ES1004 Econometrics by Example

Lecture 13: Economic Forecasting

Dr. Hany Abdel-Latif

Swansea University, UK

Gujarati textbook, second edition [chapter 16]

25th February 2017

Time Series Econometrics

13. stationary and nonstationary time series
14. cointegration and error correction models
15. asset price volatility: the ARCH and GARCH models
16. economic forecasting

- previous course on time series econometrics

ES1002 Lectures

ES1002 EViews

The Economic Society

Economic Forecasts

- Economics [GDP, unemployment, consumption, investment, interest rates]
- financial asset management [asset returns, exchange rates and commodity prices]
- financial risk management [asset return volatility]
- marketing [response of sales to different marketing schemes]
- business and government [revenue forecasts]
- crisis management [probabilities of default, currency devaluations]

Econometric Models

- based on past and current information, the objective of forecasting is to provide quantitative estimate(s) of the likelihood of the future course of the object of interest
    - e.g. personal consumption expenditure
- we develop econometric models and use one or more methods of forecasting its future course

Methods of Forecasting

- there are several methods of forecasting
- we will consider three prominent methods of forecasting in this chapter
  1. regression models,
  2. autoregressive integrated moving average (ARIMA) models [Box–Jenkins (BJ) methodology]
  3. vector autoregression (VAR) models (Sims)

Example: Consumption Function

- forecasting is probably the most important purpose of estimating regression models
- consider the following bivariate regression

$$PCE_t = \beta_1 + \beta_2 PDI_t + u_t$$

where
PCE = per capita personal consumption expenditure,
PDI = per capita personal disposable

Example: Consumption Function

$$PCE_t = \beta_1 + \beta_2 PDI_t + u_t$$

- the slope coefficient represents the MPC
- to estimate this regression, we obtained aggregate data on these variables for the US for 1960–2008
- see `Table16.1` on Piazza

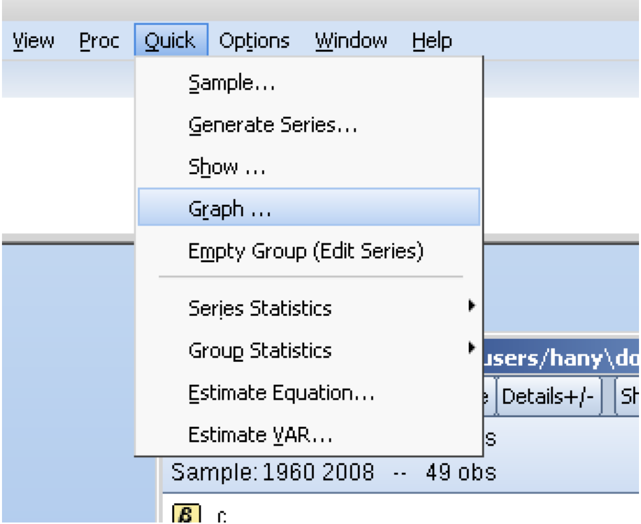
Holdover Sample

- to estimate the consumption function, initially we use the observations from 1960–2004
- save the last four observations
  - called the holdover sample
  - to evaluate the performance of the estimated model

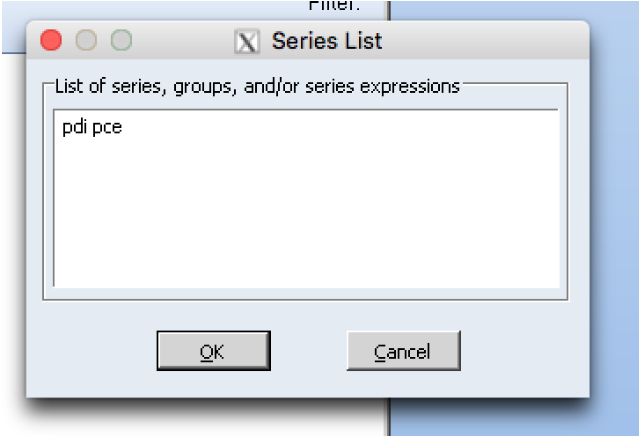
Plot the Data

We first plot the data to get some idea of the nature of the relationship between the two variables (Figure 16.1).

