Technical Assistance 8761 – Solomon Islands Economic Modeling for Policy Support

Workshop 4: Estimating from Local Data

David Roland-Holst
University of California, Berkeley

Jointly Organized by:
Economic Reform Unit, MOFT, Solomon Islands and the
Asian Development Bank
9-13 April 2018

Examples of Policy from National Data

- Population and Education – HIES
- Don't drink and Drive
- How to forecast the CPI

Population and Education

Solomon Islands Household Income and Expenditure Survey (HIES)

- Two rounds:
  - 2005-6
  - 2012-13
- A really good survey and national information resource.
- Can inform more effective policy design and targeting for many years.

High response rate – Quite Representative

Table 1.1 Response Rate by Province

Province	Sample Planned	Sample Responded	**Response Rate**
Choiseul	300	279	**93.0**
Western	540	436	**80.7**
Isabel	360	259	**71.9**
Central	360	332	**92.2**
Renbell	120	113	**94.2**
Guadalcanal	600	570	**95.0**
Malaita	780	687	**88.1**
Makira-Ulawa	420	401	**95.5**
Temotu	360	312	**86.7**
Honiara	480	433	**90.2**
Solomon Islands	**4,320**	**3,822**	88.5

Income vs Expenditure

Table 1.2 Ratio of Total Annual Expenditure over Total Annual Income by Province

	Province																					
	Choiseul		Western		Isabel		Central		Rennell - Bellona		Guadalcanal		Malaita		Makira-Ulawa		Temotu		Honiara town		Solomon Islands	
Ratio Exp/Income	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
0.00 to 0.25	0	0.0	40	0.3	0	0.0	39	0.9	0	0.0	12	0.1	32	0.1	64	0.8	25	0.6	138	1.4	348	0.4
0.26 to 0.50	77	1.5	295	2.2	152	3.3	88	2.1	5	0.7	318	2.2	637	2.9	190	2.5	88	2.0	589	5.9	2,438	2.8
0.51 to 0.75	271	5.4	673	4.9	288	6.2	188	4.5	39	5.9	529	3.6	604	2.7	472	6.3	222	5.2	907	9.1	4,193	4.8
0.76 to 1.00	603	11.9	1,937	14.2	614	13.3	483	11.5	63	9.3	1,445	9.9	3,155	14.3	1,348	17.9	835	19.4	1,594	16.0	12,076	13.9
1.01 to 1.25	1,558	30.8	2,744	20.1	1,765	38.3	1,000	23.8	181	26.9	3,204	21.9	4,769	21.6	2,308	30.7	1,230	28.6	1,473	14.7	20,231	23.3
1.26 to 1.50	838	16.6	1,942	14.2	828	17.9	961	22.8	153	22.8	2,313	15.8	4,536	20.5	1,067	14.2	974	22.7	1,152	11.5	14,764	17.0
1.51 to 1.75	424	8.4	1,108	8.1	341	7.4	492	11.7	62	9.2	1,784	12.2	2,697	12.2	671	8.9	451	10.5	1,027	10.3	9,057	10.4
1.76 to 2.00	369	7.3	1,008	7.4	193	4.2	287	6.8	82	12.2	1,338	9.2	1,866	8.4	308	4.1	191	4.5	627	6.3	6,270	7.2
2.00 +	915	18.1	3,903	28.6	433	9.4	672	16.0	88	13.1	3,669	25.1	3,819	17.3	1,096	14.6	285	6.6	2,478	24.8	17,358	20.0
Total	5,056	100.0	13,650	100.0	4,614	100.0	4,209	100.0	672	100.0	14,611	100.0	22,115	100.0	7,524	100.0	4,300	100.0	9,984	100.0	86,734	100.0

Snapshot of Income Status

Table 0.2: Average Annual Household and Per Capita Expenditure and Household Size by Area

Area	Average Annual Household Expenditure	Median Annual Household Expenditure	Average Household Size	Average Annual Per Capita Expenditure	Median Annual Per Capita Expenditure
Urban	69,935	52,726	6.9	10,215	7,701
Rural	23,366	17,668	6.0	4,871	2,927
Solomon Islands	30,069	20,035	6.2	4,887	3,256

Demographics
Fertility vs Migration

Table 2.1 Population Estimate by Sex, households and Province :1999-2005/6

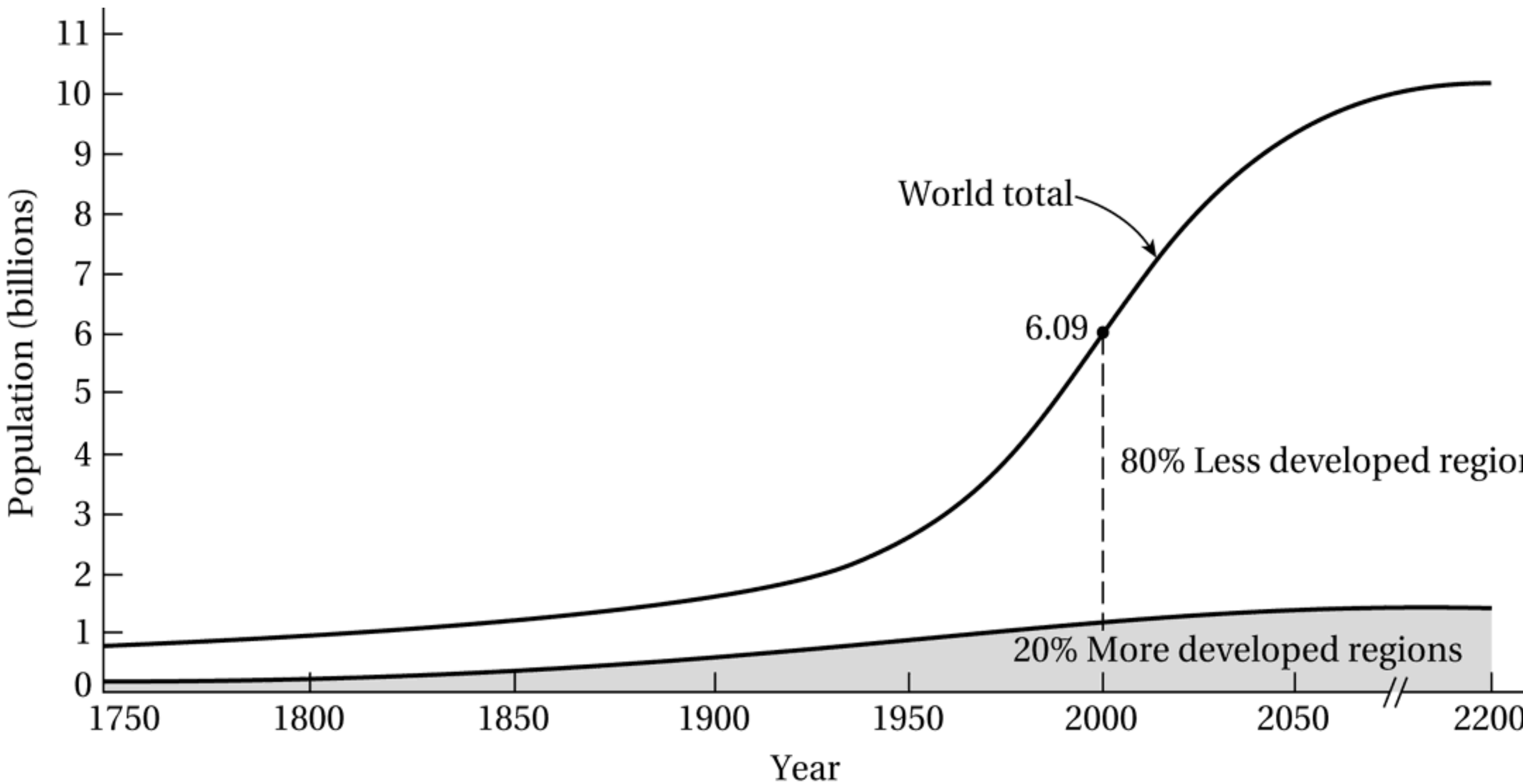
Province	2005/6					1999				
	Male	Female	Total	Private Household	Annual growth % (1999-2005)	Male	Female	Total	Private Household	Annual growth % (1986-1999)
Choiseul	16,193	15,066	31,259	5,056	7.4	10,236	9,772	20,008	3,045	3.0
Western	40,640	41,212	81,852	13,650	4.4	33,190	29,549	62,739	9,570	3.2
Isabel	11,256	12,382	23,638	4,614	2.4	10,424	9,997	20,421	3,472	2.6
Central	12,822	11,670	24,491	4,209	2.1	11,193	10,384	21,577	3,533	2.0
Rennel	2,196	2,213	4,409	672	10.3	1,230	1,147	2,377	423	2.2
Guadalcanal	44,678	39,759	84,438	14,611	5.6	31,423	28,852	60,275	10,164	1.5
Malaita	72,470	68,098	140,569	22,115	2.3	61,209	61,411	122,620	18,362	3.3
Makira	26,442	23,584	50,026	7,524	8.0	15,943	15,063	31,006	4,859	2.7
Temotu	11,736	12,064	23,800	4,300	3.8	9,146	9,766	18,912	3,335	1.9
Honiara	35,665	33,525	69,189	9,984	5.7	27,387	21,720	49,107	6,641	3.8
Grand Total	274,098	259,573	533,672	86,734	4.4	211,381	197,661	409,042	63,404	2.8

Age Distribution: Very high dependency

Table 2.3: Percentage distribution of estimated population in Age Group by Area and Sex

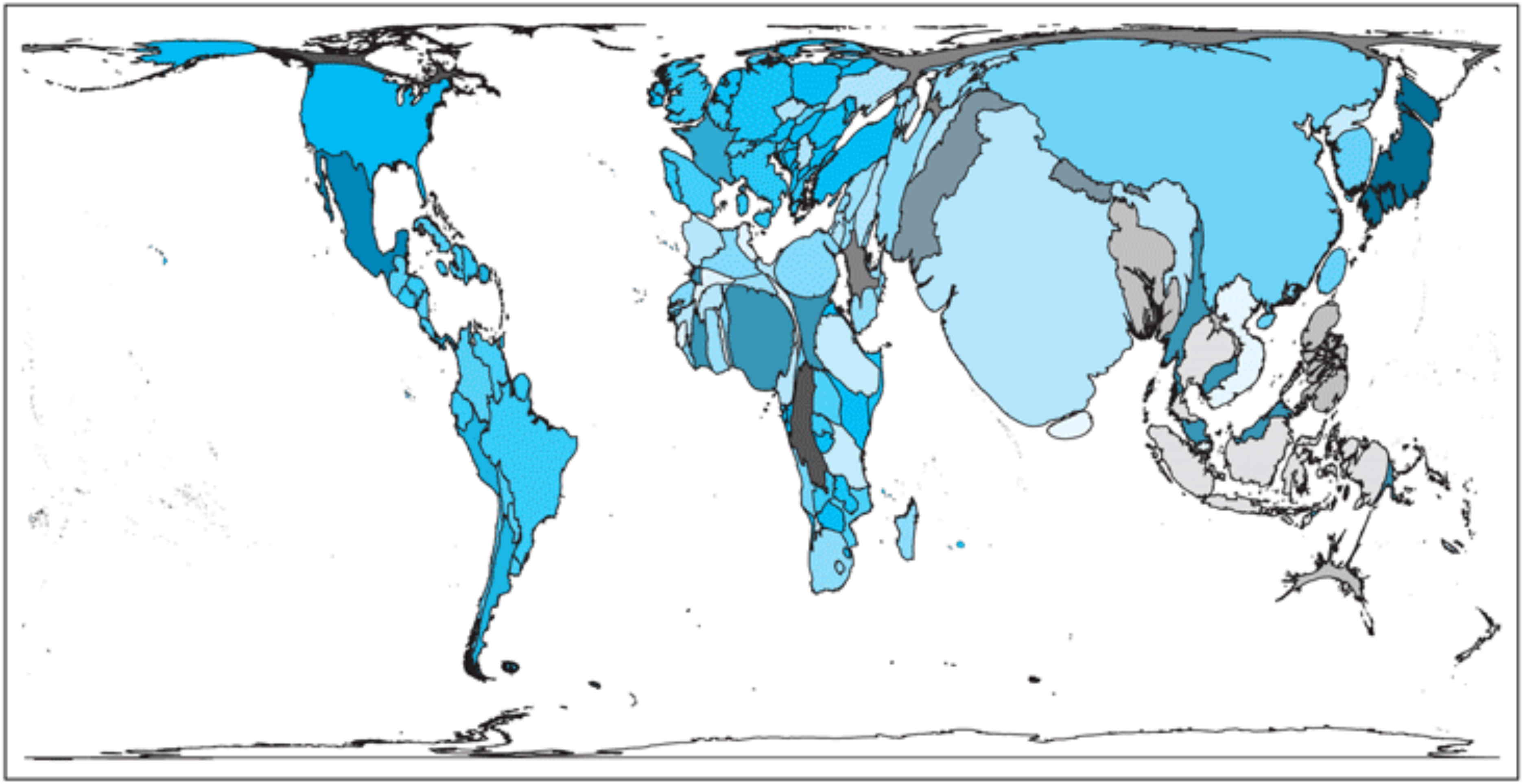
Age group	Urban			Rural			Solomon Islands		
	male	female	Total	male	female	Total	male	female	Total
<15	34.0	35.6	34.8	40.7	40.4	40.6	39.6	39.6	39.6
15-19	12.3	12.6	12.4	10.2	10.0	10.1	10.6	10.4	10.5
20-24	12.1	12.4	12.2	7.9	9.4	8.6	8.6	9.9	9.2
25-29	11.8	12.2	12.0	8.2	9.4	8.8	8.8	9.9	9.3
30-34	7.8	9.9	8.8	7.7	8.0	7.9	7.7	8.3	8.0
35-39	7.3	5.8	6.6	6.3	6.2	6.3	6.5	6.1	6.3
40-44	4.0	4.0	4.0	4.9	4.3	4.6	4.8	4.3	4.5
45-49	4.2	2.8	3.5	3.7	3.4	3.6	3.8	3.3	3.6
50-54	2.5	2.0	2.3	2.8	2.7	2.8	2.8	2.6	2.7
55-59	2.0	1.2	1.6	2.1	2.2	2.2	2.1	2.1	2.1
60-64	0.9	1.0	0.9	1.8	1.5	1.7	1.7	1.4	1.6
65+	1.0	0.6	0.8	3.0	2.0	2.5	2.7	1.7	2.2
Ns	0.1	0.0	0.1	0.5	0.3	0.4	0.5	0.3	0.4
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Population: Historical trends



(a) Absolute size

Spatial Distribution of Population

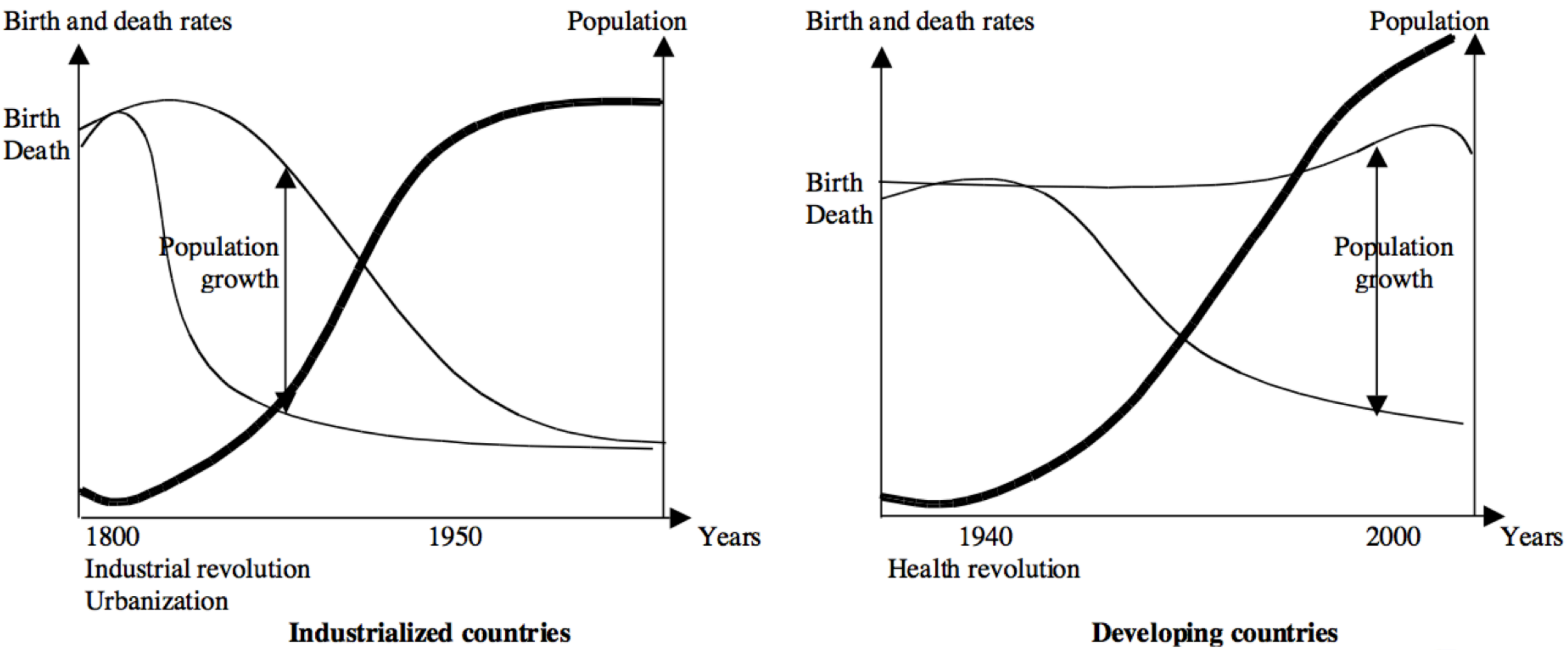


Source: worldmapper.org:http://www.worldmapper.org/display.php?selected=2).

