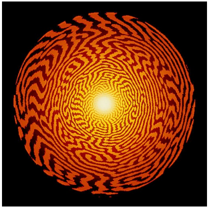
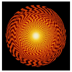
Forecasting

Francis X. Diebold
University of Pennsylvania



August 11, 2015

The painting is *Enigma*, by Glen Josselsohn, from Wikimedia Commons.



Elements of Forecasting in Business, Finance, Economics and Government

1. Forecasting in Action
    1.1 Operations planning and control
    1.2 Marketing
    1.3 Economics
    1.4 Financial speculation
    1.5 Financial risk management
    1.6 Capacity planning
    1.7 Business and government budgeting
    1.8 Demography
    1.9 Crisis management

Forecasting Methods: An Overview

Review of probability, statistics and regression

Six Considerations Basic to Successful Forecasting

1. Forecasts and decisions
2. The object to be forecast
3. Forecast types
4. The forecast horizon
5. The information set
6. Methods and complexity
   - 6.1 The parsimony principle
   - 6.2 The shrinkage principle

Statistical Graphics for Forecasting

- Why graphical analysis is important
- Simple graphical techniques
- Elements of graphical style
- Application: graphing four components of real GNP

Modeling and Forecasting Trend

- Modeling trend
- Estimating trend models
- Forecasting trend
- Selecting forecasting models using the Akaike and Schwarz criteria
- Application: forecasting retail sales

Modeling and Forecasting Seasonality

- The nature and sources of seasonality
- Modeling seasonality
- Forecasting seasonal series
- Application: forecasting housing starts

Characterizing Cycles

- Covariance stationary time series
- White noise
- The lag operator
- Wold's theorem, the general linear process, and rational distributed lags
- Estimation and inference for the mean, autocorrelation and partial autocorrelation functions
- Application: characterizing Canadian employment dynamics

Modeling Cycles: MA, AR and ARMA Models

- Moving-average (MA) models
- Autoregressive (AR) models
- Autoregressive moving average (ARMA) models
- Application: specifying and estimating models for forecasting employment

Forecasting Cycles

- Optimal forecasts
- Forecasting moving average processes
- Forecasting infinite-ordered moving averages
- Making the forecasts operational
- The chain rule of forecasting
- Application: forecasting employment

Putting it all Together: A Forecasting Model with Trend, Seasonal and Cyclical Components

- Assembling what we've learned
- Application: forecasting liquor sales
- Recursive estimation procedures for diagnosing and selecting forecasting models

Forecasting with Regression Models

- Conditional forecasting models and scenario analysis
- Accounting for parameter uncertainty in confidence intervals for conditional forecasts
- Unconditional forecasting models
- Distributed lags, polynomial distributed lags, and rational distributed lags
- Regressions with lagged dependent variables, regressions with ARMA disturbances, and transfer function models
- Vector autoregressions
- Predictive causality
- Impulse-response functions and variance decomposition
- Application: housing starts and completions

Evaluating and Combining Forecasts

- Evaluating a single forecast
- Evaluating two or more forecasts: comparing forecast accuracy
- Forecast encompassing and forecast combination
- Application: OverSea shipping volume on the Atlantic East trade lane

Unit Roots, Stochastic Trends, ARIMA Forecasting Models, and Smoothing

- Stochastic trends and forecasting
- Unit roots: estimation and testing
- Application: modeling and forecasting the yen/dollar exchange rate
- Smoothing
- Exchange rates, continued

Volatility Measurement, Modeling and Forecasting

- The basic ARCH process
- The GARCH process
- Extensions of ARCH and GARCH models
- Estimating, forecasting and diagnosing GARCH models
- Application: stock market volatility

Useful Books, Journals and Software

<u>Books</u>

Statistics review, etc.:

- Wonnacott, T.H. and Wonnacott, R.J. (1990), *Introductory Statistics*, Fifth Edition. New York: John Wiley and Sons.
- Pindyck, R.S. and Rubinfeld, D.L. (1997),*Econometric Models and Economic Forecasts*, Fourth Edition. New York: McGraw-Hill.
- Maddala, G.S. (2001), *Introduction to Econometrics*, Third Edition. New York: Macmillan.
- Kennedy, P. (1998), *A Guide to Econometrics*, Fourth Edition. Cambridge, Mass.: MIT Press.

Useful Books, Journals and Software cont.

Time series analysis:

- Chatfield, C. (1996), *The Analysis of Time Series: An Introduction*, Fifth Edition. London: Chapman and Hall.
- Granger, C.W.J. and Newbold, P. (1986), *Forecasting Economic Time Series*, Second Edition. Orlando, Florida: Academic Press.
- Harvey, A.C. (1993), *Time Series Models*, Second Edition. Cambridge, Mass.: MIT Press.
- Hamilton, J.D. (1994), *Time Series Analysis*, Princeton: Princeton University Press.

Useful Books, Journals and Software cont.

Special insights:

- Armstrong, J.S. (Ed.) (1999), *The Principles of Forecasting*. Norwell, Mass.: Kluwer Academic Forecasting.
- Makridakis, S. and Wheelwright S.C. (1997), *Forecasting: Methods and Applications*, Third Edition. New York: John Wiley.
- Bails, D.G. and Peppers, L.C. (1997), *Business Fluctuations*. Englewood Cliffs: Prentice Hall.
- Taylor, S. (1996), *Modeling Financial Time Series*, Second Edition. New York: Wiley.

Useful Books, Journals and Software cont.

Journals

- *Journal of Forecasting*
- *Journal of Business Forecasting Methods and Systems*
- *Journal of Business and Economic Statistics*
- *Review of Economics and Statistics*
- *Journal of Applied Econometrics*

Useful Books, Journals and Software cont.

Software

- General:
  - Eviews
  - S+
  - Minitab
  - SAS
  - R
  - Python
  - Many more...
- Cross-section:
  - Stata
- Open-ended:
  - Matlab

Useful Books, Journals and Software cont.

Online Information

- Resources for Economists:

A Brief Review of Probability, Statistics, and Regression for Forecasting

Topics

- Discrete Random Variable
- Discrete Probability Distribution
- Continuous Random Variable
- Probability Density Function
- Moment
- Mean, or Expected Value
- Location, or Central Tendency
- Variance
- Dispersion, or Scale
- Standard Deviation
- Skewness
- Asymmetry
- Kurtosis
- Leptokurtosis

A Brief Review of Probability, Statistics, and Regression for Forecasting

Topics cont.

- Skewness
- Asymmetry
- Kurtosis
- Leptokurtosis
- Normal, or Gaussian, Distribution
- Marginal Distribution
- Joint Distribution
- Covariance
- Correlation
- Conditional Distribution
- Conditional Moment
- Conditional Mean
- Conditional Variance

A Brief Review of Probability, Statistics, and Regression for Forecasting cont.

Topics cont.

- Population Distribution
- Sample
- Estimator
- Statistic, or Sample Statistic
- Sample Mean
- Sample Variance
- Sample Standard Deviation
- Sample Skewness
- Sample Kurtosis
- χ^2 Distribution
- t Distribution
- F Distribution
- Jarque-Bera Test

Regression as Curve Fitting

Least-squares estimation:

$$\min_{\beta} \sum_{t=1}^{T} [y_t - \beta_0 - \beta_1 x_t]^2$$

Fitted values:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$$

Residuals:

$$e_t = y_t - \hat{y}_t$$

Regression as a probabilistic model

Simple regression:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

$$\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$$

Multiple regression:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + \varepsilon_t$$

$$\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$$

Regression as a probabilistic model cont.

Mean dependent var 10.23

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

S.D. dependent var 1.49

$$SD = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \bar{y})^2}{T-1}}$$

Sum squared resid 43.70

$$SSR = \sum_{t=1}^{T} e_t^2$$

Regression as a probabilistic model cont.

F−statistic 30.89

$$\mathrm{F} = \frac{(\mathrm{SSR}_{\mathrm{res}} - \mathrm{SSR}) \;/\; (\mathrm{k} - 1)}{\mathrm{SSR} \;/\; (\mathrm{T} - \mathrm{k})}$$

S.E. of regression 0.99

$$\mathrm{s}^2 = \frac{\sum_{\mathrm{t}=1}^{\mathrm{T}} \mathrm{e}_{\mathrm{t}}^2}{\mathrm{T} - \mathrm{k}}$$

$$\mathrm{SER} = \sqrt{\mathrm{s}^2} = \sqrt{\frac{\sum_{\mathrm{t}=1}^{\mathrm{T}} \mathrm{e}_{\mathrm{t}}^2}{\mathrm{T} - \mathrm{k}}}$$

Regression as a probabilistic model cont.

R−squared 0.58

$$R^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T}(y_t - \bar{y}_t)^2}$$

or

$$R^2 = 1 - \frac{\frac{1}{T}\sum_{t=1}^{T} e_t^2}{\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y}_t)^2}$$

Adjusted R−squared 0.56

$$\bar{R}^2 = 1 - \frac{\frac{1}{T-k}\sum_{t=1}^{T} e_t^2}{\frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y}_t)^2}$$

Regression as a probabilistic model cont.

Akaike info criterion 0.03

$$\mathrm{AIC} \;=\; \mathrm{e}^{\left(\frac{2\mathrm{k}}{\mathrm{T}}\right)} \, \frac{\sum_{\mathrm{t}=1}^{\mathrm{T}} \mathrm{e}_{\mathrm{t}}^{2}}{\mathrm{T}}$$

Schwarz criterion 0.15

$$\mathrm{SIC} \;=\; \mathrm{T}^{\left(\frac{\mathrm{k}}{\mathrm{T}}\right)} \, \frac{\sum_{\mathrm{t}=1}^{\mathrm{T}} \mathrm{e}_{\mathrm{t}}^{2}}{\mathrm{T}}$$

Regression as a probabilistic model cont.

Durbin – Watson stat 1.97

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

$$v_t \overset{iid}{\sim} N(0, \sigma^2)$$

$$\varepsilon_t = \phi \varepsilon_{t-1} + v_t$$

$$DW = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

Regression of y on x and z

Regression of y on x and z

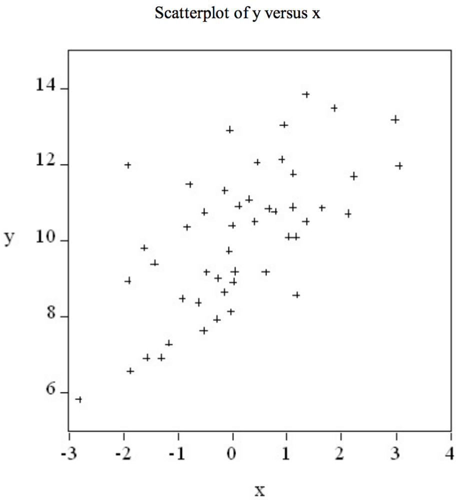
LS // Dependent Variable is Y

Sample: 1960 2007
Included observations: 48

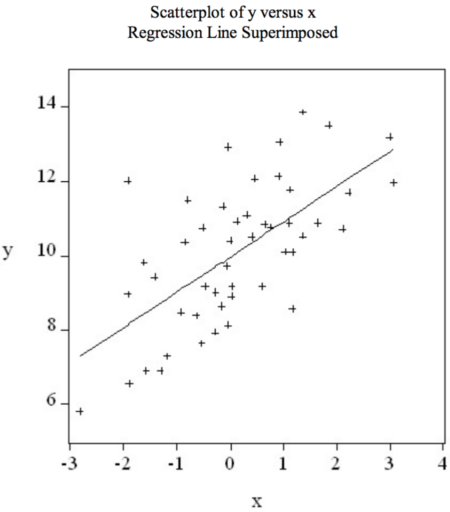
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.884732	0.190297	51.94359	0.0000
X	1.073140	0.150341	7.138031	0.0000
Z	-0.638011	0.172499	-3.698642	0.0006

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

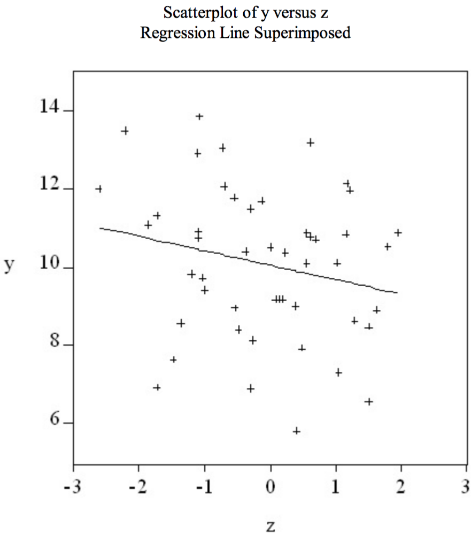
Scatterplot of y versus x



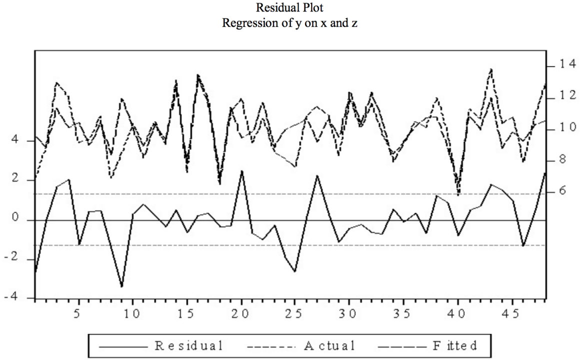
Scatterplot of y versus x – Regression Line Superimposed



Scatterplot of y versus z – Regression Line Superimposed



Residual Plot – Regression of y on x and z



Six Considerations Basic to Successful Forecasting

1. The Decision Environment and Loss Function

$$L(e) = e^2$$

$$L(e) = |e|$$

2. The Forecast Object
   - Event outcome, event timing, time series.
3. The Forecast Statement
   - Point forecast, interval forecast, density forecast, probability forecast

Six Considerations Basic to Successful Forecasting cont.

4. The Forecast Horizon
    - h-step ahead forecast
    - h-step-ahead extrapolation forecast
5. The Information Set

$$\Omega_T^{univariate} = \{y_T,\ y_{T-1},\ ...,\ y_1\}$$

$$\Omega_T^{multivariate} = \{y_T,\ x_T,\ y_{T-1},\ x_{T-1},\ ...,\ y_1,\ x_1\}$$

6. Methods and Complexity, the Parsimony Principle, and the
    - Shrinkage Principle
    - Signal vs. noise
    - Smaller is often better
    - Even incorrect restrictions can help

Six Considerations Basic to Successful Forecasting cont.

Decision Making with Symmetric Loss

	Demand High	Demand Low
Build Inventory	0	$10,000
Reduce Inventory	$10,000	0

Decision Making with Asymmetric Loss

	Demand High	Demand Low
Build Inventory	0	$10,000
Reduce Inventory	$20,000	0

Six Considerations Basic to Successful Forecasting cont.

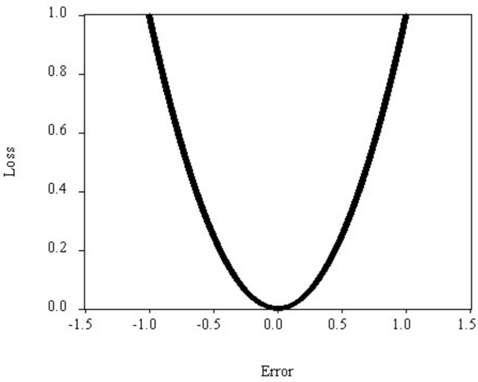
Forecasting with Symmetric Loss

	High Actual Sales	Low Actual Sales
High Forecasted Sales	0	$10,000
Low Forecasted Sales	$10,000	0

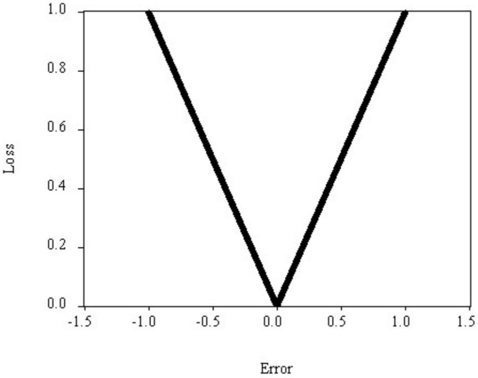
Forecasting with Asymmetric Loss

	High Actual Sales	Low Actual Sales
High Forecasted Sales	0	$10,000
Low Forecasted Sales	$20,000	0

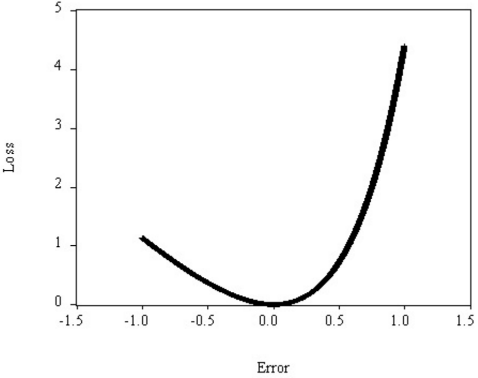
Quadratic Loss



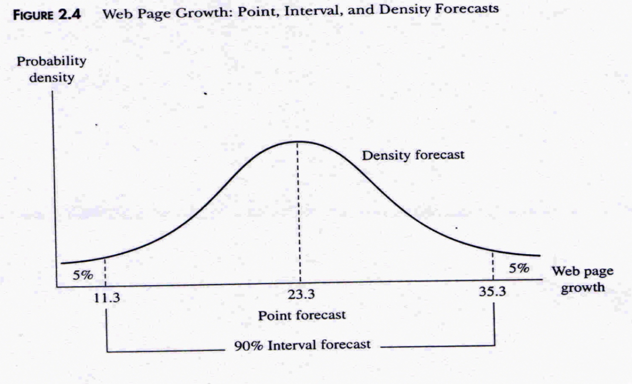
Absolute Loss



Asymmetric Loss

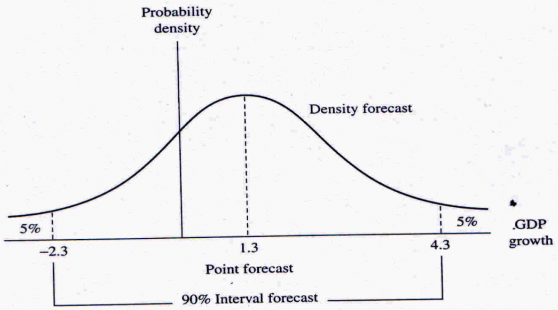


Forecast Statement



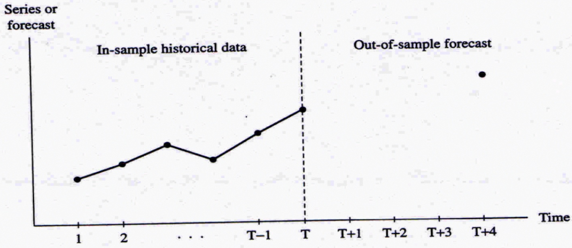
Forecast Statement cont.

FIGURE 2.5 U.S. Real GDP Growth: Point, Interval, and Density Forecasts

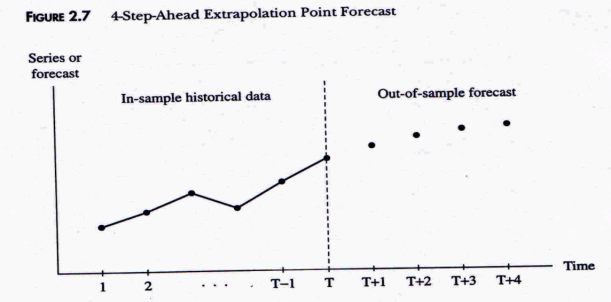


Extrapolation Forecast

FIGURE 2.6 4-Step-Ahead Point Forecast



Extrapolation Forecast cont.



Statistical Graphics For Forecasting

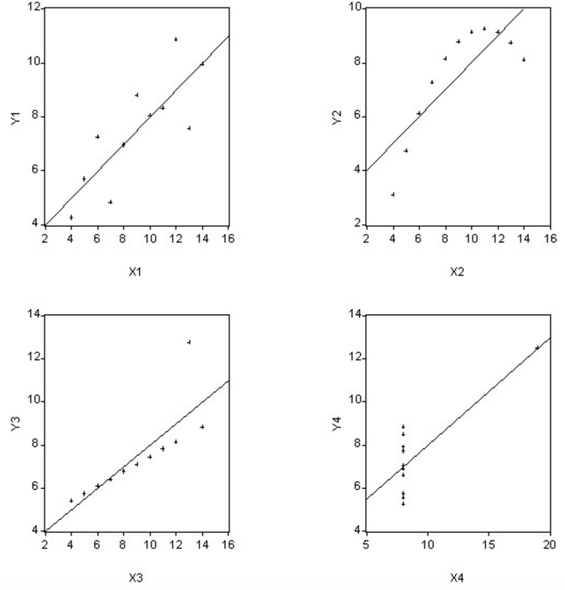
1. Why Graphical Analysis is Important
   - Graphics helps us summarize and reveal patterns in data
   - Graphics helps us identify anomalies in data
   - Graphics facilitates and encourages comparison of different pieces of data
   - Graphics enables us to present a huge amount of data in a small space, and it enables us to make huge data sets coherent
2. Simple Graphical Techniques
   - Univariate, multivariate
   - Time series vs. distributional shape
   - Relational graphics
3. Elements of Graphical Style
   - Know your audience, and know your goals.
   - Show the data, and appeal to the viewer.
   - Revise and edit, again and again.
4. Application: Graphing Four Components of Real GNP

Anscombe's Quartet

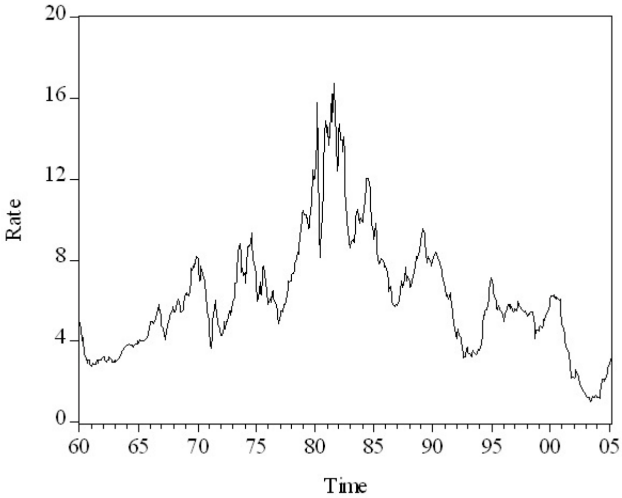
(1)		(2)		(3)		(4
x1	y1	x2	y2	x3	y3	x4
10.0	8.04	10.0	9.14	10.0	7.46	8.
8.0	6.95	8.0	8.14	8.0	6.77	8.
13.0	7.58	13.0	8.74	13.0	12.74	8.
9.0	8.81	9.0	8.77	9.0	7.11	8.
11.0	8.33	11.0	9.26	11.0	7.81	8.
14.0	9.96	14.0	8.10	14.0	8.84	8.
6.0	7.24	6.0	6.13	6.0	6.08	8.
4.0	4.26	4.0	3.10	4.0	5.39	19
12.0	10.84	12.0	9.13	12.0	8.15	8.
7.0	4.82	7.0	7.26	7.0	6.42	8.
5.0	5.68	5.0	4.74	5.0	5.73	8.

