. Specify here your variables:
   - ✓ ***log(m1)*** – dependent variable
   - ✓ ***c*** – constant
   - ✓ ***log(m1(-1))***
   - ✓ ***log(cpi)***
   - ✓ ***log(cpi(-1))***
   - ✓ ***log(cpi(-2))***



This model is formulated as follows:

$$\log(M1_t) = \beta_0 + \beta_1 \log(M1_{t-1}) + \beta_2 \log(CPI_t) + \beta_3 \log(CPI_{t-1}) + \beta_4 \log(CPI_{t-2}) + \varepsilon_t,$$
$$t = 3, \ldots, T, \varepsilon_t \sim WN(0, \sigma^2).$$

Simple Time Series Regressions: Example 4 (Part II)

- The estimation output is displayed here.

- ❑ Notice that the lagged dependent variable (***log(m1(-1))***) is close to unity and is highly significant.
- ❑ If errors are serially correlated, OLS estimates are biased and inconsistent in the presence of lagged dependent (see *User Guide* for details).

Equation: EQ03 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(M1)
Method: Least Squares
Date: 04/05/13 Time: 22:40
Sample (adjusted): 1960M03 2011M12
Included observations: 622 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.010950	0.003446	3.178026	0.0016
LOG(M1(-1))	0.994099	0.002849	348.9773	0.0000
LOG(CPI)	0.070822	0.109118	0.649043	0.5166
LOG(CPI(-1))	-0.421052	0.195764	-2.150816	0.0319
LOG(CPI(-2))	0.357229	0.110003	3.247445	0.0012

R-squared	0.999934	Mean dependent var	6.290681
Adjusted R-squared	0.999933	S.D. dependent var	0.822974
S.E. of regression	0.006723	Akaike info criterion	-7.158591
Sum squared resid	0.027887	Schwarz criterion	-7.122956
Log likelihood	2231.322	Hannan-Quinn criter.	-7.144741
F-statistic	2326285.	Durbin-Watson stat	1.486173
Prob(F-statistic)	0.000000		

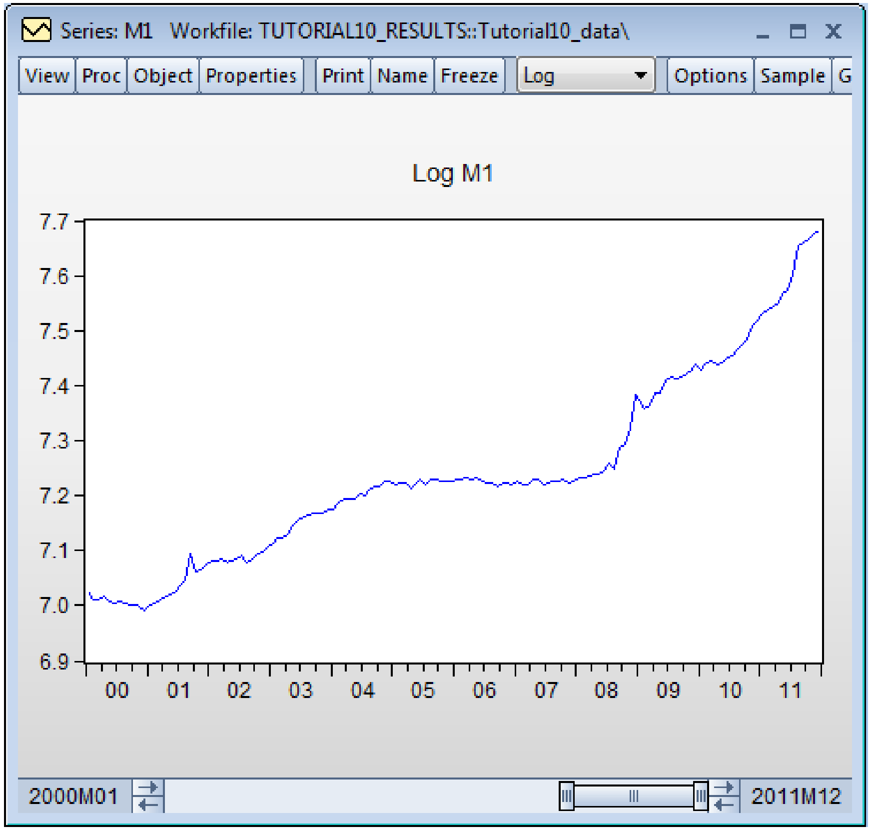
EViews: Introductory User Guide

TIME SERIES ESTIMATION: DATE FUNCTIONS

Time Series Regressions: Date Dummies (Part I)

- EViews allows you to estimate regression models using dummy variables directly in the estimation window without first having to create the dummies.

- For example, judging by the behavior of ***M1*** in the graph shown here, it appears that the series grew at a much more rapid pace since 2008, thanks to the many rounds of quantitative easing (QE) carried out by the Federal Reserve.



Time Series Regressions: Date Dummies (Part II)

- Let's check whether the period since 2008 has indeed had an outsized impact in the growth of *M1*. For more dummy operators see the previous parts.

Date Dummies:

1. On the main menu, select ***Object*** → ***New Object*** → ***Equation*** and click **OK.**
2. The ***Equation Estimation*** box opens. Specify here your variables:
   - ✓ ***log(m1)*** – dependent variable
   - ✓ ***c*** – constant
   - ✓ ***log(cpi)***
   - ✓ ***log(ip)***
   - ✓ ***tbill***
   - ✓ ***@year>2008*** - dummy variable, equal to 1 for the period after 2008 and 0 otherwise

This model is formulated as follows:

$$\log(M1_t) = \beta_0 + \beta_1 \log(CPI_t) + \beta_2 \log(IP_t) + \beta_3 TBILL_t + \beta_4 1_{[t>Y2008]} + \varepsilon_t,$$
$$t = 1, \ldots, T, \varepsilon_t \sim WN(0, \sigma^2).$$

Equation: EQ04 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(M1)
Method: Least Squares
Date: 11/19/12 Time: 23:47
Sample: 1960M01 2011M12
Included observations: 624

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.847154	0.041678	20.32641	0.0000
LOG(CPI)	1.037515	0.015235	68.10099	0.0000
LOG(IP)	0.214960	0.025293	8.498799	0.0000
TBILL	-0.018851	0.001044	-18.05355	0.0000
@YEAR>2008	0.085273	0.013727	6.211980	0.0000

R-squared	0.992995	Mean dependent var	6.286356
Adjusted R-squared	0.992950	S.D. dependent var	0.825190
S.E. of regression	0.069288	Akaike info criterion	-2.493109
Sum squared resid	2.971713	Schwarz criterion	-2.457563
Log likelihood	782.8500	Hannan-Quinn criter.	-2.479296
F-statistic	21936.49	Durbin-Watson stat	0.028152
Prob(F-statistic)	0.000000		

*Notice that the impact of the dummy variable on **M1** is large and significant. This means that post-2008, there is a sizable and significant increase in **M1**.*

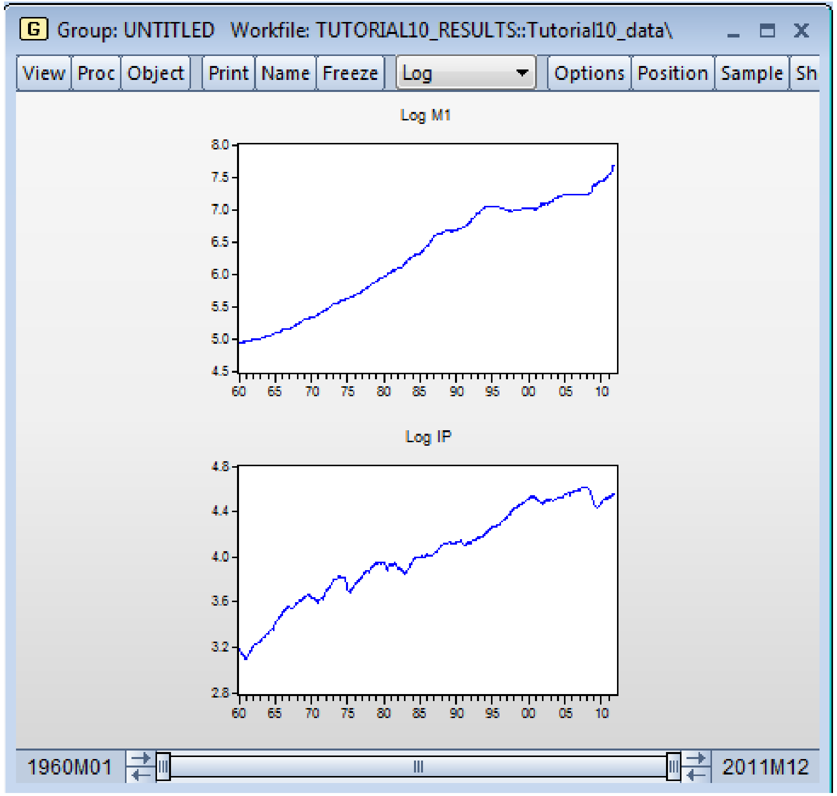
EViews: Introductory User Guide

TIME SERIES ESTIMATION: TRENDS AND SEASONALITY

Time Trend (Part I)

- Many economic time series have a tendency to grow over time.
- Ignoring the fact that two series contain a ***time trend*** (are trending together) can lead us to falsely conclude that changes in one variable actually *cause* changes in the other variable.
- Finding a relationship between two or more trending variables simply because they are growing over time is an example of a ***spurious regression problem***.
- The good news is that adding a time-trend to the regression eliminates this problem.