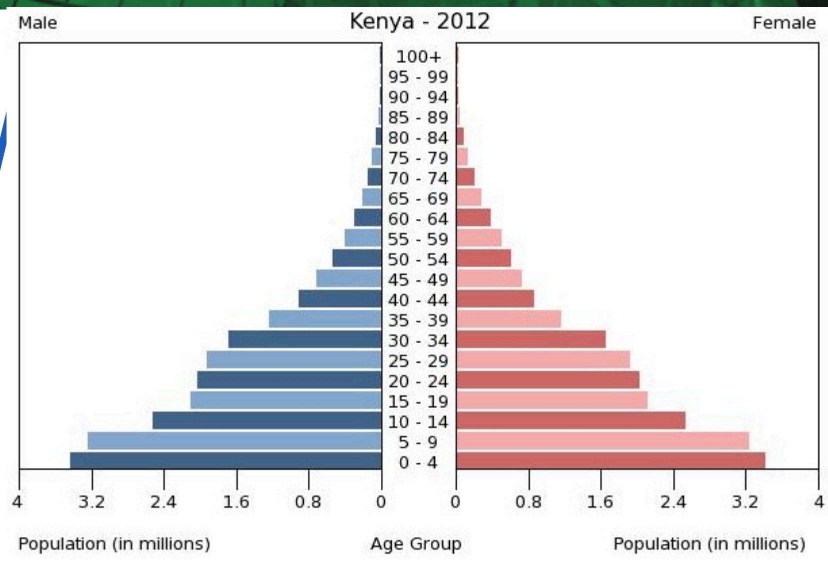
The Demographic Transition

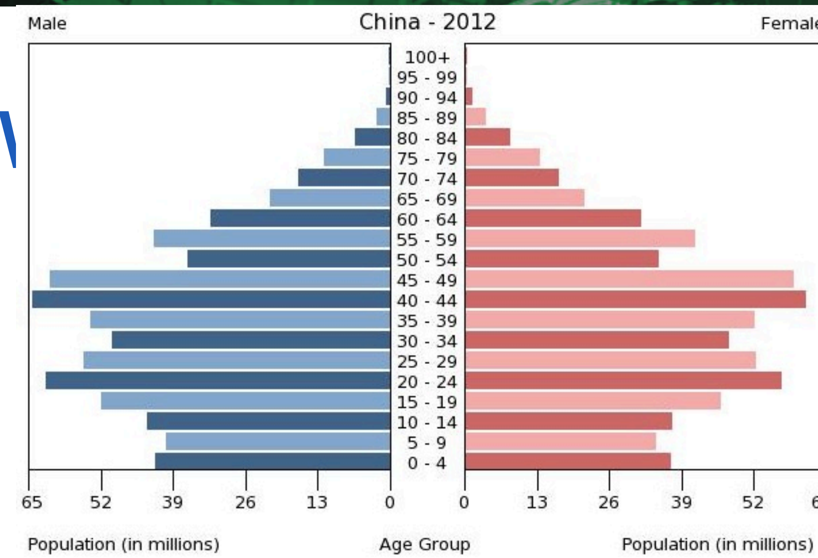
- Stage I: High birthrates and death rates
- Stage II: Continued high birthrates, declining death rates
- Stage III: Falling birthrates and death rates, eventually stabilizing

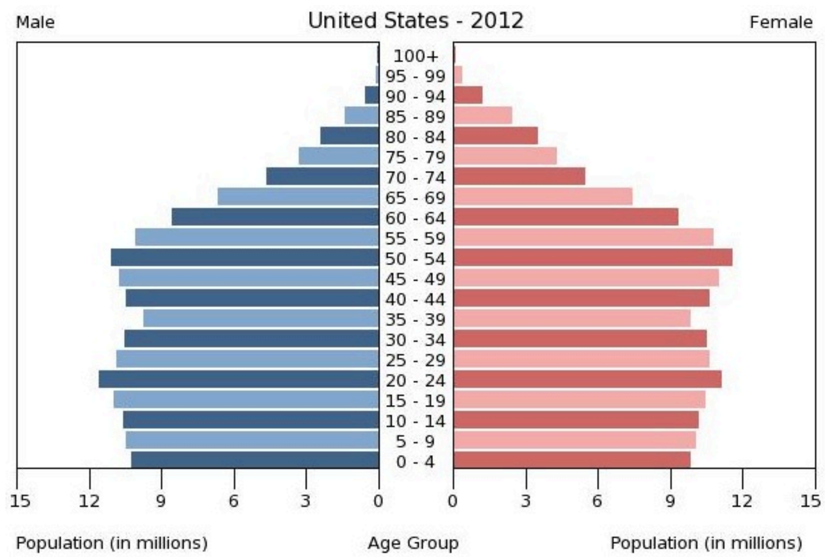
Demographic Transitions North and South

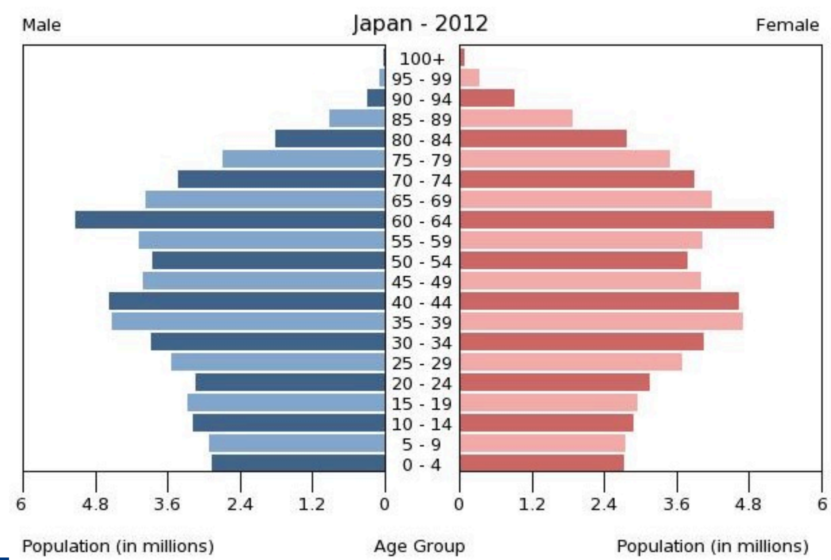


Population Structure Matters: Pyramids of Age and Gender

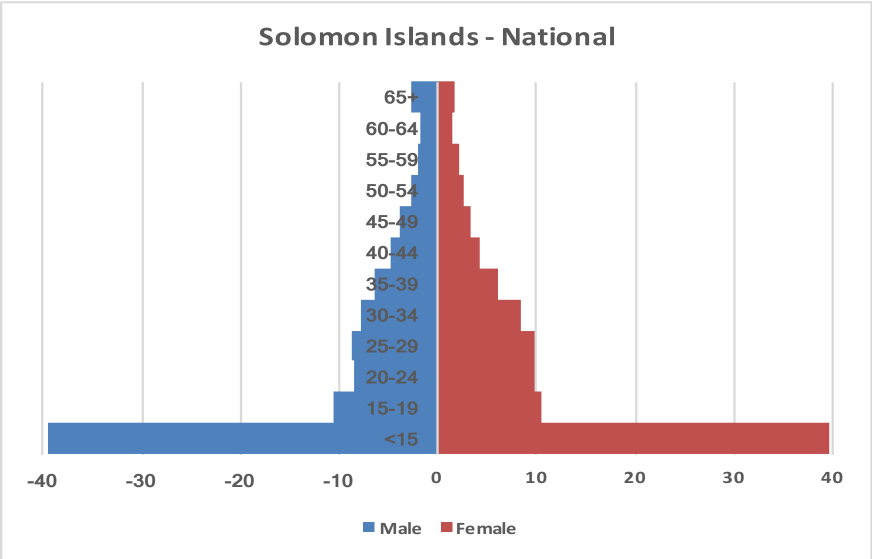


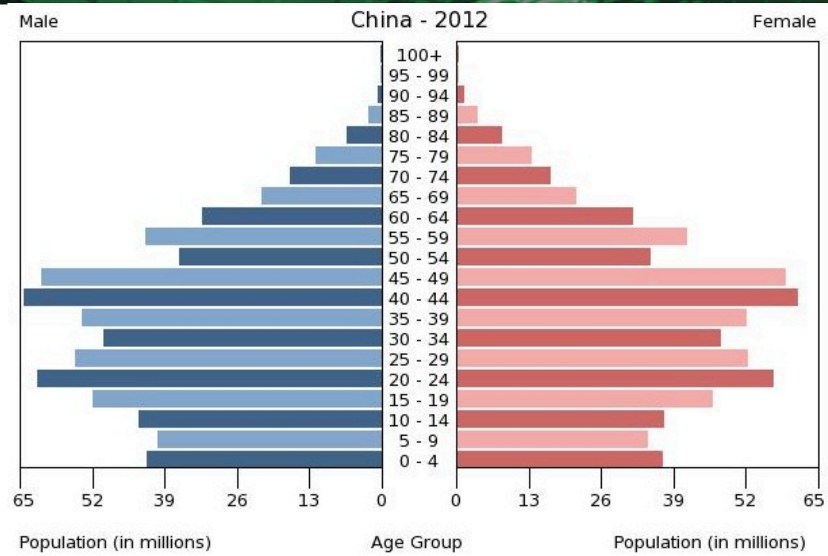


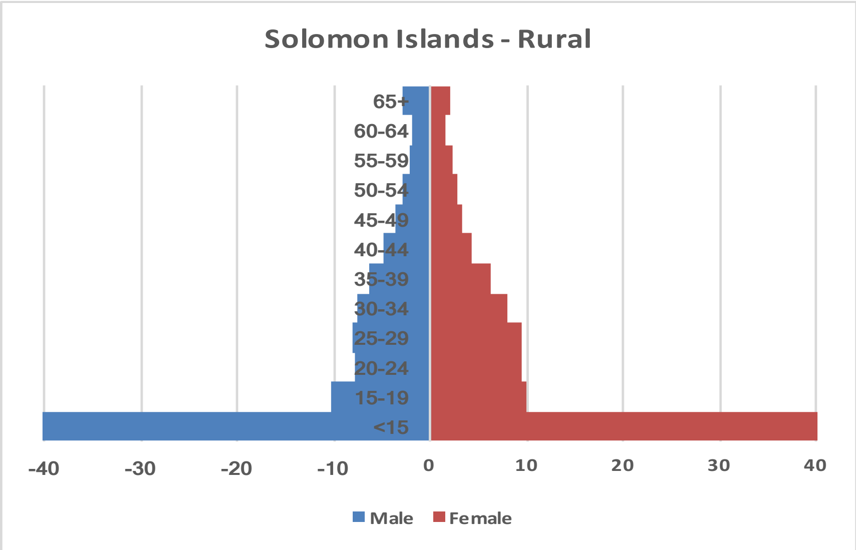


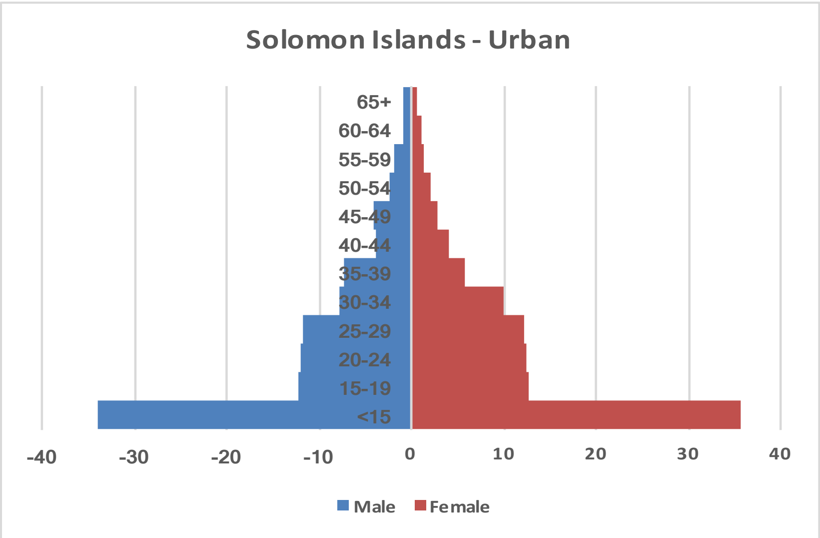


Solomons Population Pyramid: Very high youth dependency

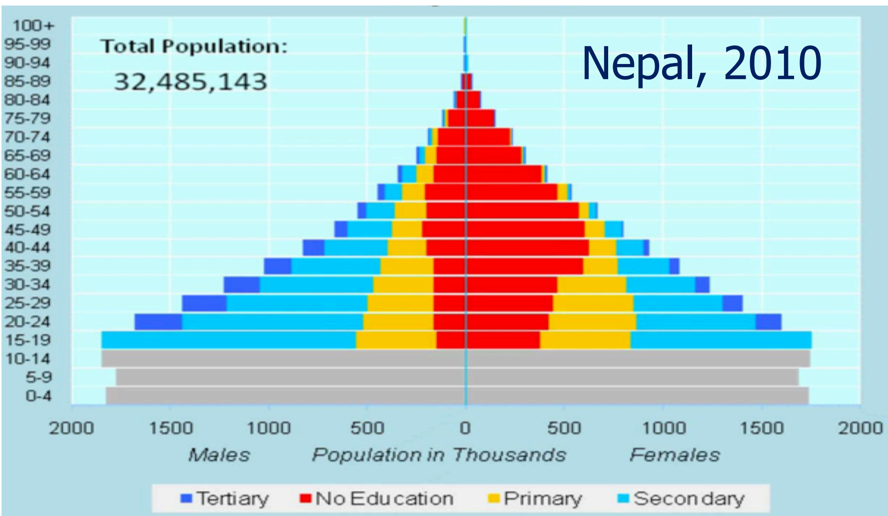


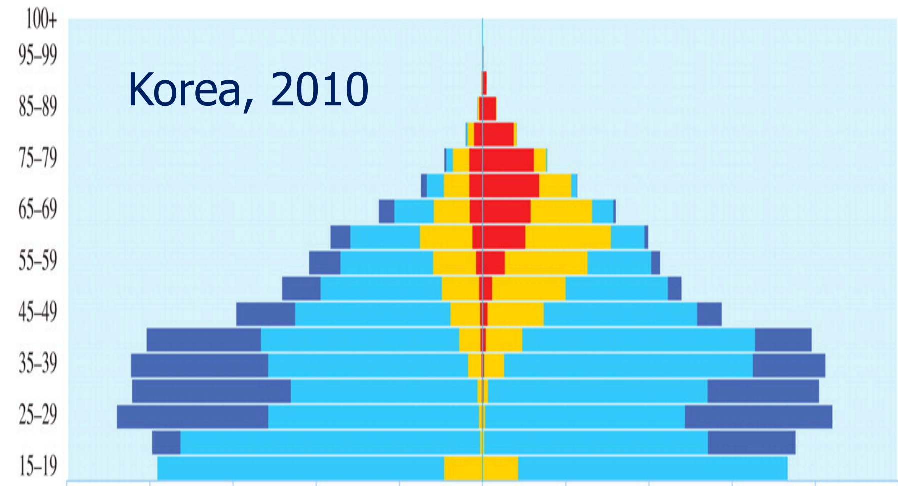


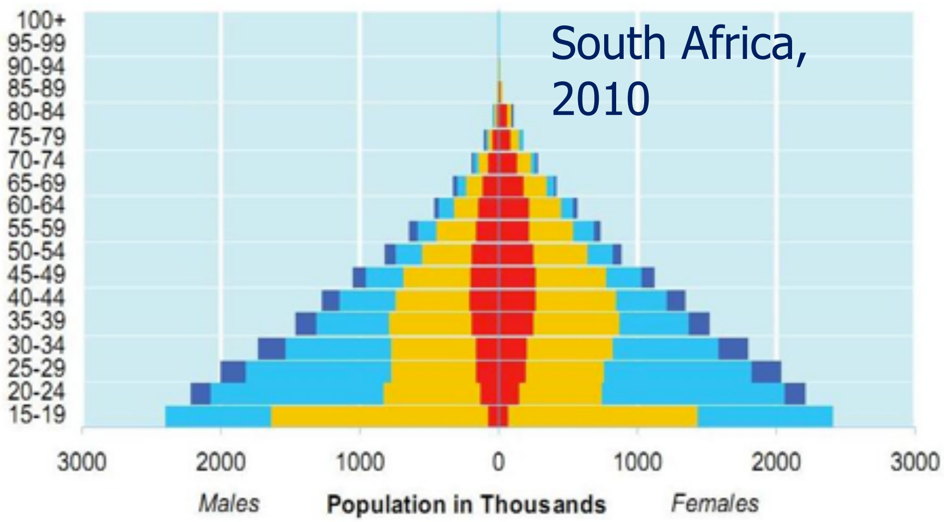


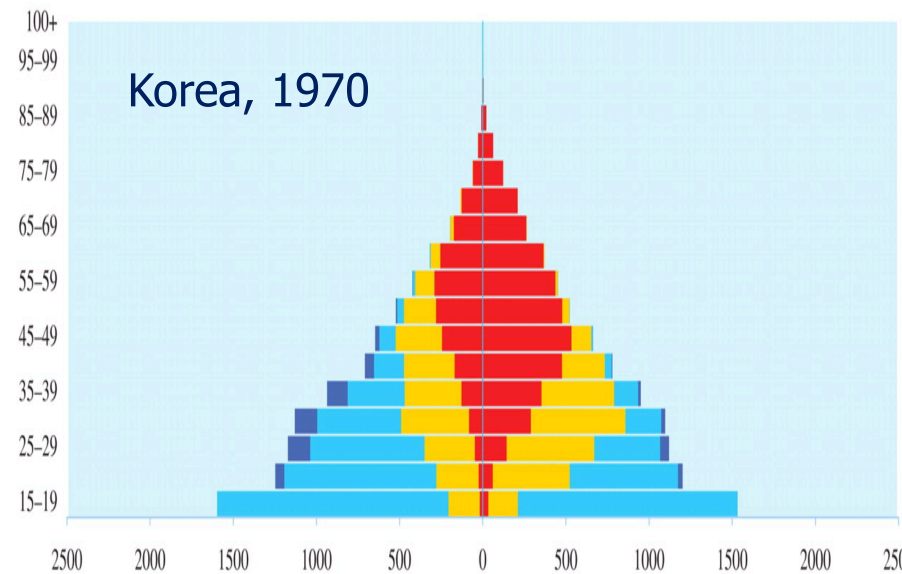


Population/Education Pyramids: Access/Attainment









What Developing Countries Can Do

- Address incentives and disincentives for having children through the principal variables influencing the demand for children
- Coercion is not a good option
- Raise the socioeconomic status of women
- Persuasion through education
- Family planning programs
- Increase employment opportunities for women (increases opportunity cost of having more children, as in microeconomic household theory)
- Help facilitate genuine and faster development of developing countries that still have high fertility rates

Don’t Drink and Drive

Social Policy and Panel Data

What are panel data?