Anscombe's Quartet

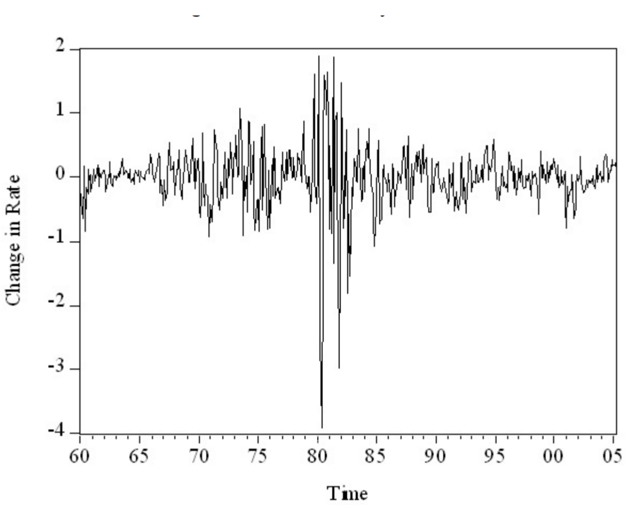
Anscombe's Quartet – Bivariate Scatterplot



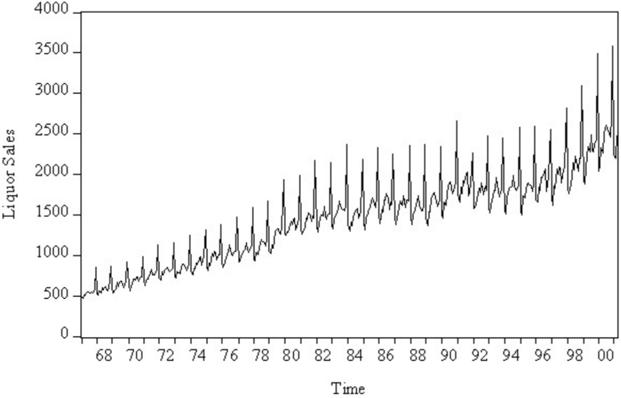
1-Year Treasury Bond Rates



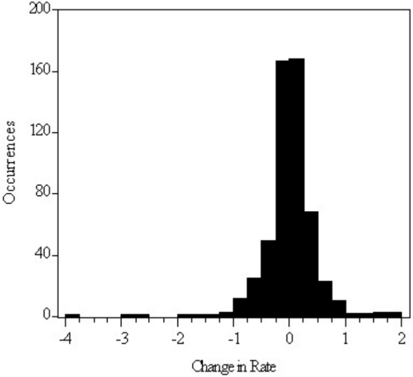
Change in 1-Year Treasury Bond Rates



Liquor Sales

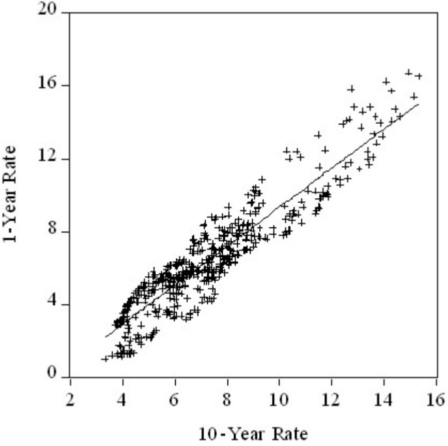


Histogram and Descriptive Statistics – Change in 1-Year Treasury Bond Rates

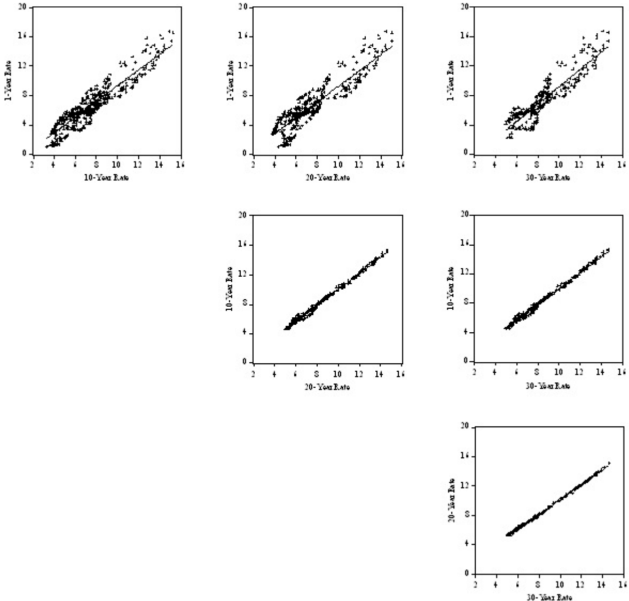


Sample 1960M01 2005M03	
Observations 543	
Mean	-0.003389
Median	0.000000
Maximum	1.900000
Minimum	-3.910000
Std. Dev.	0.470731
Skewness	-1.380580
Kurtosis	17.44767
Jarque-Bera	4895.124
Probability	0.000000

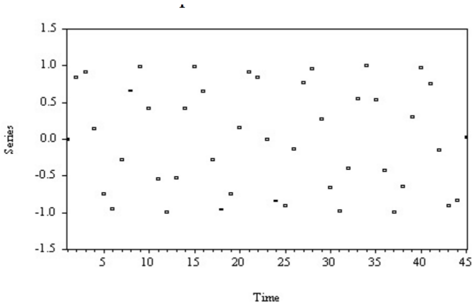
Scatterplot 1-Year vs. 10-year Treasury Bond Rates



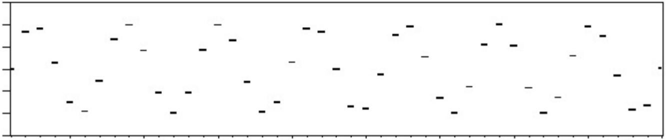
Scatterplot Matrix – 1-, 10-, 20-, and 30-Year Treasury Bond Rates



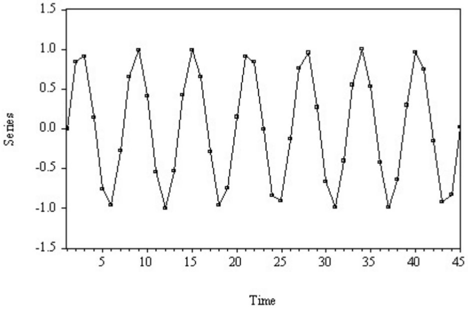
Time Series Plot – Aspect Ratio 1:1.6



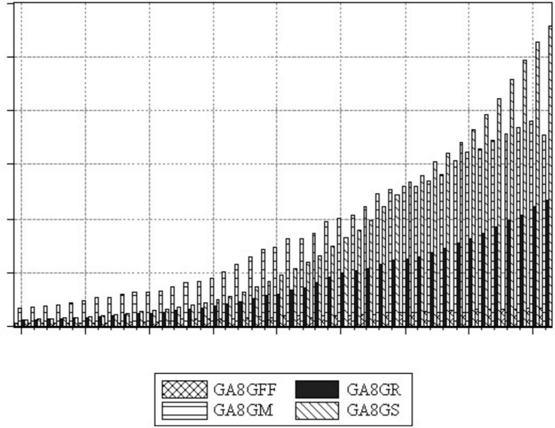
Time Series Plot – Banked to 45 Degrees



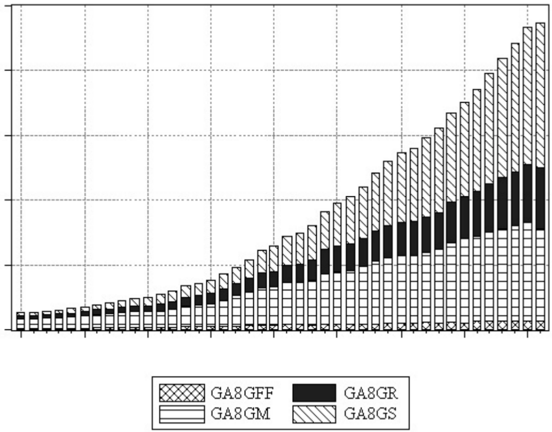
Time Series Plot – Aspect Ratio 1:1.6



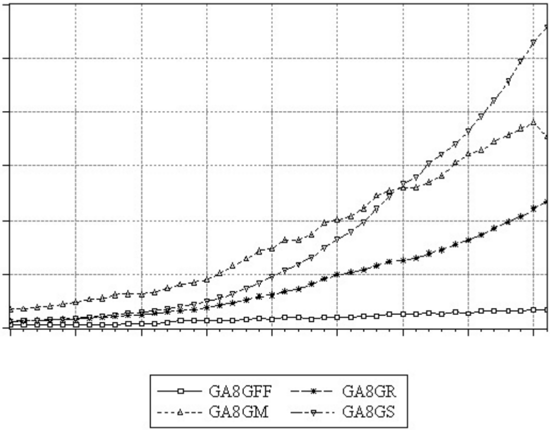
Graph



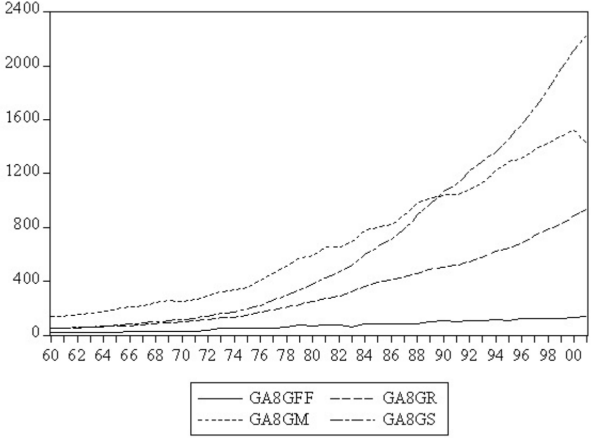
Graph



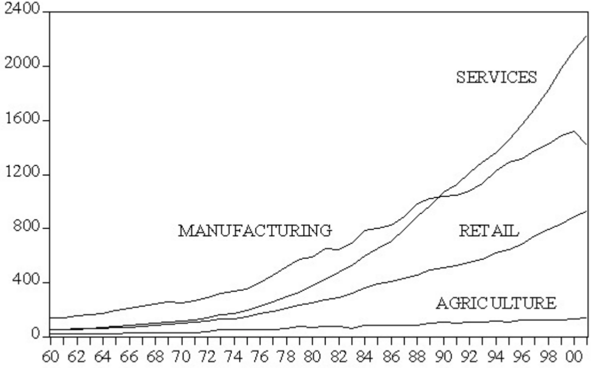
Graph



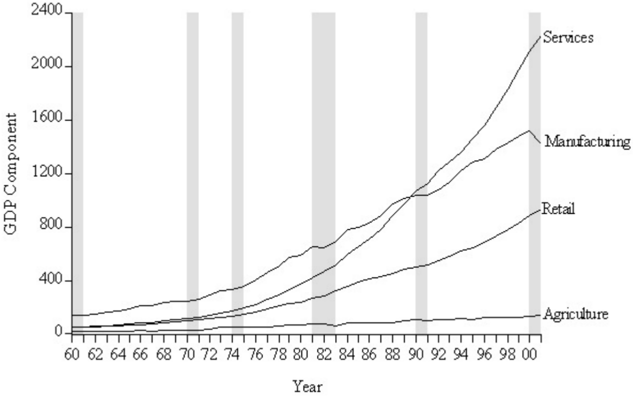
Graph


2400
2000
1600
1200
800
400
0
60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 00
GA8GFF
GA8GR
GA8GM
GA8GS

Graph


2400
2000
1600
1200
800
400
0
SERVICES
MANUFACTURING
RETAIL
AGRICULTURE
60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 00

Components of Real GDP (Millions of Current Dollars, Annual



Modeling and Forecasting Trend

1. Modeling Trend

$$T_t = \beta_0 + \beta_1 TIME_t$$

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t^2$$

$$T_t = \beta_0 e^{\beta_1 TIME_t}$$

$$\ln(T_t) = \ln(\beta_0) + \beta_1 TIME_t$$

Modeling and Forecasting Trend

2. Estimating Trend Models

$$(\hat{\beta}_0,\ \hat{\beta}_1) = \arg\min_{\beta_0,\beta_1} \sum_{t=1}^{T} \left[y_t - \beta_0 - \beta_1\,\text{TIME}_t \right]^2$$

$$(\hat{\beta}_0,\ \hat{\beta}_1,\ \hat{\beta}_2) = \arg\min_{\beta_0,\beta_1,\beta_2} \sum_{t=1}^{T} \Big[y_t - \beta_0 - \beta_1\,\text{TIME}_t - \beta_2\,\text{TIME}_t^2$$

$$(\hat{\beta}_0,\ \hat{\beta}_1) = \arg\min_{\beta_0,\beta_1} \sum_{t=1}^{T} \left[y_t - \beta_0 e^{\beta_1 \text{TIME}_t} \right]^2$$

$$(\hat{\beta}_0,\ \hat{\beta}_1) = \arg\min_{\beta_0,\beta_1} \sum_{t=1}^{T} \left[\ln y_t - \ln \beta_0 - \beta_1\,\text{TIME}_t \right]^2$$

Modeling and Forecasting Trend

3. Forecasting Trend

$$\mathrm{y_t} = \beta_0 + \beta_1 \, \mathrm{TIME_t} + \varepsilon_\mathrm{t}$$

$$\mathrm{y_{T+h}} = \beta_0 + \beta_1 \, \mathrm{TIME_{T+h}} + \varepsilon_\mathrm{T+h}$$

$$\mathrm{y_{T+h,T}} = \beta_0 + \beta_1 \, \mathrm{TIME_{T+h}}$$

$$\hat{\mathrm{y}}_\mathrm{T+h,T} = \hat{\beta}_0 + \hat{\beta}_1 \, \mathrm{TIME_{T+h}}$$

Modeling and Forecasting Trend

3. Forecasting Trend cont.

$$\mathrm{y}_{\mathrm{T+h,T}} \pm 1.96\sigma$$

$$\hat{\mathrm{y}}_{\mathrm{T+h,T}} \pm 1.96\hat{\sigma}$$

$$\mathrm{N}(\mathrm{y}_{\mathrm{T+h,T}},\ \sigma^2)$$

$$\mathrm{N}(\hat{\mathrm{y}}_{\mathrm{T+h,T}},\ \hat{\sigma}^2)$$

Modeling and Forecasting Trend

4. Selecting Forecasting Models

$$\mathrm{MSE} = \frac{\sum_{t=1}^{T} e_t^2}{T}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Modeling and Forecasting Trend

4. Selecting Forecasting Models cont.

$$s^2 = \frac{\sum_{t=1}^{T} e_t^2}{T - k}$$

$$s^2 = \left(\frac{T}{T - k}\right) \frac{\sum_{t=1}^{T} e_t^2}{T}$$

$$\bar{R}^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2 \ / \ T - k}{\sum_{t=1}^{T}(y_t - \bar{y})^2 \ / \ T - 1} = 1 - s^{\frac{2}{\sum_{t=1}^{T}(y_t - \bar{y})^2 \ / \ T-}}$$

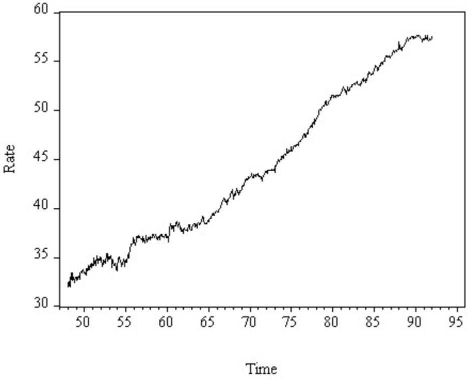
4. Selecting Forecasting Models cont.

$$\mathrm{AIC} = e^{\left(\frac{2k}{T}\right)} \frac{\sum_{t=1}^{T} e_t^2}{T}$$

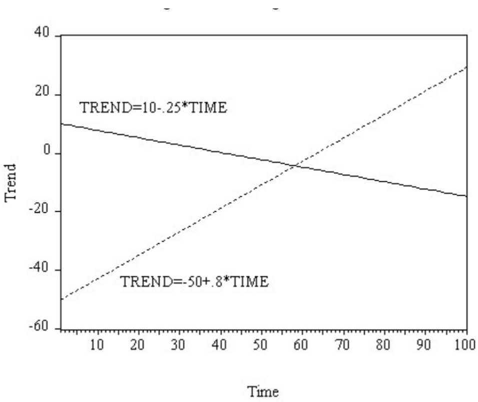
$$\mathrm{SIC} = T^{\left(\frac{k}{T}\right)} \frac{\sum_{t=1}^{T} e_t^2}{T}$$

- Consistency
- Efficiency

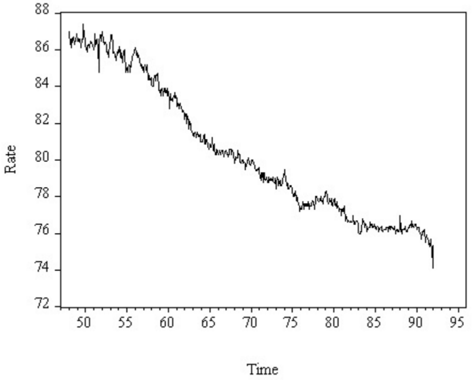
Labor Force Participation Rate



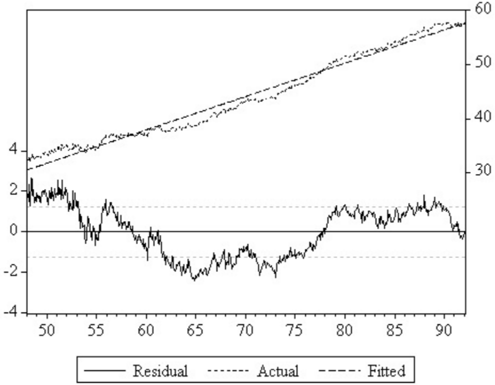
Increasing and Decreasing Labor Trends



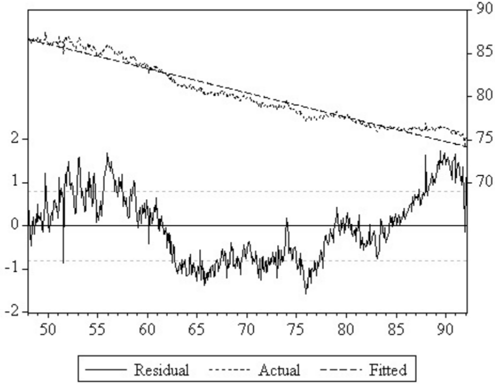
Labor Force Participation Rate



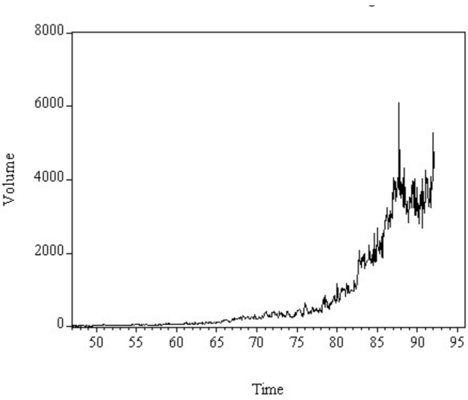
Linear Trend – Female Labor Force Participation Rate



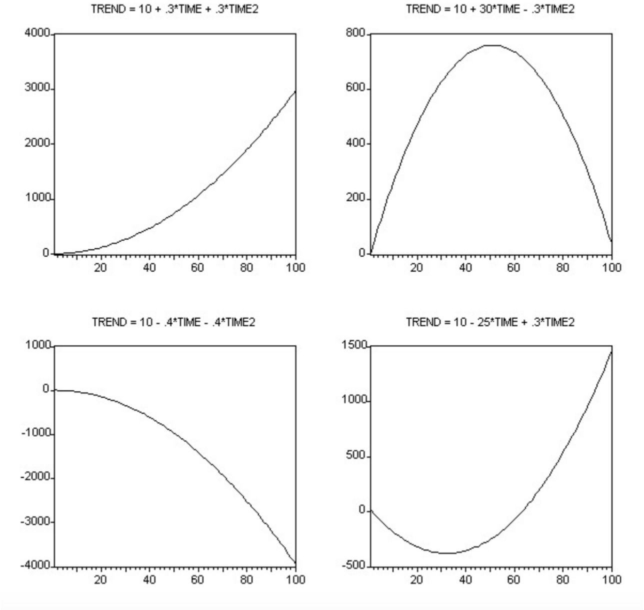
Linear Trend – Male Labor Force Participation Rate