- For example, judging from their graphs (shown below), ***M1*** and ***IP*** appear to grow together over time.



Time Trend (Part II)

- Time trends can be accommodated easily in EViews using function ***@trend***.

Time trend:

1. On the main menu, select ***Object*** → ***New Object*** → ***Equation*** and click **OK.**
2. The ***Equation Estimation*** box opens. Specify here your variables:
   - ✓ ***log(m1)*** – dependent variable
   - ✓ ***c*** – constant
   - ✓ ***log(ip)***
   - ✓ ***@trend*** – trend variable

This model is formulated as follows:

$$\log(M1_t) = \beta_0 + \beta_1 \log(IP_t) + \beta_2 t + \varepsilon_t,$$
$$t = 1, \ldots, T, \varepsilon_t \sim WN(0, \sigma^2).$$

Equation: EQ10 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(M1)
Method: Least Squares
Date: 11/25/12 Time: 23:10
Sample: 1960M01 2011M12
Included observations: 624

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.020579	0.212328	23.64536	0.0000
LOG(IP)	-0.042589	0.063964	-0.665823	0.5058
@TREND	0.004612	0.000147	31.45030	0.0000

R-squared	0.973727	Mean dependent var	6.286356
Adjusted R-squared	0.973643	S.D. dependent var	0.825190
S.E. of regression	0.133969	Akaike info criterion	-1.177614
Sum squared resid	11.14559	Schwarz criterion	-1.156286
Log likelihood	370.4156	Hannan-Quinn criter.	-1.169326
F-statistic	11507.80	Durbin-Watson stat	0.002548
Prob(F-statistic)	0.000000		

- ❑ The coefficient of ***log(ip)*** is negative and not significant. In addition, the coefficient of the ***time trend*** is positive and statistically significant.
- ❑ The R-squared is very high even though the coefficient of ***log(ip)*** is not significant. Movements in ***log(m1)*** are explained by the time trend.
- ❑ This means that omitting ***@trend*** can result in a spurious regression yielding biased estimators of the impact of ***log(ip)*** on ***log(m1)***.

Seasonality (Part I)

- Sometimes time series exhibit seasonal patterns. For example, retails sales tend to be higher in the last quarter of the year because of the Holiday Shopping season.
- Most macroeconomic series are already seasonally adjusted beforehand so there is no need to worry about seasonal issues.
- If you suspect a series displays seasonal patterns, you can include a set of dummy variables to account for the seasonality in the dependent variable.
- EViews has a built-in function that creates dummy variables corresponding to each month (or quarter, if the data is quarterly).

Seasonality (Part II)

Regression with seasonal factors:

1. On the main menu, select ***Object*** → ***New Object*** → ***Equation*** and click **OK.**
2. The ***Equation Estimation*** box opens. Specify here your variables:
   - ✓ ***log(m1)*** – dependent variable
   - ✓ ***c*** – constant
   - ✓ ***log(ip)***
   - ✓ ***log(cpi)***
   - ✓ ***@trend*** – trend variable
   - ✓ ***@seas(2), …, @seas(12)*** - seasonal dummies

This model is formulated as follows:

$$\log(M1_t) = \beta_0 + \beta_1 \log(IP_t) + \beta_2 \log(CPI_t) + \beta_3 t + \sum_{j=2}^{12} \alpha_j 1_{[t:month\sim j]} + \varepsilon_t, t = 1, \dots, T, \varepsilon_t \sim WN(0, \sigma^2).$$

Equation: EQ12 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(M1)
Method: Least Squares
Date: 11/26/12 Time: 01:50
Sample: 1960M01 2011M12
Included observations: 624

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.699636	0.148310	18.20261	0.0000
LOG(IP)	-0.136870	0.039139	-3.497040	0.0005
LOG(CPI)	0.791548	0.024327	32.53799	0.0000
@TREND	0.001846	0.000123	14.95019	0.0000
@SEAS(2)	-0.000666	0.016032	-0.041545	0.9669
@SEAS(3)	-0.000627	0.016032	-0.039090	0.9688
@SEAS(4)	-0.000722	0.016032	-0.045041	0.9641
@SEAS(5)	-0.001008	0.016032	-0.062870	0.9499
@SEAS(6)	-0.001478	0.016032	-0.092207	0.9266
@SEAS(7)	-0.001122	0.016033	-0.070004	0.9442
@SEAS(8)	-0.000531	0.016033	-0.033142	0.9736
@SEAS(9)	0.000386	0.016033	0.024079	0.9808
@SEAS(10)	0.000169	0.016033	0.010522	0.9916
@SEAS(11)	0.001791	0.016034	0.111733	0.9111
@SEAS(12)	0.003615	0.016034	0.225470	0.8217

R-squared	0.990406	Mean dependent var	6.286356
Adjusted R-squared	0.990186	S.D. dependent var	0.825190
S.E. of regression	0.081748	Akaike info criterion	-2.146599
Sum squared resid	4.069813	Schwarz criterion	-2.039961
Log likelihood	684.7390	Hannan-Quinn criter.	-2.105160
F-statistic	4490.809	Durbin-Watson stat	0.008287
Prob(F-statistic)	0.000000		

- The seasonal dummies are not individually statistically significant (you can also carry out a Wald test for joint significance and conclude the same).
- This means that the **M1** series does not display seasonal patterns (this is because the series is already adjusted for season patterns).

EViews: Introductory User Guide

TIME SERIES ESTIMATION: TESTING SERIAL CORRELATION

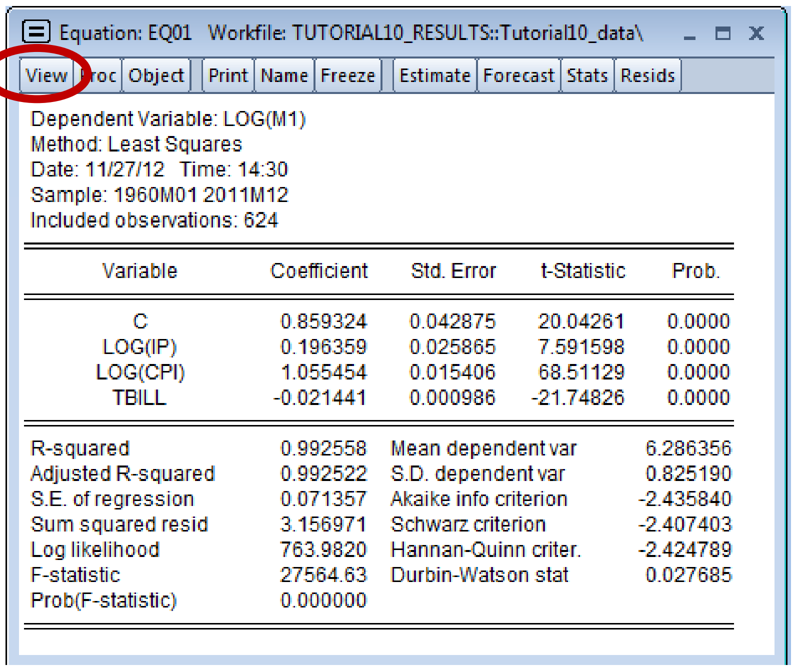
Serial Correlation

- As we have seen in the previous examples, residuals from our time series regressions appear to be correlated with their own lagged values (they display serial correlation).
- Serial correlation is a common occurrence in time series data because the data is ordered (over time); it is therefore not surprising that neighboring error terms turn out to be correlated. Serial correlation violates the standard assumption of regression theory that error terms are uncorrelated.

- If untreated, serial correlation leads to a number of issues:
  - ✓ Reported standard errors and t-statistics are invalid (even asymptotically).
  - ✓ Coefficients may be biased, though not necessarily inconsistent (if data is weakly dependent). In the presence of lagged dependent variables, OLS estimates are biased and inconsistent.
- EViews provides tools for detecting serial correlation and correcting regressions to account for its presence.

Detecting Serial Correlation: Visual Inspection (Part I)

- You can visually inspect your residuals to see if serial correlation is present.
  1. Open an equation.
  2. Click **View** on the ***Equation*** box menu bar.
  3. Select ***Actual, Fitted, Residual*** → ***Actual, Fitted, Residual Graph***.
- From the graph shown here, residuals seem to display runs of positive and negative values, providing strong visual evidence of serial correlation.