- Panel data consists of the observations on the same n entities at two or more time periods T. If the data set contains observations on the variables X and Y, then the data are denoted
    - $(X_{it}, Y_{it}), i = 1, \ldots, n \text{ and } t = 1, \ldots, T,$
- where the first subscript, i, refers to the entity being observed, and
- the second subscript, t, refers to the date at which it is observed.
- Balanced panel Vs. unbalanced panel.
    - Balanced panel: Variables are observed for each entity and
    - each time period.
    - Unbalanced panel: Some missing data for at least one time
    - period.
- We consider the analysis of balanced panel. But extension to unbalanced is straightforward.

Omitted Variables Biases

- Issue: Do alcohol taxes help decrease traffic deaths?
- Data: fatality.wf1
  - 48 U.S. states (excluding Alaska and Hawaii): N = 48.
  - 1982 -1988: T =7.
  - fatality rate = # of traffic accident deaths per 10,000 people.
  - beertax = tax per a case of beer ($).

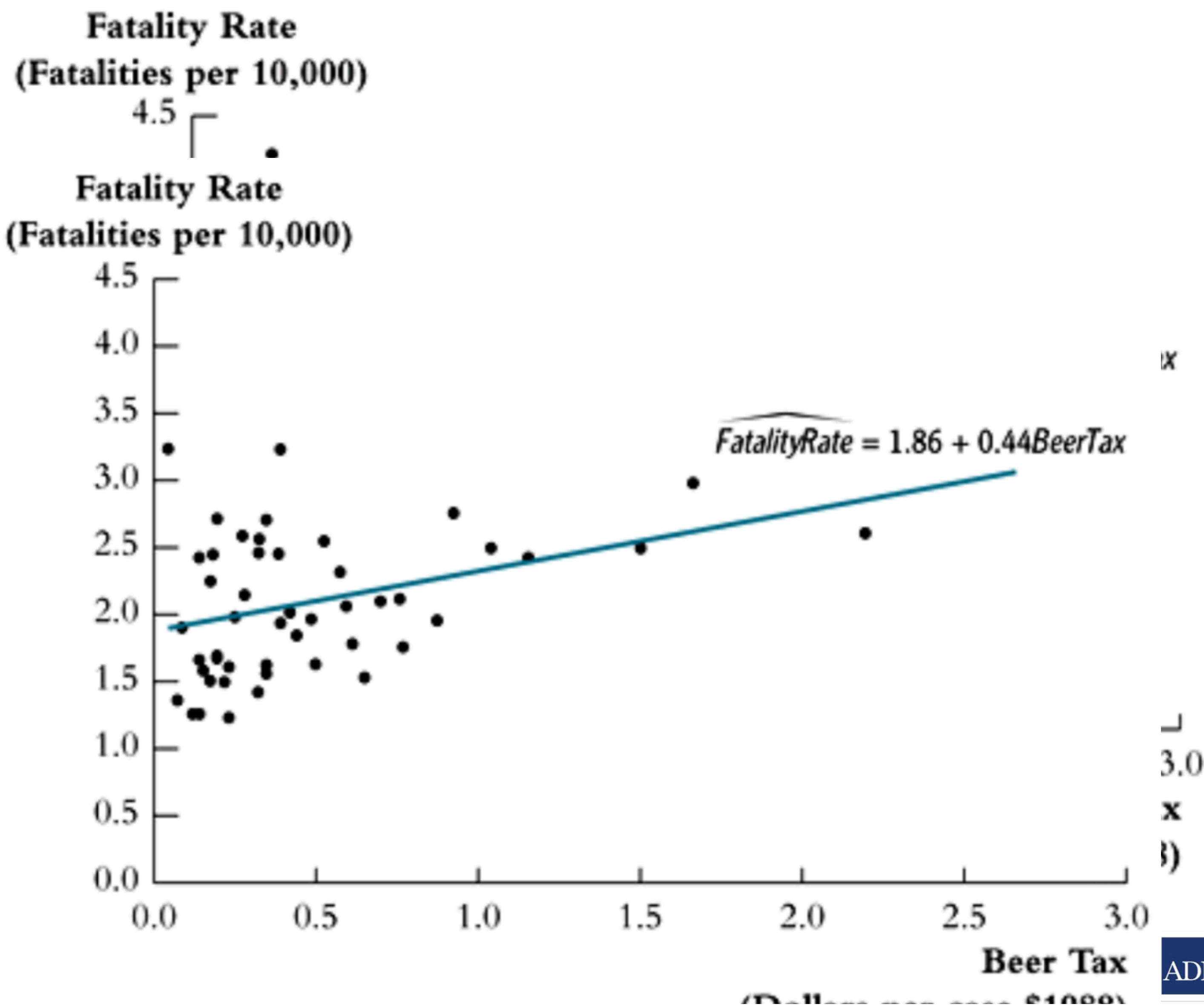
- Estimation results for the 1982 data:

$$\widehat{FatalityRate} = \underset{(0.15)}{2.01} + \underset{(0.13)}{0.15}BeerTax$$

- Estimation results for the 1988 data:

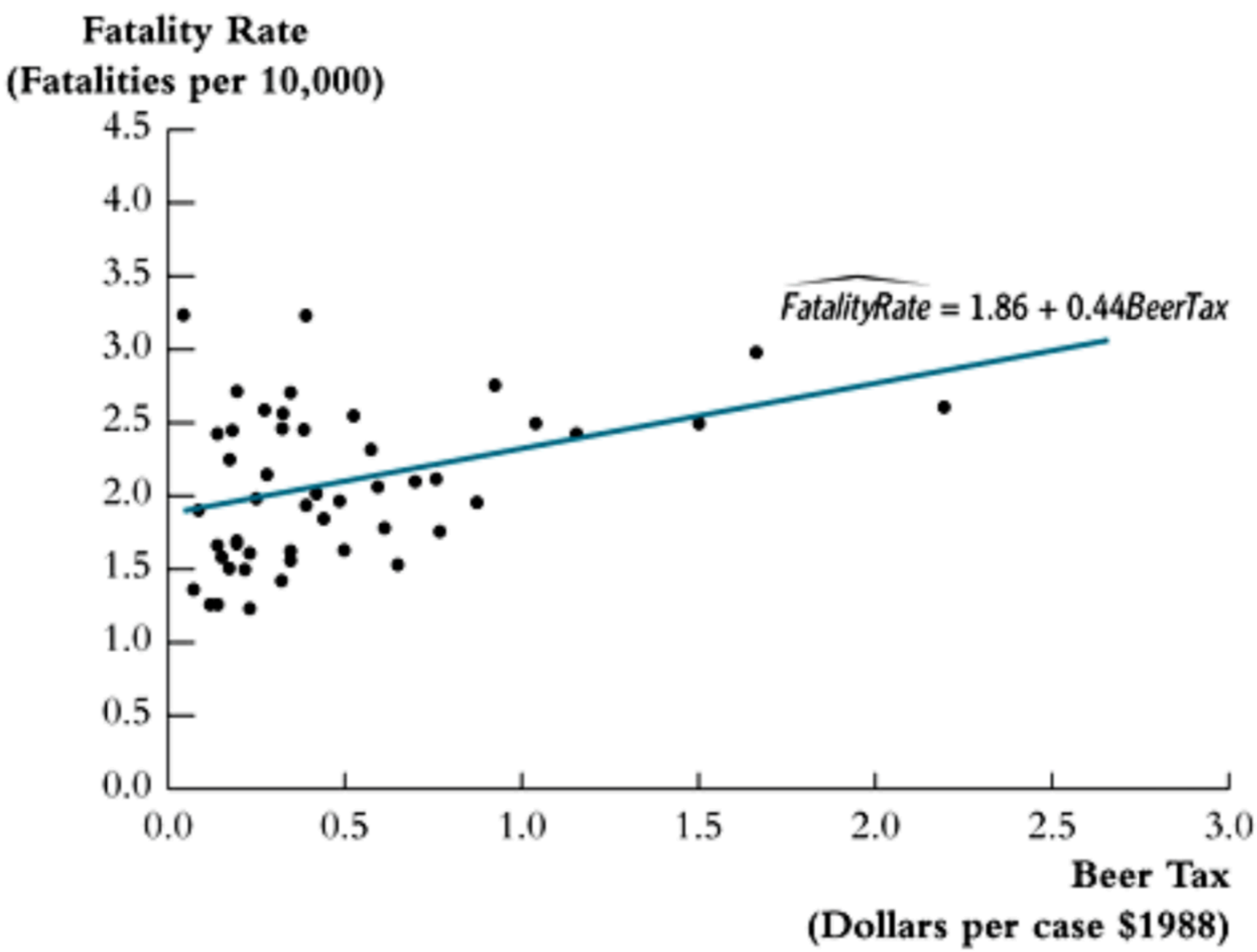
$$\widehat{FatalityRate} = \underset{(0.11)}{1.86} + \underset{(0.13)}{0.44}BeerTax$$

Traffic Fatalities and Beer Tax 1982 Data



Panel a is a scatterplot of traffic fatality rates and the real tax on a case of beer (in 1988 dollars) for 48 states in 1982. Panel b shows the data for 1988. Both plots show a positive relationship between the fatality rate and the real beer tax.

Traffic Fatalities and Beer Tax 1988 Data



Panel a is a scatterplot of traffic fatality rates and the real tax on a case of beer (in 1988 dollars) for 48 states in 1982. Panel b shows the data for 1988. Both plots show a positive relationship between the fatality rate and the real beer tax.

What is going on here?

- By looking at individual time periods, we are comparing different states at the same time,
- Not different taxes in the same state
- States will differ in many ways, but here can’t isolate the effect of changing a tax

- Two equations for 1982 and 1988:

$$FatalityRate_{i,1988} = \beta_0 + \beta_1 BeerTax_{i,1988} + \beta_2 Z_i + u_{i,1988}.$$

$$FatalityRate_{i,1982} = \beta_0 + \beta_1 BeerTax_{i,1982} + \beta_2 Zi + u_{i,1982}. \rightarrow$$

$$FatalityRate_{i,1988} - Fatality_{i,1982}$$
$$= \beta_1(BeerTax_{i,1988} - BeerTax_{i,1982}) + (u_{i,1988} - u_{i,1982})$$

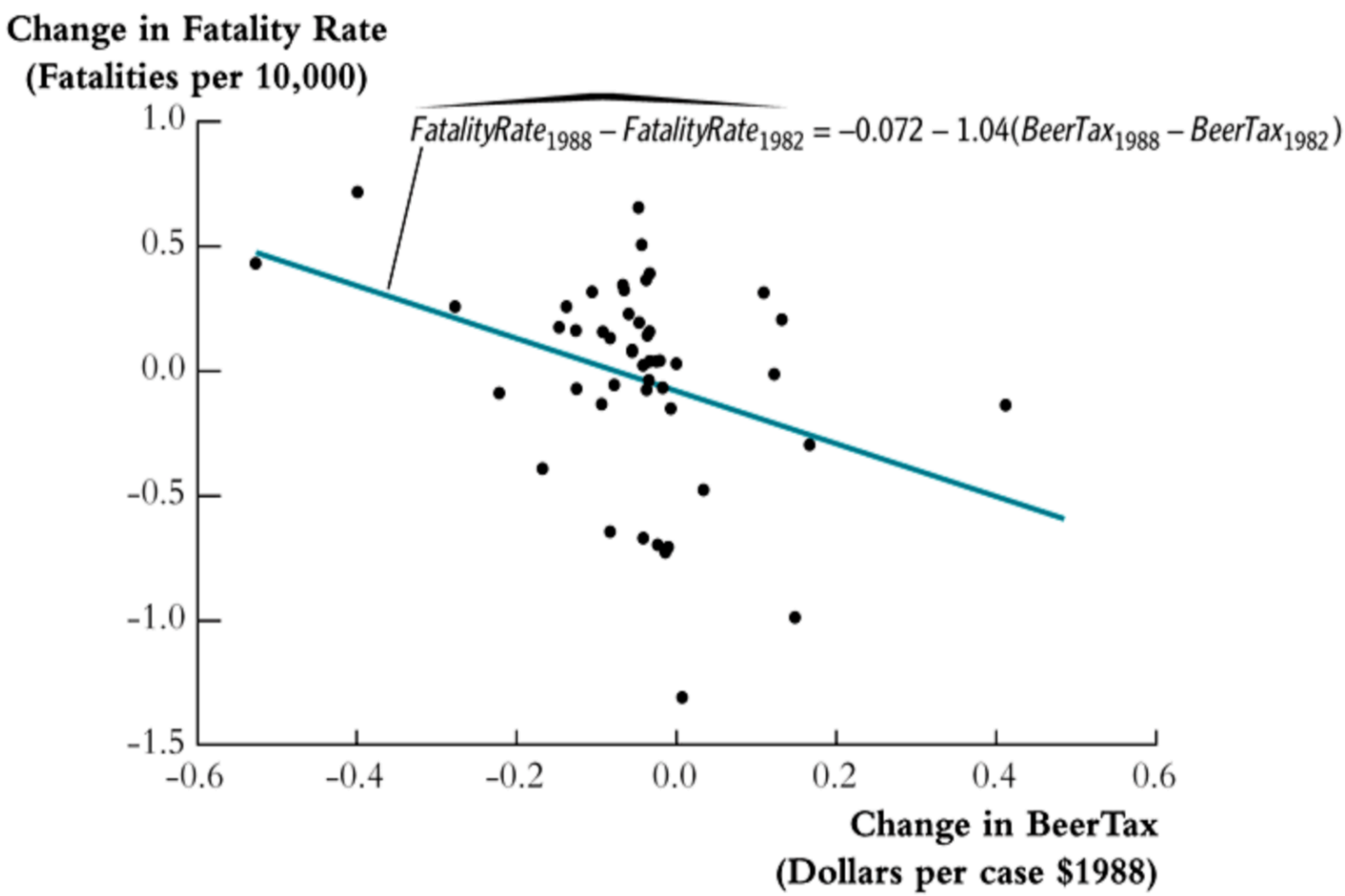
- Actual estimation results for (1):

$$FatalityRate_{i,1988} - Fatality_{i,1982}$$
$$= -0.072 - 1.04(BeerTax_{1988} - BeerTax_{1982})$$

(0.065) (0.36)

Panel Data Reveal the Expected Relationship

This is a scatterplot of the *change* in the traffic fatality rate and the *change* in real beer taxes between 1982 and 1988 for 48 states. There is a negative relationship between changes in the fatality rate and changes in the beer tax.



Comments

- As real beer tax increases by $1 per case, the traffic fatality rate falls by 1.04 deaths per 10,000 people.
- This is a **big** effect, because mean traffic fatality rate is approximately two.
- This before-and-after approach works well if $T = 2$.
- What should do if $T > 2$?

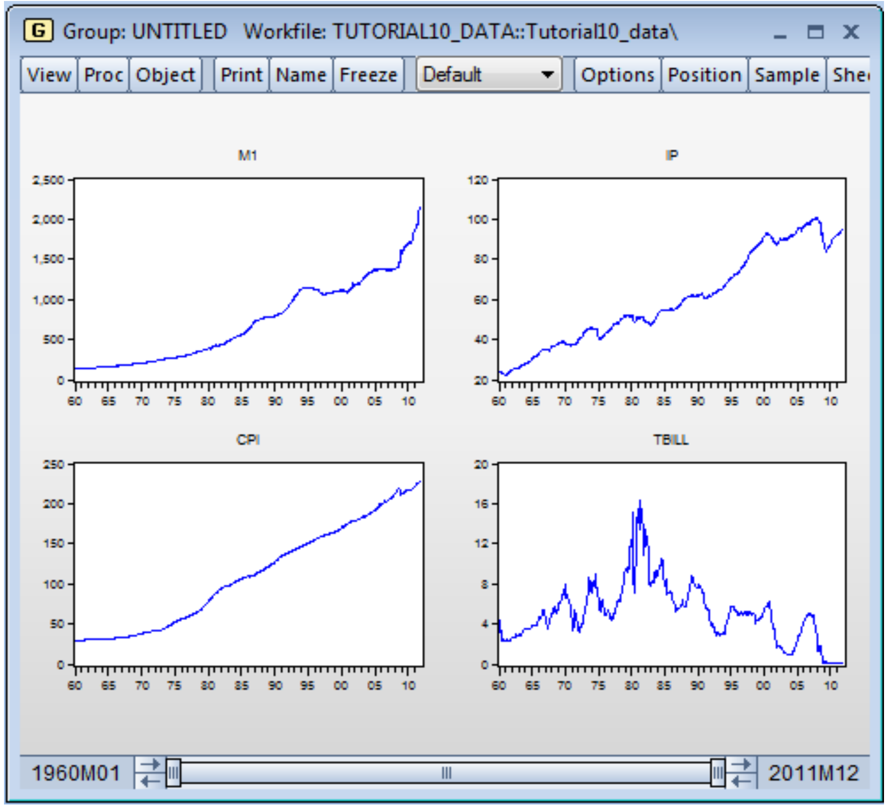
How to Forecast with the CPI

Data and Workfile Documentation

- ***WS4_Data.wf1*** and ***WS4_Data.xls*** have 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*)

Simple Time Series Regressions

- 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).