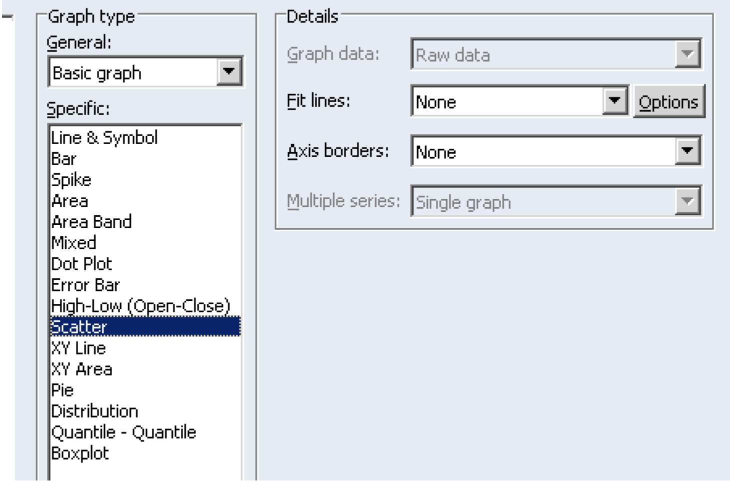
Per capita PCE & PDI, USA 1960–2004



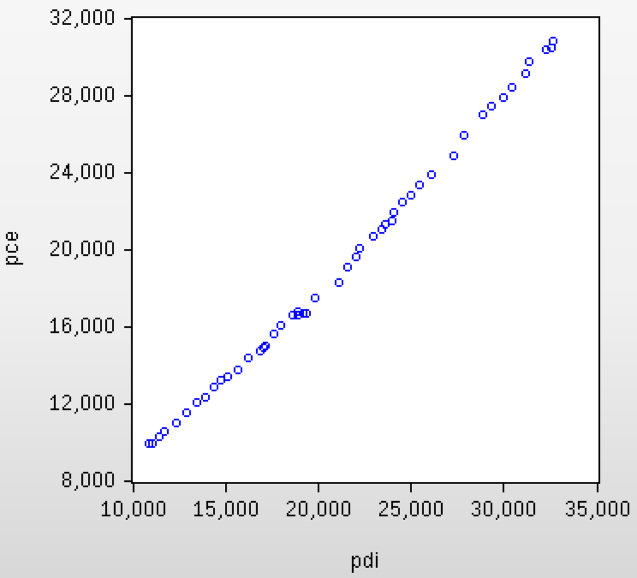
Per capita PCE & PDI, USA 1960–2004



Per capita PCE & PDI, USA 1960–2004



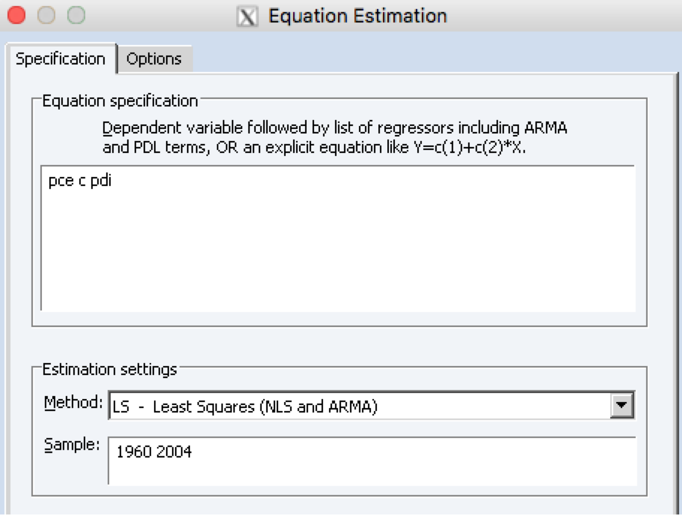
Per capita PCE & PDI, USA 1960–2004



Example: Estimation

- we plot the data to get some idea of the nature of the relationship between the two variables
- the figure shows almost a linear relationship between PCE and PDI
- we fitting a linear regression model to the data

Estimates of the consumption function, 1960–2004



Estimates of the consumption function, 1960–2004

Dependent Variable: PCE
Method: Least Squares
Date: 02/12/17 Time: 23:48
Sample: 1960 2004
Included observations: 45

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1083.978	193.9579	-5.588729	0.0000
PDI	0.953768	0.009233	103.2981	0.0000

R-squared	0.995986	Mean dependent var	18197.91
Adjusted R-squared	0.995893	S.D. dependent var	5515.914
S.E. of regression	353.4907	Akaike info criterion	14.61702
Sum squared resid	5373095.	Schwarz criterion	14.69731
Log likelihood	-326.8829	Hannan-Quinn criter.	14.64695
F-statistic	10670.51	Durbin-Watson stat	0.299775
Prob(F-statistic)	0.000000		