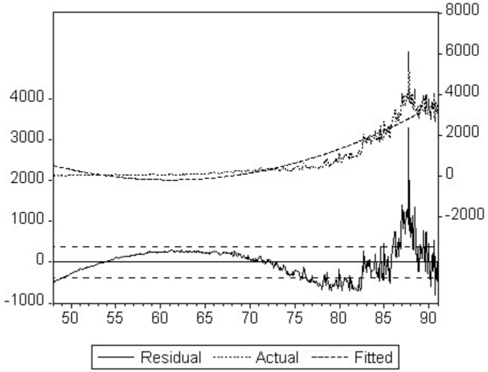
Volume on the New York Stock Exchange



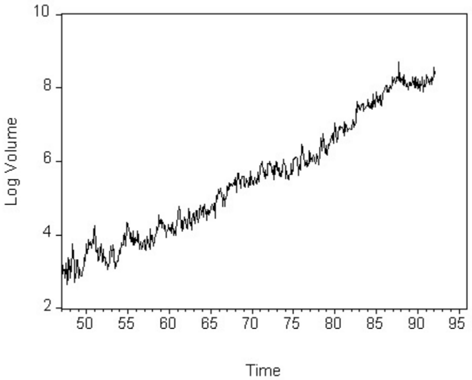
Various Shapes of Quadratic Trends



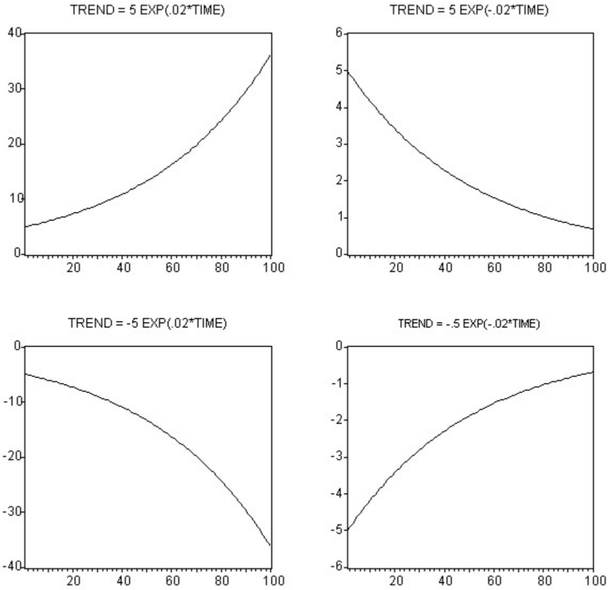
Quadratic Trend – Volume on the New York Stock Exchange



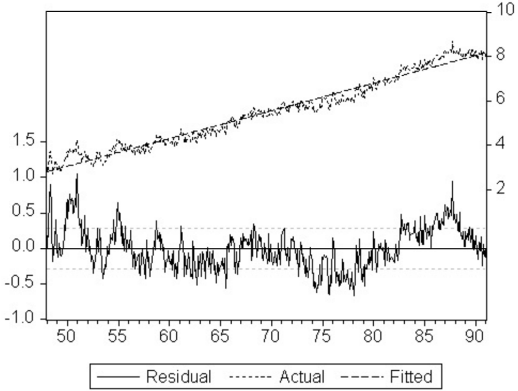
Log Volume on the New York Stock Exchange



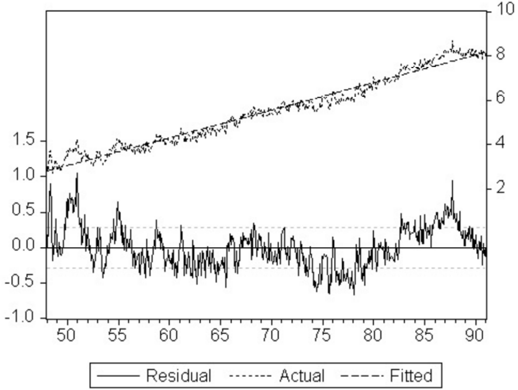
Various Shapes of Exponential Trends



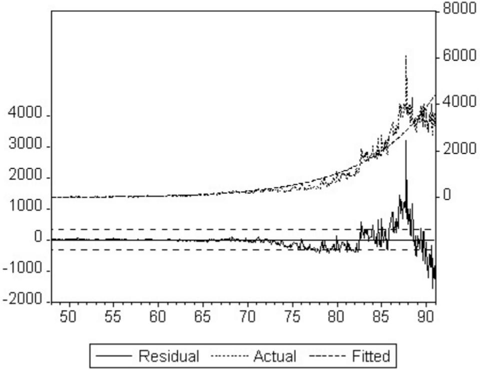
Linear Trend – Log Volume on the New York Stock Exchange



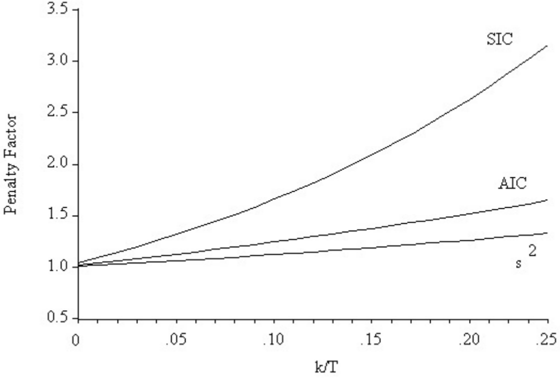
Exponential Trend – Volume on the New York Stock Exchange



Degree-of-Freedom Penalties – Various Model Selection Criteria



Retail Sales



Retail Sales – Linear Trend Regression

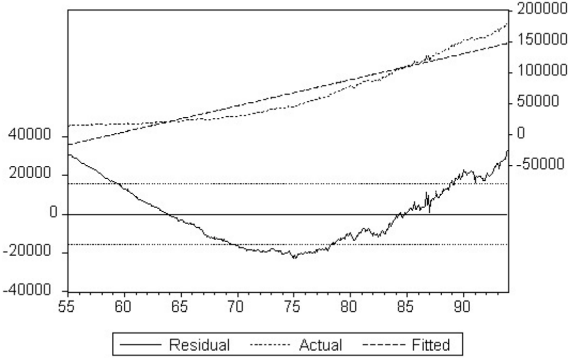
Retail Sales
Linear Trend Regression

Dependent Variable is RTRR
Sample: 1955:01 1993:12
Included observations: 468

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	-16391.25	1469.177	-11.15676	0.0000
TIME	349.7731	5.428670	64.43073	0.0000

R-squared	0.899076	Mean dependent var	65630.56
Adjusted R-squared	0.898859	S.D. dependent var	49889.26
S.E. of regression	15866.12	Akaike info criterion	19.34815
Sum squared resid	1.17E+11	Schwarz criterion	19.36587
Log likelihood	-5189.529	F-statistic	4151.319
Durbin-Watson stat	0.004682	Prob(F-statistic)	0.000000

Retail Sales – Linear Trend Residual Plot



Retail Sales – Quadratic Trend Regression

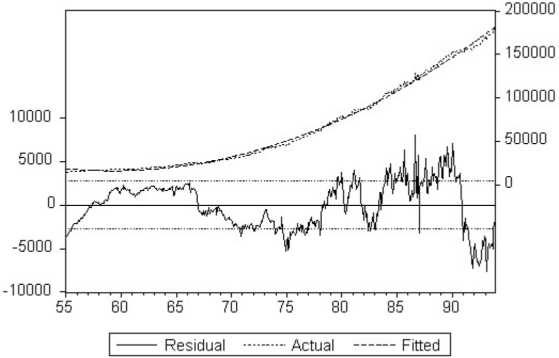
Retail Sales
Quadratic Trend Regression

Dependent Variable is RTRR
Sample: 1955:01 1993:12
Included observations: 468

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	18708.70	379.9566	49.23905	0.0000
TIME	-98.31130	3.741388	-26.27669	0.0000
TIME2	0.955404	0.007725	123.6754	0.0000

R-squared	0.997022	Mean dependent var	65630.56
Adjusted R-squared	0.997010	S.D. dependent var	49889.26
S.E. of regression	2728.205	Akaike info criterion	15.82919
Sum squared resid	3.46E+09	Schwarz criterion	15.85578
Log likelihood	-4365.093	F-statistic	77848.80
Durbin-Watson stat	0.151089	Prob(F-statistic)	0.000000

Retail Sales – Quadratic Trend Residual Plot



Retail Sales – Log Linear Trend Regression

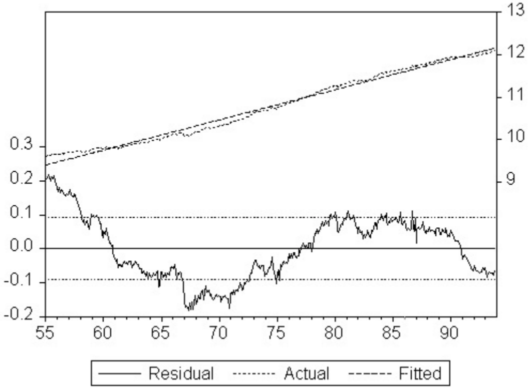
Retail Sales
Log Linear Trend Regression

Dependent Variable is LRTRR
Sample: 1955:01 1993:12
Included observations: 468

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	9.389975	0.008508	1103.684	0.0000
TIME	0.005931	3.14E-05	188.6541	0.0000

R-squared	0.987076	Mean dependent var	10.78072
Adjusted R-squared	0.987048	S.D. dependent var	0.807325
S.E. of regression	0.091879	Akaike info criterion	-4.770302
Sum squared resid	3.933853	Schwarz criterion	-4.752573
Log likelihood	454.1874	F-statistic	35590.36
Durbin-Watson stat	0.019949	Prob(F-statistic)	0.000000

Retail Sales – Log Linear Trend Residual Plot



Retail Sales – Exponential Trend Regression

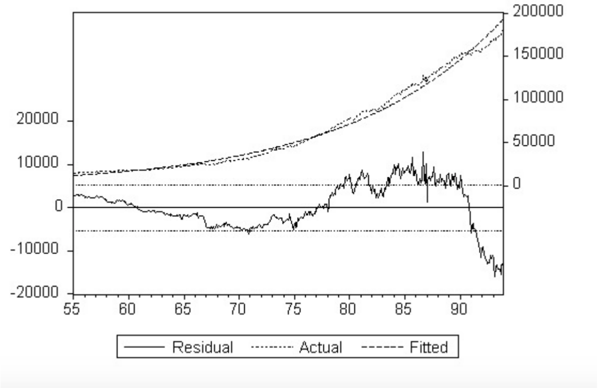
Retail Sales
Exponential Trend Regression

Dependent Variable is RTRR
Sample: 1955:01 1993:12
Included observations: 468
Convergence achieved after 1 iterations
RTRR=C(1)*EXP(C(2)*TIME)

	Coefficient	Std. Error	T-Statistic	Prob.
C(1)	11967.80	177.9598	67.25003	0.0000
C(2)	0.005944	3.77E-05	157.7469	0.0000

R-squared	0.988796	Mean dependent var	65630.56
Adjusted R-squared	0.988772	S.D. dependent var	49889.26
S.E. of regression	5286.406	Akaike info criterion	17.15005
Sum squared resid	1.30E+10	Schwarz criterion	17.16778
Log likelihood	-4675.175	F-statistic	41126.02
Durbin-Watson stat	0.040527	Prob(F-statistic)	0.000000

Retail Sales – Exponential Trend Residual Plot

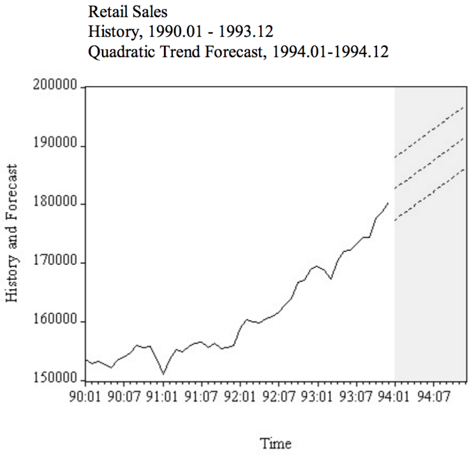


Model Selection Criteria

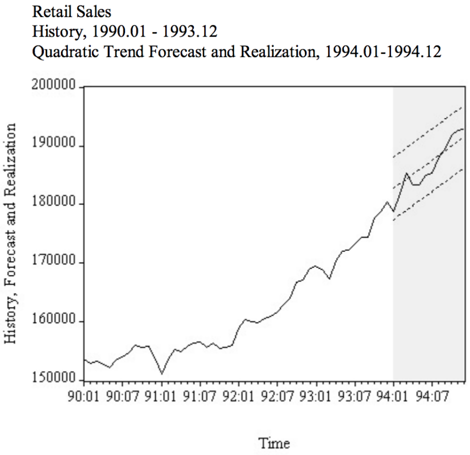
Linear, Quadratic and Exponential Trend Models

	Linear Trend	Quadratic Trend	Exponential Trend
AIC	19.35	15.83	17.15
SIC	19.37	15.86	17.17

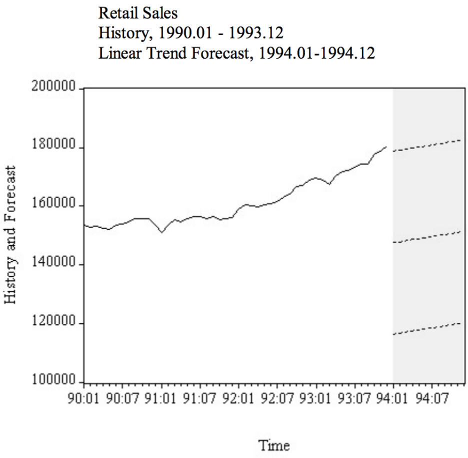
Retail Sales – History January, 1990 – December, 1994



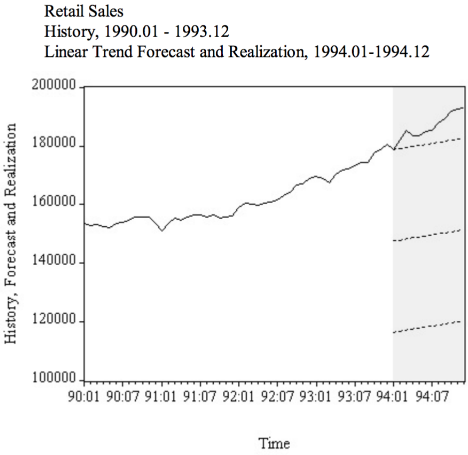
Retail Sales – History January, 1990 – December, 1994



Retail Sales – History January, 1990 – December, 1994



Retail Sales – History January, 1990 – December, 1994



Modeling and Forecasting Seasonality

1. The Nature and Sources of Seasonality
2. Modeling Seasonality

$$D_1 = (1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, \ldots)$$
$$D_2 = (0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, \ldots)$$
$$D_3 = (0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, \ldots)$$
$$D_4 = (0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, \ldots)$$

$$y_t = \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t$$

$$y_t = \beta_1 \, TIME_t + \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t$$

$$y_t = \beta_1 \, TIME_t + \sum_{i=1}^{s} \gamma_i D_{it} + \sum_{i=1}^{v_1} \delta_i^{HD} HDV_{it} + \sum_{i=1}^{v_2} \delta_i^{TD} TDV_{it}$$

Modeling and Forecasting Seasonality

3. Forecasting Seasonal Series

$$y_t \;=\; \beta_1\,\mathrm{TIME}_t \;+\; \sum_{i=1}^{s}\gamma_i D_{it} \;+\; \sum_{i=1}^{v_1}\delta_i^{HD}\mathrm{HDV}_{it} \;+\; \sum_{i=1}^{v_2}\delta_i^{TD}\mathrm{TDV}_{i}$$

$$y_{T+h} \;=\; \beta_1\,\mathrm{TIME}_{T+h} \;+\; \sum_{i=1}^{s}\gamma_i D_{i,T+h} \;+\; \sum_{i=1}^{v_1}\delta_i^{HD}\mathrm{HDV}_{i,T+h} \;+$$

$$y_{T+h,T} \;=\; \beta_1\,\mathrm{TIME}_{T+h} \;+\; \sum_{i=1}^{s}\gamma_i D_{i,T+h} \;+\; \sum_{i=1}^{v_1}\delta_i^{HD}\mathrm{HDV}_{i,T+h} \;+$$

$$\hat{y}_{T+h,T} \;=\; \hat{\beta}_1\,\mathrm{TIME}_{T+h} \;+\; \sum_{i=1}^{s\hat{\gamma}_i} D_{i,T+h} \;+\; \sum_{i=1}^{v_1}\hat{\delta}_i^{HD}\mathrm{HDV}_{i,T+h} \;+$$

Modeling and Forecasting Seasonality

3. Forecasting Seasonal Series cont.

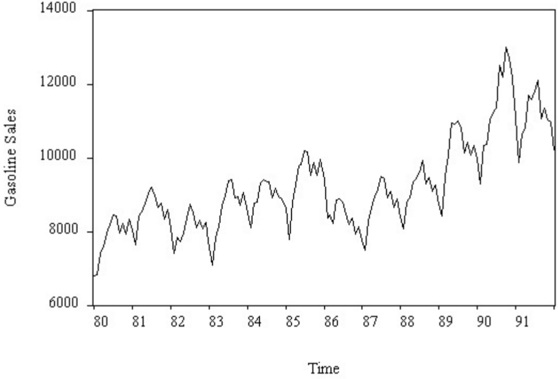
$$y_{T+h,T} \pm 1.96\sigma$$

$$\hat{y}_{T+h,T} \pm 1.96\hat{\sigma}$$

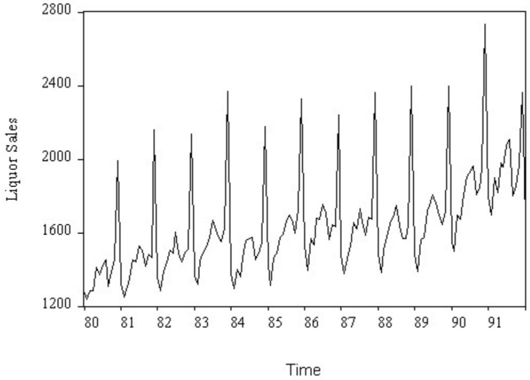
$$N(y_{T+h,T},\ \sigma^2)$$

$$N(\hat{y}_{T+h,T},\ \hat{\sigma}^2)$$

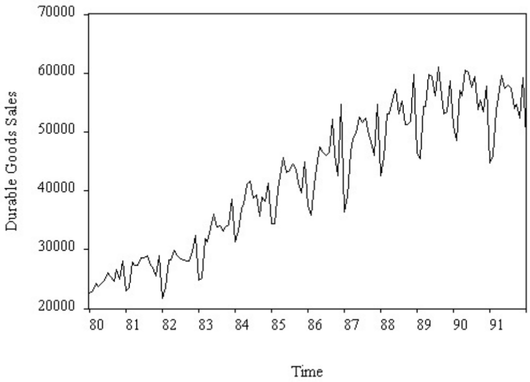
Gasoline Sales



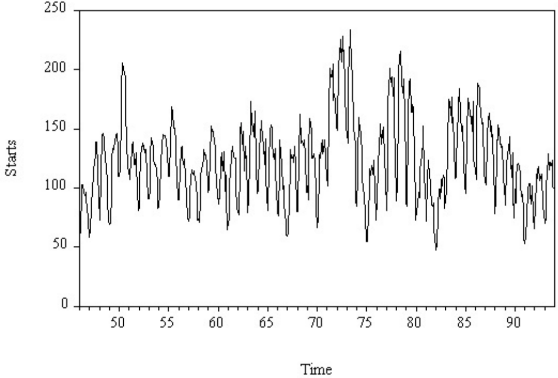
Liquor Sales



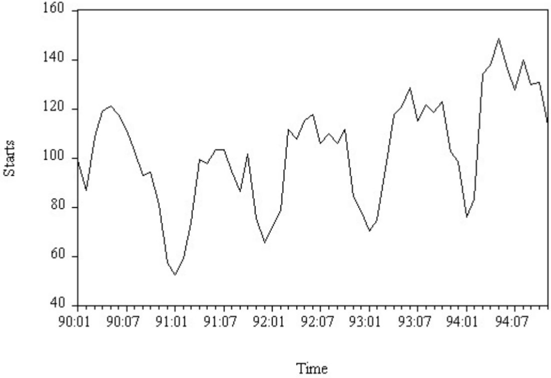
Durable Goods Sales



Housing Starts, January, 1946 – November, 1994



Housing Starts, January, 1990 – November, 1994



Housing Starts Regression Results - Seasonal Dummy Variable Model

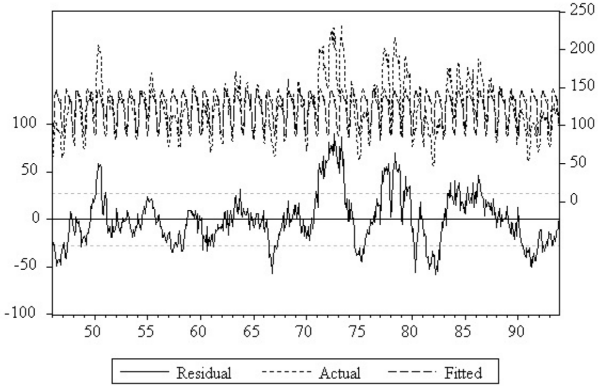
Regression Results
Seasonal Dummy Variable Model
Housing Starts

LS // Dependent Variable is STARTS
Sample: 1946:01 1993:12
Included observations: 576

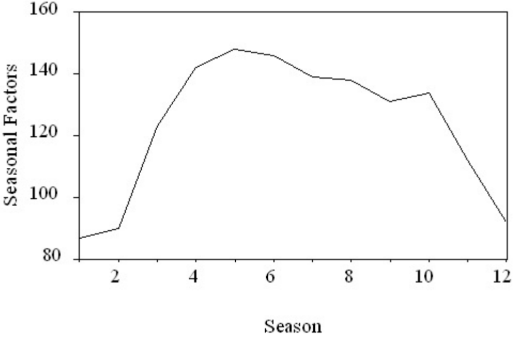
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D1	86.50417	4.029055	21.47009	0.0000
D2	89.50417	4.029055	22.21468	0.0000
D3	122.8833	4.029055	30.49929	0.0000
D4	142.1687	4.029055	35.28588	0.0000
D5	147.5000	4.029055	36.60908	0.0000
D6	145.9979	4.029055	36.23627	0.0000
D7	139.1125	4.029055	34.52733	0.0000
D8	138.4167	4.029055	34.35462	0.0000
D9	130.5625	4.029055	32.40524	0.0000
D10	134.0917	4.029055	33.28117	0.0000
D11	111.8333	4.029055	27.75671	0.0000
D12	92.15833	4.029055	22.87344	0.0000

R-squared	0.383780	Mean dependent var	123.3944
Adjusted R-squared	0.371762	S.D. dependent var	35.21775
S.E. of regression	27.91411	Akaike info criterion	6.678878
Sum squared resid	439467.5	Schwarz criterion	6.769630
Log likelihood	-2728.825	F-statistic	31.93250