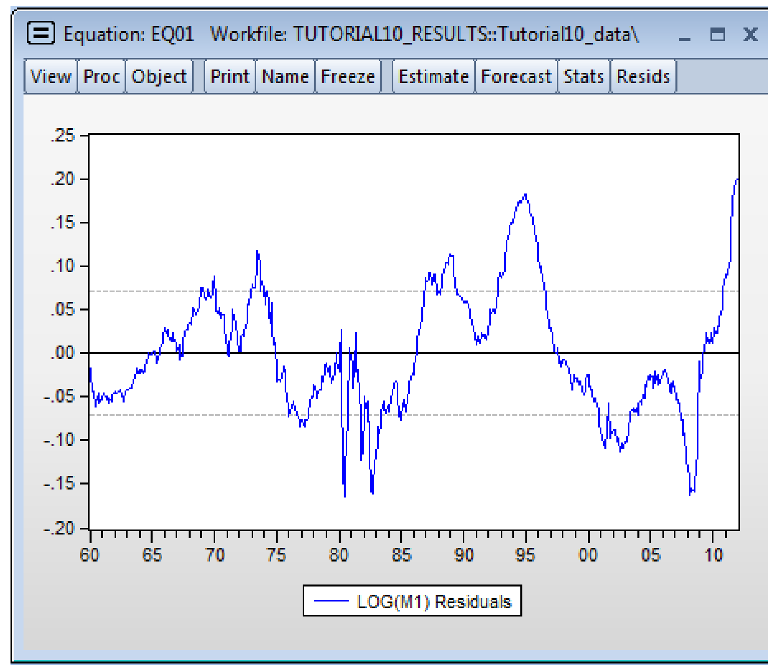

Equation: EQ01 Workfile: TUTORIAL10_RESULTS::Tutorial10_data\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOG(M1)
Method: Least Squares
Date: 11/27/12 Time: 14:30
Sample: 1960M01 2011M12
Included observations: 624

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.859324	0.042875	20.04261	0.0000
LOG(IP)	0.196359	0.025865	7.591598	0.0000
LOG(CPI)	1.055454	0.015406	68.51129	0.0000
TBILL	-0.021441	0.000986	-21.74826	0.0000

R-squared	0.992558	Mean dependent var	6.286356
Adjusted R-squared	0.992522	S.D. dependent var	0.825190
S.E. of regression	0.071357	Akaike info criterion	-2.435840
Sum squared resid	3.156971	Schwarz criterion	-2.407403
Log likelihood	763.9820	Hannan-Quinn criter.	-2.424789
F-statistic	27564.63	Durbin-Watson stat	0.027685
Prob(F-statistic)	0.000000		



Detecting Serial Correlation: Visual Inspection (Part II)

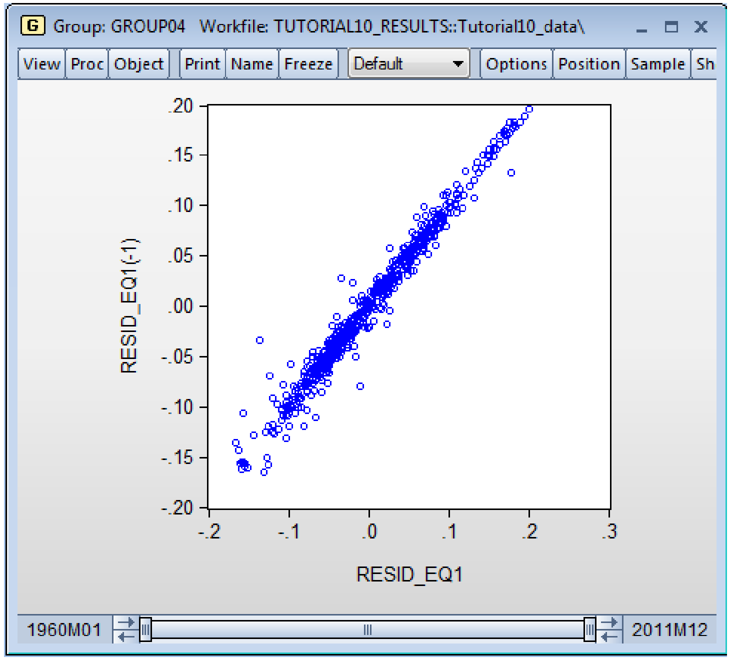
- Another way to visually inspect the residuals is to obtain a scatter plot of residuals against their lagged values.

<u>**Creating such a scatterplot:**</u>

1. Click the ***Proc*** button on the top menu of the ***Equation*** box, and select ***Make Residual Series***.
2. Name the residuals **resid_eq1**.
3. Create a group consisting of residuals and their lagged values by typing in the command window (and pressing **Enter** after typing):

```
show resid_eq1 resid_eq1(-1)
```

4. The **Group Spreadsheet** opens up. On the top menu of the Group Spreadsheet, select ***View*** → ***Graph***.
5. The **Graph Options** dialog box opens up. Select **Scatter** under ***Graph type*** → ***Specific***.



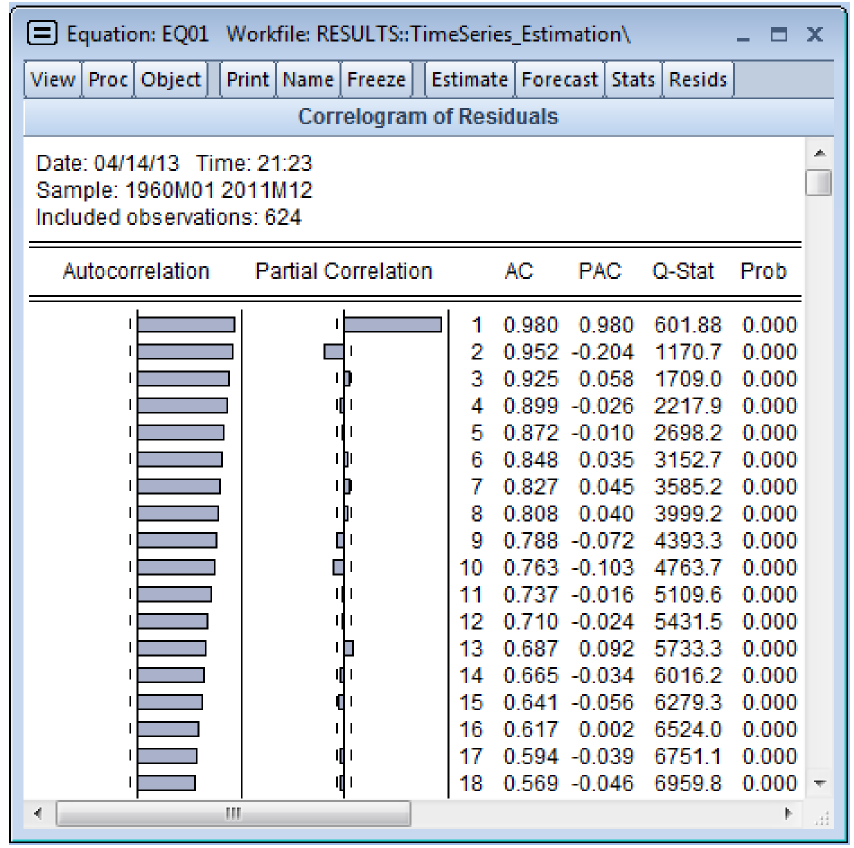
- There is strong evidence that residuals and their lagged values are positively correlated.

Detecting Serial Correlation: Visual Inspection (Part III)

- Another visual approach is to look at the ***Correlogram*** which shows the empirical pattern of correlation between residuals and their own past values.

Correlogram of residuals:

1. On the ***Equation*** box, click ***View*** → ***Residual Diagnostics*** → ***Correlogram-Q-Statistic***.
2. The ***Lag Specification*** box opens up. Select the number of lags.



	AC	PAC	Q-Stat	Prob
1	0.980	0.980	601.88	0.000
2	0.952	-0.204	1170.7	0.000
3	0.925	0.058	1709.0	0.000
4	0.899	-0.026	2217.9	0.000
5	0.872	-0.010	2698.2	0.000
6	0.848	0.035	3152.7	0.000
7	0.827	0.045	3585.2	0.000
8	0.808	0.040	3999.2	0.000
9	0.788	-0.072	4393.3	0.000
10	0.763	-0.103	4763.7	0.000
11	0.737	-0.016	5109.6	0.000
12	0.710	-0.024	5431.5	0.000
13	0.687	0.092	5733.3	0.000
14	0.665	-0.034	6016.2	0.000
15	0.641	-0.056	6279.3	0.000
16	0.617	0.002	6524.0	0.000
17	0.594	-0.039	6751.1	0.000
18	0.569	-0.046	6959.8	0.000

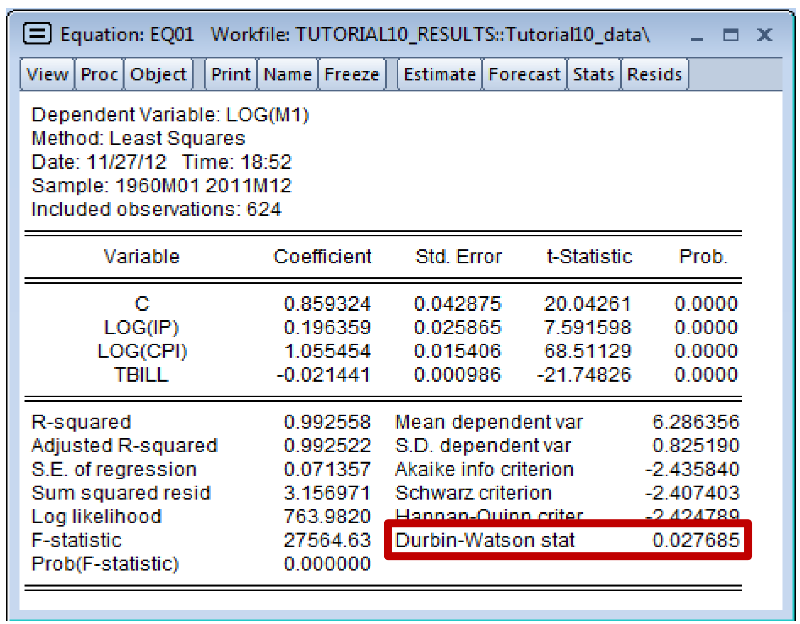
- The correlogram is shown here.
- If there is no serial correlation the AC and PAC at all lags should be near zero and all Q-statistics should be insignificant.
- Clearly, this is not the case here: the correlogram shows substantial and persistent autocorrelation in residuals.