Estimated Consumption Function

$$\hat{PCE}_t = -1083.978 + 0.9537 PDI_t$$

- we can use this regression model to forecast the future value(s) PCE
- suppose we want to find out $E(PCE_{2005}|PDI_{2005})$ $[PDI_{2005} = \$31,318]$

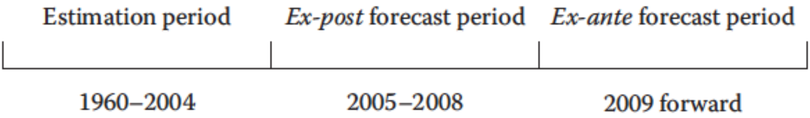
Point & Interval Forecasts

- in point forecasts we provide a single value for each forecast period
- in interval forecasts we obtain a range, or an interval, that will include the realized value with some probability
- the interval forecast provides a margin of uncertainty about the point forecast

ex post & *ex ante*

- estimation period: we have data on all the variables in the model
- *ex post* forecast period: we also know the values of the regressand and regressors
    - the holdover period - used to get some idea about the performance of the fitted model
- *ex ante* forecast we estimate the values of the depend variable beyond the estimation period but we may not know the values of the regressors with certainty

ex post & *ex ante*



Conditional & Unconditional Forecasts

- conditional forecasts: we forecast the variable of interest conditional on the assumed values of the regressors
- recall that all along we have conducted our regression analysis, conditional on the given values of the regressors.
- this type of conditional forecasting is also known as scenario analysis or contingency analysis

Conditional & Unconditional Forecasts

- in unconditional forecasts, we know the values of the regressors with certainty instead of picking some arbitrary values of them, as in conditional forecasting
- of course, that is a rarity; it actually involves the forecasting the right-hand side variables (i.e. regressors) problem
- for the present purposes we will work with conditional forecasts

Point Forecast

- estimate the point forecast of consumption expenditure for 2005, given that $PDI_{2005} = \$31,318$ billion

$$\begin{aligned}
\hat{PCE}_{2005} &= b_1 + b_2 PDI_{2005} \\
&= -1083.978 + 0.953768(31318) \\
&= 28783.998 \\
&\approx 28784
\end{aligned}$$

Forecast Error

- the best mean predicted value of PCE in 2005 is $\$28,784$ billion [given $PDI_{2005} = \$31,318$]
- in `table16.1` the actual value of PCE for 2005 was $\$29,771$ billion
- the actual value was greater than the estimated value by $987 billion
- this is the forecast error

Forecast Band I

- since the *PCE* figure is an estimate, it is subject to an error
- should estimate of the forecast error we are likely to make
- if the error term u is normally distributed, then $\hat{Y}_{2005}$ is normally distributed with mean equal to $(\beta_1 + \beta_2 X_{2005})$

$$var(\hat{Y}_{2005}) = \sigma^2 \left[\frac{1}{n} + \frac{(X_{2005} - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right]$$

$\bar{X}$ sample mean of the X values $(1960 - 2004)$,
σ^2 variance of the error term u,
n sample size

Forecast Band II

- since we do not observe the true variance u, we estimate from the sample

$$\hat{\sigma}^2 = \sum e_t^2/(n-2)$$

given the X value for 2005 we can establish a 95% confidence interval for true $E(Y_{2005})$

$$Pr[\hat{Y}_{2005} - t_{\alpha/2} se(\hat{Y}_{2005}) \leq E(Y_{2005}) \leq \hat{Y}_{2005} + t_{\alpha/2} se(\hat{Y}_{2005})] = 95\%$$

$se(\hat{Y}_{2005})$ standard error obtained from $var(\hat{Y}_{2005})$
$\alpha = 5\%$

Forecast Band III

- 95% confidence interval for $E(Y_{2005})$ is ($28,552 billion, $29,019)
- although the single best estimate is $28,784 billion
- we will have to compute such confidence interval for each $E(Y|X)$ in our sample
- if we connect such confidence intervals, we obtain confidence band