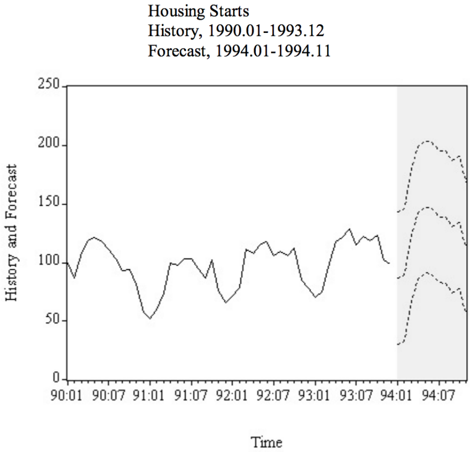
Residual Plot



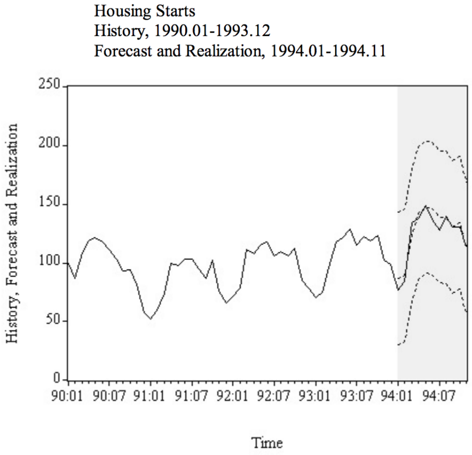
Housing Starts – Estimated Seasonal Factors



Housing Starts



Housing Starts



Characterizing Cycles

1. Covariance Stationary Time Series

- Realization
- Sample Path
- Covariance Stationary

$$Ey_t = \mu_t$$
$$Ey_t = \mu$$

$$\gamma(t, \tau) = cov(y_t,\ y_{t-\tau}) = E(y_t - \mu)(y_{t-\tau} - \mu)$$
$$\gamma(t, \tau) = \gamma(\tau)$$

$$\rho(\tau) = \frac{cov(y_t,\ y_{t-\tau})}{\sqrt{var(y_t)}\ \sqrt{var(y_{t-\tau})}} = \frac{\gamma(\tau)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} = \frac{\gamma(\tau)}{\gamma(0)}$$

1.Characterizing Cycles Cont.

$$\mathrm{corr}(\mathrm{x}, \mathrm{y}) = \frac{\mathrm{cov}(\mathrm{x}, \mathrm{y})}{\sigma_{\mathrm{x}}\ \sigma_{\mathrm{y}}}$$

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)},\ \tau = 0,\ 1,\ 2,\ \ldots.$$

$$\rho(\tau) = \frac{\mathrm{cov}(\mathrm{y}_{\mathrm{t}},\ \mathrm{y}_{\mathrm{t}-\tau})}{\sqrt{\mathrm{var}(\mathrm{y}_{\mathrm{t}})}\ \sqrt{\mathrm{var}(\mathrm{y}_{\mathrm{t}-\tau})}} = \frac{\gamma(\tau)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} = \frac{\gamma(\tau)}{\gamma(0)}$$

- $\mathrm{p}(\tau)$ regression of y_{t} on $\mathrm{y}_{\mathrm{t}-1}$, ..., $\mathrm{y}_{\mathrm{t}-\tau}$

2.White Noise

$$y_t \sim WN(0, \sigma^2)$$

$$y_t \overset{iid}{\sim} (0, \sigma^2)$$

$$y_t \overset{iid}{\sim} N(0, \sigma^2)$$

2.White Noise Cont.

$$E(y_t) = 0$$
$$\text{var}(y_t) = \sigma^2$$

$$E(y_t|\Omega_{t-1}) = 0$$
$$\text{var}(y_t|\Omega_{t-1}) = E[(y_t - E(y_t|\Omega_{t-1}))^2|\Omega_{t-1}] = \sigma^2$$

3.The Lag Operator

$$
\begin{aligned}
L\, y_t &= y_{t-1} \\
L^2 y_t &= L(L(y_t)) = L(y_{t-1}) = y_{t-2} \\
B(L) &= b_0 + b_1 L + b_2 L^2 + \ldots\, b_m L^m \\
L^m\, y_t &= y_{t-m} \\
\Delta y_t &= (1 - L) y_t = y_t - y_{t-1} \\
(1 + .9L + .6L^2) y_t &= y_t + .9 y_{t-1} + .6 y_{t-2} \\
B(L) &= b_0 + b_1 L + b_2 L^2 + \ldots = \sum_{i=0}^{\infty} b_i L^i \\
B(L)\, \varepsilon_t &= b_0\, \varepsilon_t + b_1\, \varepsilon_{t-1} + b_2\, \varepsilon_{t-2} + \ldots = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}
\end{aligned}
$$

4.Wold's Theorem, the General Linear Process, and Rational Distributed Lags

Wold's Theorem

Let $\{y_t\}$ be any zero-mean covariance-stationary process. Then:

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

where

$$b_0 = 1$$

and

$$\sum_{i=0}^{\infty} b_i^2 < \infty$$

The General Linear Process

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$$

$$\varepsilon_t \overset{iid}{\sim} WN(0, \sigma^2),$$

where $b_0 = 1$ and

$$\sum_{i=0}^{\infty} b_i^2 < \infty$$

The General Linear Process Cont.

$$E(y_t) = E(\sum_{i=0}^{\infty} b_i \varepsilon_{t-i}) = \sum_{i=0}^{\infty} b_i E\varepsilon_{t-i} = \sum_{i=0}^{\infty} b_i \bullet 0 = 0$$

$$var(y_t) = var(\sum_{i=0}^{\infty} b_i \varepsilon_{t-i}) = \sum_{i=0}^{\infty} b_i^2 var(\varepsilon_{t-i}) = \sum_{i=0}^{\infty} b_i^2 \sigma^2 = \sigma^2 \sum_{i=0}^{\infty} b$$

$$E(y_t|\Omega_{t-1}) = E(\varepsilon_t|\Omega_{t-1}) + b_1 E(\varepsilon_{t-1}|\Omega_{t-1}) + b_2 E(\varepsilon_{t-2}|\Omega_{t-1}) + ... =$$

$$var(y_t|\Omega_{t-1}) = E[(y_t - E(y_t|\Omega_{t-1}))^2|\Omega_{t-1}] = E(\varepsilon_t^2|\Omega_{t-1}) = E(\varepsilon_t^2)$$

Rational Distributed Lags

$$B(L) = \frac{\Theta\ (L)}{\Phi\ (L)}$$

$$\Theta(L) = \sum_{i=0}^{q} \theta_i L^i$$

$$\Phi(L) = \sum_{i=0}^{p} \phi_i L^i$$

$$B(L) \approx \frac{\Theta\ (L)}{\Phi\ (L)}$$

5.Estimation and Inference for the Mean, Auto Correlation and Partial Autocorrelation Functions

$$\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$$

$$\rho(\tau) = \frac{E\ [(y_t - \mu)\ (y_{t-\tau} - \mu)]}{E[(y_t - \mu)^2]}$$

$$\hat{\rho}(\tau) = \frac{\frac{1}{T} \sum_{t=\tau+1}^{T} [(y_t - \bar{y})(y_{t-\tau} - \bar{y})]}{\frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})^2} = \frac{\sum_{t=\tau+1}^{T} [(y_t - \bar{y})}{\sum_{t=1}^{T} (y_t -}$$

$$\hat{y}_t = \hat{c} + \hat{\beta}_1\ y_{t-1} + \ldots + \hat{\beta}_\tau\ y_{t-\tau}$$

$$\hat{p}(\tau) \equiv \hat{\beta}_\tau$$

$$\hat{\rho}(\tau),\ \hat{p}(\tau) \sim N\left(0, \frac{1}{T}\right)$$

5.Estimation and Inference for the Mean, Auto Correlation and Partial Autocorrelation Functions Cont.

$$\hat{\rho}(\tau) \;\sim\; \mathrm{N}(0,\; \frac{1}{\mathrm{T}})$$

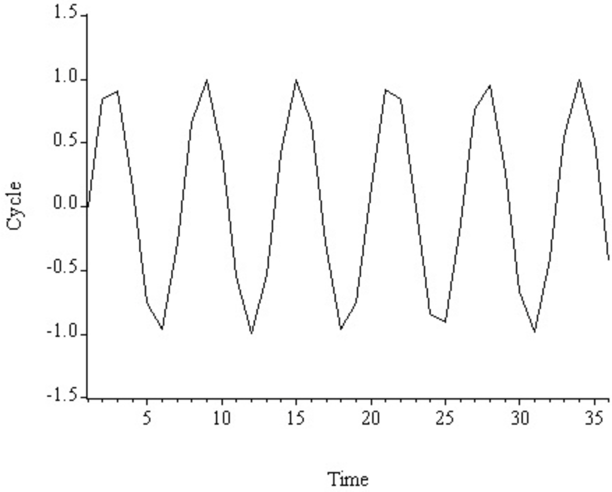
$$\mathrm{rootT}\hat{\rho}(\tau) \;\sim\; \mathrm{N}(0,\; 1)$$

$$\mathrm{T}\;\hat{\rho}^2(\tau) \;\sim\; \chi^2_1$$

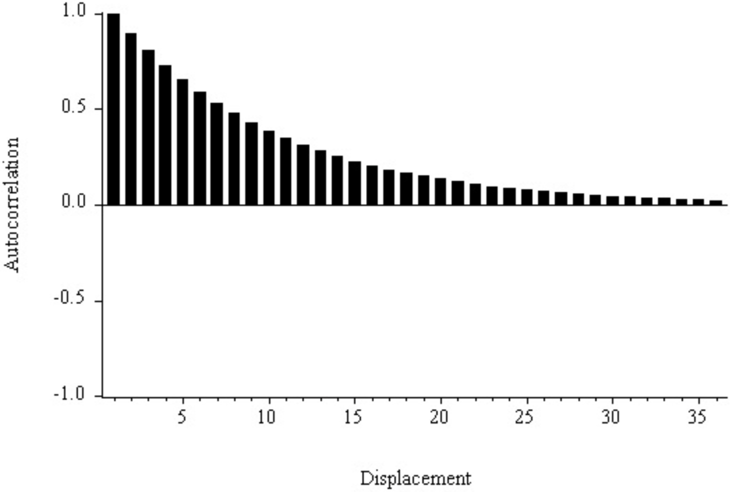
$$\mathrm{Q_{BP}} \;=\; \mathrm{T} \sum_{\tau=1}^{\mathrm{m}} \hat{\rho}^2(\tau)$$

$$\mathrm{Q_{LB}} \;=\; \mathrm{T}\,(\mathrm{T}+2) \sum_{\tau=1}^{\mathrm{m}} \left(\frac{1}{\mathrm{T}-\tau}\right) \;\hat{\rho}^2(\tau)$$

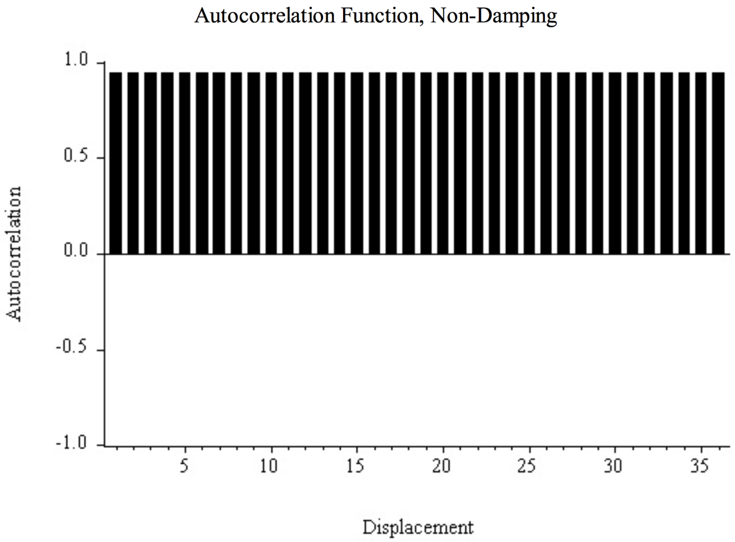
A Rigid Cycle Pattern



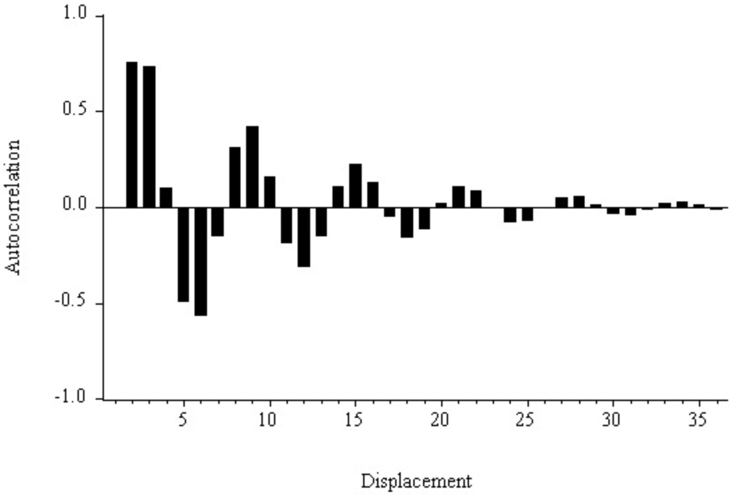
Autocorrelation Function, One-Sided Gradual Damping



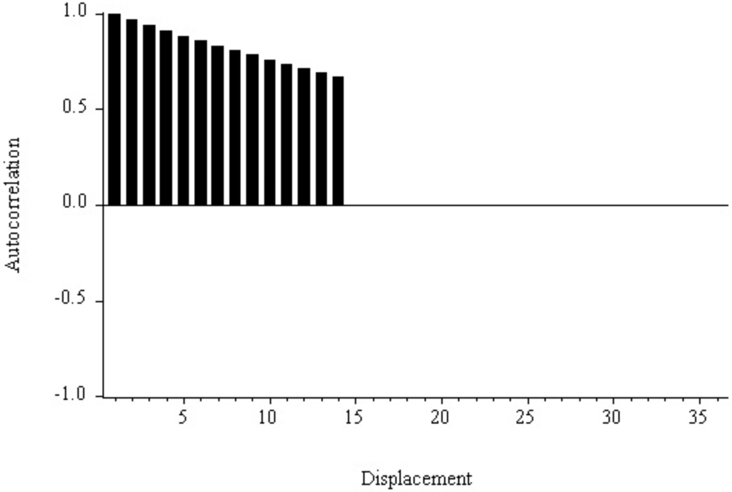
Autocorrelation Function, Non-Damping



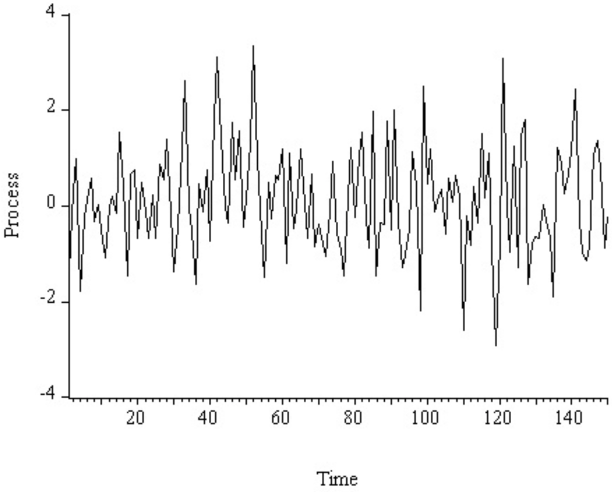
Autocorrelation Function, Gradual Damped Oscillation



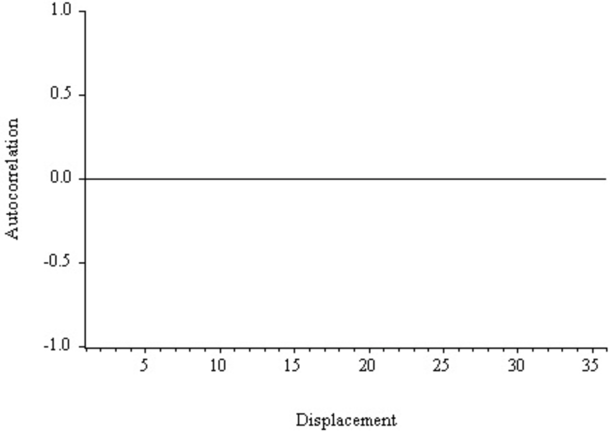
Autocorrelation Function, Sharp Cutoff



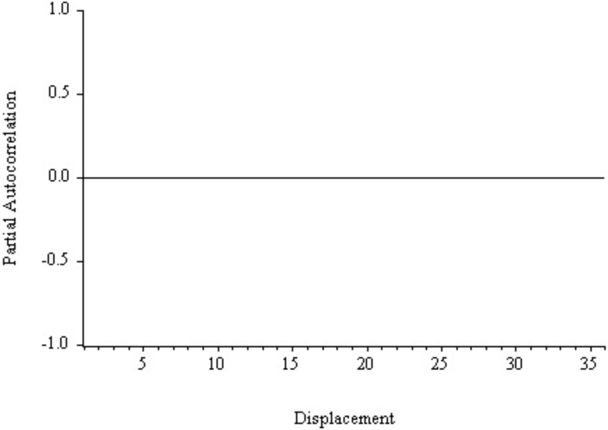
Realization of White Noise Process



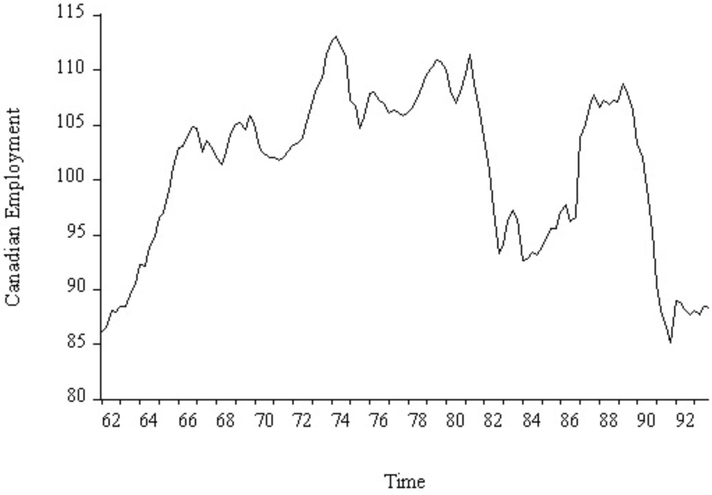
Population Autocorrelation Function of White Noise Process



Population Partialautocorrelation Function of White Noise Process



Canadian Employment Index

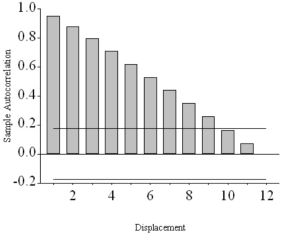


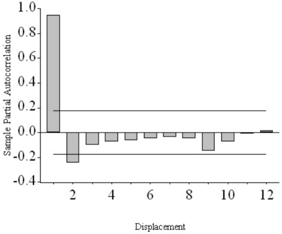
Canadian Employment Index Correlogram

Sample: 1962:1 1993:4
Included observations: 128

	Acorr.	P. Acorr.	Std. Error	Ljung-Box	p-v
v					
1	0.949	0.949	.088	118.07	0.000
2	0.877	-0.244	.088	219.66	0.000
3	0.795	-0.101	.088	303.72	0.000
4	0.707	-0.070	.088	370.82	0.000
5	0.617	-0.063	.088	422.27	0.000
6	0.526	-0.048	.088	460.00	0.000
7	0.438	-0.033	.088	486.32	0.000
8	0.351	-0.049	.088	503.41	0.000
9	0.258	-0.149	.088	512.70	0.000
10	0.163	-0.070	.088	516.43	0.000
11	0.073	-0.011	.088	517.20	0.000
12	-0.005	0.016	.088	517.21	0.000

Canadian Employment Index, Sample Autocorrelation and Partial Autocorrelation Functions





Modeling Cycles: MA,AR, and ARMA Models

The MA(1) Process

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1} = (1 + \theta L)\varepsilon_t$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

If invertible:

$$y_t = \varepsilon_t + \theta y_{t-1} - \theta^2 y_{t-2} + \theta^3 y_{t-3} - \ldots$$

Modeling Cycles: MA,AR, and ARMA Models Cont.

$$\mathrm{E}y_t = \mathrm{E}(\varepsilon_t) + \theta \mathrm{E}(\varepsilon_{t-1}) = 0$$

$$\mathrm{var}(y_t) = \mathrm{var}(\varepsilon_t) + \theta^2 \mathrm{var}(\varepsilon_{t-1}) = \sigma^2 + \theta^2\sigma^2 = \sigma^2(1 + \theta^2)$$

$$\mathrm{E}(y_t|\Omega_{t-1}) = \mathrm{E}((\varepsilon_t + \theta\varepsilon_{t-1})|\Omega_{t-1}) = \mathrm{E}(\varepsilon_t|\Omega_{t-1}) + \theta \mathrm{E}(\varepsilon_{t-1}|\Omega_{t-1}$$

$$\mathrm{var}(y_t|\Omega_{t-1}) = \mathrm{E}[(y_t - \mathrm{E}(y_t|\Omega_{t-1}))^2|\Omega_{t-1}] = \mathrm{E}(\varepsilon_t^2|\Omega_{t-1}) = \mathrm{E}(\varepsilon_t^2)$$

The MA(q) Process

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} = \Theta(L)\varepsilon_t$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

where

$$\Theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$$

The AR(1) Process

$$y_t = \phi y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

If covariance stationary:

$$y_t = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + ...$$

Moment Structure

$$E(y_t) \; = \; E(\varepsilon_t \; + \; \phi\varepsilon_{t-1} \; + \; \phi^2\varepsilon_{t-2} \; + \; ...)$$

$$= \; E(\varepsilon_t) \; + \; \phi E(\varepsilon_{t-1}) \; + \; \phi^2 E(\varepsilon_{t-2}) \; + \; ...$$

$$= \; 0$$

$$var(y_t) \; = \; var(\varepsilon_t \; + \; \phi\varepsilon_{t-1} \; + \; \phi^2\varepsilon_{t-2} \; + \; ...)$$

$$= \; \sigma^2 \; + \; \phi^2\sigma^2 \; + \; \phi^4\sigma^2 \; + \; ...$$

$$= \; \sigma^2 \sum_{i=0}^{\infty} \phi^{2i}$$

$$= \; \sigma^{\frac{2}{1-\phi^2}}$$

Moment Structure Cont.

$$
\begin{aligned}
E(y_t|y_{t-1}) &= E((\phi y_{t-1} + \varepsilon_t) \mid y_{t-1}) \\
&= \phi E(y_{t-1}|y_{t-1}) + E(\varepsilon_t|y_{t-1}) \\
&= \phi y_{t-1} + 0 \\
&= \phi y_{t-1}
\end{aligned}
$$

$$
\begin{aligned}
var(y_t|y_{t-1}) &= var((\phi y_{t-1} + \varepsilon_t) \mid y_{t-1}) \\
&= \phi^2 var(y_{t-1}|y_{t-1}) + var(\varepsilon_t|y_{t-1}) \\
&= 0 + \sigma^2 \\
&= \sigma^2
\end{aligned}
$$

Moment Structure Cont.