Testing Serial Correlation

- Visual checks provide important information, but we may want to carry out formal tests for serial correlation.
- EViews provides three test statistics:
    1. **Durbin-Watson**
    2. **Breusch-Godfrey**
    3. **Ljung-Box Q-Statistic** (see the previous slide)

Testing Serial Correlation: Durbin-Watson Statistic

- EViews automatically computes the *DW statistic* and includes it in every equation object.
  - ✓ To test the hypothesis of no serial correlation, compare the reported *DW statistic* to a table of critical values. Note that EViews does not compute *p*-values for the DW statistic.
  - ✓ In this case, the *DW=0.02768*; it means we reject the null of no serial correlation.


Equation: EQ01 Workfile: TUTORIAL10_RESULTS::Tutorial10_data\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(M1)
Method: Least Squares
Date: 11/27/12 Time: 18:52
Sample: 1960M01 2011M12
Included observations: 624

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.859324	0.042875	20.04261	0.0000
LOG(IP)	0.196359	0.025865	7.591598	0.0000
LOG(CPI)	1.055454	0.015406	68.51129	0.0000
TBILL	-0.021441	0.000986	-21.74826	0.0000

R-squared	0.992558	Mean dependent var	6.286356
Adjusted R-squared	0.992522	S.D. dependent var	0.825190
S.E. of regression	0.071357	Akaike info criterion	-2.435840
Sum squared resid	3.156971	Schwarz criterion	-2.407403
Log likelihood	763.9820	Hannan-Quinn criter.	-2.424789
F-statistic	27564.63	Durbin-Watson stat	0.027685
Prob(F-statistic)	0.000000		

Testing Serial Correlation: Breusch-Godfrey Test

- A more general test for serial correlation is the Breusch-Godfrey test.

Breusch- Godfrey Test:

1. On the ***Equation*** box menu, click ***View→Residual Diagnostics→ Serial Correlation LM test***.
2. The ***Lag Specification*** box opens up. Here you need to specify the highest order of serial correlation you would like to test. If testing for first order serial correlation, specify ***lags=1***.

   The null hypothesis is that there is no serial correlation in the residuals up to the specified order.

- The top panel reports the test statistics in two versions: the F-statistic and the Chi-squared statistics. The associated *p-values* are also shown next to each statistic.
- The bottom panel provides additional information of the auxiliary regression that is carried out to create the test statistic.
- The null hypothesis of no serial correlation is easily rejected, corroborating our previous findings.

Equation: EQ01 Workfile: TUTORIAL10_RESULTS::Tutorial10_data\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	22044.02	Prob. F(1,619)	0.0000
Obs*R-squared	606.9566	Prob. Chi-Square(1)	0.0000

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 11/27/12 Time: 19:28
Sample: 1960M01 2011M12
Included observations: 624
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.009233	0.007092	1.301957	0.1934
LOG(IP)	-0.005533	0.004278	-1.293263	0.1964
LOG(CPI)	0.003444	0.002548	1.351658	0.1770
TBILL	-0.000429	0.000163	-2.629070	0.0088
RESID(-1)	0.992783	0.006687	148.4723	0.0000

R-squared	0.972687	Mean dependent var	-1.26E-15
Adjusted R-squared	0.972510	S.D. dependent var	0.071185
S.E. of regression	0.011803	Akaike info criterion	-6.033019
Sum squared resid	0.086227	Schwarz criterion	-5.997473
Log likelihood	1887.302	Hannan-Quinn criter.	-6.019206
F-statistic	5511.005	Durbin-Watson stat	1.443291
Prob(F-statistic)	0.000000		

EViews: Introductory User Guide

TIME SERIES ESTIMATION: CORRECTING SERIAL CORR

Correcting Serial Correlation

- If you detect serial correlation, a specific action needs to be done.
- Serial correlation in the error term may be evidence of a serious problem of model misspecification.
- If your goal is to estimate a model with complete dynamics, you need to respecify the model.
- If you do not wish to estimate a fully dynamic model, but would like to carry out statistical inference, then you need to account for serial correlation so that test statistics are valid.
- EViews has built-in features to correct for either autoregressive (*AR(p))* or moving average (*MA(q)*) errors, or both.

Correcting Serial Correlation: AR Example (Part I)

- Let's illustrate with a new model how to:
  - ✓ Check for serial correlation.
  - ✓ Correct for serial correlation.

<u>Assume the following model:</u>

$$TBILL_t = \beta_0 + \beta_1 \log(M1_t) + \beta_2 \log(CPI_t) + \beta_3 IP_t + \beta_4 t + \varepsilon_t, t = 1, \dots, T, \varepsilon_t \sim WN(0, \sigma^2).$$

- There is plenty of evidence that errors are serially correlated based on:
  - ✓ *DW*-statistic (which is near 0 here)
  - ✓ *Residual* plot (next slide)
  - ✓ *Breusch-Godfrey* test (next slide).

Equation: EQ13 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

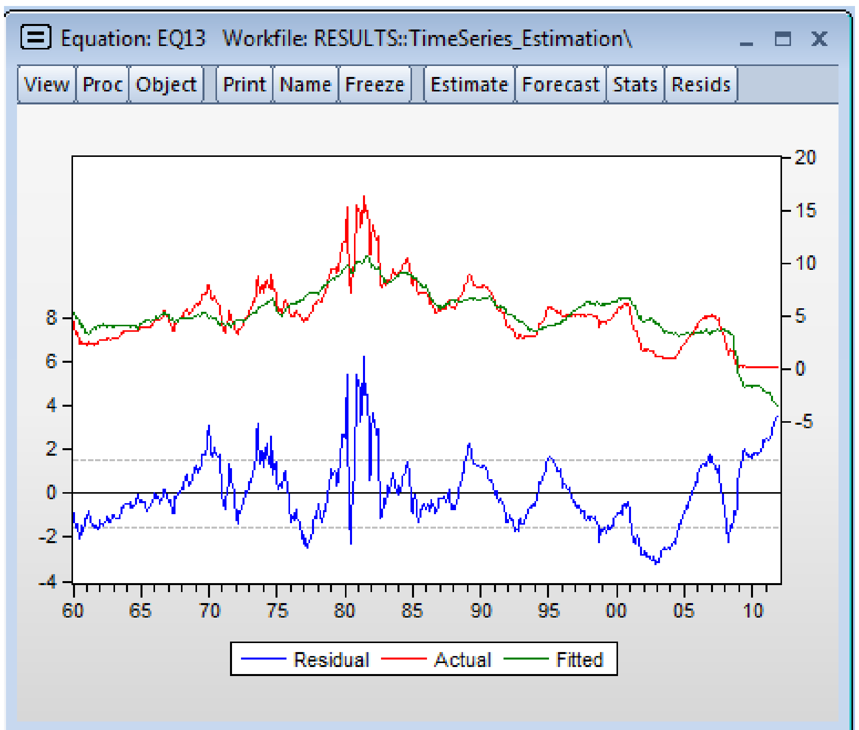
Dependent Variable: TBILL
Method: Least Squares
Date: 12/02/12 Time: 23:35
Sample: 1960M01 2011M12
Included observations: 624

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-59.43675	3.474332	-17.10739	0.0000
LOG(M1)	-10.24916	0.765226	-13.39365	0.0000
LOG(CPI)	20.57633	0.760181	27.06767	0.0000
LOG(IP)	14.37032	0.746583	19.24812	0.0000
@TREND	-0.068292	0.002726	-25.05265	0.0000

R-squared	0.729276	Mean dependent var	5.127516
Adjusted R-squared	0.727527	S.D. dependent var	2.957858
S.E. of regression	1.543971	Akaike info criterion	3.714573
Sum squared resid	1475.602	Schwarz criterion	3.750119
Log likelihood	-1153.947	Hannan-Quinn criter.	3.728386
F-statistic	416.8657	Durbin-Watson stat	0.082899
Prob(F-statistic)	0.000000		

Correcting Serial Correlation: AR Example (Part II)

- **Residual Plot**
  1. Click **View** on ***Equation*** box.
  2. Select ***Actual, Fitted, Residual→ Actual, Fitted Residual Graph***.



- **Breusch-Godfrey Test**
  1. Click **View** on ***Equation*** box.
  2. Select ***Residual Diagnostic→Serial Correlation LM Test*** (select 2 lags).