Forecast Band V

Equation: UNTITLED Workfile: TABLE16_1::Table16_1\

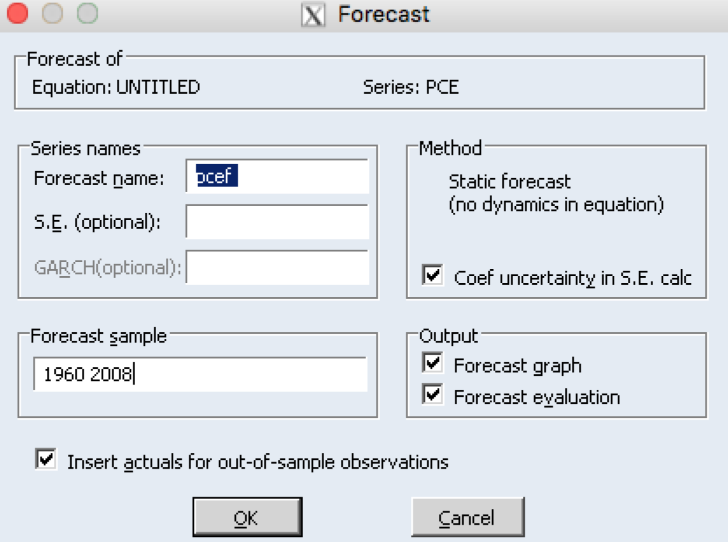
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PCE
Method: Least Squares
Date: 02/25/17 Time: 08:15
Sample: 1960 2004
Included observations: 45

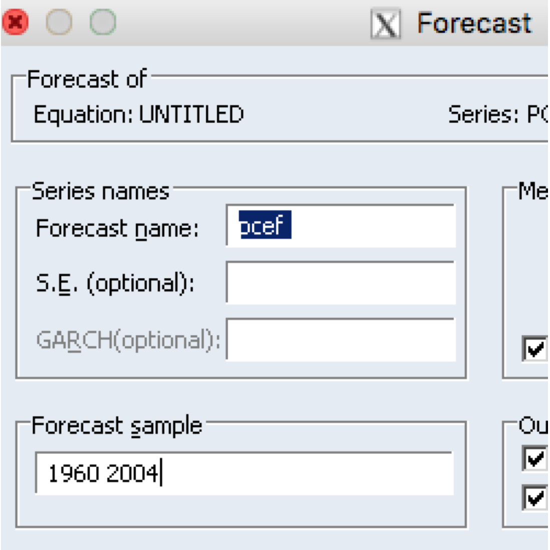
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1083.978	193.9579	-5.588729	0.0000
PDI	0.953768	0.009233	103.2981	0.0000

R-squared	0.995986	Mean dependent var	18197.91
Adjusted R-squared	0.995893	S.D. dependent var	5515.914
S.E. of regression	353.4907	Akaike info criterion	14.61702
Sum squared resid	5373095.	Schwarz criterion	14.69731
Log likelihood	-326.8829	Hannan-Quinn criter.	14.64695
F-statistic	10670.51	Durbin-Watson stat	0.299775
Prob(F-statistic)	0.000000		

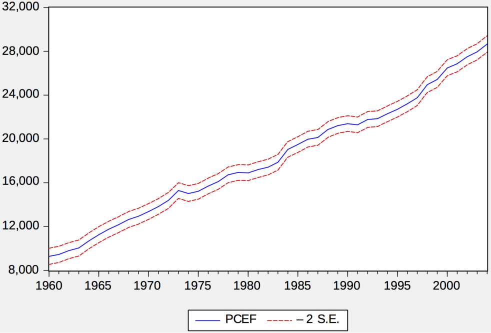
Forecast Band V



Forecast Band V



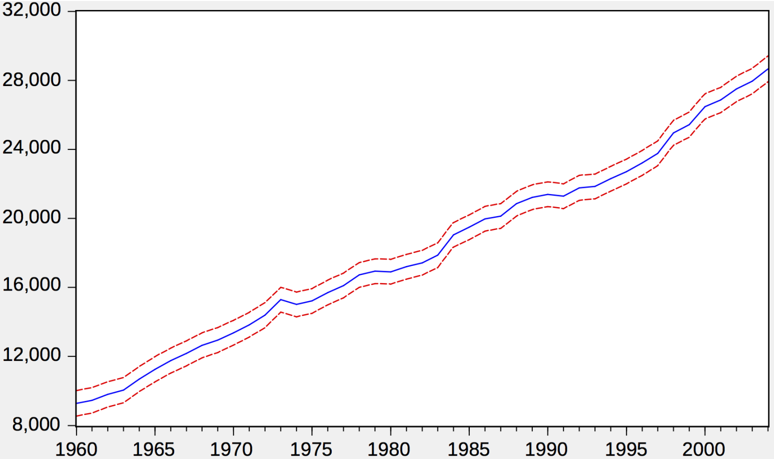
Forecast Band V



Forecast: PCEF
Actual: PCE
Forecast sample: 1960 2004
Included observations: 45

Root Mean Squared Error	345.5461
Mean Absolute Error	291.6878
Mean Abs. Percent Error	1.795941
Theil Inequality Coefficient	0.009095
Bias Proportion	0.000000
Variance Proportion	0.001005
Covariance Proportion	0.998995
Theil U2 Coefficient	0.715652
Symmetric MAPE	1.801425