Autocovariances and autocorrelations:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

$$y_t y_{t-\tau} = \phi y_{t-1} y_{t-\tau} + \varepsilon_t y_{t-\tau}$$

For

$$\tau \geq 1$$

,

$$\gamma(\tau) = \phi \gamma(\tau - 1).$$

(Yule-Walker equation) But

$$\gamma(0) = \sigma^{\frac{2}{1-\phi^2}}$$

Moment Structure Cont.

. Thus

$$\gamma(\tau) \ = \ \phi^{\tau} \ \sigma^{\frac{2}{1-\phi^2}}, \ \tau \ = \ 0, \ 1, \ 2, \ ....$$

and

$$\rho(\tau) \ = \ \phi^{\tau}, \ \tau \ = \ 0, \ 1, \ 2, \ ....$$

Partial autocorrelations:

The AR(p) Process

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

The ARMA(1,1) Process

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

MA representation if invertible:

$$y_t = \frac{(1 + \theta L)}{(1 - \phi L)} \varepsilon_t$$

AR representation of covariance stationary:

$$\frac{(1 - \phi L)}{(1 + \theta L)} y_t = \varepsilon_t$$

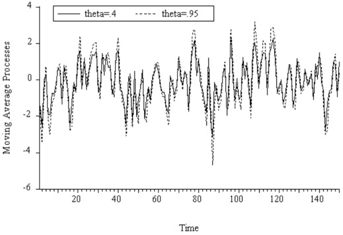
The ARMA(p,q) Process

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

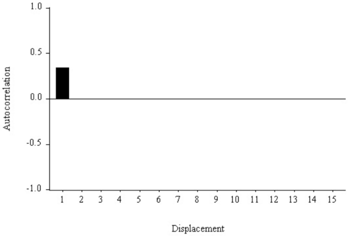
$$\Phi(L) y_t = \Theta(L) \varepsilon_t$$

Realization of Two MA(1) Processes



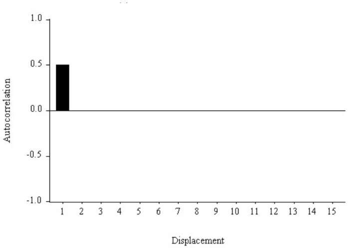
Population Autocorrelation Function MA(1) Process

$$\theta = .4$$



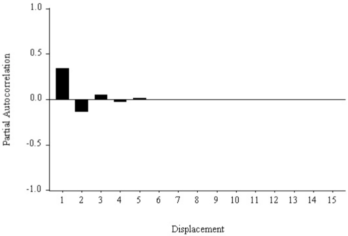
Population Autocorrelation Function MA(1) Process

$$\theta = .95$$



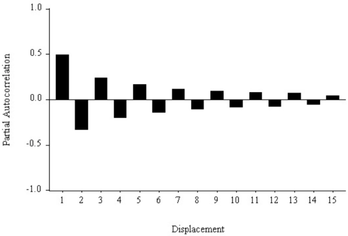
Population Partial Autocorrelation Function MA(1) Process

$$\theta = .4$$

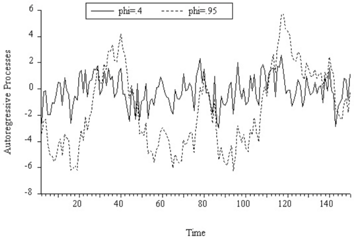


Population Partial Autocorrelation Function MA(1) Process

$$\theta = .95$$

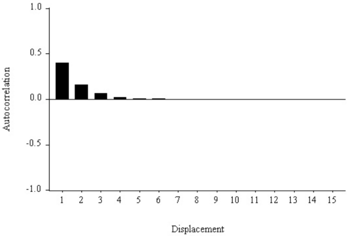


Realization of Two AR(1) Processes



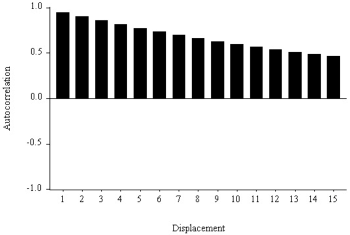
Population Autocorrelation Function AR(1) Process

$$\phi = .4$$



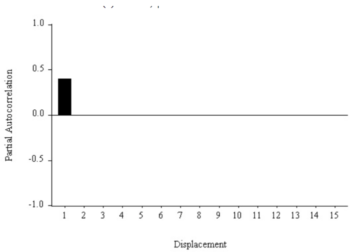
Population Autocorrelation Function AR(1) Process

$$\phi = .95$$



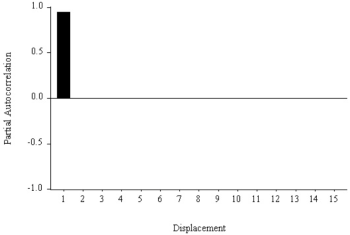
Population Partial Autocorrelation Function AR(1) Process

$$\phi = .4$$

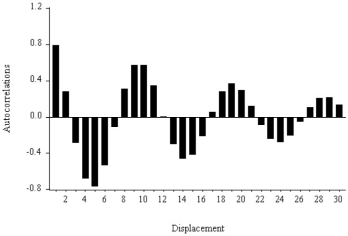


Population Partial Autocorrelation Function AR(1) Process

$$\phi = .95$$



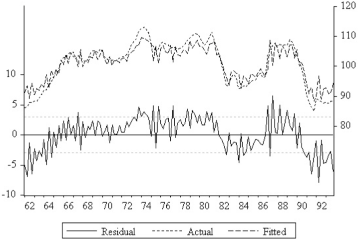
Population Autocorrelation Function AR(2) Process with Complex Roots



Employment: MA(4) Model

Sample: 1962:1 1993:4
Included observations: 128

Employment MA(4) Residual Plot



Employment: MA(4) Model

Residual Correlogram Sample: 1962:1 1993:4

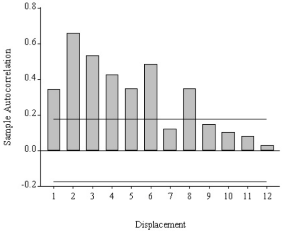
Included observations: 128

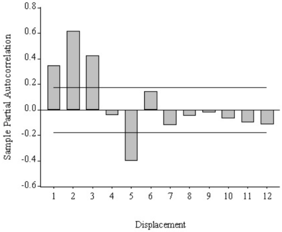
Q−statistic probabilities adjusted for 4 ARMA term(s)

	Acorr.	P. Acorr.	Std. Error	Ljung-Box	p-v
1	0.345	0.345	.088	15.614	
2	0.660	0.614	.088	73.089	
3	0.534	0.426	.088	111.01	
4	0.427	−0.042	.088	135.49	
5	0.347	-0.398	.088	151.79	0.000
6	0.484	0.145	.088	183.70	0.000
7	0.121	−0.118	.088	185.71	0.000
8	0.348	−0.048	.088	202.46	0.000
9	0.148	−0.019	.088	205.50	0.000
10	0.102	−0.066	.088	206.96	0.000
11	0.081	−0.098	.088	207.89	0.000
12	0.029	−0.113	.088	208.01	0.000

Employment: MA(4) Model

Residual Sample Autocorrelation and Partial Autocorrelation Functions, With Plus or Minus Two Standard Error Bands

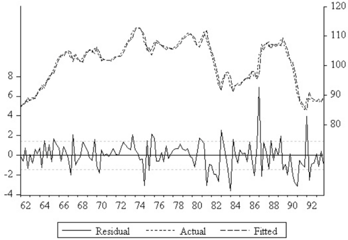




Employment: AR(2) Model

LS // Dependent Variable is CANEMP
Sample: 1962:1 1993:4
Included observations: 128
Convergence achieved after 3 iterations
Variable Coefficient Std. Error t−Statistic Prob.

Employment AR(2) Model Residual Plot



Employment AIC Values of Various ARMA Models

				MA Order		
		0	1	2	3	4
	0		2.86	2.32	2.47	2.20
	1	1.01	.83	.79	.80	.81
AR Order	2	.762	.77	.78	.80	.80
	3	.77	.761	.77	.78	.79
	4	.79	.79	.77	.79	.80

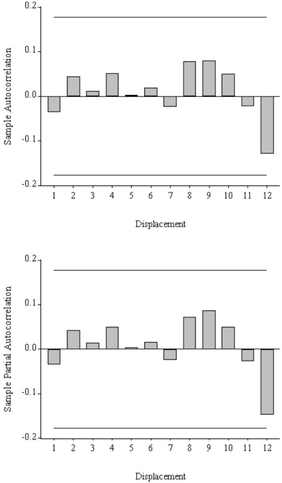
Employment SIC Values of Various ARMA Models

				MA Order		
		0	1	2	3	4
	0		2.91	2.38	2.56	2.31
	1	1.05	.90	.88	.91	.94
AR Order	2	.83	.86	.89	.92	.96
	3	.86	.87	.90	.94	.96
	4	.90	.92	.93	.97	1.00

Employment: ARMA(3,1) Model

LS // Dependent Variable is CANEMP
Sample: 1962:1 1993:4
Included observations: 128
Convergence achieved after 17 iterations
Variable Coefficient Std. Error t−Statistic Prob.

Employment ARMA(3) Model Residual Plot



Employment: ARMA(3,1) Model Residual Correlogram

Sample: 1962:1 1993:4

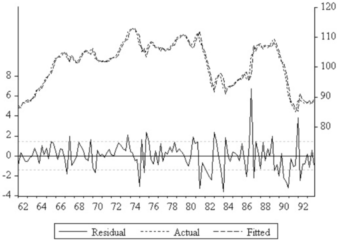
Included observations: 128

Q−statistic probabilities adjusted for 4 ARMA term(s)

	Acorr.	P. Acorr.	Std. Error	Ljung-Box	p-v
1	−0.032	−0.032	.09	0.1376	
2	0.041	0.040	.09	0.3643	
3	0.014	0.017	.09	0.3904	
4	0.048	0.047	.09	0.6970	
5	0.006	0.007	.09	0.7013	0.402
6	0.013	0.009	.09	0.7246	0.696
7	−0.017	−0.019	.09	0.7650	0.858
8	0.064	0.060	.09	1.3384	0.855
9	0.092	0.097	.09	2.5182	0.774
10	0.039	0.040	.09	2.7276	0.842
11	−0.016	−0.022	.09	2.7659	0.90
12	−0.137	−0.153	.09	5.4415	0.71

Employment: ARMA(3) Model

Residual Sample Autocorrelation and Partial Autocorrelation Functions, With Plus or Minus Two Standard Error Bands



Forecasting Cycles

$$\Omega_T = \{y_T, y_{T-1}, y_{T-2}, ...\},$$

$$\Omega_T = \{\varepsilon_T, \varepsilon_{T-1}, \varepsilon_{T-2}, ...\}.$$

Optimal Point Forecasts for Infinite-Order Moving Averages

$$y_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i},$$

where

$$\varepsilon_t \sim WN(0, \sigma^2)$$

,

Forecasting Cycles Cont.

$$b_0 = 1$$

, and

$$\sigma^2 \sum_{i=0}^{\infty} b_i^2 < \infty$$

.

$$y_{T+h} = \varepsilon_{T+h} + b_1 \varepsilon_{T+h-1} + ... + b_h \varepsilon_T + b_{h+1} \varepsilon_{T-1} + ...$$

$$y_{T+h,T} = b_h \varepsilon_T + b_{h+1} \varepsilon_{T-1} + ...$$

$$e_{T+h,T} = (y_{T+h} - y_{T+h,T}) = \sum_{i=0}^{h-1} b_i \varepsilon_{T+h-i},$$

$$\sigma_h^2 = \sigma^2 \sum_{i=0}^{h-1} b_i^2.$$

Interval and Density Forecasts

$$y_{T+h} = y_{T+h,T} + e_{T+h,T}.$$

95% h-step-ahead interval forecast:

$$y_{T+h,T} \pm 1.96\sigma_h$$

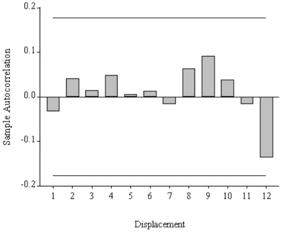
h-step-ahead density forecast:

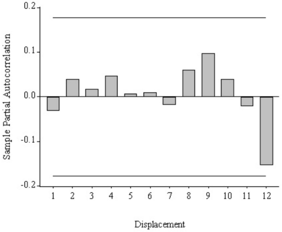
$$N(y_{T+h,T},\ \sigma_h^2)$$

Making the Forecasts Operational

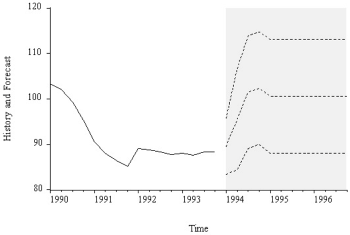
The Chain Rule of Forecasting

Employment History and Forecast MA(4) Model

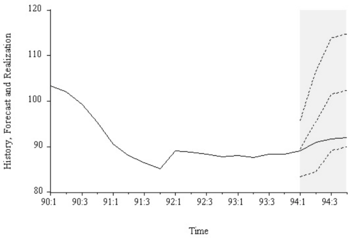




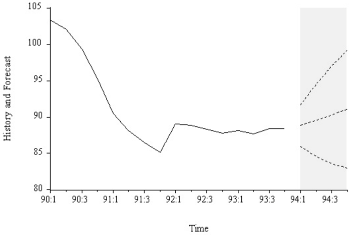
Employment History and Long-Horizon Forecast MA(4) Model



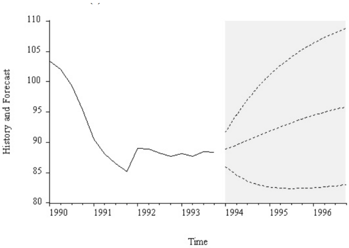
Employment History, Forecast and Realization MA(4) Model



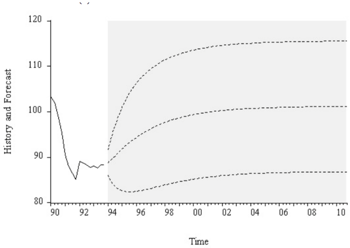
Employment History and Forecast AR(2) Model



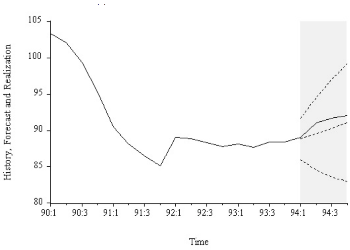
Employment History and Long-Horizon Forecast AR(2) Model



Employment History and Very Long-Horizon Forecast AR(2) Model



Employment History, Forecast and Realization AR(2) Model



Putting it all Together

A Forecast Model with Trend, Seasonal and Cyclical Components

The full model:

$$y_t = T_t(\theta) + \sum_{i=1}^{s} \gamma_i D_{it} + \sum_{i=1}^{v_1} \delta_i^{HD} HDV_{it} + \sum_{i=1}^{v_2} \delta_i^{TD} TDV_{it} + \varepsilon_t$$

$$\Phi(L)\varepsilon_t = \Theta(L)v_t$$

$$\Phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$$

$$\Theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$$

$$v_t \sim WN(0, \sigma^2).$$

Point Forecasting

$$y_{T+h} = T_{T+h}(\theta) + \sum_{i=1}^{s} \gamma_i D_{i,T+h} + \sum_{i=1}^{v_1} \delta_i^{HD} HDV_{i,T+h} + \sum_{i=1}^{v_2} \delta_i^{TD} T$$

$$y_{T+h,T} = T_{T+h}(\theta) + \sum_{i=1}^{s} \gamma_i D_{i,T+h} + \sum_{i=1}^{v_1} \delta_i^{HD} HDV_{i,T+h} + \sum_{i=1}^{v_2} \delta_i^{TD}$$

$$\hat{y}_{T+h,T} = T_{T+h}(\hat{\theta}) + \sum_{i=1}^{s\hat{\gamma}_i} D_{i,T+h} + \sum_{i=1}^{v_1} \hat{\delta}_i^{HD} HDV_{i,T+h} + \sum_{i=1}^{v_2} \hat{\delta}_i^{TD} T$$

Interval Forecasting and Density Forecasting

Interval Forecasting:

$$\hat{y}_{T+h,T} \ \pm \ z_{\alpha/2}\hat{\sigma}_h$$

e.g.: (95% interval)

$$\hat{y}_{T+h,T} \ \pm \ 1.96\hat{\sigma}_h$$

Density Forecasting:

$$N(\hat{y}_{T+h,T}, \ \hat{\sigma}_h^2)$$

Recursive Estimation

$$y_t = \sum_{k=1}^{K} \beta_k x_{kt} + \varepsilon_t$$

$$\varepsilon_t \sim iidN(0, \sigma^2),$$

$$t = 1, ..., T.$$

OLS estimation uses the full sample, $t = 1, ..., T$.

Recursive least squares uses an expanding sample.
Begin with the first K observations and estimate the model.
Then estimate using the first $K + 1$ observations, and so on.
At the end we have a set of recursive parameter estimates:
$\hat{\beta}_{k,t}$, for $k = 1, ..., K$ and $t = K, ..., T$.

Recursive Residuals

At each t, $t = K, ..., T-1$, compute a 1-step forecast,

$$\hat{y}_{t+1,t} = \sum_{k=1}^{K} \hat{\beta}_{kt} x_{k,t+1}.$$

The corresponding forecast errors, or recursive residuals, are

$$\hat{e}_{t+1,t} = y_{t+1} - \hat{y}_{t+1,t}.$$

$$\hat{e}_{t+1,t} \sim N(0, \sigma^2 r_t)$$

where $r_t > 1$ for all t

Standardized Recursive Residuals and CUSUM

$$w_{t+1,t} \equiv \frac{\hat{e}_{t+1,t}}{\sigma\sqrt{r_t}},$$

$$t = K, ..., T-1.$$

Under the maintained assumptions,

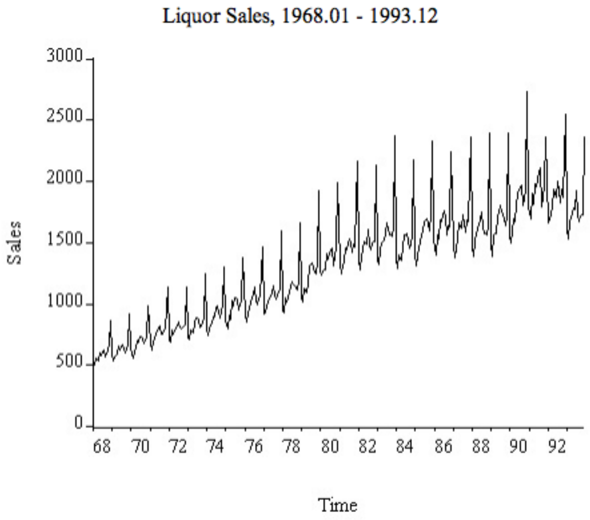
$$w_{t+1,t} \sim iidN(0,1).$$

Then

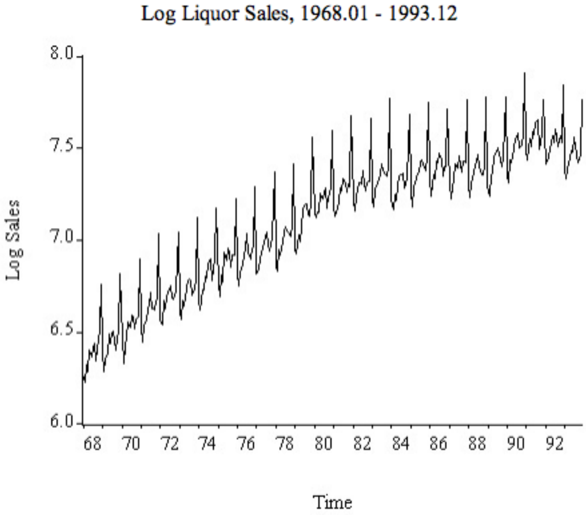
$$CUSUM_{t*} \equiv \sum_{t=K}^{t^*} w_{t+1,t}, \;\; t^* = K, ..., T-1$$

is just a sum of *iid* $N(0,1)$'s.