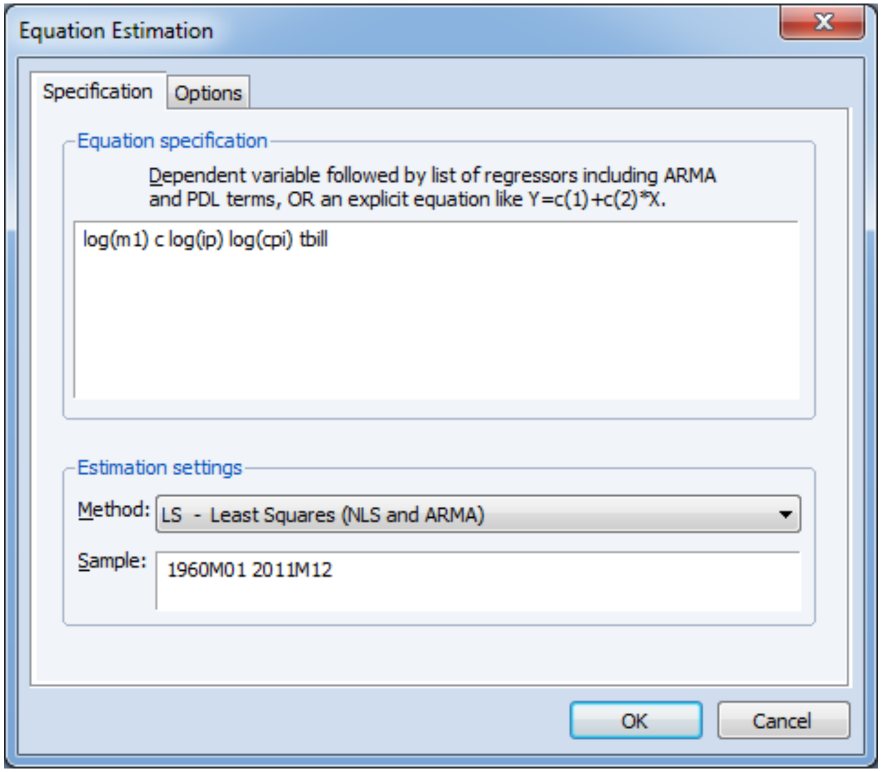
Simple Time Series Regressions: Example 1

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

Estimation: Example 1

1. Open ***Data.wf1*** 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
4. Click **OK**.

*__Note:__ This equation is saved as ***eq01*** in ***WS4_Results.wf1*** workfile.



Simple Time Series Regressions: Example 1 (cont'd)

- The estimation Output is shown here.

<u>A number of things stand out:</u>

✓ 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

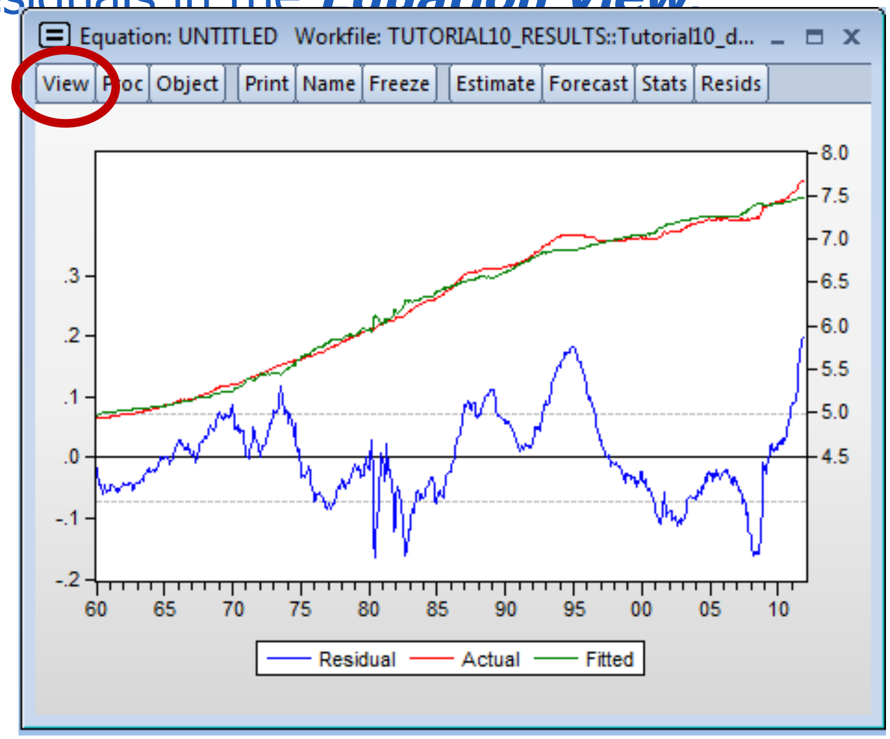
Dependent Variable: LOG(M1)
Method: Least Squares
Date: 03/25/13 Time: 02:57
Sample: 1960M01 2011M12
Included observations: 624

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 (cont'd)

- Let's take a look at the behavior of residuals in the ***Equation View***



<u>Examining Residuals:</u>

1. Open ***eq01*** in ***WS4_Results.wf1*** workfile.
2. Click **View** on the ***Equation Object*** menu bar.
3. Select **Actual, Fitted, Residual** → **Actual, Fitted, Residual Graph.**

- 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

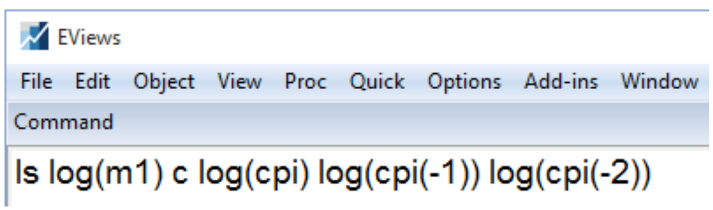
- Distributed lag models are easy to specify in EViews.
- In these models, one or more variables affects the dependent variable with a lag.
- Lags of dependent or independent variables can be specified directly in the equation box in EViews (see *Data Function* tutorial for more details).

Estimation: Example 2

1. Suppose you want to examine how ***M1*** is affected by ***CPI*** and its first two lags in log form. Type in the command window:

```
ls log(m1) c log(cpi) log(cpi(-1)) log(cpi(-2))
```

2. Press **Enter.**



Note that command ***ls***, denotes that the method of estimation is least squares.

Note:** This equation is saved as ***eq02 in ***WS4_Results.wf1*** workfile.

Simple Time Series Regressions: Example 2 (cont'd)

- The estimation Output is shown here.

A few quick notes:

- ✓ 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 (cont'd)

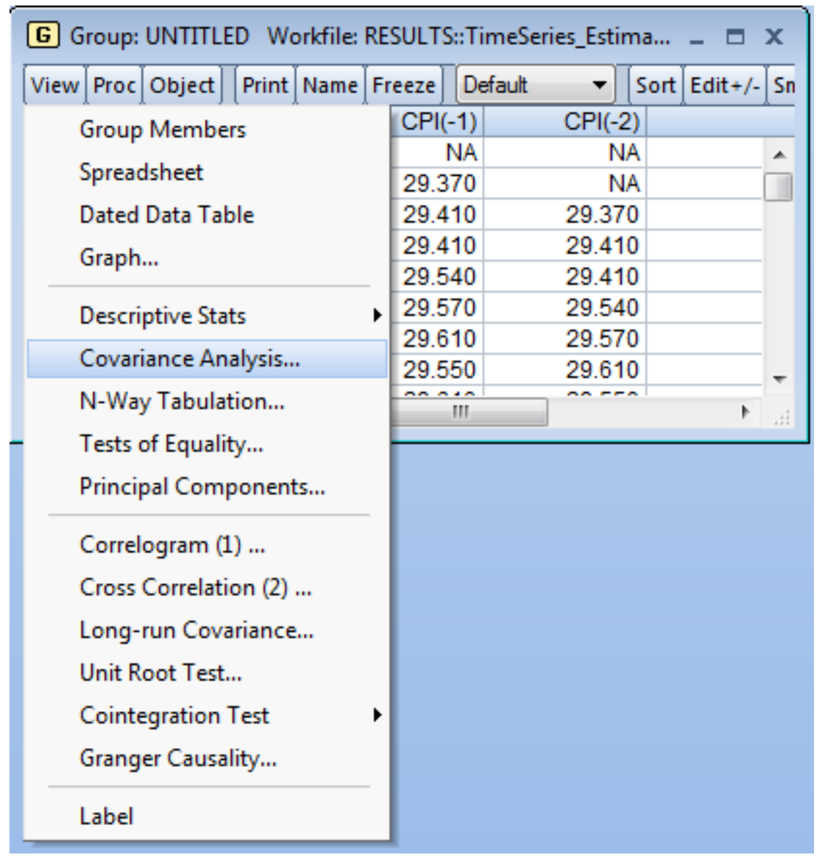
✓ You can look at the correlation matrix of ***CPI*** and its first two lags. To create the correlation matrix, follow these steps:

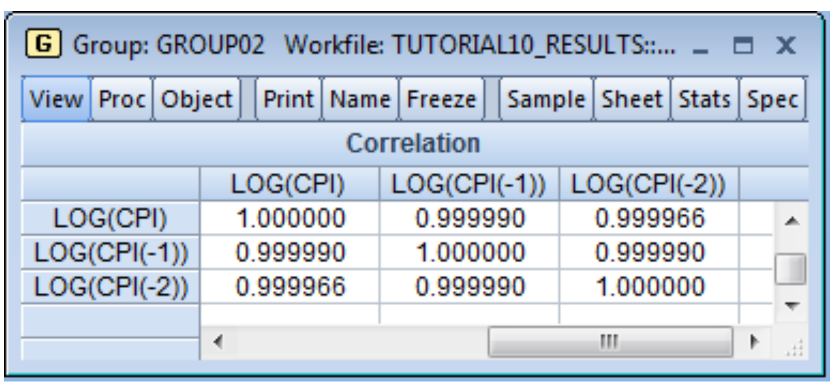
1. Type in the command window:

   ```
   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***.
4. Click **OK**.





Correlation

	LOG(CPI)	LOG(CPI(-1))	LOG(CPI(-2))
LOG(CPI)	1.000000	0.999990	0.999966
LOG(CPI(-1))	0.999990	1.000000	0.999990
LOG(CPI(-2))	0.999966	0.999990	1.000000

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.

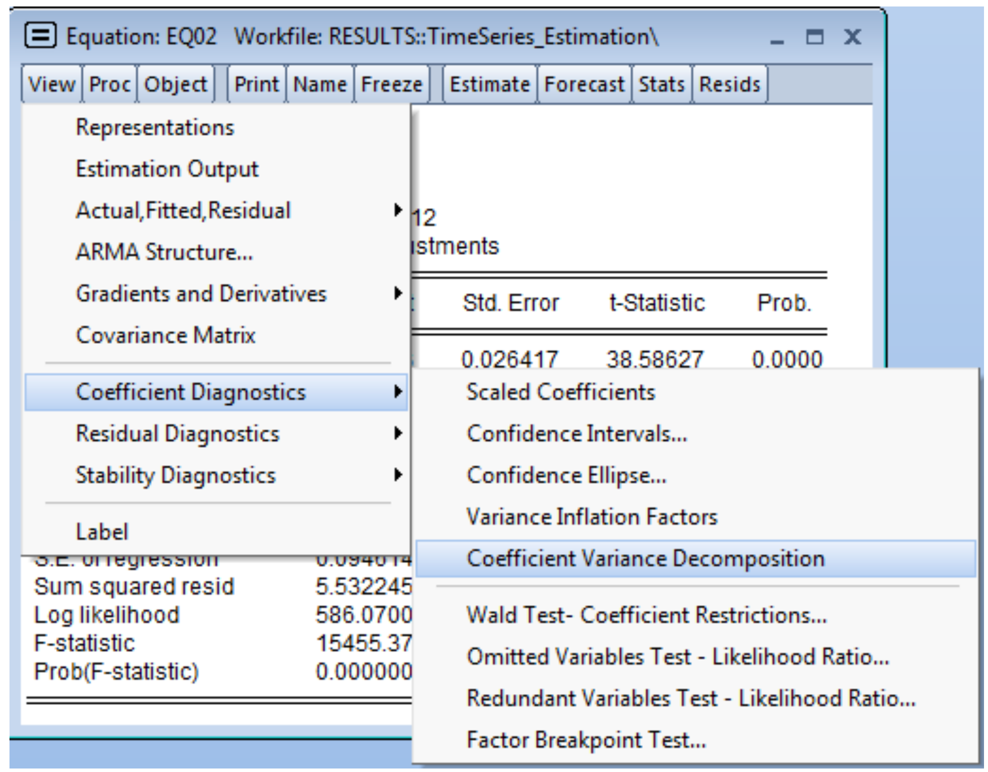
Simple Time Series Regressions: Example 2 (cont'd)

- 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 (i.e., $X_1 = X_2 + 3X_3$) which the simple correlation matrix may fail to detect.

<u>Coefficient Variance Decomposition:</u>

1. Click on ***eq02*** to open it.
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 (cont'd)

- The Equation View changes and displays a table showing Eigenvalues, Condition Numbers, Variance Decomposition Proportions and Eigenvectors, as shown here.
  - ✓ 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

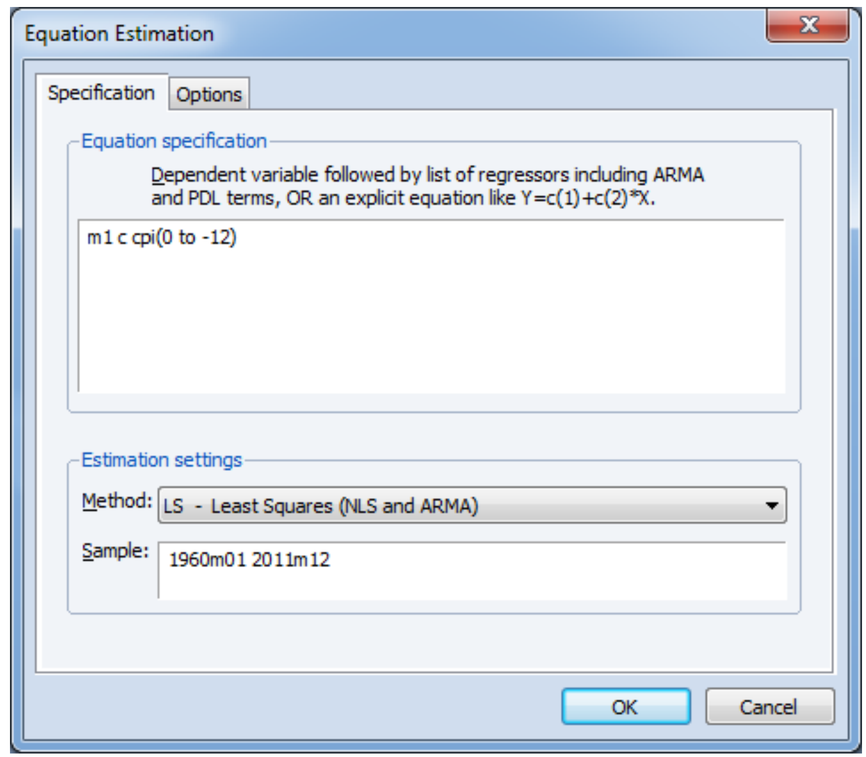
	Associated Eigenvalue			
Variable	1	2	3	4
C	-6.22E-05	0.008903	0.992187	-0.124440
LOG(CPI)	0.408399	-0.707309	-0.065526	-0.573263
LOG(CPI(-1))	-0.816492	-0.000485	-0.071895	-0.572863
LOG(CPI(-2))	0.408106	0.706848	-0.078115	-0.572462

Simple Time Series Regressions: Example 3

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

Estimation: Example 3

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.
4. Click **OK.**



Note:** This equation is saved as ***eq02a in ***WS4_Results.wf1*** workfile.

Simple Time Series Regressions: Example 3 (cont'd)

- The estimation output is shown here.

A few quick notes:

- ✓ 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

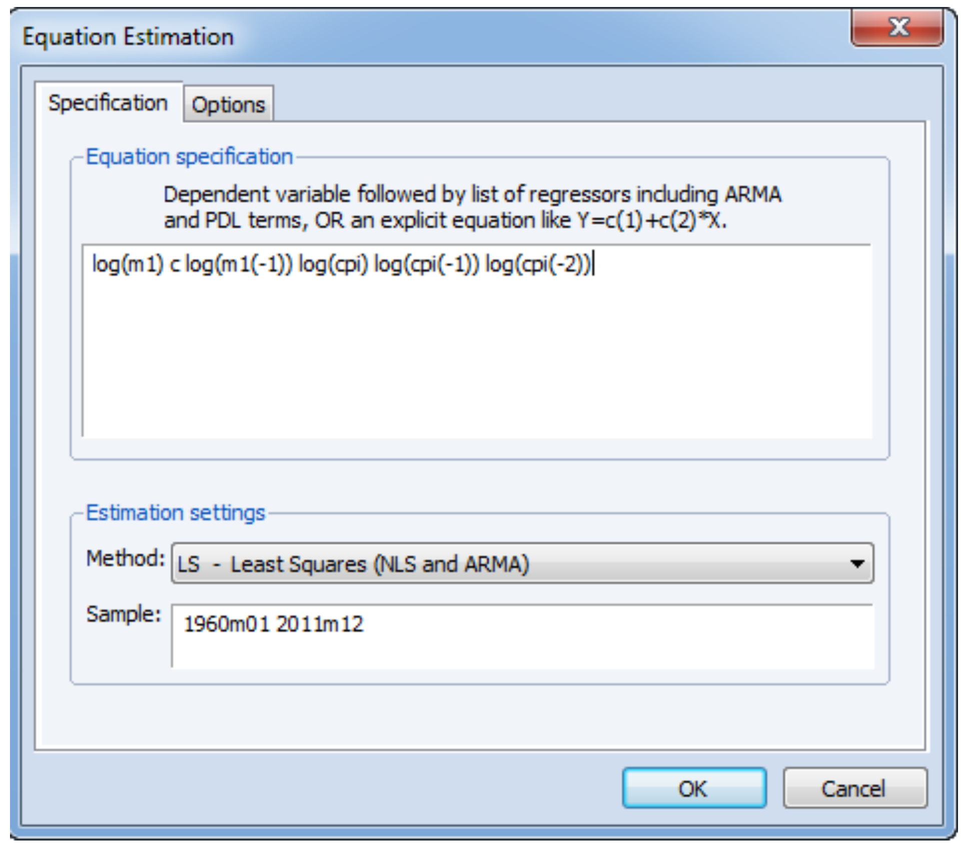
R-squared	0.968625	Mean dependent var	736.1742
Adjusted R-squared	0.967943	S.D. dependent var	501.4292
S.E. of regression	89.77775	Akaike info criterion	11.85516
Sum squared resid	4819906.	Schwarz criterion	11.95620
Log likelihood	3613.679	Hannan-Quinn criter.	11.89446
F-statistic	1420.155	Durbin-Watson stat	0.027281
Prob(F-statistic)	0.000000		

Simple Time Series Regressions Example 4

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

<u>Estimation: Example 4</u>

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))***
3. Click **OK.**



Note:** This equation is saved as ***eq03 in ***WS4_Results.wf1*** workfile.

Simple Time Series Regressions Example 4 (cont'd)

- The estimation output is shown here

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		

A few quick notes:

✓ Notice that the lagged dependent variable (***log(m1(-1))***) is close to unity and is highly significant.

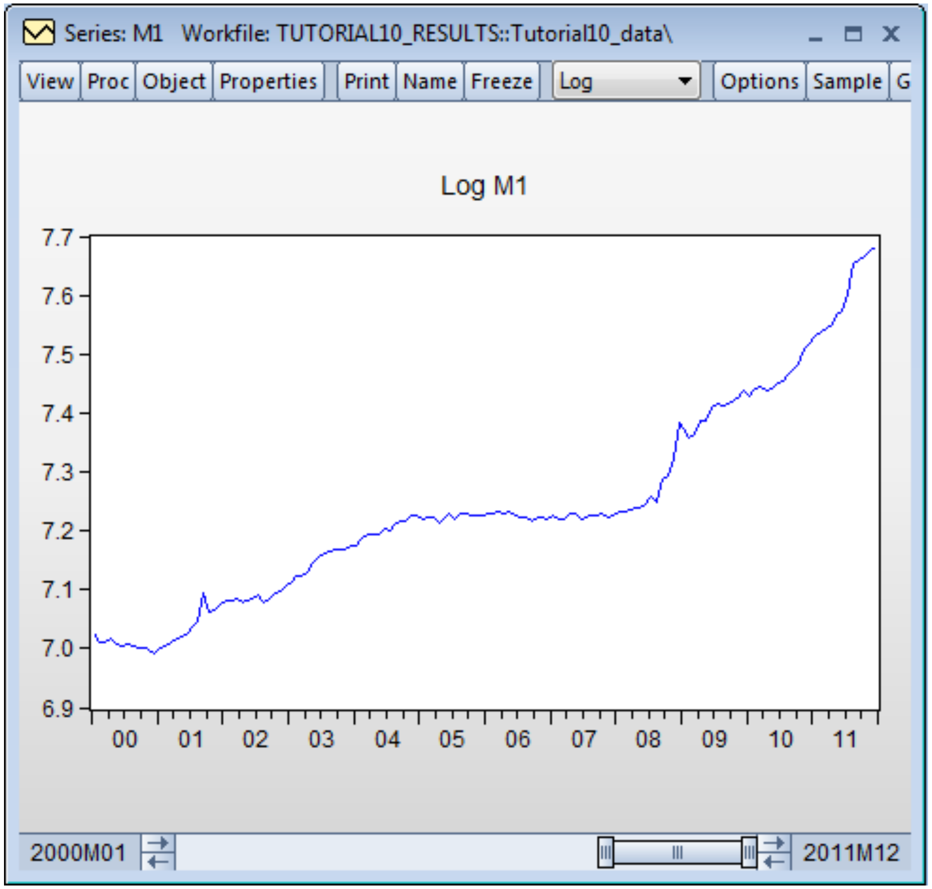
✓ If errors are serially correlated, OLS estimates are biased and inconsistent in the presence of lagged dependent (see *Appendix to Basic Time Series Estimation* for details).

Time Series Models and Date Functions

Time Series Regressions and Date Dummies

- 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 and Date Dummies: Example 1

- Let's check whether the period since 2008 has indeed had an outsized impact in the growth of M1.

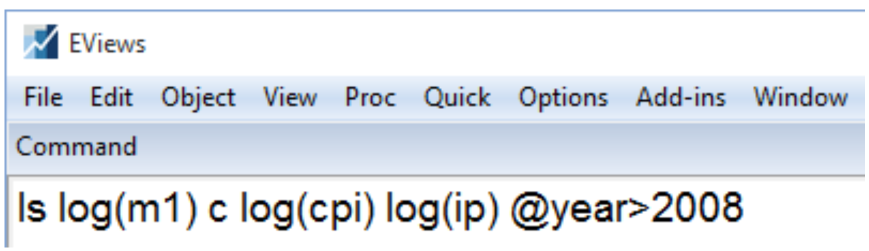
Date Dummies: Example 1

1. Specify the equation by typing in the command window:

```
ls log(m1) c log(cpi) log(ip) @year>2008
```

Note that command ***@year>2008*** creates a dummy variable, equal to 1 for the period after 2008 and 0 otherwise.

2. Press **Enter**.



Note:** This equation is saved as ***eq04 in ***WS4_Results.wf1*** workfile.

Time Series Regressions and Date Dummies: Example 1 (cont'd)

- The estimation output is shown here.
- 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***.

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		

Time Series Regressions and Date Dummies: Example 2

- You can also create a dummy for each year post-2008, to determine their individual impact in the growth of ***M1***.

Date Dummies: Example 2

1. Enter in the command window:

```
ls log(m1) c log(cpi) log(ip)
tbill @year=2008 @year=2009
@year=2010 @year=2011
```

2. Press **Enter** (please type the command in one line).

- Note that command ***@year=2008*** creates a dummy variable equal to 1 for all months in 2008 and 0 otherwise, and so on.

.

Note:** This equation is saved as ***eq05 in ***WS4_Results.wf1*** workfile.

EViews

File Edit Object View Proc Quick Options Add-ins Window Help

Command

ls log(m1) c log(cpi) log(ip) tbill @year=2008 @year=2009 @year=2010 @year=2011

Equation: EQ05 Workfile: RESULTS::TimeSeries_Estimation\

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

Dependent Variable: LOG(M1)
Method: Least Squares
Date: 11/20/12 Time: 00:18
Sample: 1960M01 2011M12
Included observations: 624

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.642998	0.049480	12.99519	0.0000
LOG(IP)	0.271078	0.030692	8.832175	0.0000
LOG(CPI)	1.010272	0.018492	54.63376	0.0000
@YEAR=2008	-0.040880	0.025191	-1.622787	0.1051
@YEAR=2009	0.127980	0.025234	5.071650	0.0000
@YEAR=2010	0.158954	0.025173	6.314363	0.0000
@YEAR=2011	0.258509	0.025194	10.26084	0.0000

R-squared	0.989614	Mean dependent var	6.286356
Adjusted R-squared	0.989513	S.D. dependent var	0.825190
S.E. of regression	0.084504	Akaike info criterion	-2.092891
Sum squared resid	4.405907	Schwarz criterion	-2.043127
Log likelihood	659.9821	Hannan-Quinn criter.	-2.073553
F-statistic	9798.527	Durbin-Watson stat	0.018788
Prob(F-statistic)	0.000000		

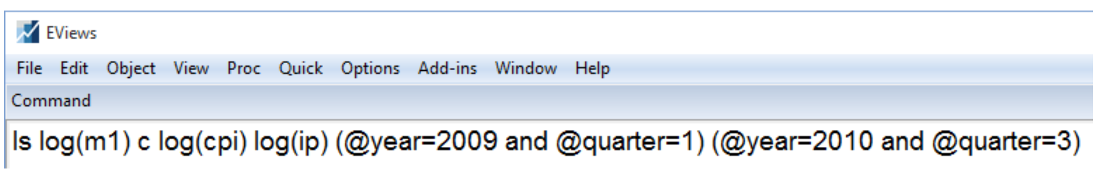
Time Series Regressions and Date Dummies: Example 3

- We can examine the direct impact of QE on ***M1***, by specifying the year and the quarter (or month) when it was announced.
- QE 1 largely took place in the first quarter of 2009, whereas QE 2 was announced in the third quarter of 2010.

Date Dummies: Example 3

1. Type in the command window:

```
ls log(m1) c log(cpi) log(ip)
(@year=2009 and @quarter=1) (@year=2010
and @quarter=3)
```



2. Press **Enter** (please type the command in one line).

- Note that command ***@year=2009*** and ***@quarter=1*** creates a dummy variable, equal to 1 for all months in the first quarter of 2009, and 0 otherwise.

Time Series Regressions and Date Dummies: Example 3 (cont'd)

- Estimation output is shown here.
- Notice that while the announcement of QE1 is not statistically significant, QE2 is.

Note:** This equation is saved as ***eq06 in ***WS4_Results.wf1*** workfile.

Equation: EQ06 Workfile: RESULTS::TimeSeries_Estimation\

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

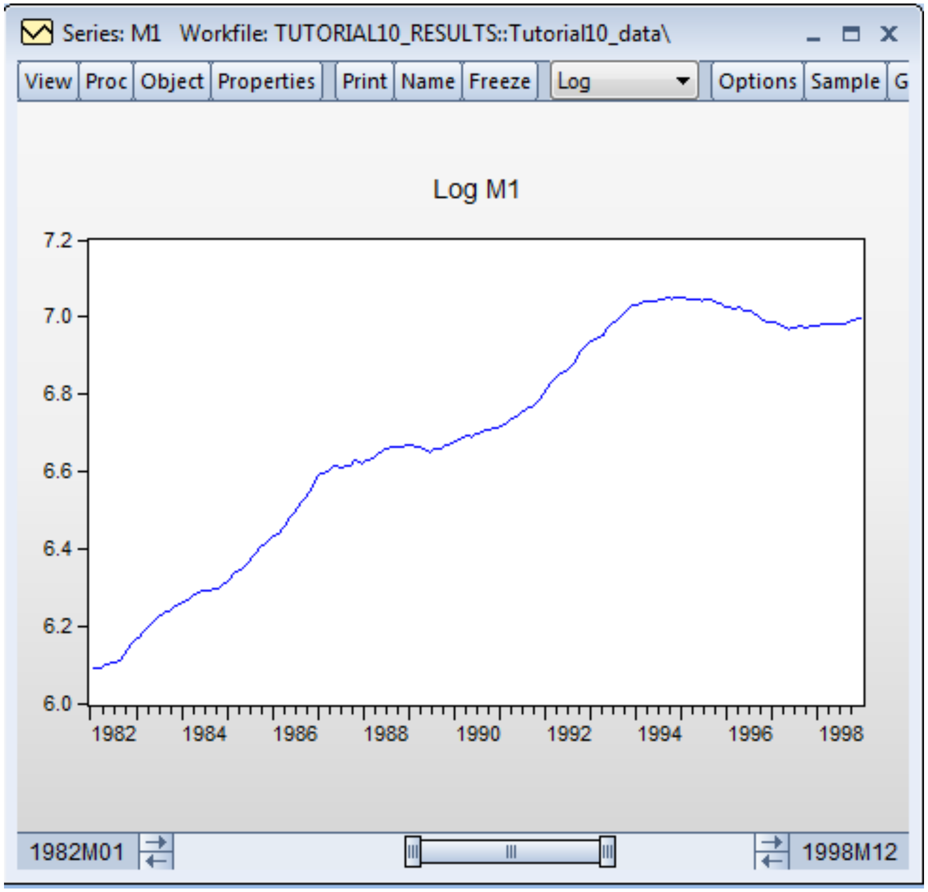
Dependent Variable: LOG(M1)
Method: Least Squares
Date: 11/25/12 Time: 21:43
Sample: 1960M01 2011M12
Included observations: 624

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.608298	0.054477	11.16609	0.0000
LOG(CPI)	1.038984	0.020366	51.01572	0.0000
LOG(IP)	0.249711	0.034031	7.337778	0.0000
(@YEAR=2009 AND @QUARTER=1)	0.074946	0.054770	1.368360	0.1717
(@YEAR=2010 AND @QUARTER=3)	0.133944	0.054759	2.446066	0.0147

R-squared	0.987044	Mean dependent var	6.286356
Adjusted R-squared	0.986960	S.D. dependent var	0.825190
S.E. of regression	0.094231	Akaike info criterion	-1.878160
Sum squared resid	5.496367	Schwarz criterion	-1.842614
Log likelihood	590.9860	Hannan-Quinn criter.	-1.864347
F-statistic	11789.29	Durbin-Watson stat	0.017239
Prob(F-statistic)	0.000000		

Time Series Regressions and Date Dummies (cont'd)

- You can also create a dummy variable by specifying a date value using the ***@during*** function.
- For example, based on the graph shown here, it also seems that the growth of ***M1*** was unusually high in the period between 1988 and 1993.
- Let's create a dummy variable for this period and examine its impact on the growth of ***M1***.



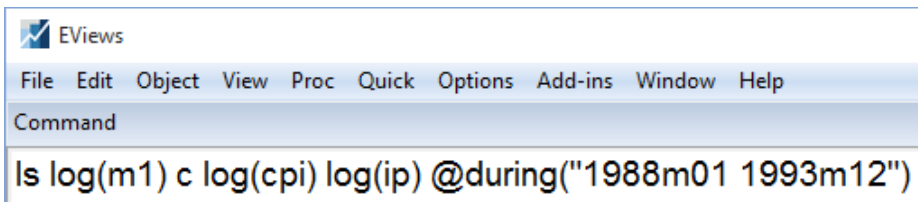
Time Series Regressions and Date Dummies: Example 4

Date Dummies: Example 4

1. Specify the equation by typing in the command window:

```
ls log(m1) c log(cpi) log(ip) @during("1988m01 1993m12")
```

2. Press **Enter**.



Note:** This equation is saved as ***eq07
in ***WS4_Results.wf1*** workfile.

Alternatively, you can also type in the command line:

```
ls log(m1) c log(cpi) log(ip) (@date>=@dateval("1988m01") and
@date<=@dateval("1993m12"))
```

Time Series Regressions and Date Dummies: Example 4 (cont'd)

- Estimation output is shown here.
- Note that command: `@during("1988m01 1993m12")` creates a dummy variable equal to 1 from January 1980 until December 1993, and 0 otherwise.

Equation: EQ07 Workfile: RESULTS::TimeSeries_Estimation\

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

Dependent Variable: LOG(M1)
Method: Least Squares
Date: 04/26/13 Time: 02:04
Sample: 1960M01 2011M12
Included observations: 624

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.526554	0.054927	9.586371	0.0000
LOG(IP)	0.317464	0.035378	8.973494	0.0000
LOG(CPI)	0.995043	0.021463	46.36088	0.0000
@DURING("1988m01 1993m12")	0.073181	0.012535	5.838146	0.0000

R-squared	0.987565	Mean dependent var	6.286356
Adjusted R-squared	0.987505	S.D. dependent var	0.825190
S.E. of regression	0.092242	Akaike info criterion	-1.922406
Sum squared resid	5.275359	Schwarz criterion	-1.893969
Log likelihood	603.7907	Hannan-Quinn criter.	-1.911356
F-statistic	16412.71	Durbin-Watson stat	0.009846
Prob(F-statistic)	0.000000		

Time Series Regressions and Date Dummies: Example 5

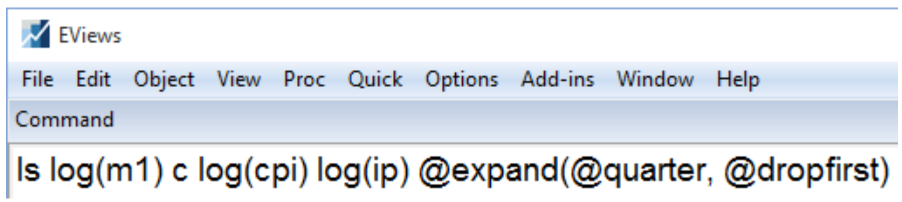
- You can also use ***@expand*** function to create a set of dummy variables.

Date Dummies: Example 5

1. Suppose you want to check the effect of each quarter of the year in the growth of ***M1***. Type in the command window:

```
ls log(m1) c log(cpi) log(ip) @expand(@quarter, @dropfirst)
```

2. Press ***Enter***.



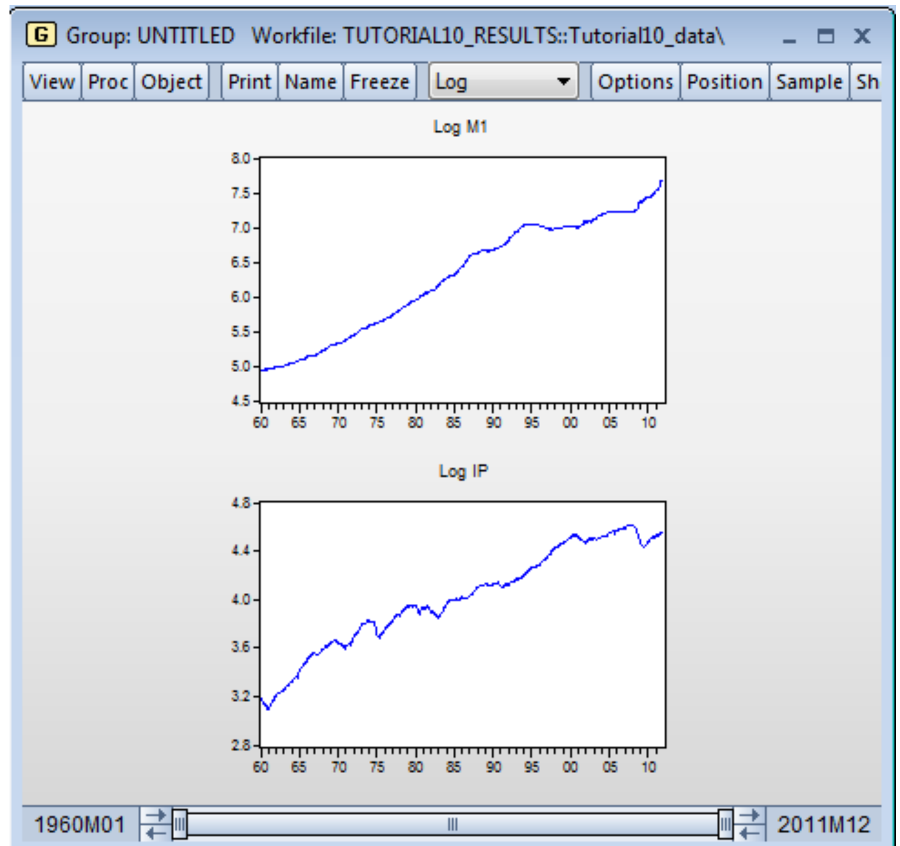
- Note that command ***@expand(@quarter)*** creates a set of dummy variables: ***@quarter=1*** is a dummy variable equal to 1 for the first three months of each year and 0 otherwise; ***@quarter=2*** equals 1 for months in the second quarter, etc.
- Note also that since here we have included a constant, we need to drop one of the dummy variables in order not fall in the dummy variable trap. This can be accomplished by using ***@dropfirst*** command.

Trends and Seasonality

Time Trends

- 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 Trends (cont'd)

- Time trends can be accommodated easily in EViews using function ***@trend***.
- Let's illustrate the impact of time trend in the simple regression of ***IP*** on ***M1***.

Estimation with no Time Trends:

1. First, let's specify a simple regression without time trends. Type in the command window:

```
ls log(m1) c log(ip)
```

2. Press **Enter.**

Equation: EQ09 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	-1.450611	0.084318	-17.20412	0.0000
LOG(IP)	1.927221	0.020893	92.24404	0.0000

R-squared	0.931880	Mean dependent var	6.286356
Adjusted R-squared	0.931771	S.D. dependent var	0.825190
S.E. of regression	0.215546	Akaike info criterion	-0.228085
Sum squared resid	28.89815	Schwarz criterion	-0.213867
Log likelihood	73.16265	Hannan-Quinn criter.	-0.222560
F-statistic	8508.963	Durbin-Watson stat	0.006200
Prob(F-statistic)	0.000000		

- As seen, ***log(ip)*** has a highly statistically significant coefficient. The R-squared is also very high, indicating -- at first look --a very good fit for the model.

Note:** This equation is saved as ***eq09 in ***WS4_Results.wf1*** workfile.

Time Trends: Example 1

- Let's include a time trend and see if anything changes.

Time Trends: Example 1

1. Type in the command window:

```
ls log(m1) c log(ip) @trend
```

2. Press **Enter.**

Results are very different now:

- ✓ First, the coefficient of ***log(ip)*** is now 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).***

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		

Note:** This equation is saved as ***eq10 in ***WS4_Results.wf1*** workfile.

Time Trends: Example 2

- Quadratic time trend are also very easy to specify in EViews.

  Time Trends: Example 2

  1. Type in the command window:

  ```
  ls log(m1) c @trend @trend^2
  ```

  2. Press **Enter.**

- For a more detailed discussion on time trends, interpretation, and de-trending, see the companion to this tutorial: *Appendix: Time Series Models.*

Note:** This equation is saved as ***eq11 in ***WS4_Results.wf1*** workfile.

Equation: EQ11 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: 23:27
Sample: 1960M01 2011M12
Included observations: 624

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.695590	0.012616	372.1939	0.0000
@TREND	0.006290	9.35E-05	67.23958	0.0000
@TREND^2	-2.85E-06	1.45E-07	-19.57715	0.0000

R-squared	0.983742	Mean dependent var	6.286356
Adjusted R-squared	0.983690	S.D. dependent var	0.825190
S.E. of regression	0.105386	Akaike info criterion	-1.657580
Sum squared resid	6.896939	Schwarz criterion	-1.636253
Log likelihood	520.1651	Hannan-Quinn criter.	-1.649292
F-statistic	18788.10	Durbin-Watson stat	0.004268
Prob(F-statistic)	0.000000		

Seasonality

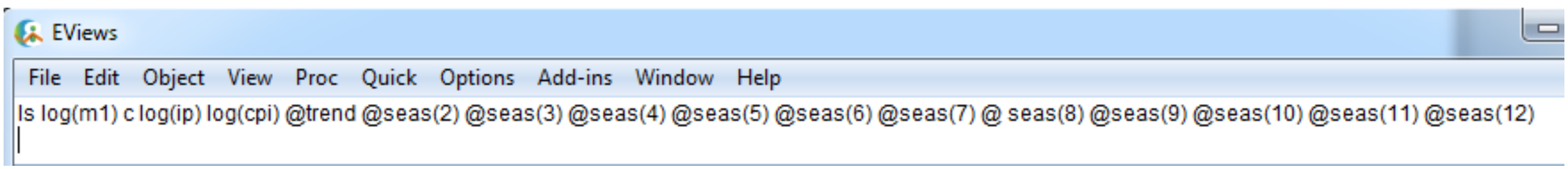
- Sometimes time series exhibit seasonal patterns. For example, retails sales tend to be higher in the last quarter of the year because of the Holiday Shopping season.
- Most macroeconomic series are already seasonally adjusted beforehand so there is no need to worry about seasonal issues.
- If you suspect 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 (cont'd)

Regression with seasonal factors:

1. Type in the command window:

```
ls log(m1) c log(ip) log(cpi) @trend @seas(2) @seas(3) @seas(4) @seas(5)
@seas(6) @seas(7) @ seas(8) @seas(9) @seas(10) @seas(11) @seas(12)
```



2. Press **Enter** (please type the command in one line).

Note:** This equation is saved as ***eq12 in ***WS4_Results.wf1*** workfile.

Seasonality (cont'd)

- 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).

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		

Testing for 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.

Serial Correlation (cont'd)

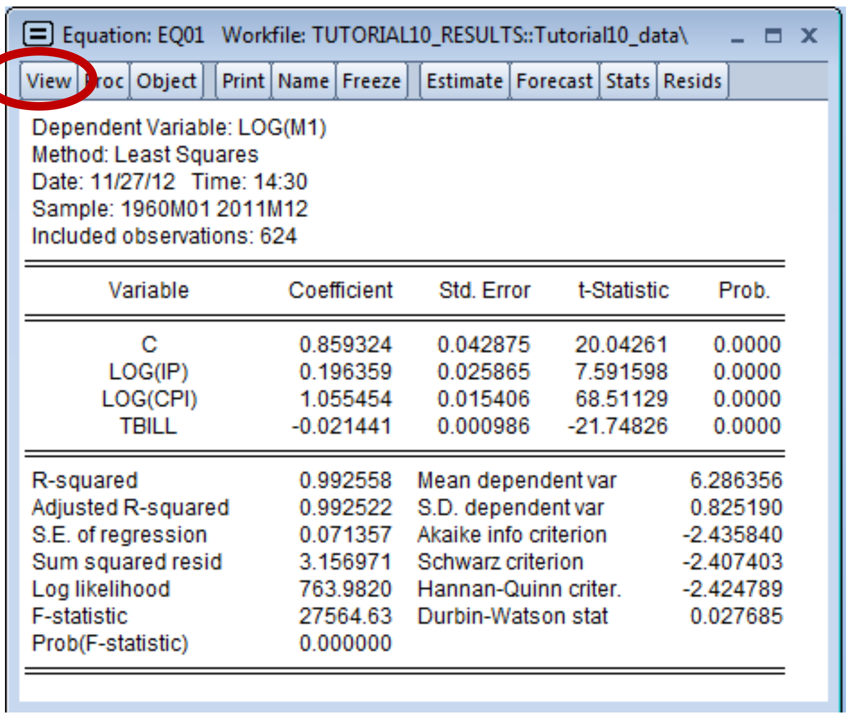
- If untreated, serial correlation leads to a number of issues:
  - ✓ Reported standard errors and t-statistics are invalid (even asymptotically).
  - ✓ Coefficients may be biased, though not necessarily inconsistent (if data is weakly dependent).
  - ✓ In the presence of lagged dependent variables,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

- Fortunately, EViews provides tools for detecting serial correlation and correcting regressions to account for its presence.

Note: in the following examples we will treat serial correlation as a statistical issue. However, serial correlation is sometimes a hint of model misspecification. Though we don't investigate this here, misspecification is likely an issue in the following examples.

Detecting Serial Correlation Visual Inspection: Example 1

- You can visually inspect your residuals to see if serial correlation is present.
  1. Open ***eq01*** in the ***Results.wf1*** workfile.
  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.



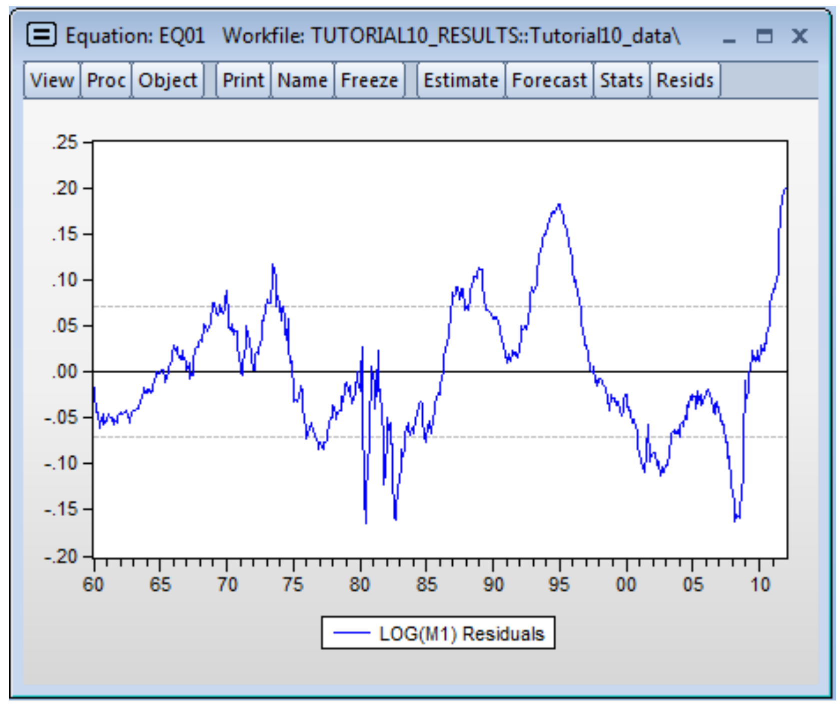
Equation: EQ01 Workfile: TUTORIAL10_RESULTS::Tutorial10_data\

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

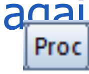
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.859324	0.042875	20.04261	0.0000
LOG(IP)	0.196359	0.025865	7.591598	0.0000
LOG(CPI)	1.055454	0.015406	68.51129	0.0000
TBILL	-0.021441	0.000986	-21.74826	0.0000

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: Example 2

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

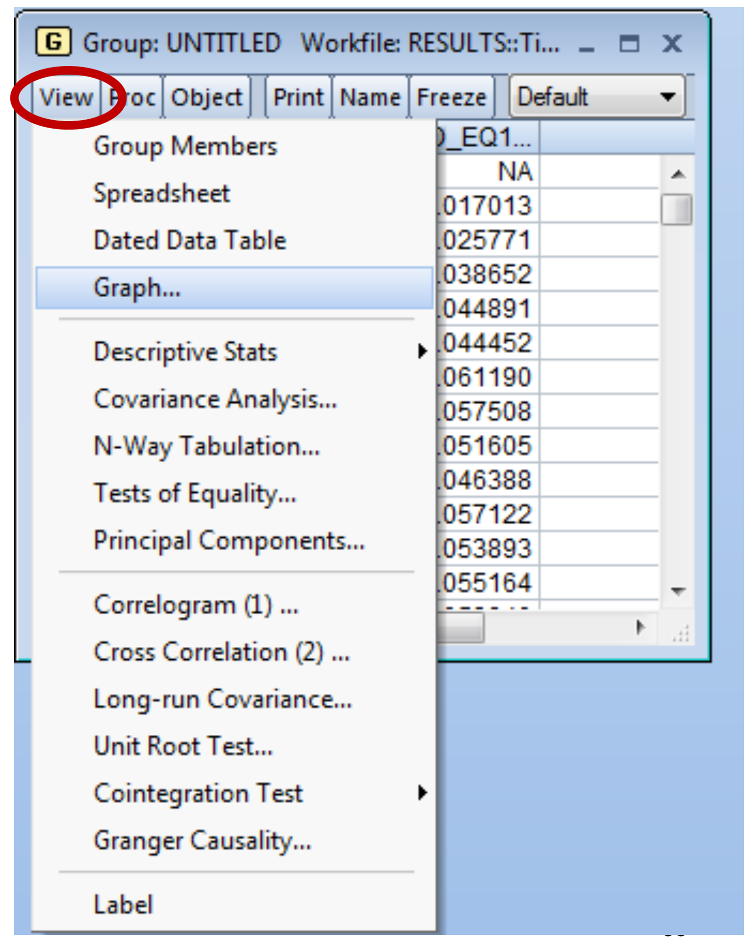


<u>Scatterplot of residuals against lagged values:</u>

1. Click on ***eq01*** to open it. Click the 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):

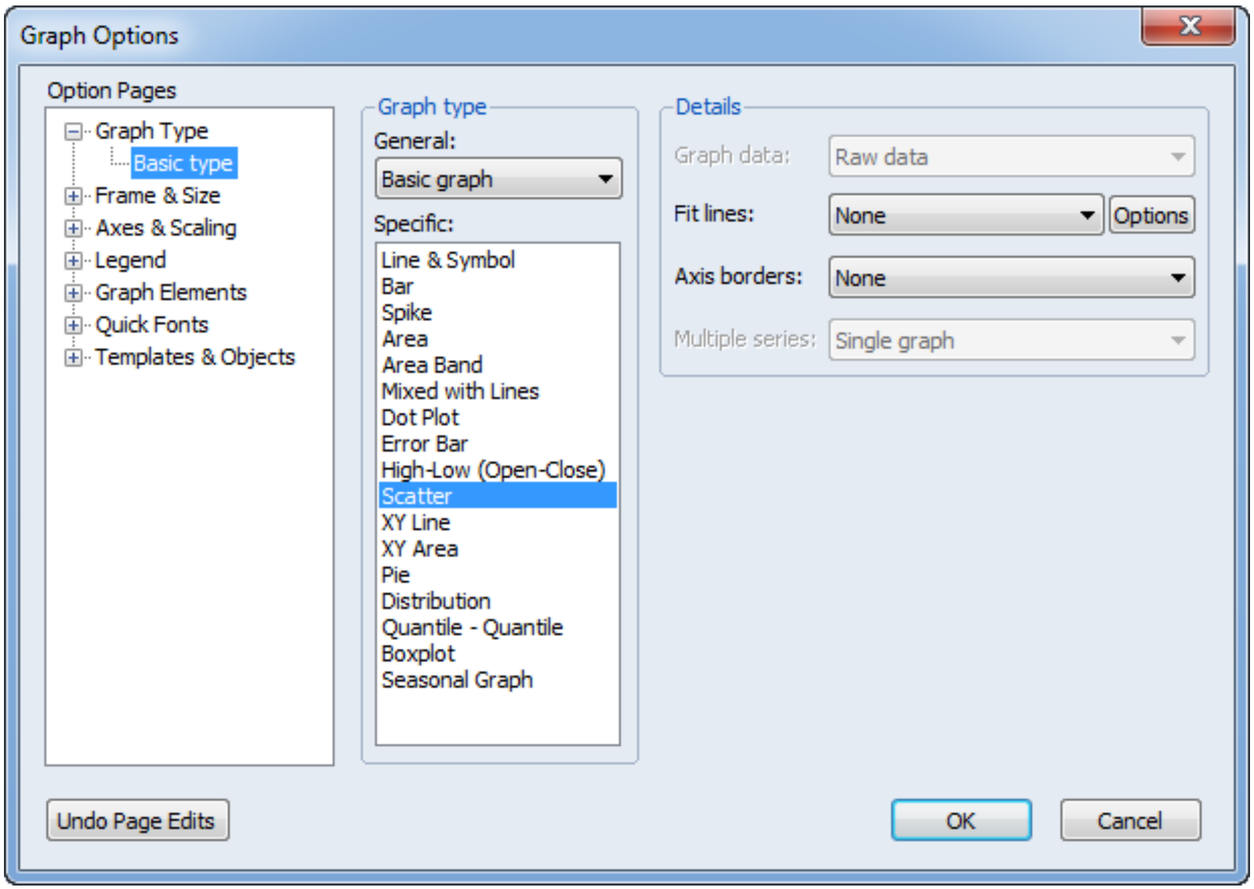
```
show resid_eq1 resid_eq1(-1)
```

4. The **Group Spreadsheet** opens up. On the top menu of the Group Spreadsheet, select **View → Graph**



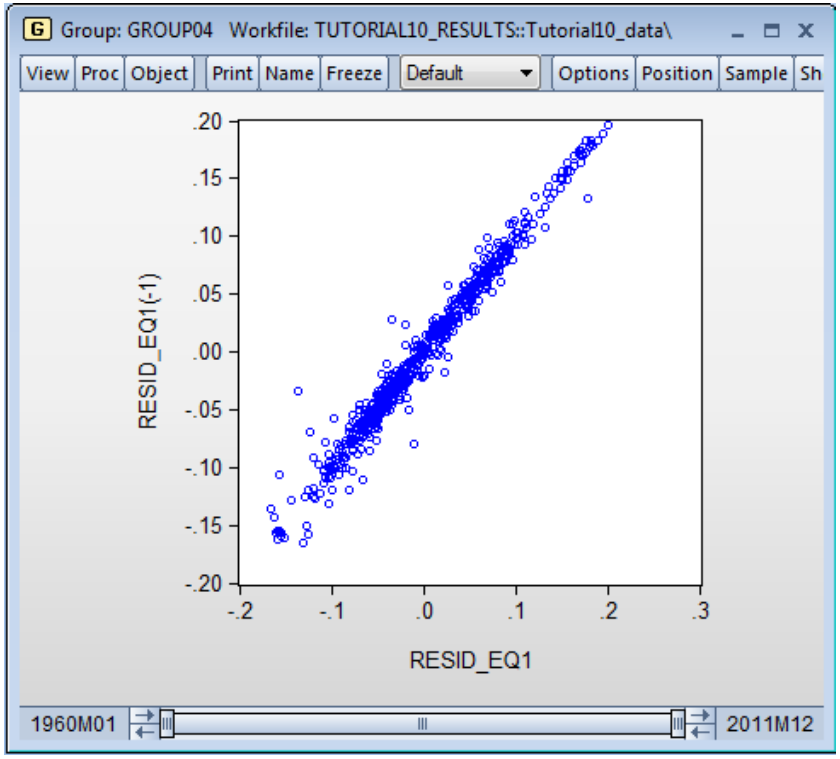
Detecting Serial Correlation
Visual Inspection: Example 2 (cont'd)

5. The **Graph Options** dialog box opens up. Select **Scatter** under ***Graph type* → *Specific*.**
6. Click **OK.**



Detecting Serial Correlation Visual Inspections: Example 2 (cont'd)

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

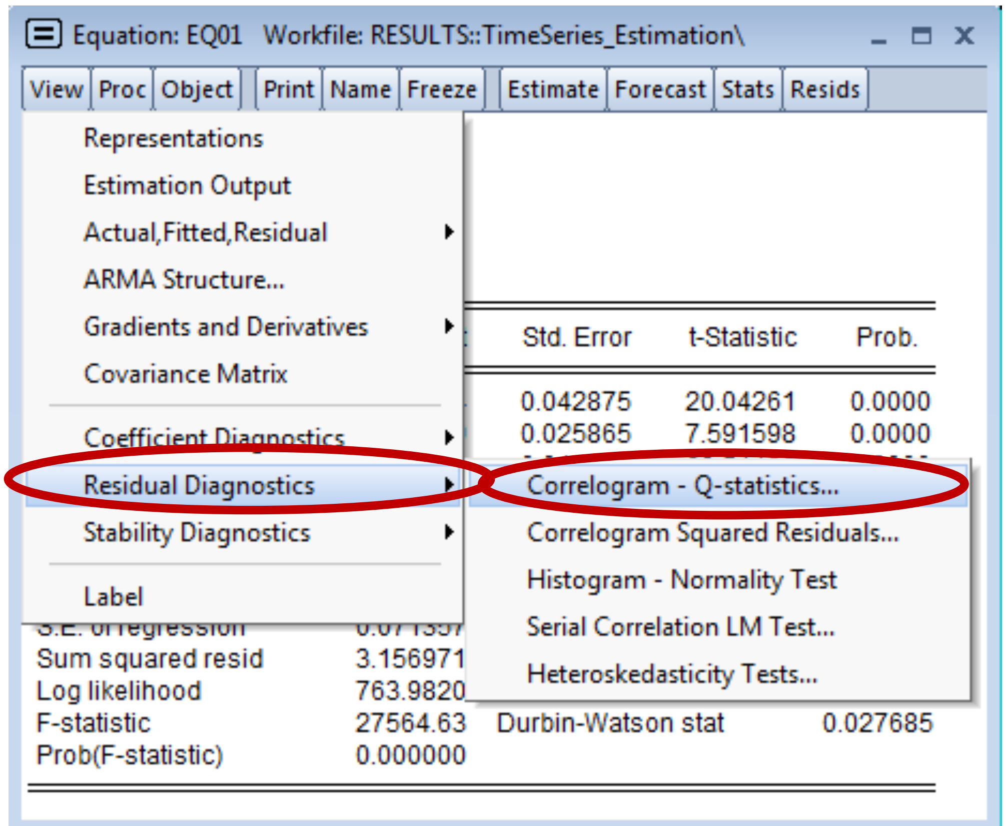


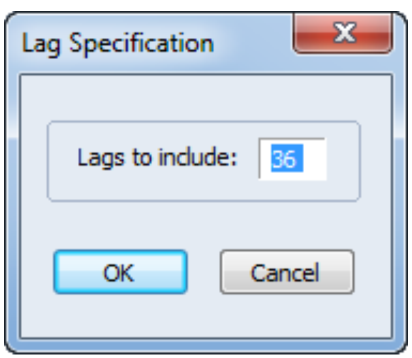
Detecting Serial Correlation Visual Inspection: Example 3

- 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. Click on ***eq01*** to open it. On the ***Equation*** box, click **View → Residual Diagnostics → Correlogram-Q-Statistic.**
2. The ***Lag Specification*** box opens up. Select the number of lags (the default, 36, is fine).
3. Click **OK.**

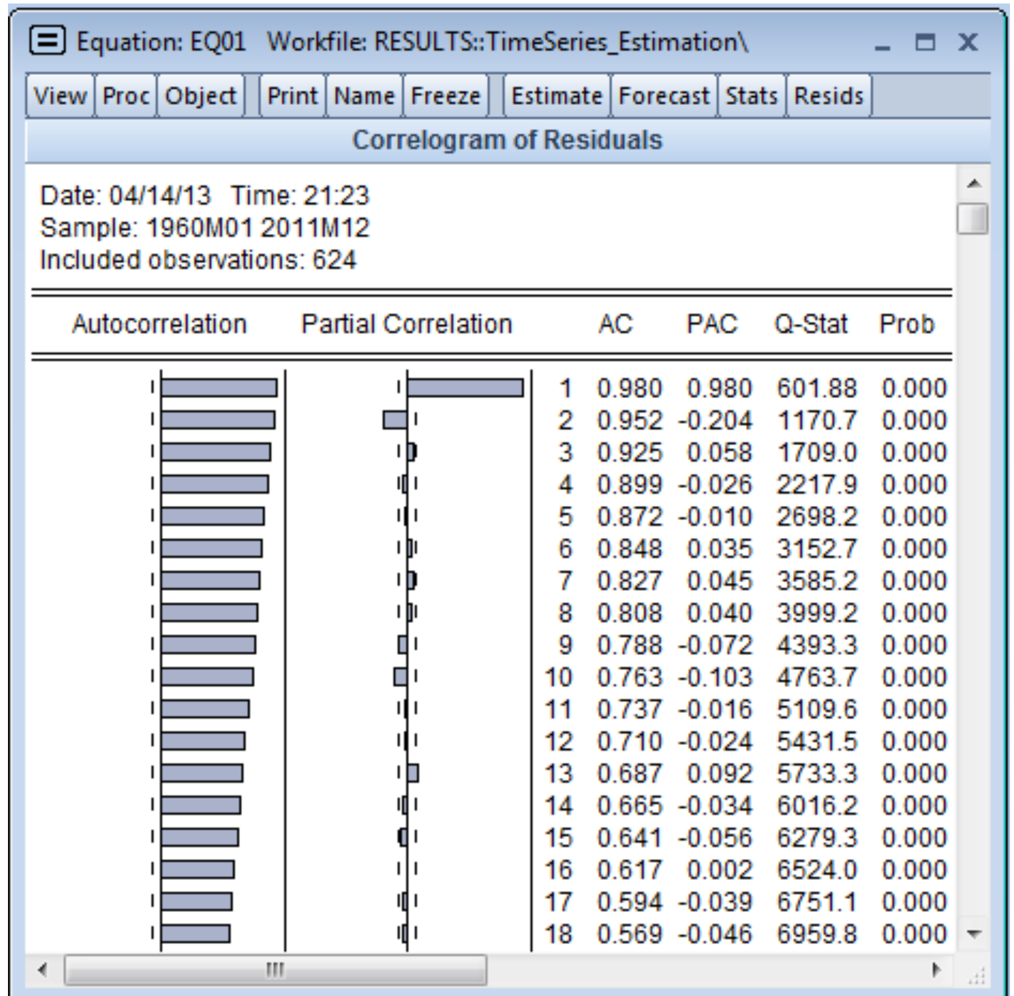




Detecting Serial Correlation Visual Inspection: Example 3

- 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.

Note:** This output is saved as ***table01 in ***WS4_Results.wf1*** workfile.



	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

Testing for Serial Correlation in EViews

- 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 (or the correlogram; explained above).

Testing for 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. Notice that EViews does not compute p-values for the DW statistic.
  - ✓ In our case, the *DW=0.02768*, which means we soundly 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		

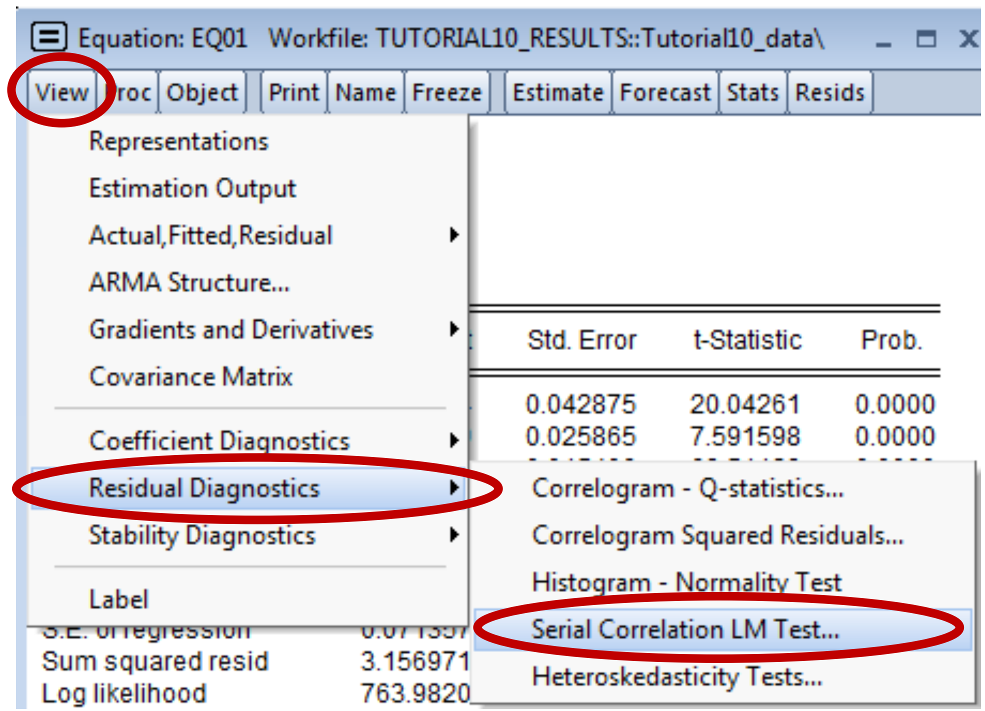
Testing for Serial Correlation
Breusch-Godfrey Test: Example 1

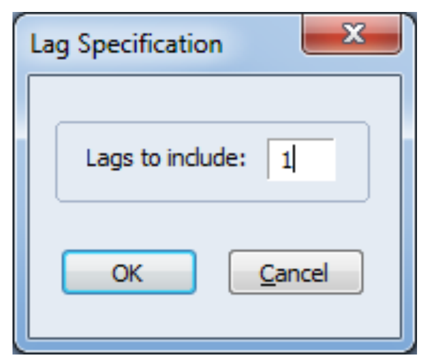
- A more general test for serial correlation is the Breusch-Godfrey test.

  Breusch- Godfrey Test: Example 1

  1. Click on ***eq01*** to open it. 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***.
  3. Click **OK.**

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





Testing for Serial Correlation Breusch-Godfrey Test: Example 1 (cont'd)

- Results are shown here.
  - ✓ The top panel reports the test statistics in two versions: the F-statistic and the Chi-squared statistics (either one is fine). 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.
  - ✓ Since we are testing for first order serial correlation, there is only one residual lag in the auxiliary regression.
  - ✓ 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		

Note:** This output is saved as ***table02 in ***WS4_Results.wf1*** workfile.

Correcting for Serial Correlation: ARMA Models

Correcting for Serial Correlation

- If you detect serial correlation, something 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 re-specify 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 for Serial Correlation (cont'd)

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

<u>Estimate the model:</u>

1. Enter in the command window:

```
ls tbill c log(m1) log(cpi)
log(ip) @trend
```

2. Click **OK** (please type the command in one line).

- There is plenty of evidence that errors are serially correlated based on:
  - ✓ DW-statistic (which is very small)
  - ✓ 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

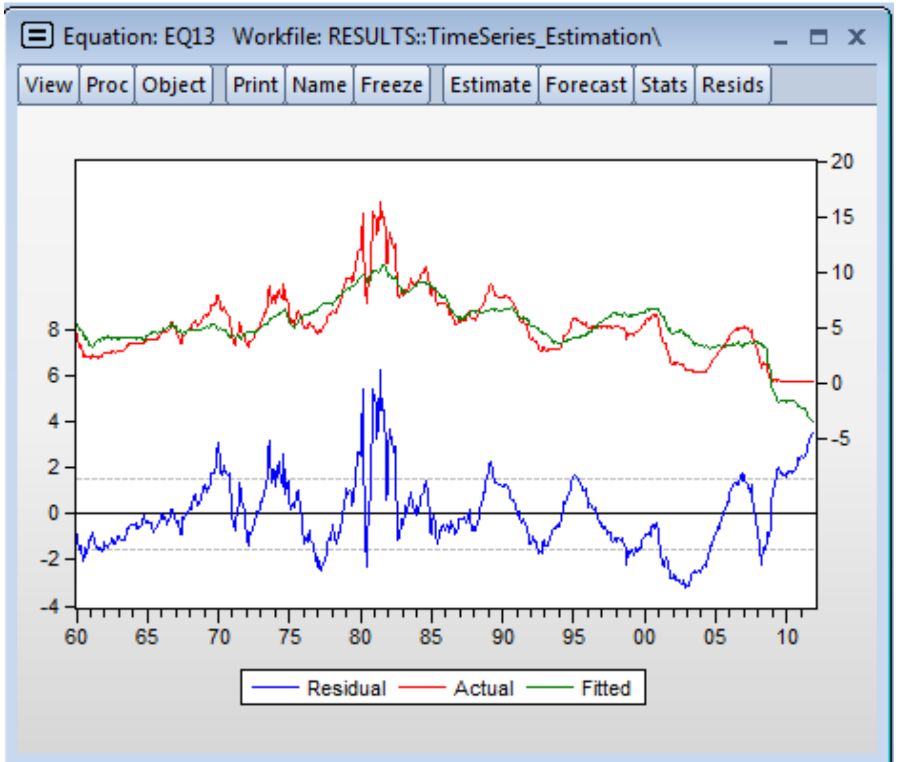
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-59.43675	3.474332	-17.10739	0.0000
LOG(M1)	-10.24916	0.765226	-13.39365	0.0000
LOG(CPI)	20.57633	0.760181	27.06767	0.0000
LOG(IP)	14.37032	0.746583	19.24812	0.0000
@TREND	-0.068292	0.002726	-25.05265	0.0000

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	$\boxed{\text{Akaike info criterion}}$	$\boxed{3.714573}$
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		

Note:** This equation is saved as ***eq13 in ***WS4_Results.wf1*** workfile.

Correcting for Serial Correlation (cont'd)

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



Note:** This output is saved as ***table04 in ***WS4_Results.wf1*** workfile.

- 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\

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 for Serial Correlation: AR(p) models: Example 1

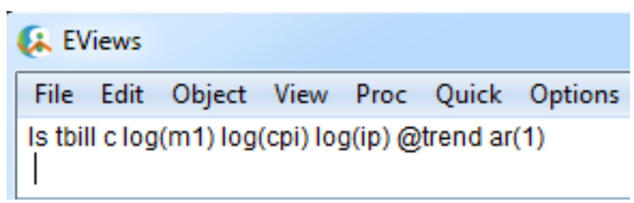
- Now let's correct for serial correlation in this model. Suppose you suspect the error term in the previous model follow an AR(1) process.
- You can specify an AR(1) model in EViews very easily.

Example 1: *AR(1)* Model:

1. Type in the command window the same equation as before, adding the term ***AR(1)*** as if it were another variable.