Forecast Band V



Forecast Band VI

- the variance of the estimated mean values increases as X value moves further away from its mean value
- this means that forecast error will increase as we move further away from the mean value of the regressor
- forecasting $E(Y|X)$ for X values much greater than $\bar{X}$ will lead to substantial forecast errors

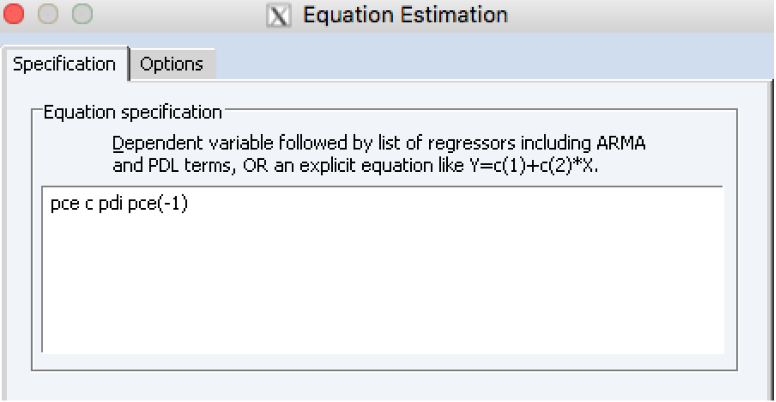
Forecast Band VII

- in the graph, measures of the quality of the forecast
  - root mean square, mean absolute error, mean absolute percentage error
  - the Theil inequality coefficient which lies between 0 and 1; the closer it is to zero, the better is the model
- these are useful if we are comparing two or more methods of forecasting

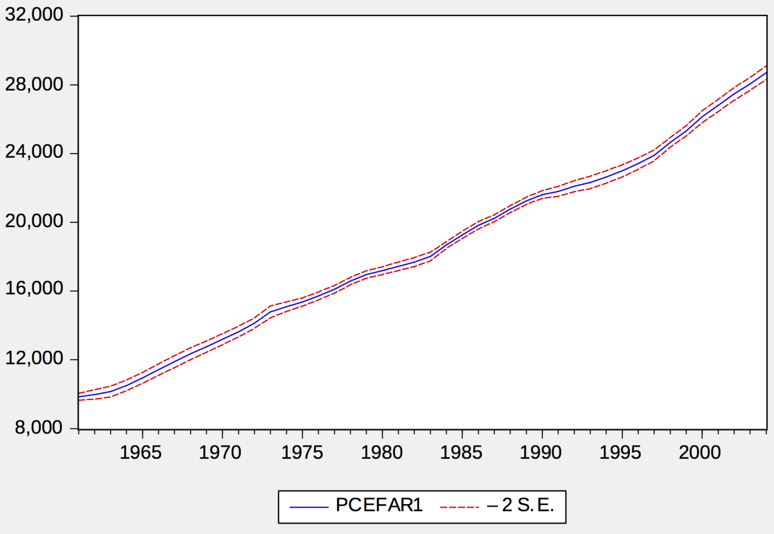
Autocorrelation

- DW statistic in regression results suggests that the error term suffers from first-order positive serial correlation
- if we can take into account serial correlation, the forecast error could be made smaller

Autocorrelation



Autocorrelation



The Box-Jenkins Methodology

- BJ forecasting methodology analyses the probabilistic, or stochastic, properties of economic time series on their own
- "let the data speak for themselves"
- allows for Y_t to be explained by
  - past, lagged, values of Y_t itself, and
  - current and lagged values of white noise u_t
- assumes that Y_t is stationary

Autoregressive AR Model

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + u_t$$

- u_t a white noise error term
- autoregressive model of order p, $AR(p)$
- the value of p determined empirically using some criterion e.g., AIC

Moving Average MA Model

$$Y_t = C_0 + C_1 u_t + C_2 u_{t-1} + \cdots + C_q u_{t-q}$$

- we express Y_t as a weighted, or moving, average of the current and past white noise error terms
- known as $MA(q)$ model
- the value of q determined empirically

ARMA Model

- we can combine the AR and MA models
- called $ARMA(p, q)$
- p autoregressive terms, q moving average terms
- the values of p and q determined empirically