Liquor Sales, 1968.1-1993.12



Log Liquor Sales, 1968.01 - 1993.12



Log Liquor Sales: Quadratic Trend Regression

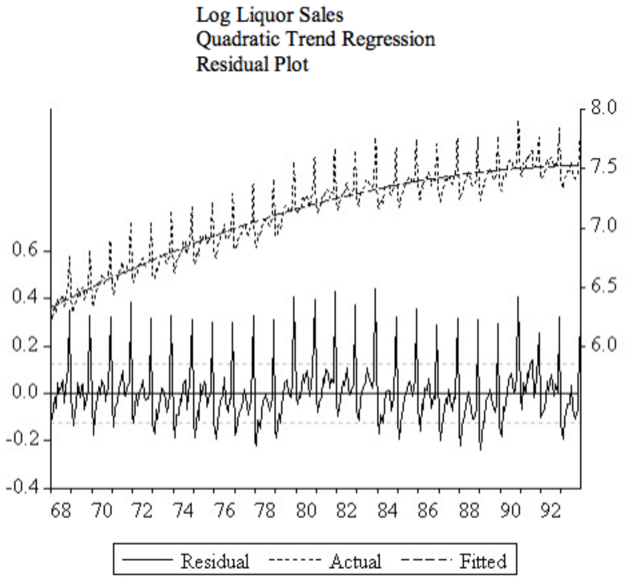
Log Liquor Sales
Quadratic Trend Regression

LS // Dependent Variable is LSALES
Sample: 1968:01 1993:12
Included observations: 312

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.237356	0.024496	254.6267	0.0000
TIME	0.007690	0.000336	22.91552	0.0000
TIME2	-1.14E-05	9.74E-07	-11.72695	0.0000

R-squared	0.892394	Mean dependent var	7.112383
Adjusted R-squared	0.891698	S.D. dependent var	0.379308
S.E. of regression	0.124828	Akaike info criterion	-4.152073
Sum squared resid	4.814823	Schwarz criterion	-4.116083
Log likelihood	208.0146	F-statistic	1281.296
Durbin-Watson stat	1.752858	Prob(F-statistic)	0.000000

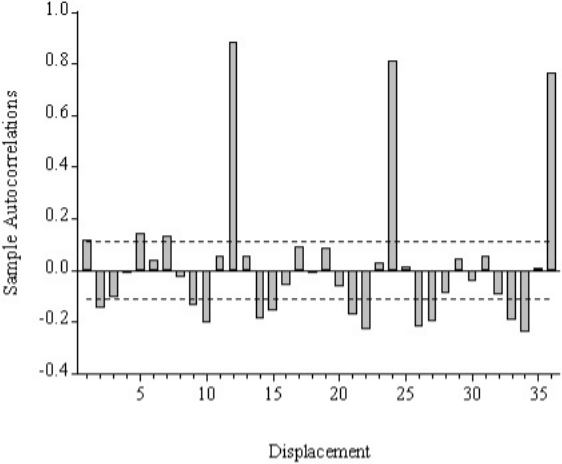
Liquor Sales Quadratic Trend Regression Residual Plot



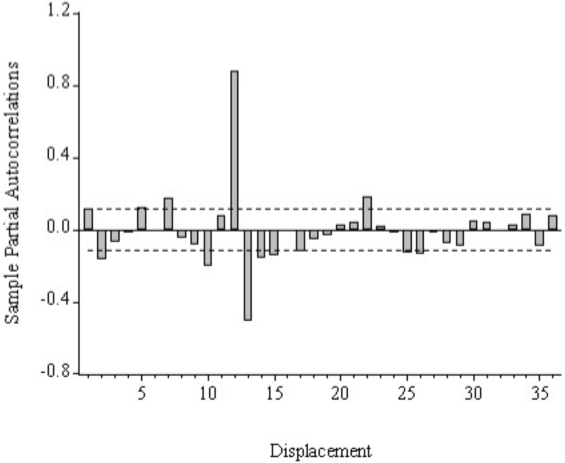
Liquor Sales Quadratic Trend Regression Residual Correlogram

	Acorr.	P. Acorr.	Std. Error	Ljung-Box
1	0.117	0.117	.056	4.3158 0.0
2	−0.149	−0.165	.056	11.365 0.0
3	−0.106	−0.069	.056	14.943 0.0
4	−0.014	−0.017	.056	15.007 0.0
5	0.142	0.125	.056	21.449 0.0
6	0.041	−0.004	.056	21.979 0.0
7	0.134	0.175	.056	27.708 0.0
8	−0.029	−0.046	.056	27.975 0.0
9	−0.136	−0.080	.056	33.944 0.0
10	−0.205	−0.206	.056	47.611 0.0
11	0.056	0.080	.056	48.632 0.0
12	0.888	0.879	.056	306.26 0.0
13	0.055	−0.507	.056	307.[illegible]
14	−0.187	−0.159	.056	318.79 0.0
15	−0.159	−0.144	.056	327.17 0.0
16	[illegible]	[illegible]	.056	[illegible]

Liquor Sales Quadratic Trend Regression Residual Sample Autocorrelation Functions



Liquor Sales Quadratic Trend Regression Residual Partial Autocorrelation Functions



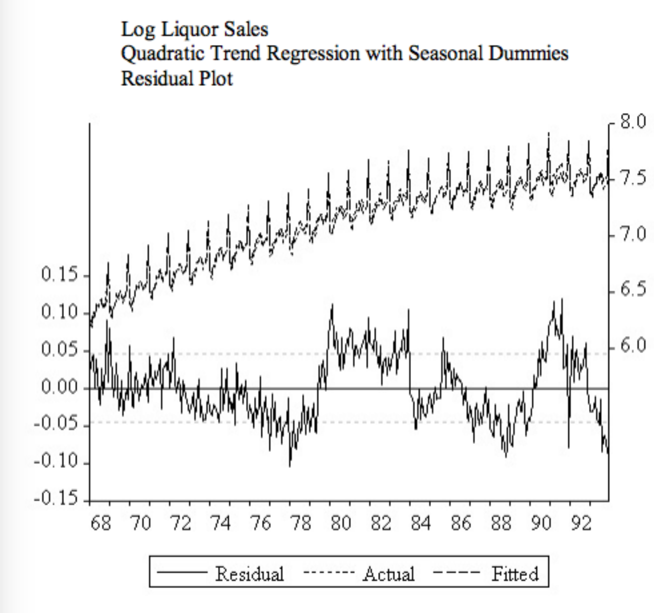
Log Liquor Sales: Quadratic Trend Regression With Seasonal Dummies and AR(3) Disturbances

LS // Dependent Variable is LSALES
Sample: 1968:01 1993:12
Included observations: 312

Variable	Coefficient	Std. Error	t-Statistic	Prob.
TIME	0.007656	0.000123	62.35882	0.0000
TIME2	-1.14E-05	3.56E-07	-32.06823	0.0000
D1	6.147456	0.012340	498.1699	0.0000
D2	6.088653	0.012353	492.8890	0.0000
D3	6.174127	0.012366	499.3008	0.0000
D4	6.175220	0.012378	498.8970	0.0000
D5	6.246086	0.012390	504.1398	0.0000
D6	6.250387	0.012401	504.0194	0.0000
D7	6.295979	0.012412	507.2402	0.0000
D8	6.268043	0.012423	504.5509	0.0000
D9	6.203832	0.012433	498.9630	0.0000
D10	6.229197	0.012444	500.5968	0.0000
D11	6.259770	0.012453	502.6602	0.0000
D12	6.580068	0.012463	527.9819	0.0000

R-squared	0.986111	Mean dependent var	7.112383
Adjusted R-squared	0.985505	S.D. dependent var	0.379308
S.E. of regression	0.045666	Akaike info criterion	-6.128963
Sum squared resid	0.621448	Schwarz criterion	-5.961008
Log likelihood	527.4094	F-statistic	1627.567
Durbin-Watson stat	0.586187	Prob(F-statistic)	0.000000

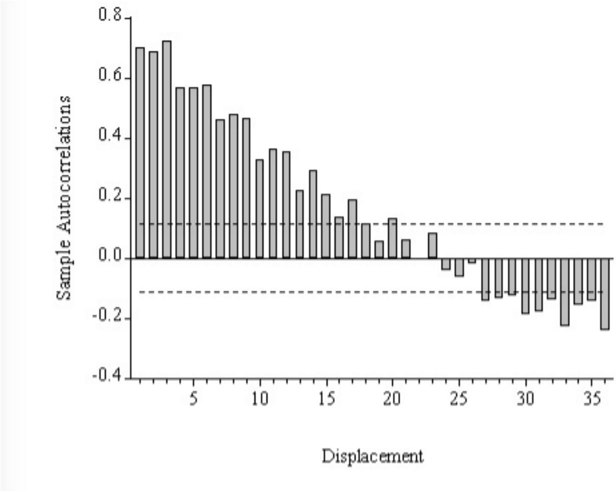
Liquor Sales Quadratic Trend Regression with Seasonal Dummies Residual Plot



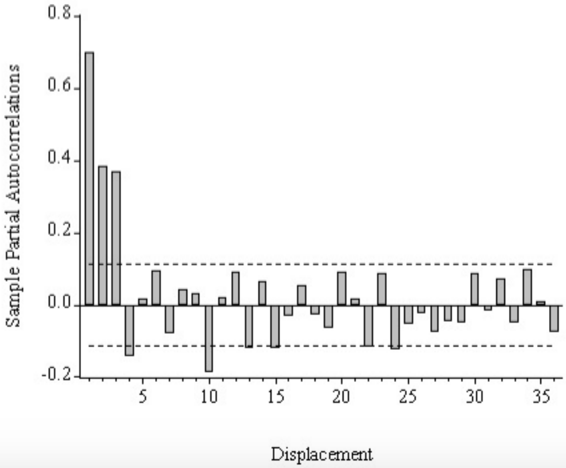
Liquor Sales Quadratic Trend Regression with Seasonal Dummies Residual Correlogram

	Acorr.	P. Acorr.	Std.Error	Ljung-Box	p-value
1	0.700	0.700	.056	154.34	0.000
2	0.686	0.383	.056	302.86	0.000
3	0.725	0.369	.056	469.36	0.000
4	0.569	−0.141	.056	572.36	0.000
5	0.569	0.017	.056	675.58	0.000
6	0.577	0.093	.056	782.19	0.000
7	0.460	−0.078	.056	850.06	0.000
8	0.480	0.043	.056	924.38	0.000
9	0.466	0.030	.056	994.46	0.000
10	0.327	−0.188	.056	1029.1	0.000
11	0.364	0.019	.056	1072.1	0.000
12	0.355	0.089	.056	1113.3	0.000
13	0.225	−0.119	.056	1129.9	0.000
14	0.291	0.065	.056	1157.8	0.000
15	0.211	−0.119	.056	1172.4	0.000
16	0.138	−0.031	.056	1178.7	0.000

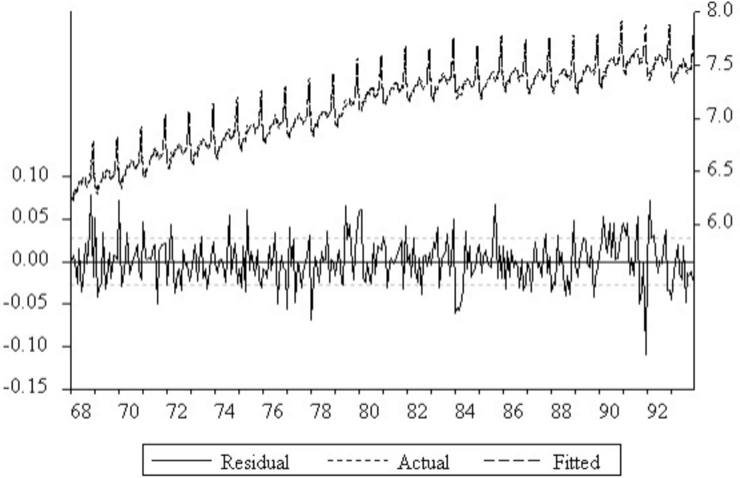
Liquor Sales Quadratic Trend Regression with Seasonal Dummies Residual Sample Autocorrelation Functions



Liquor Sales Quadratic Trend Regression with Seasonal Dummies Residual Sample Partial Autocorrelation Functions



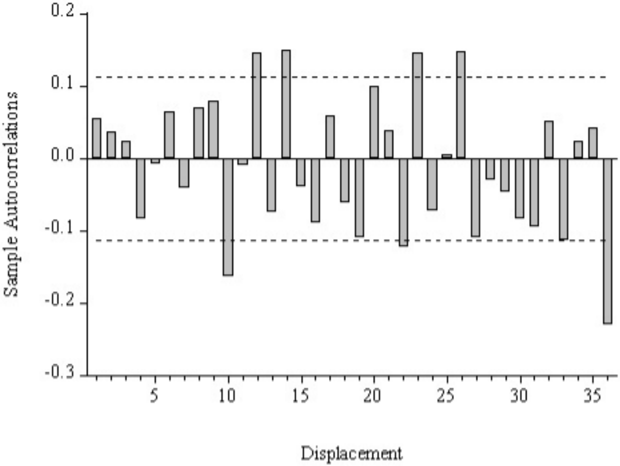
Liquor Sales Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances Residual Plot



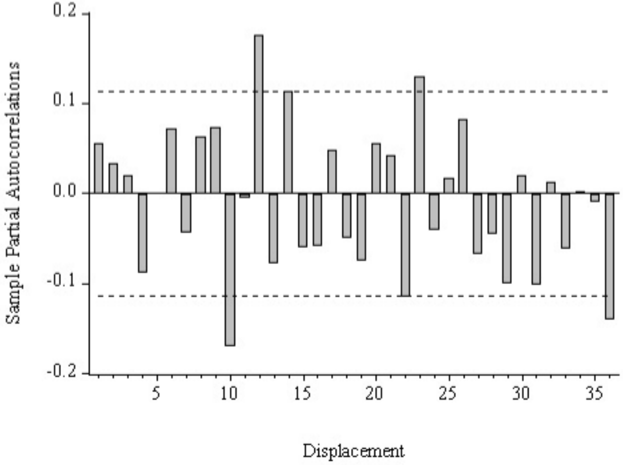
Liquor Sales Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances Residual Correlogram

	Acorr.	P. Acorr.	Std. Error	Ljung-Box	p-v
1	0.056	0.056	.056	0.9779	0.323
2	0.037	0.034	.056	1.4194	0.492
3	0.024	0.020	.056	1.6032	0.659
4	−0.084	−0.088	.056	3.8256	0.430
5	−0.007	0.001	.056	3.8415	0.572
6	0.065	0.072	.056	5.1985	0.519
7	−0.041	−0.044	.056	5.7288	0.572
8	0.069	0.063	.056	7.2828	0.506
9	0.080	0.074	.056	9.3527	0.405
10	−0.163	−0.169	.056	18.019	0.055
11	−0.009	−0.005	.056	18.045	0.081
12	0.145	0.175	.056	24.938	0.015
13	−0.074	−0.078	.056	26.750	0.01[illegible]
14	0.149	0.113	.056	34.034	0.00[illegible]
15	−0.039	−0.060	.056	34.532	0.003
16	−0.080	−0.058	.056	37.126	0.002

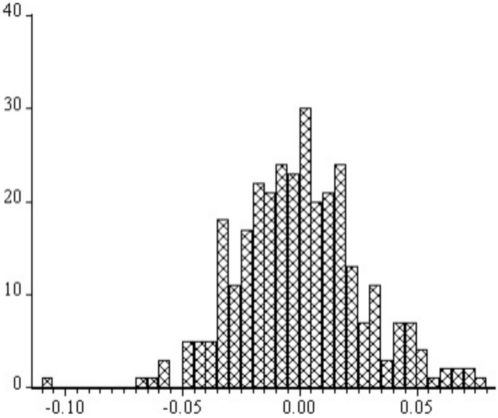
Liquor Sales Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances Residual Sample Autocorrelation Functions



Liquor Sales Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances Residual Sample Partial Autocorrelation Functions



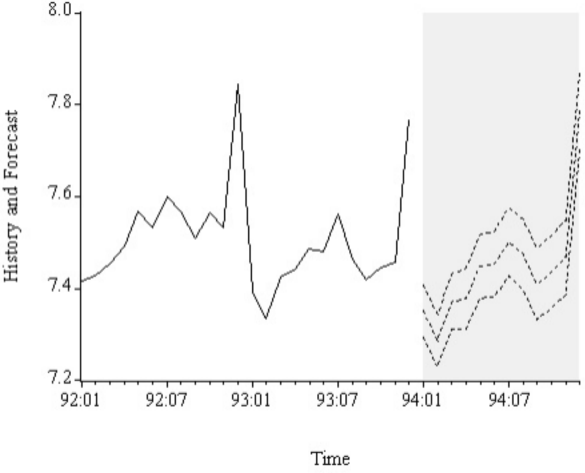
Liquor Sales Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances Residual Histogram and Normality Test



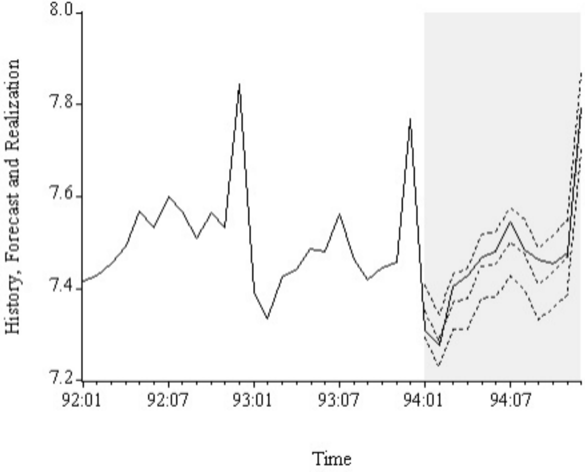
Series: Residuals
Sample 1968:01 1993:12
Observations 312

Mean	3.77E-16
Median	-0.000160
Maximum	0.078468
Minimum	-0.109856
Std. Dev.	0.026635
Skewness	0.077911
Kurtosis	3.740378
Jarque-Bera	7.441714
Probability	0.024213

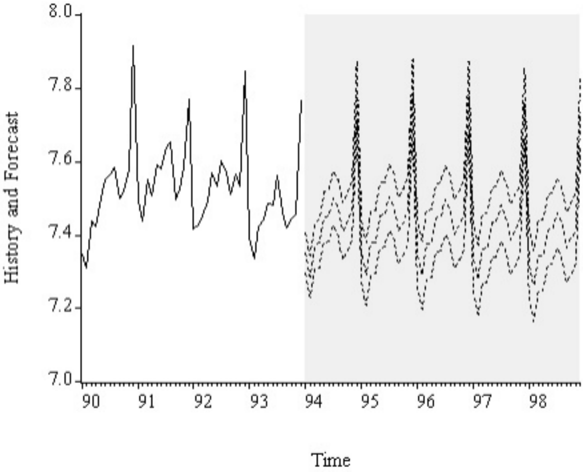
Log Liquor Sales History and 12-Month-Ahead Forecast



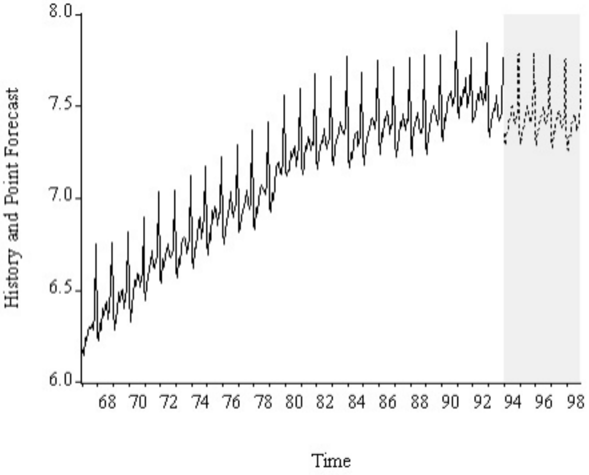
Log Liquor Sales History, 12-Month-Ahead Forecast, and Realization



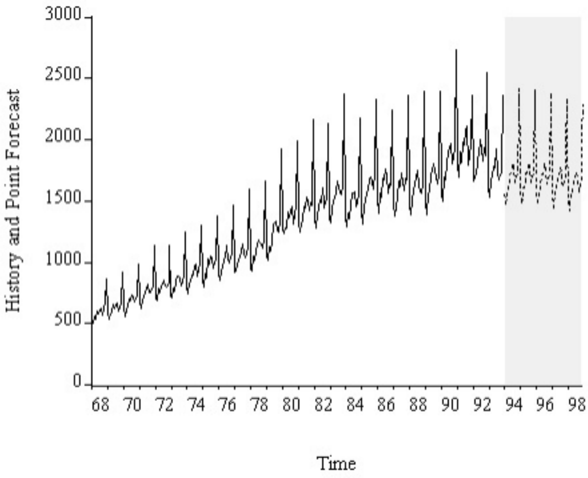
Log Liquor Sales History and 60-Month-Ahead Forecast



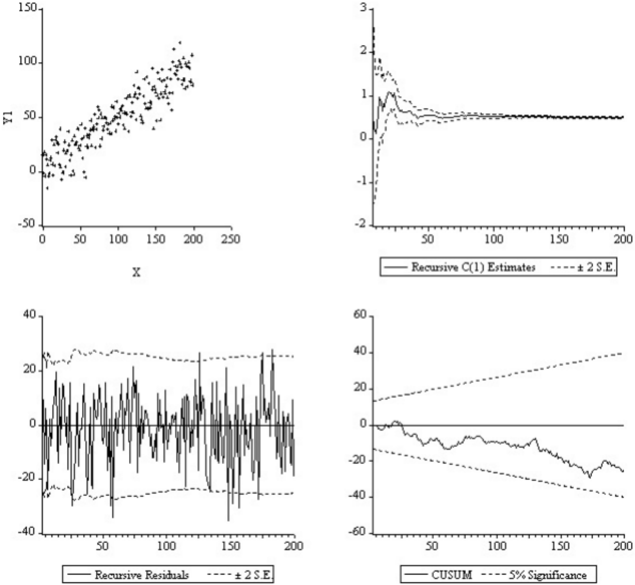
Log Liquor Sales Long History and 60-Month-Ahead Forecast



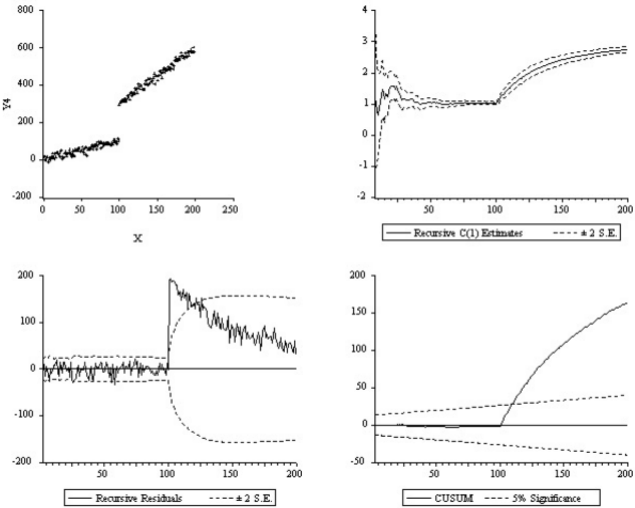
Liquor Sales Long History and 60-Month-Ahead Forecast