```
ls tbill c log(m1) log(cpi) log(ip) @trend ar(1)
```

2. Press **Enter**.



Note:** This equation is saved as ***eq14 in ***WS4_Results.wf1*** workfile.

Correcting for Serial Correlation AR(p) models: Example 1 (cont'd)

- Results are shown here.

A few notes:

- ✓ 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.

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 for Serial Correlation AR(p) models: Example 1 (cont'd)

A few notes (cont'd):

- ✓ 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. The results above are for the default ML method.
- ✓ You can change the ARMA estimation method using the Options tab of the estimation dialog.
- ✓ The results of the same estimation in versions prior to EViews 9 are shown below:

Equation: EQ14 Workfile: RESULTS::TimeSeries_Estimation\

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

Dependent Variable: TBILL
Method: ARMA Conditional Least Squares (Marquardt - EViews legacy)
Date: 07/29/15 Time: 11:48
Sample (adjusted): 1960M02 2011M12
Included observations: 623 after adjustments
Convergence achieved after 14 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-77.52357	15.12180	-5.126610	0.0000
LOG(M1)	-1.530652	2.423761	-0.631519	0.5279
LOG(CPI)	13.77027	3.447891	3.993825	0.0001
LOG(IP)	13.83459	2.207270	6.267738	0.0000
@TREND	-0.080538	0.014372	-5.603997	0.0000
AR(1)	0.967406	0.011118	87.01414	0.0000

R-squared	0.978478	Mean dependent var	5.128764
Adjusted R-squared	0.978303	S.D. dependent var	2.960070
S.E. of regression	0.436012	Akaike info criterion	1.187291
Sum squared resid	117.2957	Schwarz criterion	1.229999
Log likelihood	-363.8410	Hannan-Quinn criter.	1.203888
F-statistic	5610.198	Durbin-Watson stat	1.468803
Prob(F-statistic)	0.000000		

Inverted AR Roots .97

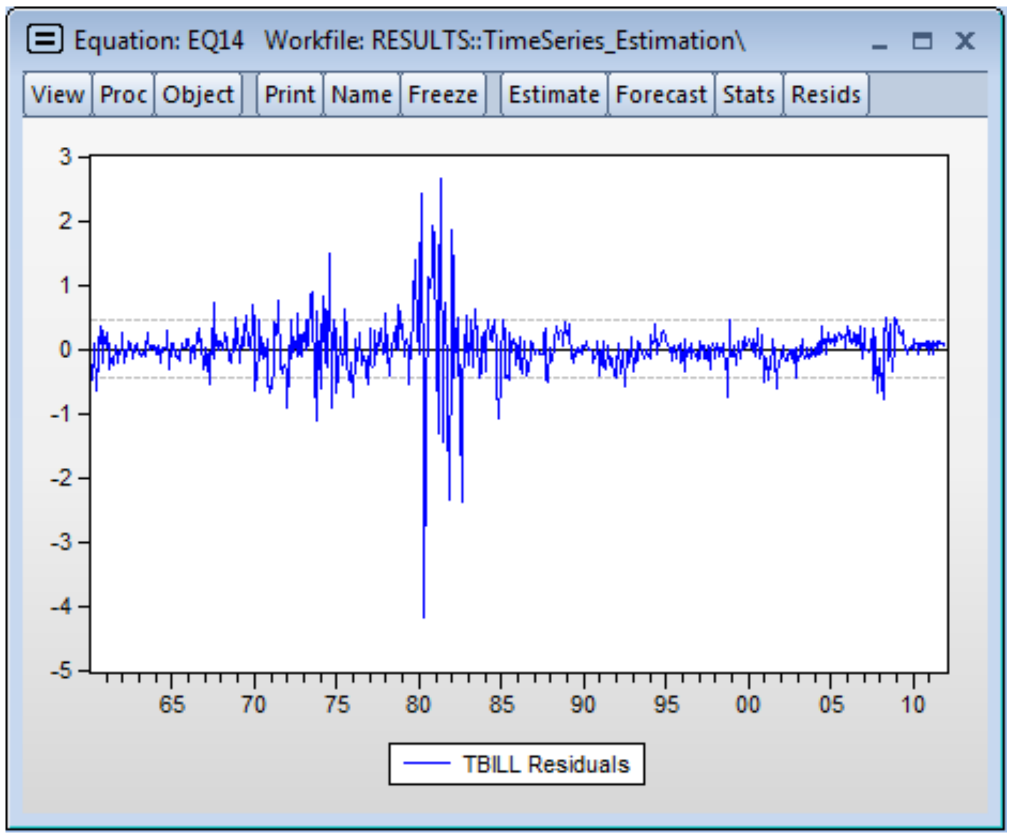
Correcting for Serial Correlation AR(p) models: Example 1 (cont'd)

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

<u>Residual Plot:</u>

1. Click on ***eq14*** to open it. 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 for Serial Correlation AR(p) models: Example 1 (cont'd)

- Breusch-Godfrey Test:
  1. Re-estimate ***eq14*** 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***.
  4. 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 good specification (it does not fully address serial correlation).

Note:** This output is saved as ***table05 in ***WS4_Results.wf1*** workfile.

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 for Serial Correlation AR(p) models: Example 2

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

Example 2*: AR(p)* Model

1. If you suspect that the error term follows an AR(2) process, then add terms *AR(1)* and *AR(2)* to your regression command:

```
ls tbill c log(m1) log(cpi)
log(ip) @trend   ar(1) ar(2)
```

2. Press **Enter** (please type the command in one line).

- Notice that now both *AR(1)* and *AR(2)* terms are estimated. They are both statistically significant.

Note:** This equation is saved as ***eq15 in ***WS4_Results.wf1*** workfile.

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 for Serial Correlation: AR(p) models: Example 3

- You can just as easily specify non-contiguous ***AR(p)*** terms (i.e., *ar(1), ar(4),* etc.)

Example 3*: AR(p)* Model

1. You can add non-contiguous AR terms by typing them directly in the command window (or equation box):

```
ls tbill c log(m1) log(cpi)
log(ip) @trend ar(1) ar(3)
ar(5)
```

2. Press **Enter** (please type the command in one line).

- Notice that all *AR* terms are estimated.

Note:** This equation is saved as ***eq16 in ***WS4_Results.wf1*** workfile.

Equation: EQ16 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: 11:53
Sample: 1960M01 2011M12
Included observations: 624
Convergence achieved after 14 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-70.40118	17.99168	-3.912985	0.0001
LOG(M1)	-2.009392	2.776238	-0.723782	0.4695
LOG(CPI)	13.68657	4.277505	3.199661	0.0014
LOG(IP)	12.38062	2.072393	5.974068	0.0000
@TREND	-0.074182	0.018526	-4.004183	0.0001
AR(1)	1.020976	0.016580	61.57763	0.0000
AR(3)	-0.083437	0.018295	-4.560583	0.0000
AR(5)	0.025958	0.014411	1.801253	0.0722
SIGMASQ	0.186274	0.004584	40.63559	0.0000

R-squared	0.978675	Mean dependent var	5.127516
Adjusted R-squared	0.978397	S.D. dependent var	2.957858
S.E. of regression	0.434742	Akaike info criterion	1.190544
Sum squared resid	116.2353	Schwarz criterion	1.254527
Log likelihood	-362.4498	Hannan-Quinn criter.	1.215408
F-statistic	3527.996	Durbin-Watson stat	1.565888
Prob(F-statistic)	0.000000		

Inverted AR Roots	.96	.33+.30i	.33-.30i	-.30+.22i
	-.30-.22i			

Correcting for Serial Correlation: AR(p) models (cont'd)

A couple of additional notes:

- ✓ As shown above, EViews allows you to include non-contiguous ***AR*** terms (***ar(2) ar(4),*** etc.)
- ✓ 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)*** (or, in EViews 8, ***ar(1 to 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.

Correcting for Serial Correlation MA(q) models: Example 1

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

Example 1: MA(1) Model

1. If you suspect your errors follow an MA(1) process, then add term *MA(1)* in your regression command:

```
ls tbill c log(m1)log(cpi)
log(ip) @trend ma(1)
```

2. Press **Enter** (please type the command in one line).

Note:** This equation is saved as ***eq18 in ***WS4_Results.wf1*** workfile.

Equation: EQ18 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: 11:53
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	-58.29338	4.560925	-12.78104	0.0000
LOG(M1)	-10.01354	0.891650	-11.23036	0.0000
LOG(CPI)	20.34622	0.686036	29.65768	0.0000
LOG(IP)	13.90759	0.817128	17.02009	0.0000
@TREND	-0.067423	0.003613	-18.65948	0.0000
MA(1)	0.906602	0.016854	53.79295	0.0000
SIGMASQ	0.745412	0.043239	17.23938	0.0000

R-squared	0.914663	Mean dependent var	5.127516
Adjusted R-squared	0.913833	S.D. dependent var	2.957858
S.E. of regression	0.868256	Akaike info criterion	2.569260
Sum squared resid	465.1369	Schwarz criterion	2.619024
Log likelihood	-794.6090	Hannan-Quinn criter.	2.588598
F-statistic	1102.191	Durbin-Watson stat	0.492218
Prob(F-statistic)	0.000000		

Inverted MA Roots -.91

Correcting for Serial Correlation MA(q) models: Example 2

- You can just as easily specify higher order or non-contiguous MA terms.

Example 2: MA(q) Model

1. If you want to specify an MA(3) process, type in the command window (note: this is a feature of EViews 8):

```
ls tbill c log(m1) log(cpi) log(ip) @trend ma(1 to 3)
```

2. Press **Enter** (please type the command in one line).

- Note: For non-contiguous MA terms specify all MA terms in your equation.

*__Note:__ This equation is saved as ***eq19*** in ***WS4_Results.wf1*** workfile.

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

Correcting for Serial Correlation: MA(q) models (cont'd)

Some additional notes:

- ✓ As it must be obvious by now, if your errors follow an MA(2) process, you need to include both ***ma(1)*** and ***ma(2)*** 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 for Serial Correlation ARMA(p,q) Models: Example 1

- ARMA models (autoregressive and moving average errors) are also estimated very easily in EViews.

Example 1: ARMA(1,1) Model

1. Suppose you suspect your errors follow an ARMA(1,1) process. To specify this, add AR and MA terms in your regression command:

   ```
   ls tbill c log(m1) log(cpi)
   log(ip) @trend ar(1) ma(1)
   ```

2. Press **Enter** (please type the command in one line).

- Results show both AR and MA terms.

*__Note:__ This equation is saved as ***eq20*** in ***WS4_Results.wf1*** workfile.

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 8 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-54.95901	16.83223	-3.265105	0.0012
LOG(M1)	-2.766148	2.515193	-1.099776	0.2719
LOG(CPI)	13.33536	4.284563	3.112420	0.0019
LOG(IP)	9.209863	2.037330	4.520555	0.0000
@TREND	-0.062635	0.017744	-3.530000	0.0004
AR(1)	0.944015	0.015368	61.42789	0.0000
MA(1)	0.421720	0.014887	28.32834	0.0000
SIGMASQ	0.165917	0.004893	33.91041	0.0000

R-squared	0.981005	Mean dependent var	5.127516
Adjusted R-squared	0.980789	S.D. dependent var	2.957858
S.E. of regression	0.409965	Akaike info criterion	1.072191
Sum squared resid	103.5320	Schwarz criterion	1.129065
Log likelihood	-326.5237	Hannan-Quinn criter.	1.094292
F-statistic	4544.873	Durbin-Watson stat	2.054379
Prob(F-statistic)	0.000000		

Inverted AR Roots	.94
Inverted MA Roots	-.42

Correcting for Serial Correlation ARMA(p,q) Models: Example 2

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

Example 2: ARMA(p,q) Model

1. Suppose you suspect your errors follow an ARMA(2,1) process. To specify this, type in the command window:

```
ls tbill c log(m1) log(cpi)
log(ip) @trend ar(1 to 2) ma(1)
```

2. Press **Enter** (please type the command in one line).

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	

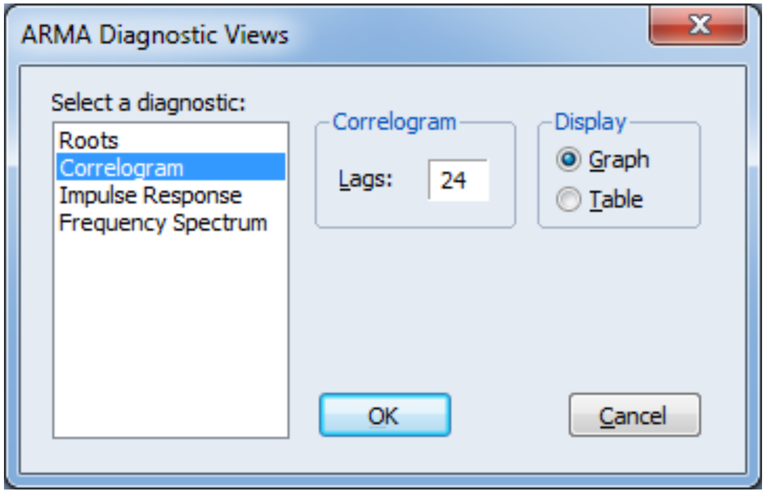
Note:** This equation is saved as ***eq21 in ***WS4_Results.wf1*** workfile.

ARMA Equation Diagnostics: Correlogram

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

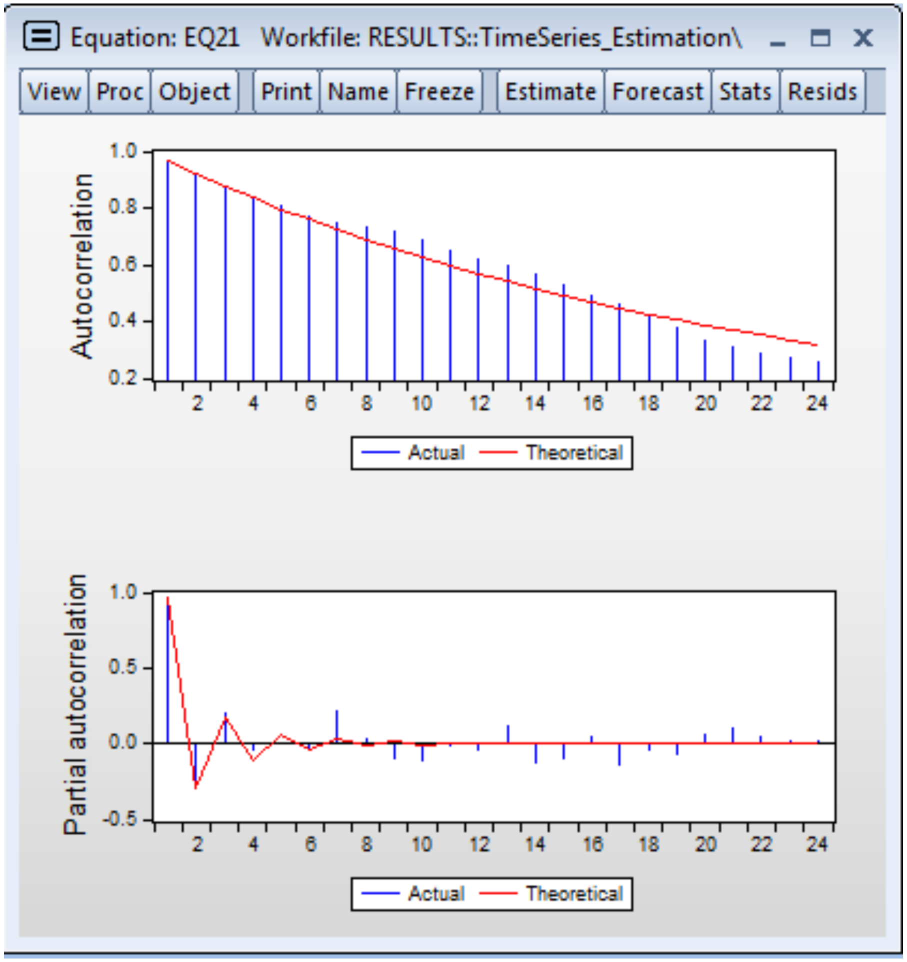
ARMA Correlogram:

1. Click on ***eq21*** to open it. 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).



ARMA Equation Diagnostics: Correlogram (cont'd)

- The graph shows autocorrelations and partial correlations for:
  - ✓ Theoretical correlogram (red line) – this corresponds to ARMA terms.
  - ✓ Empirical correlogram of residuals (blue spikes) -- this corresponds 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

- 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 (cont'd)

- Let's first estimate the following model in levels.

  Estimating in levels:

  1. Type in the command window

     ```
     ls tbill c log(ip)
     tbill(-1) @trend
     ```

  2. Click **OK**.

- 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		

Note:** This equation is saved as ***eq22 in ***WS4_Results.wf1*** workfile.

Differencing and Serial Correlation (cont'd)

- Now let's estimate the same model in first differences.
- Remember that EViews allows you to difference the data very easily by typing ***d()*** before the name of the variable or ***dlog*** (if you want to take log differences).

Estimating the model in first-differences:

1. Type in the command window:

```
ls d(tbill) c dlog(ip) d(tbill(-1))@trend
```

2. Press **Enter**.

*__Note:__ This equation is saved as ***eq23*** in ***WS4_Results.wf1*** workfile.

Differencing and Serial Correlation (cont'd)

- Results are shown here.

  <u>A few things:</u>

  - ✓ 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 (compared to levels).

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		