ARIMA Model

- BJ methodology assumes that Y_t is stationary, or can be made stationary by differencing
- known as $ARIMA(p, d, q)$ model, where d denotes the number of times Y_t has to be differenced to be stationary
- in most applications $d = 1$
- if Y_t is already stationary, then an $ARIMA(p, d, q)$ becomes an $ARMA(p, q)$ model

BJ Methodology in Four Steps

1. identification
   - determine the appropriate values of p, d, and q
   - the main tool in this search are the correlogram and partial correlogram
2. estimation
   - estimate the parameters of the chosen model
   - use OLS and sometimes use nonlinear (in parameter) estimation methods

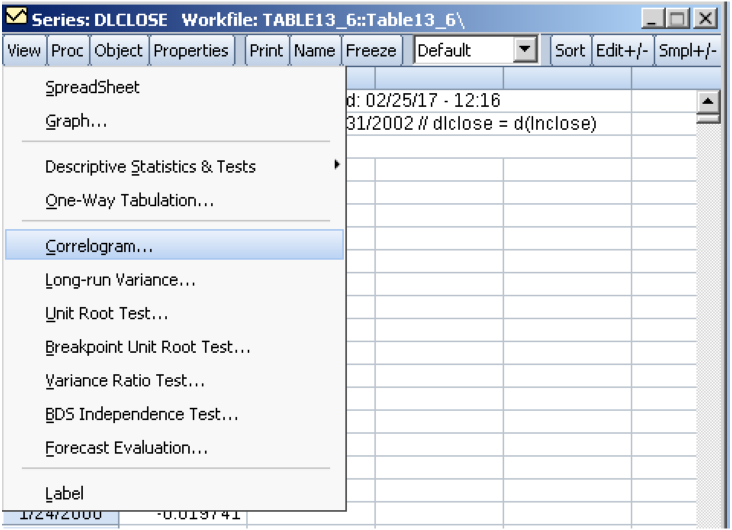
BJ Methodology in Four Steps

3. diagnostic checking
   - BJ *ARIMA* modelling is more an art than science
   - not absolutely sure the chosen model is the correct one
   - check the residuals are white noise otherwise start afresh (an iterative process)
4. forecasting
   - the ultimate test of a successful *ARIMA* is its forecasting performance, within the sample period as well as outside the sample period

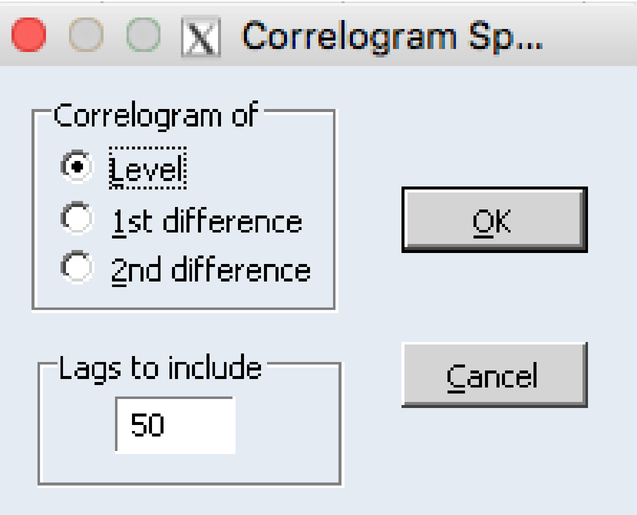
Example: IBM daily closing prices

- *LCLOSE* is nonstationary but *DLCLOSE* is stationary
- which *ARMA* model fits *DLCLOSE*
- correlgoram up to 50 lags

Example


Series: DLCLOSE Workfile: TABLE13_6::Table13_6\
View
Proc
Object
Properties
Print
Name
Freeze
Default
Sort
Edit+/-
Smpl+/-
SpreadSheet
Graph...
Descriptive Statistics & Tests
One-Way Tabulation...
Correlogram...
Long-run Variance...
Unit Root Test...
Breakpoint Unit Root Test...
Variance Ratio Test...
BDS Independence Test...
Forecast Evaluation...
Label
d: 02/25/17 - 12:16
31/2002 // dlclose = d(lnclose)
1/24/2000
-0.019741

Example


Correlogram Sp...
Correlogram of
Level
1st difference
2nd difference
OK
Lags to include
50
Cancel