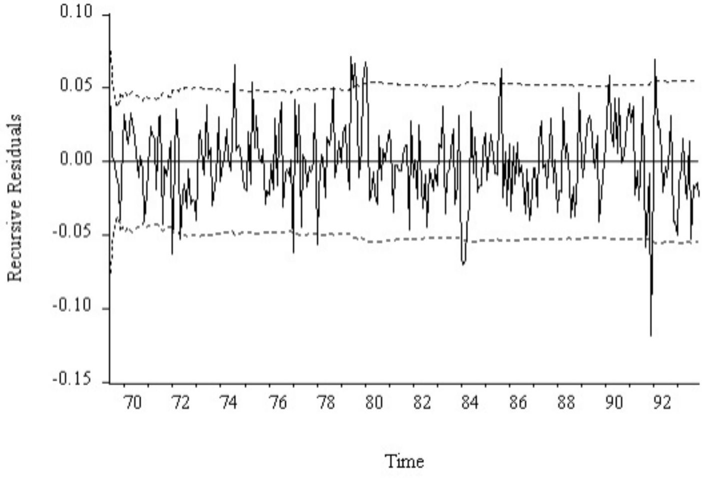
Recursive Analysis Constant Parameter Model



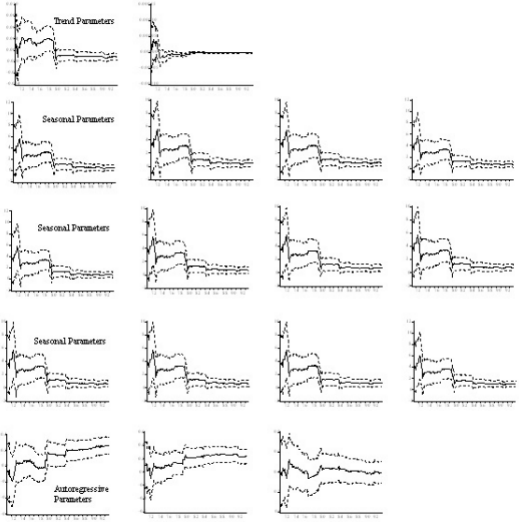
Recursive Analysis Breaking Parameter Model



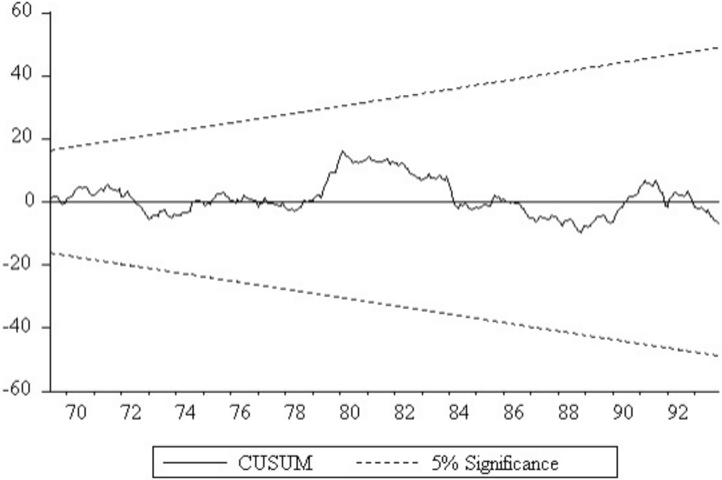
Log Liquor Sales: Quadratic Trend Regression with Seasonal Dummies and AR(3) Residuals and Two Standard Errors Bands



Log Liquor Sales: Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances Recursive Parameter Estimates



Log Liquor Sales: Quadratic Trend Regression with Seasonal Dummies and AR(3) Disturbances CUMSUM Analysis



Forecasting with Regression Models

Conditional Forecasting Models and Scenario Analysis

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

$$y_{T+h,T} | x^*_{T+h} = \beta_0 + \beta_1 x^*_{T+h}$$

Density forecast:

$$N(y_{T+h,T} | x^*_{T+h}, \sigma^2)$$

- "Scenario analysis," "contingency analysis"
- No "forecasting the RHS variables problem"

Unconditional Forecasting Models

$$y_{T+h,T} = \beta_0 + \beta_1 \, x_{T+h,T}$$

- "Forecasting the RHS variables problem"
- Could fit a model to x (e.g., an autoregressive model)
- Preferably, regress y on

$$x_{t-h}, \; x_{t-h-1}, \; ...$$

- No problem in trend and seasonal models

Distributed Lags

Start with unconditional forecast model:

$$y_t = \beta_0 + \delta x_{t-1} + \varepsilon_t$$

Generalize to

$$y_t = \beta_0 + \sum_{i=1}^{N_x} \delta_i x_{t-i} + \varepsilon_t$$

- "distributed lag model"
- "lag weights"
- "lag distribution"

Polynomial Distributed Lags

$$\min_{\beta_0, \delta_i} \sum_{t = N_x+1}^{T} \left[y_t - \beta_0 - \sum_{i=1}^{N_x} \delta_i x_{t-i} \right]^2$$

subject to

$$\delta_i = P(i) = a + bi + ci^2, \; i = 1, ..., N_x$$

- Lag weights constrained to lie on low-order polynomial
- Additional constraints can be imposed, such as

$$P(N_x) = 0$$

- Smooth lag distribution
- Parsimonious

Rational Distributed Lags

$$y_t = \frac{A(L)}{B(L)} x_t + \varepsilon_t$$

Equivalently,

$$B(L) y_t = A(L) x_t + B(L) \varepsilon_t$$

- Lags of x *and* y included
- Important to allow for lags of y, one way or another

Another way:

distributed lag regression with lagged dependent variables

$$y_t = \beta_0 + \sum_{i=1}^{N_y} \alpha_i y_{t-i} + \sum_{j=1}^{N_x} \delta_j x_{t-j} + \varepsilon_t$$

Another way:

distributed lag regression with ARMA disturbances

$$y_t = \beta_0 + \sum_{i=1}^{N_x} \delta_i x_{t-i} + \varepsilon_t$$

$$\varepsilon_t = \frac{\Theta(L)}{\Phi(L)} v_t$$

$$v_t \sim WN(0, \sigma^2)$$

Another Way: The Transfer function Model and Various Special Cases

Univariate ARMA

$$y_t = \frac{C(L)}{D(L)} \varepsilon_t$$

A(

Distributed Lag with

$$B(L) y_t = A(L) x_t + \varepsilon_t$$

, or

Lagged Dep. Variables

$$y_t = \frac{A(L)}{B(L)} x_t + \frac{1}{B(L)} \varepsilon_t$$

Distributed Lag with

Vector Autoregressions

e.g., bivariate VAR(1)

$$y_{1,t} = \phi_{11} y_{1,t-1} + \phi_{12} y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = \phi_{21} y_{1,t-1} + \phi_{22} y_{2,t-1} + \varepsilon_{2,t}$$

$$\varepsilon_{1,t} \sim WN(0, \sigma_1^2)$$

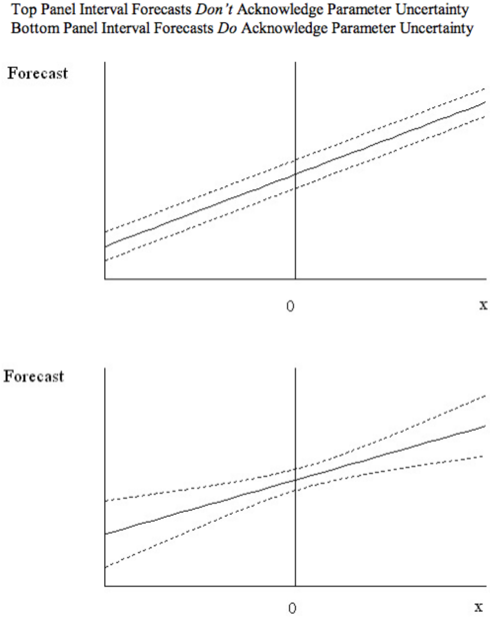
- Estimation by OLS

$$\varepsilon_{2,t} \sim WN(0, \sigma_2^2)$$

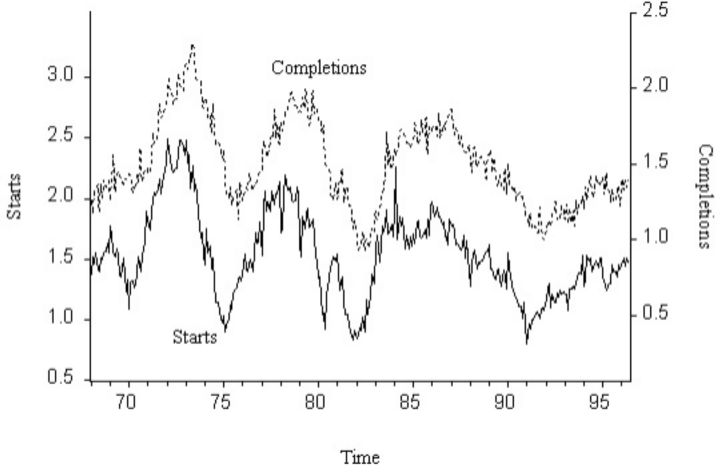
$$\text{cov}(\varepsilon_{1,t}, \varepsilon_{2,t}) = \sigma_{12}$$

- Order selection by information criteria
- Impulse-response functions, variance decompositions, predictive causality
- Forecasts via Wold's chain rule

Point and Interval Forecast



U.S. Housing Starts and Completions, 1968.01-1996.06



Notes to figure: The left scale is starts, and the right scale is completions.

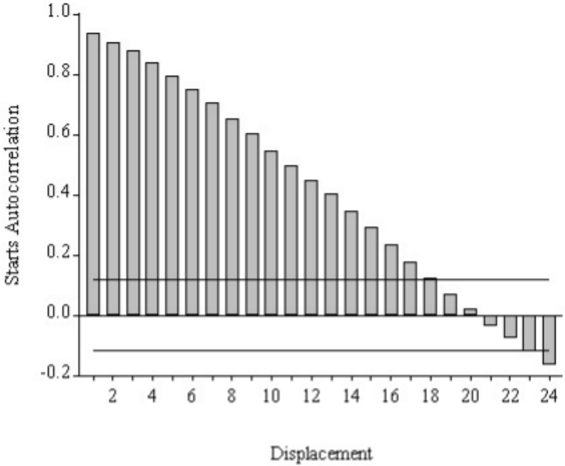
Starts Correlogram

Sample: 1968:01 1991:12

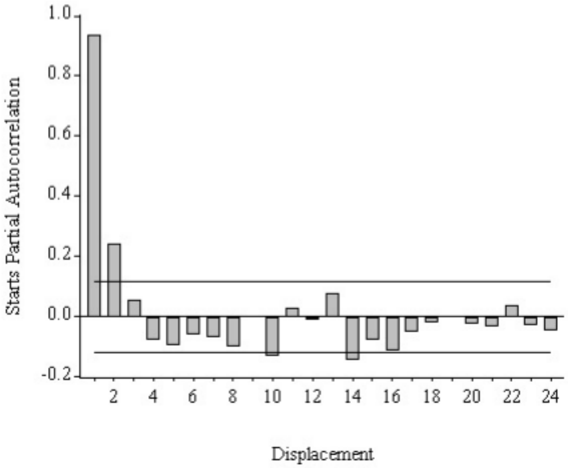
Included observations: 288

Acorr.	P. Acorr.	Std. Error	Ljung-Box	p-v
1 0.937	0.937	0.059	255.24	0.000
2 0.907	0.244	0.059	495.53	0.000
3 0.877	0.054	0.059	720.95	0.000
4 0.838	−0.077	0.059	927.39	0.000
5 0.795	−0.096	0.059	1113.7	0.000
6 0.751	−0.058	0.059	1280.9	0.000
7 0.704	−0.067	0.059	1428.2	0.000
8 0.650	−0.098	0.059	1554.4	0.000
9 0.604	0.004	0.059	1663.8	0.000
10 0.544	−0.129	0.059	1752.6	0.000
11 0.496	0.029	0.059	1826.7	0.000
12 0.446	−0.008	0.059	1886.8	0.00
13 0.405	0.076	0.059	1936.8	0.00
14 0.346	−0.144	0.059	1973.3	0.000

Starts Sample Autocorrelations



Starts Sample Partial Autocorrelations



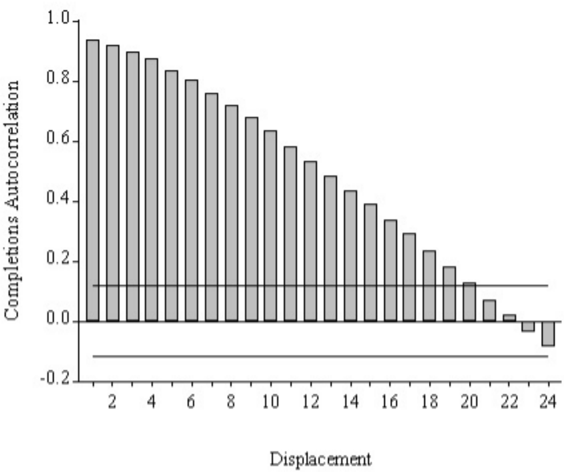
Completions Correlogram

Completions Correlogram

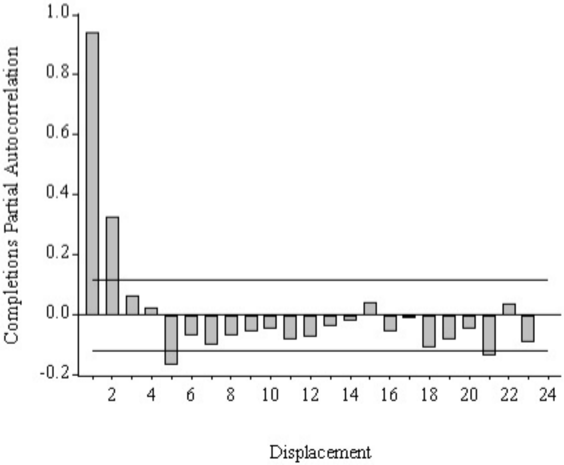
Sample: 1968:01 1991:12
Included observations: 288

Acorr.	P. Acorr.	Std. Error	Ljung-Box	p-v
1 0.939	0.939	0.059	256.61 0.000	
2 0.920	0.328	0.059	504.05 0.000	
3 0.896	0.066	0.059	739.19 0.000	
4 0.874	0.023	0.059	963.73 0.000	
5 0.834	−0.165	0.059	1168.9 0.000	
6 0.802	−0.067	0.059	1359.2 0.000	
7 0.761	−0.100	0.059	1531.2 0.000	
8 0.721	−0.070	0.059	1686.1 0.000	
9 0.677	−0.055	0.059	1823.2 0.000	
10 0.633	−0.047	0.059	1943.7 0.000	
11 0.583	−0.080	0.059	2046.3 0.000	
12 0.533	−0.073	0.059	2132.2 0.000	
13 0.483	−0.038	0.059	2203.2 0.000	

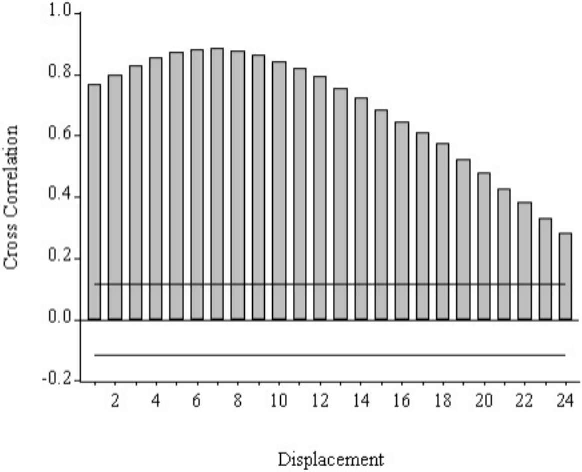
Completions Sample Autocorrelations



Completions Partial Autocorrelations

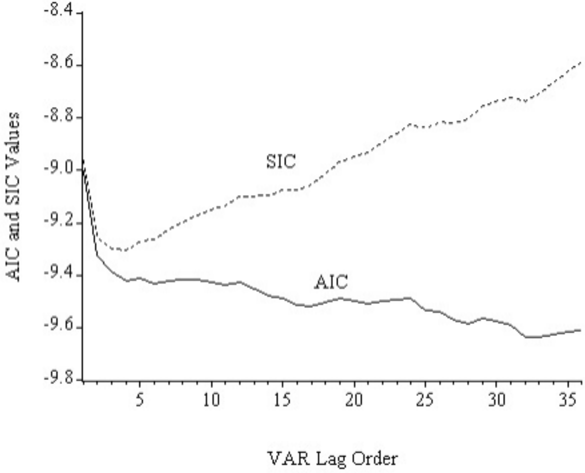


Starts and Completions: Sample Cross Correlations



Notes to figure: We graph the sample correlation between completions at time t and starts at time t-i, i = 1, 2, ..., 24.

VAR Order Selection with AIC and SIC



VAR Starts Equation

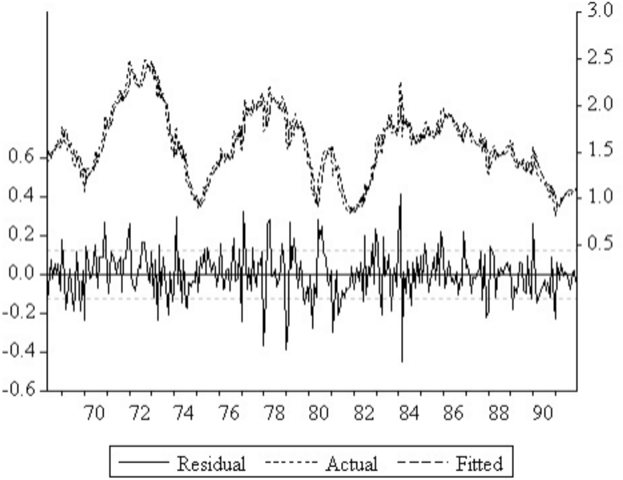
VAR Starts Equation

LS // Dependent Variable is STARTS
Sample(adjusted): 1968:05 1991:12
Included observations: 284 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.146871	0.044235	3.320264	0.0010
STARTS(-1)	0.659939	0.061242	10.77587	0.0000
STARTS(-2)	0.229632	0.072724	3.157587	0.0018
STARTS(-3)	0.142859	0.072655	1.966281	0.0503
STARTS(-4)	0.007806	0.066032	0.118217	0.9060
COMPS(-1)	0.031611	0.102712	0.307759	0.7585
COMPS(-2)	-0.120781	0.103847	-1.163069	0.2458
COMPS(-3)	-0.020601	0.100946	-0.204078	0.8384
COMPS(-4)	-0.027404	0.094569	-0.289779	0.7722

R-squared	0.895566	Mean dependent var	1.574771
Adjusted R-squared	0.892528	S.D. dependent var	0.382362
S.E. of regression	0.125350	Akaike info criterion	-4.122118
Sum squared resid	4.320952	Schwarz criterion	-4.006482
Log likelihood	191.3622	F-statistic	294.7796
Durbin-Watson stat	1.991908	Prob(F-statistic)	0.000000

VAR Start Equation Residual Plot



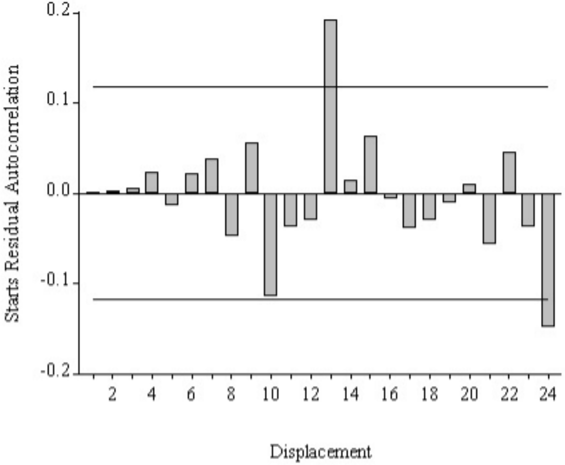
VAR Starts Equation Residual Correlogram

Sample: 1968:01 1991:12

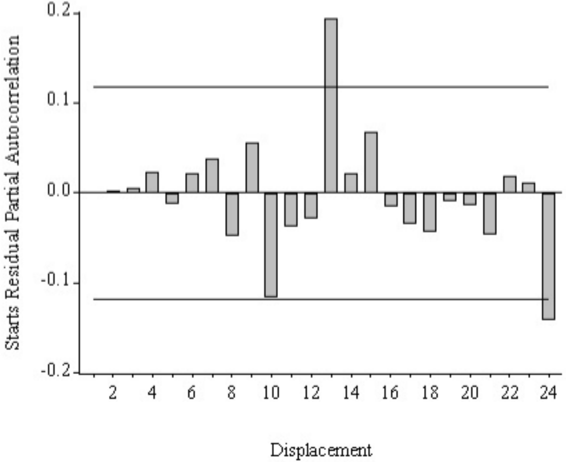
Included observations: 284

Acorr.		P. Acorr.	Std. Error	Ljung-Box		p-v
1	0.001	0.001	0.059	0.0004	0.985	
2	0.003	0.003	0.059	0.0029	0.999	
3	0.006	0.006	0.059	0.0119	1.000	
4	0.023	0.023	0.059	0.1650	0.997	
5	−0.013	−0.013	0.059	0.2108	0.999	
6	0.022	0.021	0.059	0.3463	0.999	
7	0.038	0.038	0.059	0.7646	0.998	
8	−0.048	−0.048	0.059	1.4362	0.994	
9	0.056	0.056	0.059	2.3528	0.985	
10	−0.114	−0.116	0.059	6.1868	0.799	
11	−0.038	−0.038	0.059	6.6096	0.830	
12	−0.030	−0.028	0.059	6.8763	0.86[illegible]	
13	0.192	0.193	0.059	17.947	0.16[illegible]	
14	0.014	0.021	0.059	18.010	0.206	

VAR Starts Equation Residual Sample Autocorrelations



Var Starts Equation Residual Sample Partial Autocorrelations



Evaluating and Combining Forecasts

Evaluating a single forecast Process:

$$y_t = \mu + \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \ldots$$

$$\varepsilon_t \sim WN(0, \ \sigma^2),$$

h-step-ahead linear least-squares forecast:

$$y_{t+h,t} = \mu + b_h\varepsilon_t + b_{h+1}\varepsilon_{t-1} + \ldots$$

Corresponding h-step-ahead forecast error:

$$e_{t+h,t} = y_{t+h} - y_{t+h,t} = \varepsilon_{t+h} + b_1\varepsilon_{t+h-1} + \ldots + b_{h-1}\varepsilon_{t+1}$$

with variance

$$\sigma_h^2 = \sigma^2(1 + \sum_{i=1}^{h-1} b_i^2)$$

Evaluating and Combining Forecasts

So, four key properties of optimal forecasts:
a. Optimal forecasts are unbiased
b. Optimal forecasts have 1-step-ahead errors that are white noise
c. Optimal forecasts have h-step-ahead errors that are at most MA(h-1)
d. Optimal forecasts have h-step-ahead errors with variances that are non-decreasing in h and that converge to the unconditional variance of the process

1. All are easily checked. How?

Assessing optimality with respect to an information set

Unforecastability principle: The errors from good forecasts are not be forecastable!