Equation: EQ13 Workfile: RESULTS::TimeSeries_Estimation\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	3789.435	Prob. F(2,617)	0.0000
Obs*R-squared	577.0241	Prob. Chi-Square(2)	0.0000

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 04/05/13 Time: 17:10
Sample: 1960M01 2011M12
Included observations: 624
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.526637	0.955624	-0.551093	0.5818
LOG(M1)	0.339897	0.211124	1.609935	0.1079
LOG(CPI)	-0.389158	0.210080	-1.852433	0.0644
LOG(IP)	0.046833	0.205278	0.228145	0.8196
@TREND	-0.000139	0.000749	-0.185794	0.8527
RESID(-1)	1.220008	0.038747	31.48633	0.0000
RESID(-2)	-0.269668	0.038962	-6.921374	0.0000

R-squared	0.924718	Mean dependent var	-4.30E-14
Adjusted R-squared	0.923986	S.D. dependent var	1.539007
S.E. of regression	0.424314	Akaike info criterion	1.134467
Sum squared resid	111.0860	Schwarz criterion	1.184232
Log likelihood	-346.9537	Hannan-Quinn criter.	1.153805
F-statistic	1263.145	Durbin-Watson stat	1.876415
Prob(F-statistic)	0.000000		

Correcting Serial Correlation: AR Example (Part III)

- Now let's correct for serial correlation in this model. Suppose you suspect the error term in the previous model follow an AR(1) process.

Correcting example 1: *AR(1)* Model

$$TBILL_t = \beta_0 + \beta_1 \log(M1_t) + \beta_2 \log(CPI_t) + \beta_3 IP_t + \beta_4 t + \varepsilon_t, t = 2, \ldots, T, \varepsilon_t = \varphi_1 \varepsilon_{t-1} + u_t, u_t \sim WN(0, \sigma^2).$$

- ❑ The coefficient estimate for **AR(1)** is shown in the middle panel. This is the serial correlation coefficient and in our case it is large and statistically significant.
- ❑ "***Inverted AR roots***" are shown at the bottom of the equation box. Stationarity requires that inverted roots lie inside the unit circle. There is no particular issue if roots are imaginary, but EViews issues a warning if the process is non-stationary.
- ❑ Notice that the number of observations has declined by 1; EViews adjusts the sample to free up the pre-sample observation needed for estimation of an AR model.
- ❑ The summary statistics at the bottom of the table are now based on the one-period-ahead forecast errors (which includes the information from lagged residuals) and not on the unconditional residuals.
- ❑ Versions of EViews prior to EViews 9 used constrained least squares as an estimation method for ARMA models. EViews 9 introduced Maximum Likelihood (ML) and Generalised Least Squares (GLS) estimation of ARMA models. You can change the ARMA estimation method (*Options* tab of the est. dialog).

Equation: EQ14 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: TBILL
Method: ARMA Maximum Likelihood (BFGS)
Date: 07/29/15 Time: 11:45
Sample: 1960M01 2011M12
Included observations: 624
Convergence achieved after 6 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-77.71905	17.99783	-4.318245	0.0000
LOG(M1)	-1.659585	2.650218	-0.626207	0.5314
LOG(CPI)	13.70239	4.115248	3.329662	0.0009
LOG(IP)	14.04829	1.860123	7.552343	0.0000
@TREND	-0.079405	0.018511	-4.289689	0.0000
AR(1)	0.966300	0.012059	80.13358	0.0000
SIGMASQ	0.188071	0.004441	42.35236	0.0000

R-squared	0.978469	Mean dependent var	5.127516
Adjusted R-squared	0.978260	S.D. dependent var	2.957858
S.E. of regression	0.436125	Akaike info criterion	1.193729
Sum squared resid	117.3566	Schwarz criterion	1.243493
Log likelihood	-365.4434	Hannan-Quinn criter.	1.213067
F-statistic	4673.219	Durbin-Watson stat	1.468547
Prob(F-statistic)	0.000000		

Inverted AR Roots .97

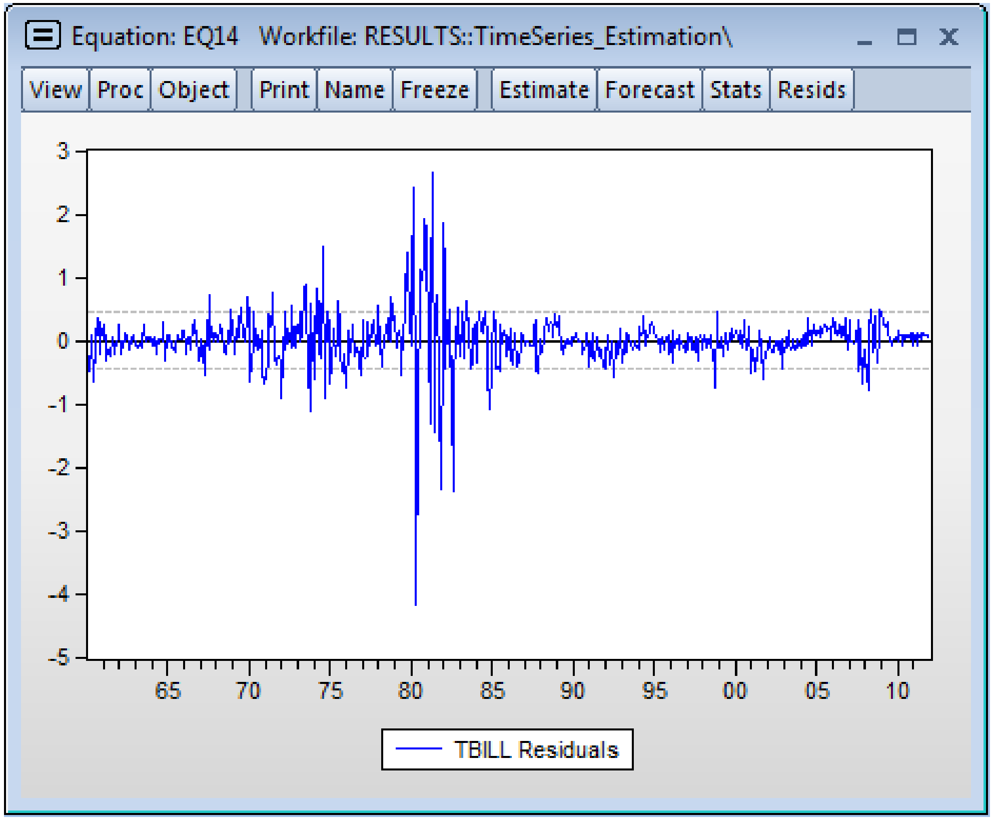
Correcting Serial Correlation: AR Example (Part IV)

- You can also check for serial correlation now by inspecting residuals and carrying out the Breusch-Godfrey Test.

Residual Plot:

1. On the ***Equation*** box menu, click on **View**.
2. Select ***Actual, Fitted, Residual*** → ***Residual Graph***.

❑ As you can see, residuals behave better now compared to the original model, though there are still some concerns regarding serial correlation.



Correcting Serial Correlation: AR Example (Part V)

Breusch-Godfrey Test:

1. Re-estimate the equation using CLS ARMA estimation method. (***Options tab***)
2. On the equation toolbar, click ***View→Residual Diagnostics→ Serial Correlation LM test***.
3. The ***Lag Specification*** box opens up. Specify ***lags=2***. Click **OK.**

❑ As you can see from the test results, we fail to reject the presence of serial correlation after including an ***AR(1)*** term. This means the ***AR(1)*** model is not a suitable specification (it does not fully address serial correlation).

Equation: EQ14 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	42.28100	Prob. F(2,615)	0.0000
Obs*R-squared	75.30730	Prob. Chi-Square(2)	0.0000

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 04/05/13 Time: 19:15
Sample: 1960M02 2011M12
Included observations: 623
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.659192	14.21049	0.046388	0.9630
LOG(M1)	0.656229	2.285937	0.287072	0.7742
LOG(CPI)	-0.056669	3.238314	-0.017500	0.9860
LOG(IP)	-1.090581	2.114127	-0.515854	0.6061
@TREND	-0.000515	0.013497	-0.038163	0.9696
AR(1)	-0.003684	0.011136	-0.330822	0.7409
RESID(-1)	0.332979	0.040428	8.236314	0.0000
RESID(-2)	-0.224751	0.041039	-5.476527	0.0000

R-squared	0.120878	Mean dependent var	1.45E-13
Adjusted R-squared	0.110872	S.D. dependent var	0.434256
S.E. of regression	0.409476	Akaike info criterion	1.064879
Sum squared resid	103.1172	Schwarz criterion	1.121823
Log likelihood	-323.7098	Hannan-Quinn criter.	1.087009
F-statistic	12.08028	Durbin-Watson stat	1.981309
Prob(F-statistic)	0.000000		

Correcting Serial Correlation: AR Example (Part VI)

- EViews allows you to estimate higher order (*AR(p)*) models just as easily.
- This should help you address issues of higher-order serial correlation.

Correcting example 1: *AR(2)* Model

$$TBILL_t = \beta_0 + \beta_1 \log(M1_t) + \beta_2 \log(CPI_t) + \beta_3 IP_t + \beta_4 t + \varepsilon_t, t = 3, \ldots, T,$$
$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + u_t, u_t \sim WN(0, \sigma^2).$$

- Notice that now both *AR(1)* and *AR(2)* terms are estimated. They are both statistically significant. The adequacy of this model can be verified as before.
- EViews allows you to include non-contiguous ***AR*** terms.
- The downside to this is that if you are estimating a higher-order AR process, EViews requires you to include all lower-order terms. For example, to estimate an *AR(3)* model, you need to include: ***ar(1) ar(2) a(3)***. If you simply type ***ar(3)*** and omit other terms, this forces the estimate of ***ar(1)*** and ***ar(2)*** to zero. You may want this on rare occasions (for example, when dealing with seasonal components), but not on a routine basis.