Example

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.058	-0.058	2.3316	0.127
		2	-0.069	-0.072	5.5701	0.062
		3	-0.010	-0.019	5.6380	0.131
		4	0.098	0.092	12.323	0.015
		5	-0.026	-0.017	12.802	0.025
		6	0.024	0.035	13.195	0.040
		7	0.016	0.019	13.364	0.064
		8	-0.034	-0.039	14.192	0.077
		9	0.005	0.008	14.211	0.115
		10	0.040	0.031	15.353	0.120
		11	-0.059	-0.058	17.822	0.086
		12	-0.016	-0.011	17.993	0.116
		13	0.033	0.022	18.779	0.130
		14	0.023	0.018	19.148	0.159
		15	0.023	0.044	19.521	0.191
		16	0.046	0.052	21.043	0.177
		17	0.050	0.058	22.789	0.156

Example: AR

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000894	0.000955	-0.936224	0.3495
AR(4)	0.091282	0.035438	2.575853	0.0102
AR(18)	-0.080021	0.038859	-2.059258	0.0399
AR(22)	-0.090111	0.036529	-2.466820	0.0139
AR(35)	-0.058962	0.035395	-1.665816	0.0962
AR(43)	0.052808	0.045020	1.172988	0.2412

Example: AR

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000892	0.000954	-0.935387	0.3499
AR(4)	0.094017	0.035462	2.651254	0.0082
AR(18)	-0.086559	0.038586	-2.243281	0.0252
AR(22)	-0.091846	0.036372	-2.525200	0.0118
SIGMASQ	0.000676	2.52E-05	26.84957	0.0000

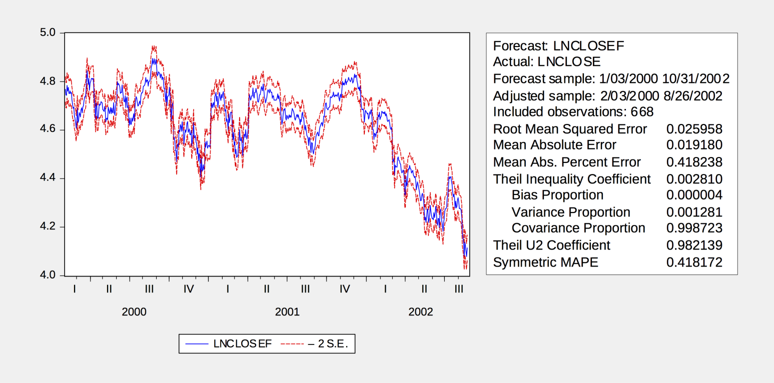
Example MA

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000889	0.000916	-0.970893	0.3319
MA(4)	0.085634	0.035251	2.429256	0.0154
MA(18)	-0.094356	0.040780	-2.313758	0.0210
MA(22)	-0.105041	0.037340	-2.813108	0.0050

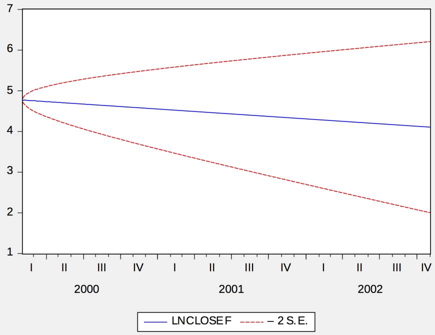
Example ARMA (4,22)(4,22)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000930	0.001053	-0.882798	0.3777
AR(4)	-0.294281	0.084685	-3.475008	0.0005
AR(22)	-0.604657	0.101289	-5.969596	0.0000
MA(4)	0.424877	0.085889	4.946830	0.0000
MA(22)	0.565924	0.091823	6.163217	0.0000