Equation: EQ15 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: TBILL
Method: ARMA Maximum Likelihood (BFGS)
Date: 07/29/15 Time: 11:51
Sample: 1960M01 2011M12
Included observations: 624
Convergence achieved after 6 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-58.13466	16.61489	-3.498950	0.0005
LOG(M1)	-2.710568	2.648626	-1.023386	0.3065
LOG(CPI)	13.74863	4.244990	3.238791	0.0013
LOG(IP)	9.676880	2.098578	4.611159	0.0000
@TREND	-0.065524	0.017272	-3.793582	0.0002
AR(1)	1.252827	0.018392	68.11769	0.0000
AR(2)	-0.294976	0.016812	-17.54522	0.0000
SIGMASQ	0.172832	0.004475	38.62024	0.0000

R-squared	0.980214	Mean dependent var	5.127516
Adjusted R-squared	0.979989	S.D. dependent var	2.957858
S.E. of regression	0.418422	Akaike info criterion	1.112781
Sum squared resid	107.8473	Schwarz criterion	1.169654
Log likelihood	-339.1875	Hannan-Quinn criter.	1.134881
F-statistic	4359.498	Durbin-Watson stat	1.868507
Prob(F-statistic)	0.000000		

Inverted AR Roots	.94	.31

Correcting Serial Correlation: MA Example

- You can correct for serial correlation, when errors follow an MA process.

<u>Correcting example 1: *MA(3)* Model</u>

$$TBILL_t = \beta_0 + \beta_1 \log(M1_t) + \beta_2 \log(CPI_t) + \beta_3 IP_t + \beta_4 t + \varepsilon_t, t = 4, \dots, T,$$
$$\varepsilon_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}, + \theta_3 u_{t-3}, u_t \sim WN(0, \sigma^2).$$

Equation: EQ19 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: TBILL
Method: ARMA Maximum Likelihood (BFGS)
Date: 07/29/15 Time: 11:55
Sample: 1960M01 2011M12
Included observations: 624
Convergence achieved after 10 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-54.99223	6.419515	-8.566415	0.0000
LOG(M1)	-8.913111	1.181876	-7.541494	0.0000
LOG(CPI)	19.38389	0.852676	22.73300	0.0000
LOG(IP)	12.25341	1.110789	11.03127	0.0000
@TREND	-0.065006	0.005198	-12.50512	0.0000
MA(1)	1.494225	0.017448	85.63869	0.0000
MA(2)	1.245272	0.028195	44.16674	0.0000
MA(3)	0.586943	0.023470	25.00781	0.0000
SIGMASQ	0.272124	0.011602	23.45436	0.0000

R-squared	0.968846	Mean dependent var	5.127516
Adjusted R-squared	0.968441	S.D. dependent var	2.957858
S.E. of regression	0.525458	Akaike info criterion	1.569772
Sum squared resid	169.8053	Schwarz criterion	1.633755
Log likelihood	-480.7690	Hannan-Quinn criter.	1.594636
F-statistic	2390.734	Durbin-Watson stat	1.588388
Prob(F-statistic)	0.000000		

Inverted MA Roots	-.33+.77i	-.33-.77i	-.84

- ❑ As it must be obvious by now, if your errors follow an MA(3) process, you need to include both ***ma(1)***, ***ma(2)*** and ***ma(3)*** terms in the regression.
- ❑ Unlike nearly all other EViews estimation procedures, ***MA*** models require a continuous sample. If your sample includes a break or has missing data (***NA*** values), EViews will give an error message.
- ❑ Notice that in general, MA models are notoriously difficult to estimate. In particular, higher order MA terms should be avoided unless absolutely required for your model.

Correcting Serial Correlation: ARMA Example (Part I)

- You can just as easily specify higher order ARMA(p,q) models.

Correcting example 1: *ARMA(2,1)* Model

$$TBILL_t = \beta_0 + \beta_1 \log(M1_t) + \beta_2 \log(CPI_t) + \beta_3 IP_t + \beta_4 t + \varepsilon_t, t = 4, \dots, T,$$
$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}, +\theta_3 u_{t-3}, u_t \sim WN(0, \sigma^2).$$

Equation: EQ20 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: TBILL
Method: ARMA Maximum Likelihood (BFGS)
Date: 07/29/15 Time: 11:56
Sample: 1960M01 2011M12
Included observations: 624
Convergence achieved after 13 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-56.92676	17.34743	-3.281567	0.0011
LOG(M1)	-2.435588	2.563601	-0.950065	0.3425
LOG(CPI)	12.80266	4.398146	2.910921	0.0037
LOG(IP)	9.844700	2.029720	4.850275	0.0000
@TREND	-0.063479	0.018378	-3.454037	0.0006
AR(1)	0.713765	0.043304	16.48259	0.0000
AR(2)	0.227503	0.044395	5.124579	0.0000
MA(1)	0.618351	0.042066	14.69954	0.0000
SIGMASQ	0.164751	0.005299	31.09019	0.0000

R-squared	0.981139	Mean dependent var	5.127516
Adjusted R-squared	0.980893	S.D. dependent var	2.957858
S.E. of regression	0.408855	Akaike info criterion	1.068400
Sum squared resid	102.8048	Schwarz criterion	1.132383
Log likelihood	-324.3407	Hannan-Quinn criter.	1.093263
F-statistic	3998.940	Durbin-Watson stat	1.994871
Prob(F-statistic)	0.000000		

Inverted AR Roots	.95	-.24
Inverted MA Roots	-.62	

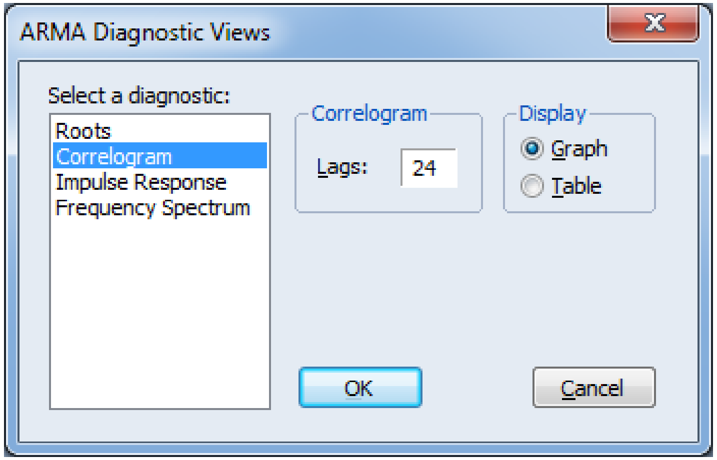
Correcting Serial Correlation: ARMA Example (Part II) - Correlogram

- EViews provides access to several diagnostic views to help you assess the ARMA terms. One of the most useful tools is the ***correlogram***.

ARMA Correlogram:

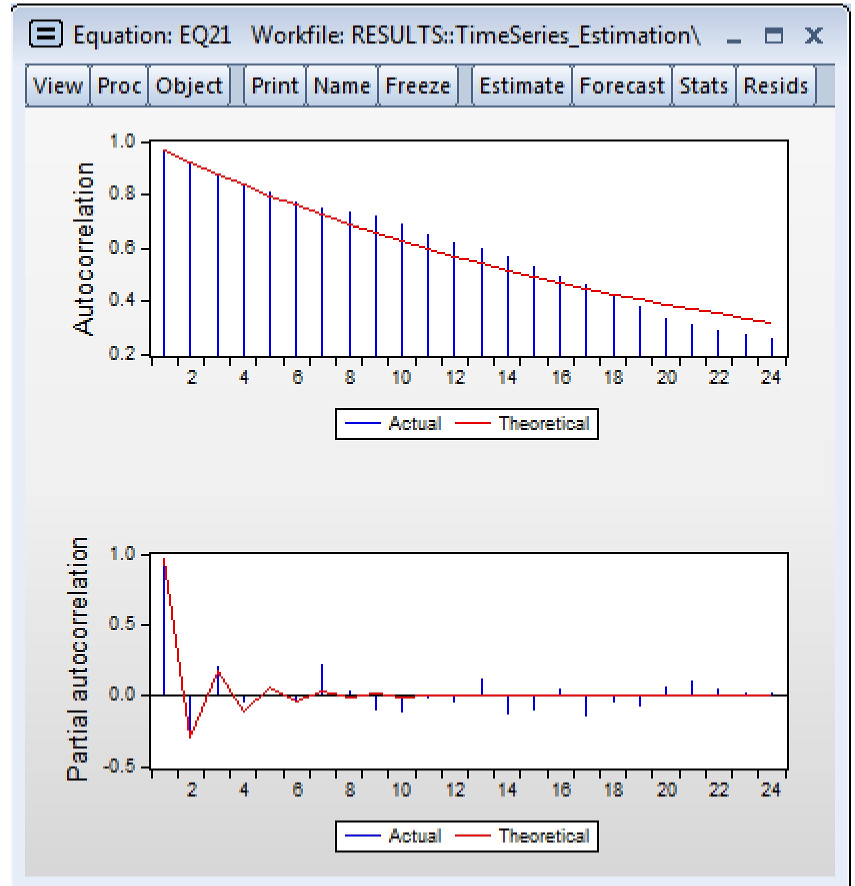
1. On the top menu of the ***Equation*** box, click ***View*** → ***ARMA Structure***.
2. The ***ARMA Diagnostic Views*** dialog box opens up. Select ***Correlogram***, the number of lags (24 here) and click **Graph** (if you want to see a graph).

Note: For other diagnostic tools, see User Guide.



Correcting Serial Correlation: ARMA Example (Part III) - Correlogram

- The graph shows autocorrelations (ACF) and partial correlations (PACF) for:
  - Theoretical correlogram (red line) corresponding to ARMA terms.
  - Empirical correlogram of residuals (blue spikes) corresponding to original residuals with no ARMA terms.
- If the model is properly specified, the blue spikes and red line should be "close".
- Note that if the ARMA model is non-stationary, EViews shows only the sample structural residual autocorrelation patterns.



Differencing and Serial Correlation (Part I)

- An alternative way to deal with serial correlation is to difference the data.
- In fact, differencing the data (e.g., taking first-order differences) addresses a number of issues that arise in time series data:
  - ✓ It eliminates most (perhaps not all) serial correlation
  - ✓ It de-trends the data
  - ✓ It transforms an I(1) process to an I(0).

Differencing and Serial Correlation (Part II)

- Let us first estimate the following model in levels:

$$TBILL_t = \beta_0 + \beta_1 \log(IP_t) + \beta_2 TBILL_{t-1} + \beta_4 t + \varepsilon_t, t = 2, \ldots, T, \varepsilon_t \sim WN(0, \sigma^2).$$

❑ There is evidence that serial correlation is present in this model. *DW-statistic* is low, and the Breusch-Godfrey test (not shown here) detects the presence of serial correlation.

Equation: EQ22 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: TBILL
Method: Least Squares
Date: 12/03/12 Time: 00:22
Sample (adjusted): 1960M02 2011M12
Included observations: 623 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.792855	0.797777	-3.500795	0.0005
LOG(IP)	0.899720	0.247337	3.637627	0.0003
TBILL(-1)	0.974549	0.007347	132.6380	0.0000
@TREND	-0.002233	0.000582	-3.836297	0.0001

R-squared	0.977187	Mean dependent var	5.128764
Adjusted R-squared	0.977077	S.D. dependent var	2.960070
S.E. of regression	0.448168	Akaike info criterion	1.239103
Sum squared resid	124.3291	Schwarz criterion	1.267576
Log likelihood	-381.9807	Hannan-Quinn criter.	1.250168
F-statistic	8838.302	Durbin-Watson stat	1.336255
Prob(F-statistic)	0.000000		

Differencing and Serial Correlation (Part III)

- Now let us estimate the same model in first differences.

  [Note that EViews allows to difference the data very easily by typing ***d()*** or ***dlog()*** before the name of the variable.]
- Let us estimate the following model:

$$\Delta TBILL_t = \beta_0 + \beta_1 \Delta \log(IP_t) + \beta_2 \Delta TBILL_{t-1} + \beta_4 t + \varepsilon_t, t = 3, \ldots, T, \varepsilon_t \sim WN(0, \sigma^2).$$

Equation: EQ23 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: D(TBILL)
Method: Least Squares
Date: 04/05/13 Time: 22:20
Sample (adjusted): 1960M03 2011M12
Included observations: 622 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.018652	0.034911	-0.534280	0.5933
DLOG(IP)	9.958048	2.257113	4.411852	0.0000
D(TBILL(-1))	0.295673	0.038230	7.733967	0.0000
@TREND	-2.52E-05	9.48E-05	-0.265719	0.7905

R-squared	0.138837	Mean dependent var	-0.006350
Adjusted R-squared	0.134656	S.D. dependent var	0.453535
S.E. of regression	0.421895	Akaike info criterion	1.118292
Sum squared resid	110.0014	Schwarz criterion	1.146799
Log likelihood	-343.7887	Hannan-Quinn criter.	1.129371
F-statistic	33.21131	Durbin-Watson stat	1.889591
Prob(F-statistic)	0.000000		

- The *DW-statistic* is now a lot closer to 2, suggesting that we have eliminated some of the serial correlation in the error term (not all disappears; BG test shows errors are serially correlated, but the problem is less severe now).
- The time trend is now not significant: taking first-differences has de-trended the data.
- The R-squared value is much lower now reflecting the fact that it is harder to fit differenced data.

EViews: Introductory User Guide

TIME SERIES ESTIMATION: HETEROSKEDASTICITY AND AUTOCORRELATION

Heteroskedasticity and Autocorrelation in Time Series (Part I)

- Nothing rules out the possibility that both heteroskedasticity and serial correlation are present in a regression model.
  - ✓ Serial correlation has a larger impact on standard errors and efficiency of estimators than heteroskedasticity
  - ✓ However, heteroskedasticity may be of concern especially in small samples.
- In addition, in many financial time series, the conditional variance of the error term depends on past values of the error term. This is also known as autoregressive conditional heteroskedasticity (ARCH).
- In this section, we demonstrate the following:
  - ✓ Testing for heteroskedasticity in time series models
  - ✓ Testing for ARCH terms
  - ✓ HAC standard errors

Heteroskedasticity and Autocorrelation in Time Series (Part II)

- Testing for Heteroskedasticity in time series data is very similar to cross section data (see ***Part 9***):
  - ✓ The one caveat is that, when testing for heteroskedasticity, residuals should not be serially correlated.
  - ✓ Any serial correlation will generally invalidate tests for heteroskedasticity.
  - ✓ It thus makes sense to test for serial correlation first, correct for serial correlation, and then test for heteroskedasticity.
  - ✓ Most commonly, you can correct for both heteroskedasticity and autocorrelation of unknown form using the HAC Consistent Covariance (Newey-West).

Testing for Heteroskedasticity in Time Series Models (Part I)

- Suppose you want to see whether the regression shown here suffers from heteroskedasticity.

- Let us perform the White test.
  - ✓ Recall, the White test is a test of the null hypothesis of no heteroskedasticity, against heteroskedasticity of unknown, general form.

Equation: EQ24 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: TBILL
Method: Least Squares
Date: 04/11/13 Time: 19:05
Sample (adjusted): 1960M04 2011M12
Included observations: 621 after adjustments

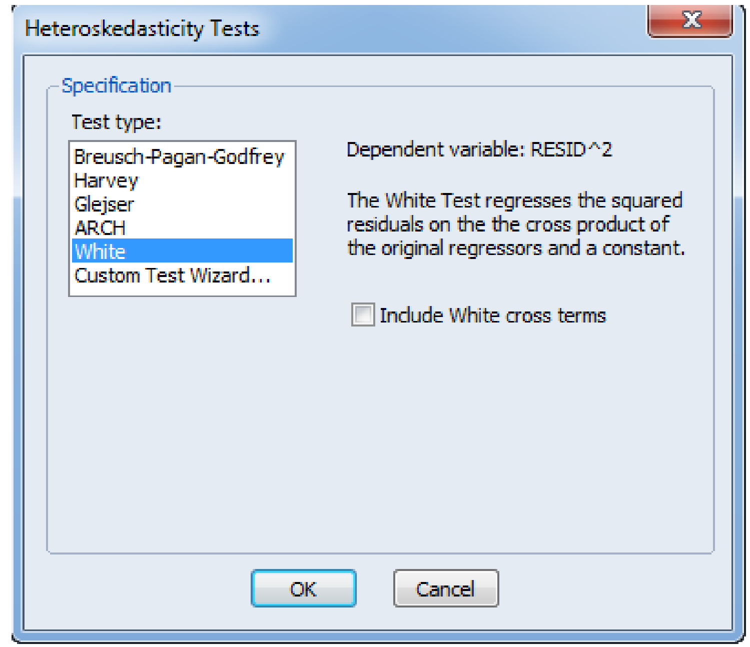
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.358020	0.742286	-3.176699	0.0016
LOG(IP)	0.767407	0.229862	3.338551	0.0009
@TREND	-0.001928	0.000540	-3.569973	0.0004
TBILL(-1)	1.373592	0.039437	34.82998	0.0000
TBILL(-2)	-0.608719	0.063402	-9.600992	0.0000
TBILL(-3)	0.210453	0.039091	5.383677	0.0000

R-squared	0.980699	Mean dependent var	5.133575
Adjusted R-squared	0.980542	S.D. dependent var	2.963565
S.E. of regression	0.413389	Akaike info criterion	1.080757
Sum squared resid	105.0974	Schwarz criterion	1.123572
Log likelihood	-329.5751	Hannan-Quinn criter.	1.097398
F-statistic	6249.850	Durbin-Watson stat	1.985804
Prob(F-statistic)	0.000000		

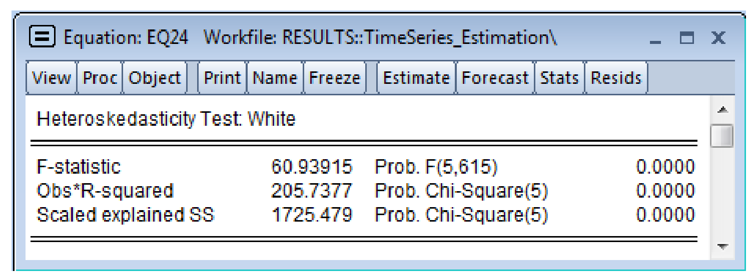
Testing for Heteroskedasticity in Time Series Models (Part II)

White Test:

1. Open an equation. On the top menu of the ***Equation*** box, select ***View*** → ***Residual Diagnostics*** → ***Heteroskedasticity Tests***.
2. The **Heteroskedasticity Tests** window opens up. Select **White** under the drop-down menu.
3. You may chose to include or exclude the cross terms. If you do not wish to include the cross term, uncheck the box "*Include White cross terms*" (as we do here). The test will simply be carried out with only the squared terms. Click ***OK***.