Forecast: ARMA (4,22)(4,22) dynamic

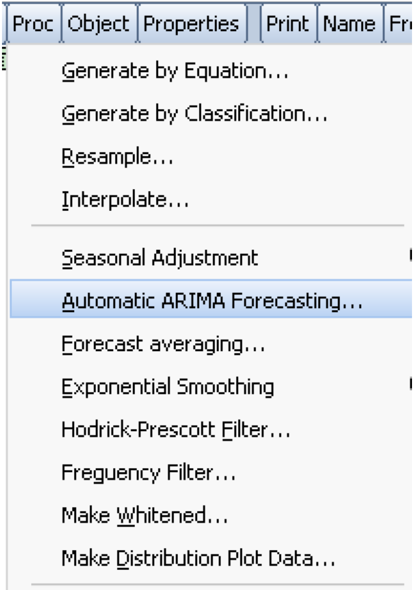


Forecast: ARMA (4,22)(4,22) static

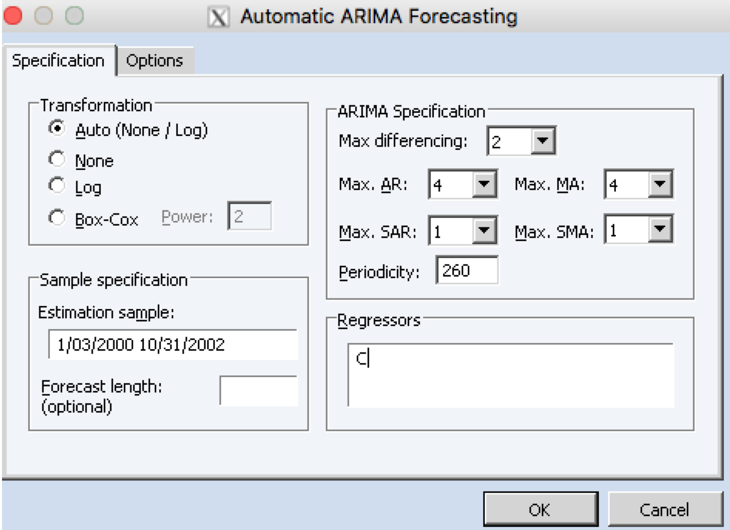


Forecast: LNCLOSEF	
Actual: LNCLOSE	
Forecast sample: 1/03/2000 10/31/2002	
Adjusted sample: 2/03/2000 10/31/2002	
Included observations: 716	
Root Mean Squared Error	0.214929
Mean Absolute Error	0.170716
Mean Abs. Percent Error	3.661897
Theil Inequality Coefficient	0.023658
Bias Proportion	0.513064
Variance Proportion	0.002187
Covariance Proportion	0.484749
Theil U2 Coefficient	7.923331
Symmetric MAPE	3.767376

EViews: Automatic ARIMA Selection



EViews: Automatic ARIMA Selection



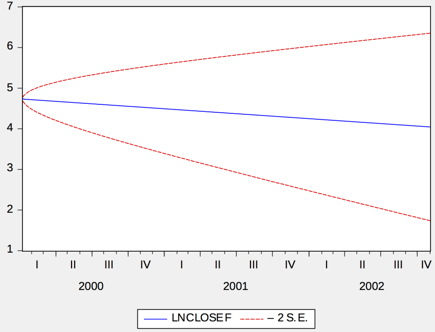
EViews: Automatic ARIMA Selection

⊟ Summary

Automatic ARIMA Forecasting
Selected dependent variable: DLOG(LNCLOSE)
Date: 02/25/17 Time: 13:55
Sample: 1/03/2000 10/31/2002
Included observations: 686
Forecast length: 0

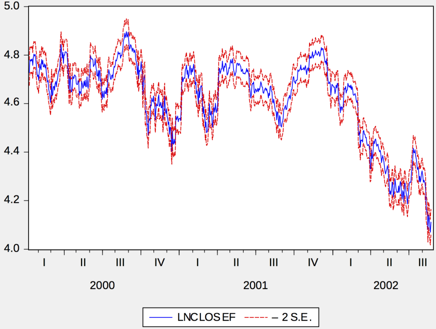
Number of estimated ARMA models: 100
Number of non-converged estimations: 0
Selected ARMA model: (0,4)(0,0)
AIC value: -6.93918150424

EViews: Automatic ARIMA Selection



Forecast: LNCLOSEF	
Actual: LNCLOSE	
Forecast sample: 1/03/2000 10/31/2002	
Adjusted sample: 1/10/2000 10/31/2002	
Included observations: 734	
Root Mean Squared Error	0.257808
Mean Absolute Error	0.212656
Mean Abs. Percent Error	4.569595
Theil Inequality Coefficient	0.028526
Bias Proportion	0.657006
Variance Proportion	0.003990
Covariance Proportion	0.339004
Theil U2 Coefficient	9.589423
Symmetric MAPE	4.726765

EViews: Automatic ARIMA Selection



Forecast: LNCLOSEF
Actual: LNCLOSE
Forecast sample: 1/03/2000 10/31/2002
Adjusted sample: 1/10/2000 8/26/2002
Included observations: 686

Root Mean Squared Error	0.026157
Mean Absolute Error	0.019260
Mean Abs. Percent Error	0.419634
Theil Inequality Coefficient	0.002829
Bias Proportion	0.000001
Variance Proportion	0.000946
Covariance Proportion	0.999033
Theil U2 Coefficient	0.993962
Symmetric MAPE	0.419594