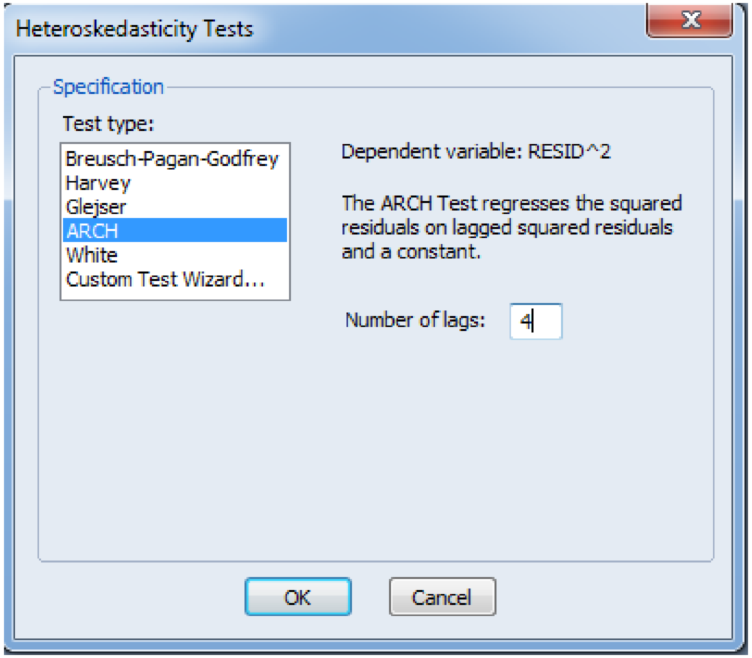


- Based on the test statistics, we reject the null of homoskedascity, which means that the error term is heteroskedastic and standard errors should be adjusted.



Testing for ARCH Terms (Part I)

- It is also possible that the previous regression has ARCH terms.
- To test for this, let's perform an ARCH LM test.
  - ✓ The null hypothesis is that there is no ARCH up to order q in the residuals.



<u>**ARCH LM Test:**</u>

1. On the equation box, select ***View*** → ***Residual Diagnostics*** → ***Heteroskedasticity Tests***.
2. The **Heteroskedasticity Tests** window opens up. Select **ARCH** under the drop-down menu.
3. Select the number of lags (4 in this case). Click ***OK***.

Testing for ARCH Terms (Part II)

- The top panel shows the results of the ARCH LM test, while the bottom panel shows the auxiliary regression used to compute the test statistics.
- We reject the null of no ARCH, which means that residuals suffer from this specific form of heteroskedasticity.

Equation: EQ24 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: ARCH

F-statistic	82.84362	Prob. F(4,612)	0.0000
Obs*R-squared	216.7305	Prob. Chi-Square(4)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/11/13 Time: 19:22
Sample (adjusted): 1960M08 2011M12
Included observations: 617 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.064720	0.024071	2.688665	0.0074
RESID^2(-1)	0.589941	0.039760	14.83749	0.0000
RESID^2(-2)	0.026327	0.045741	0.575570	0.5651
RESID^2(-3)	-0.176965	0.045725	-3.870162	0.0001
RESID^2(-4)	0.177584	0.039723	4.470537	0.0000

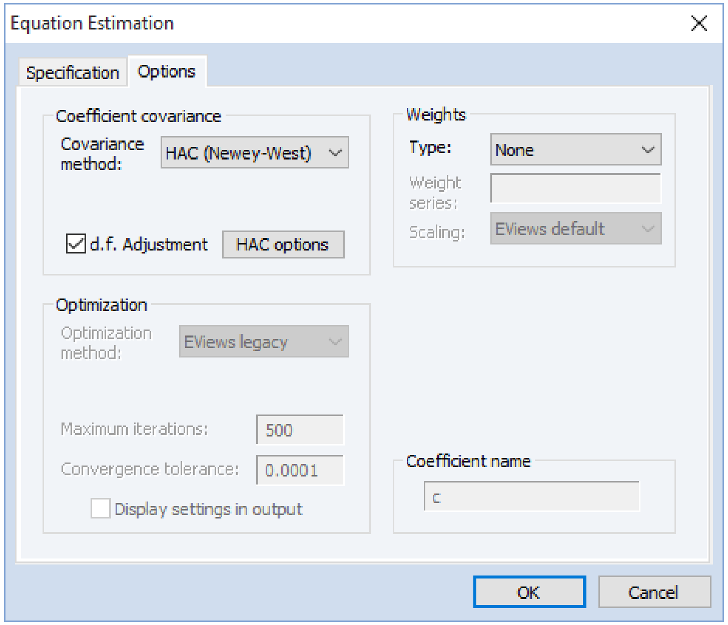
R-squared	0.351265	Mean dependent var	0.169116
Adjusted R-squared	0.347025	S.D. dependent var	0.702323
S.E. of regression	0.567526	Akaike info criterion	1.713009
Sum squared resid	197.1162	Schwarz criterion	1.748867
Log likelihood	-523.4633	Hannan-Quinn criter.	1.726950
F-statistic	82.84362	Durbin-Watson stat	2.069907
Prob(F-statistic)	0.000000		

Addressing Heteroskedasticity and Autocorrelation: Robust Std Errors (Part I)

- EViews provides built-in tools that allow you to adjust standard errors for the presence of both ***heteroskedasticity* and *autocorrelation*** of unknown form (HAC –Newey-West).

HAC (Newey-West) standard errors:

1. Click ***Estimate*** on the equation box.
2. The ***Equation Estimation*** box opens up. Click ***Options***.
3. Under the *Coefficient Covariance matrix* drop-down menu, choose **HAC (Newey-West)**. Click ***OK***.



Addressing Heteroskedasticity and Autocorrelation: Robust Std Errors (Part II)

- EViews re-estimates the equation, this time adjusting the standard errors for heteroskedasticity and autocorrelation of unknown form.
- For purpose of comparisons, we also show results with unadjusted standard errors.
- As expected, the estimated coefficient values do not change. But, the adjusted standard errors (and associated t-statistics) are different from the original regression.

Equation: EQ24A Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: TBILL
Method: Least Squares
Date: 04/11/13 Time: 19:30
Sample (adjusted): 1960M04 2011M12
Included observations: 621 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 7.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.358020	0.768225	-3.069439	0.0022
LOG(IP)	0.767407	0.249015	3.081764	0.0021
@TREND	-0.001928	0.000603	-3.194121	0.0015
TBILL(-1)	1.373592	0.115009	11.94333	0.0000
TBILL(-2)	-0.608719	0.205489	-2.962301	0.0032
TBILL(-3)	0.210453	0.112398	1.872393	0.0616

R-squared	0.980699	Mean dependent var	5.133575
Adjusted R-squared	0.980542	S.D. dependent var	2.963565
S.E. of regression	0.413389	Akaike info criterion	1.080757
Sum squared resid	105.0974	Schwarz criterion	1.123572
Log likelihood	-329.5751	Hannan-Quinn criter.	1.097398
F-statistic	6249.850	Durbin-Watson stat	1.985804
Prob(F-statistic)	0.000000	Wald F-statistic	4245.777
Prob(Wald F-statistic)	0.000000		

Equation: EQ24 Workfile: RESULTS::TimeSeries_Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: TBILL
Method: Least Squares
Date: 04/11/13 Time: 19:05
Sample (adjusted): 1960M04 2011M12
Included observations: 621 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.358020	0.742286	-3.176699	0.0016
LOG(IP)	0.767407	0.229862	3.338551	0.0009
@TREND	-0.001928	0.000540	-3.569973	0.0004
TBILL(-1)	1.373592	0.039437	34.82998	0.0000
TBILL(-2)	-0.608719	0.063402	-9.600992	0.0000
TBILL(-3)	0.210453	0.039091	5.383677	0.0000

R-squared	0.980699	Mean dependent var	5.133575
Adjusted R-squared	0.980542	S.D. dependent var	2.963565
S.E. of regression	0.413389	Akaike info criterion	1.080757
Sum squared resid	105.0974	Schwarz criterion	1.123572
Log likelihood	-329.5751	Hannan-Quinn criter.	1.097398
F-statistic	6249.850	Durbin-Watson stat	1.985804
Prob(F-statistic)	0.000000		

Radek Hendrych

Charles University

Faculty of Mathematics and Physics

Department of Probability and Mathematical Statistics

Sokolovská 83, 186 75 Prague, Czech Republic

E-mail: hendrych@karlin.mff.cuni.cz