Regression:

$$e_{t+h,t} = \alpha + \sum_{i=1}^{k-1} \alpha_i x_{it} + u_t$$

1. Test whether $\alpha_0, ..., \alpha_{k-1}$ *are* 0

Important case:

$$e_{t+h,t} = \alpha + \alpha_1 y_{t+h,t} + u_t$$

1. Test whether $(\alpha_0, \alpha_1) = (0, 0)$

Equivalently,

$$y_{t+h,t} = \beta + \beta_1 y_{t+h,t} + u_t$$

1. Test whether $(\beta_0, \beta_1) = (0, 1)$

Evaluating multiple forecasts: comparing forecast accuracy

Forecast errors, $e_{t+h,t} = y_{t+h} - y_{t+h,t}$ Forecast percent errors, $p_{t+h,t} = \frac{y_{t+h} - y_{t+h,t}}{y_{t+h}}$

$$\mathrm{ME} = \frac{1}{T}\sum_{t=1}^{T} e_{t+h,t}$$

$$\mathrm{EV} = \frac{1}{T}\sum_{1}^{T} (e_{t+h,t} - \mathrm{ME})^2$$

$$\mathrm{MSE} = \frac{1}{T}\sum_{t=1}^{T} e_{t+h,t}^2$$

$$\mathrm{MSPE} = \frac{1}{T}\sum_{t=1}^{T} p_{t+h,t}^2$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_{t+h,t}^2}$$

Forecast encompassing

$$y_{t+h} \;=\; \beta_a y^a_{t+h,t} + \beta_b y^b_{t+h,t} + \varepsilon_{t+h,t}$$

1. If $(\beta_a, \beta_b) = (1, 0)$, model a forecast-encompasses model b
2. If $(\beta_a, \beta_b) = (0, 1)$, model b forecast-encompasses model a
3. Otherwise, neither model encompasses the other

Alternative approach:

$$(y_{t+h} - y_t \;=\; \beta(y^a_{t+h,t} - y_t) + \beta_b(y^b_{t+h,t} - y_t) + \varepsilon_{t+h,t}$$

1. Useful in I(1) situations

Variance-covaiance forecast combination

Composite formed from two unbiased forecasts:

$$\mathrm{y}^{\mathrm{c}}_{\mathrm{t+h}} = \omega \mathrm{y}^{\mathrm{a}}_{\mathrm{t+h,t}} + (1-\omega)\mathrm{y}^{\mathrm{b}}_{\mathrm{t+h,t}}$$

$$\mathrm{e}^{\mathrm{c}}_{\mathrm{t+h}} = \omega \mathrm{e}^{\mathrm{a}}_{\mathrm{t+h,t}} + (1-\omega)\mathrm{e}^{\mathrm{b}}_{\mathrm{t+h,t}}$$

$$\sigma^2_{\mathrm{c}} = \omega^2\sigma^2_{\mathrm{aa}} + (1-\omega)^2\sigma^2_{\mathrm{bb}} + 2\omega(1-\omega)\sigma^2_{\mathrm{ab}}$$

$$\omega^* = \frac{\sigma^2_{\mathrm{bb}} - \sigma^2_{\mathrm{ab}}}{\sigma^2_{\mathrm{bb}} + \sigma^2_{\mathrm{aa}} - 2\mathrm{sigma}^2_{\mathrm{ab}}}$$

$$\hat{\omega}^* = \frac{\hat{\sigma}^2_{\mathrm{bb}} - \hat{\sigma}^2_{\mathrm{ab}}}{\hat{\sigma}^2_{\mathrm{bb}} + \hat{\sigma}^2_{\mathrm{aa}} - 2\hat{\sigma}^2_{\mathrm{ab}}}$$

Regression-based forecast combination

$$y_{t+h} = \beta_0 + \beta_1 y^a_{t+h,t} + \beta_2 y^b_{t+h,t} + \varepsilon_{t+h,t}$$

1. Equivalent to variance-covariance combination if weights sum to unity and intercept is excluded
2. Easy extension to include more than two forecasts
3. Time-varying combining weights
4. Dynamic combining regressions
5. Shrinkage of combining weights toward equality
6. Nonlinear combining regressions

Unit Roots, Stochastic Trends, ARIMA Forecasting Models, and Smoothing

1. Stochastic Trends and Forecasting

$$\Phi(L)y_t = \Theta(L)\varepsilon_t$$

$$\Phi(L) = \Phi'(L)(1 - L)$$

$$\Phi'(L)(1 - L)y_t = \Theta(L)\varepsilon_t$$

$$\Phi'(L)\Delta y_t = \Theta(L)\varepsilon_t$$

I(0) vs I(1) processes

Unit Roots, Stochastic Trends, ARIMA Forecasting Models, and Smoothing Cont.

- Random Walk

$$y_t = y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim \mathrm{WN}(0,\ \sigma^2),$$

- Random walk with drift

$$y_t = \delta + y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim \mathrm{WN}(0,\ \sigma^2),$$

Stochastic trend vs deterministic trend

Properties of random walks

$$y_t = y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0, \sigma^2),$$

With time 0 value y_0:

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

$$E(y_t) = y_0$$

$$var(y_t) = t\sigma^2$$

$$\lim_{t \to \infty} var(y_t) = \infty$$

Random walk with drift

Random walk with drift

$$y_t = \delta + y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0,\ \sigma^2),$$

Assuming time 0 value y_0:

$$y_t = t\delta + y_0 + \sum_{i=1}^{t} \varepsilon_i$$

$$E(y_t) = y_0 + t\delta$$

$$var(y_t) = t\sigma^2$$

$$\lim_{t \to \infty} var(y_t) = \infty$$

ARIMA(p,1,q) model

$$\Phi(L)(1-L)y_t \;=\; c+\Theta(L)\varepsilon_t$$

or

$$(1-L)y_t \;=\; c\Phi^{-1}(L)+\Phi^{-1}(L)\Theta(L)\varepsilon_t$$

where

$$\Phi(L) = 1-\Phi_1 L - \ldots - \Phi_p L^p$$

$$\Theta(L) \;=\; 1-\Theta_1 L - \ldots - \Theta_q L^q$$

and all the roots of both lag operator polynomials are outside the unit circle.

ARIMA(p,d,q) model

$$\Phi(L)(1-L)^d y_t = c + \Theta(L)\varepsilon_t$$

or

$$(1-L)^d y_t = c\Phi^{-1}(L) + \Phi^{-1}(L)\Theta(L)\varepsilon_t$$

where

$$\Phi(L) = 1 - \Phi_1 L - ... - \Phi_p L^p$$

$$\Theta(L) = 1 - \Theta_1 L - ... - \Theta_q L^q$$

and all the roots of both lag operator polynomials are outside the unit circle.

Properties of ARIMA(p,1,q) processes

- Appropriately made stationary by differencing
- Shocks have permanent effects
  - Forecasts don't revert to a mean
- Variance grows without bound as time progresses
  - Interval forecasts widen without bound as horizon grows

Random walk example

<u>Point forecast</u>

Recall that for the AR(1) process,

$$\mathrm{y_t} = \phi \mathrm{y_{t-1}} + \varepsilon_\mathrm{t}$$

$$\mathrm{y_t} \sim \mathrm{WN}(0,\ \sigma^2)$$

the optimal forecast is

$$\mathrm{y_{T+h,T}} = \phi^\mathrm{h} \mathrm{y_T}$$

Thus in the random walk case,

$$\mathrm{y_{T+h,T}} = \mathrm{y_T},\ \text{for all h}$$

Random walk example Cont.

Interval and density forecasts

Recall error associated with optimal AR(1) forecast:

$$e_{T+h,T} = (y_{T+h} - y_{T+h,T}) = \varepsilon_{T+h} + \phi\varepsilon_{T+h-1} + ... + \phi^{h-1}\varepsilon_{T+1}$$

with variance

$$\sigma_h^2 = \sigma^2 \sum_{i=0}^{h-1} \phi^{2i}$$

Thus in the random walk case,

$$e_{T+h,T} = \sum_{i=0}^{h-1} \varepsilon_{T+h-i}$$

$$\sigma_h^2 = h\sigma^2$$

$$h - \text{step} - \text{ahead 95\% interval}: \ y_T \pm 1.96\sigma\sqrt{h}$$

$$h - \text{step} - \text{ahead density forecast}: \ N(y_T, h\sigma^2)$$

Effects of Unit Roots

- Sample autocorrelation function "fails to damp"
- Sample partial autocorrelation function near 1 for $\tau = 1$, and then damps quickly
- Properties of estimators change

e.g., least-squares autoregression with unit roots
True process:

$$y_t = y_{t-1} + \varepsilon_t$$

Estimated model:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

Superconsistency: $T(\hat{\phi}_{LS} - 1)$ stabilizes as sample size grows
Bias: $E(\hat{\phi}_{LS}) < 1$

- Offsetting effects of bias and superconsistency

Unit Root Tests

$$y_t = \phi y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \overset{iid}{\sim} N(0,\ \sigma^2)$$

$$\hat{\tau} = \frac{\hat{\phi} - 1}{s\sqrt{\frac{1}{\sum_{t=2}^{T} y_{t-1}^2}}}$$

"Dickey-Fuller $\hat{\tau}$ distribution"
Trick regression:

$$y_t - y_t = (\phi - 1)y_{t-1} + \varepsilon_t$$

Allowing for nonzero mean under the alternative

Basic model:

$$(y_t - \mu) = \phi(y_{t-1} - \mu) + \varepsilon_t$$

which we rewrite as

$$y_t = \alpha + \phi y_{t-1} + \varepsilon_t$$

where

$$\alpha = \mu(1 - \phi)$$

- α vanishes when $\phi = 1$ (null)
- α is nevertheless present under the alternative,

  so we include an intercept in the regression

Dickey-Fuller $\hat{\tau}_\mu$ distribution

Allowing for deterministic linear trend under the alternative

Basic model:

$$(y_t - a - bTIME_t) = \phi(y_{t-1} - a - b\ TIME_{t-1}) + \varepsilon_t$$

or

$$y_t = \alpha + \beta TIME_t + \phi y_{t-1} + \varepsilon_t$$

where $\alpha = a(1-\phi) + b\phi$ and $\beta = b(1-\phi)$.

- Under the null hypothesis we have a random walk with drift,

$$y_t = b + y_{t-1} + \varepsilon_t$$

- Under the deterministic-trend alternative hypothesis both the intercept and the trend enter and so are included in the regression.

Allowing for higher-order autoregressive dynamics

AR(p) process:

$$y_t + \sum_{j=1}^{p} \phi_j y_{t-j} \; = \; \varepsilon_t$$

Rewrite:

$$y_t \; = \; \rho_1 y_{t-1} + \sum_{j=2}^{p} \rho_j (y_{t-j+1} - y_{t-j}) + \varepsilon_t$$

where $p \geq 2$, $\rho_1 = -\sum_{j=1}^{p} \phi_j$, and $\rho_i = \sum_{j=1}^{p} \phi_j$, $i = 2, ..., p$.

Unit root: $\rho_1 \; = \; 1$ ($AR(p-1)$ *in first differences*)

$\hat{\tau}$ distribution holds asymptotically.

Allowing for a nonzero mean in the AR(p) case

$$(y_t - \mu) + \sum_{j=1}^{p} \phi_j (y_{t-j} - \mu) \; = \; \varepsilon_t$$

or

$$y_t \; = \; \alpha + \rho y_{t-1} + \sum_{j=2}^{p} \rho_j (y_{t-j+1} - y_{t-j}) + \varepsilon_t,$$

where $\alpha \; = \; \mu(1 + \sum_{j=1}^{p} \phi_j)$, and the other parameters are as above. In the unit root case, the intercept vanishes, because $\sum_{j=1}^{p} \phi_j = -1$. $\hat{\tau}_\mu$ distribution holds asymptotically.

Allowing for trend under the alternative

$$(\mathrm{y_t - a - bTIME_t}) + \sum_{\mathrm{j=1}}^{\mathrm{p}} \phi_{\mathrm{j}}(\mathrm{y_{t-j} - a - bTIME_{t-j}}) \;=\; \varepsilon_{\mathrm{t}}$$

or

$$\mathrm{y_t} \;=\; \mathrm{k_1 + k_2 TIME_t} + \rho_1 \mathrm{y_{t-1}} + \sum_{\mathrm{j=2}}^{\mathrm{p}} \rho_{\mathrm{j}}(\mathrm{y_{t-j+1} - y_{t-j}}) + \varepsilon_{\mathrm{t}}$$

where

$$\mathrm{k_1} = \mathrm{a}(1 + \sum_{\mathrm{i=1}}^{\mathrm{p}} \phi_{\mathrm{i}}) - \mathrm{b} \sum_{\mathrm{i=1}}^{\mathrm{p}} \mathrm{i}\phi_{\mathrm{i}}$$

and

$$\mathrm{k_2} = \mathrm{bTIME_t}(1 + \sum_{\mathrm{i=1}}^{\mathrm{p}} \phi_{\mathrm{i}})$$

In the unit root case, $k_1 = -b\sum_{i=1}^{p} i\phi_i$ *and* $k_2 = 0$.

$\hat{\tau}_\tau$ distribution holds asymptotically.

General ARMA representations: augmented Dickey-Fuller tests

$$y_t = \rho_1 y_{t-1} + \sum_{j=2}^{k-1} \rho_j (y_{t-j+1} - y_{t-j}) + \varepsilon_t$$

$$y_t = \alpha + \rho_1 y_{t-1} + \sum_{j=2}^{k-1} \rho_j (y_{t-j+1} - y_{t-j}) + \varepsilon_t$$

$$y_t = k_1 + k_2 \mathrm{TIME}_t + \rho_1 y_{t-1} + \sum_{j=2}^{k-1} \rho_j (y_{t-j+1} - y_{t-j}) + \varepsilon_t$$

- k-1 augmentation lags have been included
- $\hat{\tau}, \hat{\tau}_\mu,$ *and* $\hat{\tau}_\tau$ hold asymptotically under the null

Simple moving average smoothing

1. Original data: $\{y_t\}_{t=1}^{T}$
2. Smoothed data: $\{\bar{y}_t\}$
3. Two-sided moving average is $\bar{y}_t = (2m+1)^{-1}\sum_{i=-m}^{m} y_{t-i}$
4. One-sided moving average is $\bar{y}_t = (m+1)^{-1}\sum_{i=0}^{m} y_{t-i}$
5. One-sided weighted moving average is $\bar{y}_t = \sum_{i=0}^{m} w_i y_{t-i}$

- Must choose smoothing parameter, m

Exponential Smoothing

Local level model:

$$y_t = c_{0t} + \varepsilon_t$$

$$c_{0t} = c_{0,t-1} + \eta_t$$

$$\eta_t \sim WN(0,\ \sigma_\eta^2)$$

- Exponential smoothing can construct the optimal estimate of c_0 - and hence the optimal forecast of any future value of y - on the basis of current and past y
- What if the model is misspecified?

Exponential smoothing algorithm

- Observed series, $\{y_t\}_{t=1}^{T}$
- Smoothed series, $\{\bar{y}_t\}_{t=1}^{T}$ (estimate of the local level)
- Forecasts, $\hat{y}_{T+h,T}$
  1. Initialize at t = 1: $\bar{y}_1 \;=\; y_1$
  2. Update: $\bar{y}_t \;=\; \alpha y_t + (1-\alpha)\bar{y}_{t-1}, t = 2, \ldots T$
  3. Forecast: $\hat{y}_T + h, T = \bar{y}_T$
- Smoothing parameter $\alpha \in [0, 1]$

Demonstration that the weights are exponential

Start:

$$\bar{y}_t = \alpha y_t + (1-\alpha)\bar{y}_{t-1}$$

Substitute backward for $\bar{y}$:

$$\bar{y}_t = \sum_{j=0}^{t-1} w_j y_{t-j}$$

where

$$w_j = \alpha(1-\alpha)^j$$

- Exponential weighting, as claimed
- Convenient recursive structure

Holt-Winters Smoothing

$$
\begin{aligned}
\mathrm{y_t} &= \mathrm{c_{0t}} + \mathrm{c_{1t}}\mathrm{TIME_t} + \varepsilon_\mathrm{t} \\
\mathrm{c_{0t}} &= \mathrm{c_{0,t-1}} + \eta_\mathrm{t} \\
\\
\mathrm{c_{1t}} &= \mathrm{c_{1,t-1}} + \nu_\mathrm{t}
\end{aligned}
$$

- Local level and slope model
- Holt-Winters smoothing can construct optimal estimates of c_0 and c_1 - hence the optimal forecast of any future value of y by extrapolating the trend - on the basis of current and past y

Holt-Winters smoothing algorithm

1. Initialize at $t = 2$:

$$\bar{y}_2 = y_2$$

$$F_2 = y_2 - y_1$$

2. Update:

$$\bar{y}_t = \alpha y_t + (1-\alpha)(\bar{y}_{t-1} + F_{t-1}), \ 0 < \alpha < 1$$

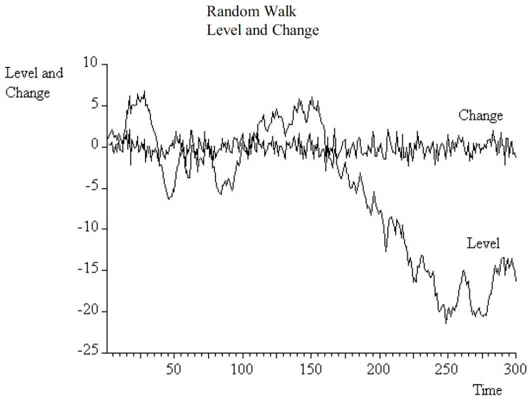
$$F_t = \beta(\bar{y}_t - \bar{y}_{t-1}) + (1-\beta)F_{t-1}, \ 0 < \beta < 1$$

$$t = 3, 4, ..., T.$$

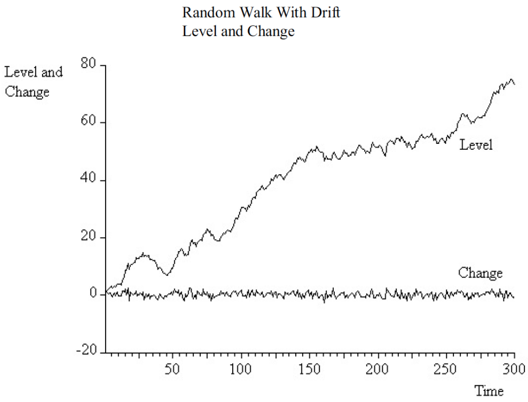
3. Forecast: $\hat{y}_{T+h,T} = \bar{y}_T + hF_T$

- *$\bar{y}_t$ is the estimated level at time t*
- *F_t is the estimated slope at time t*

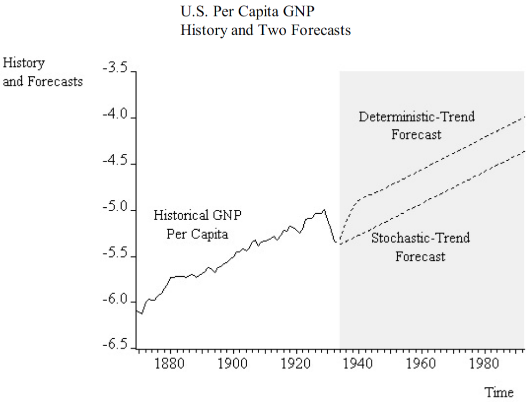
Random Walk – Level and Change



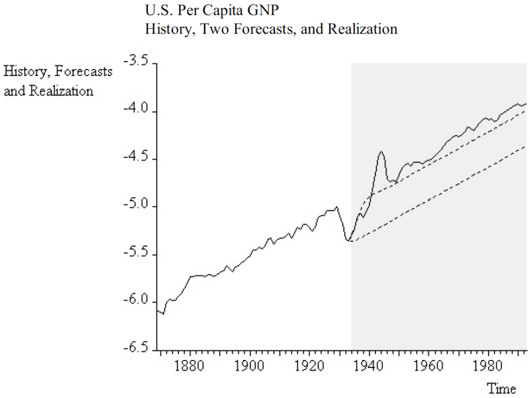
Random Walk With Drift – Level and Change



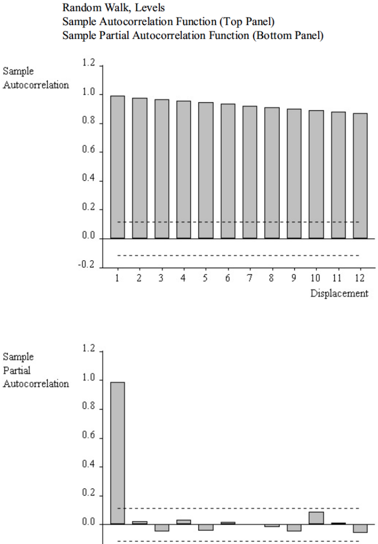
U.S. Per Capita GNP – History and Two Forecasts



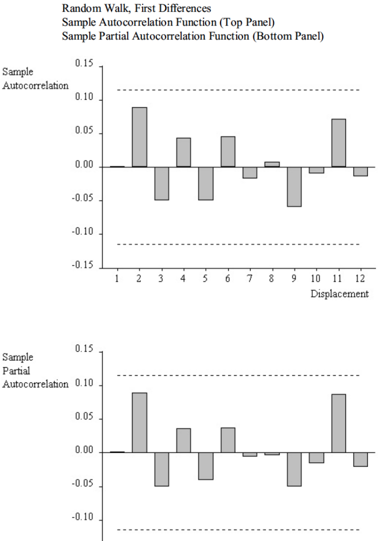
U.S. Per Capita GNP – History, Two Forecasts, and Realization



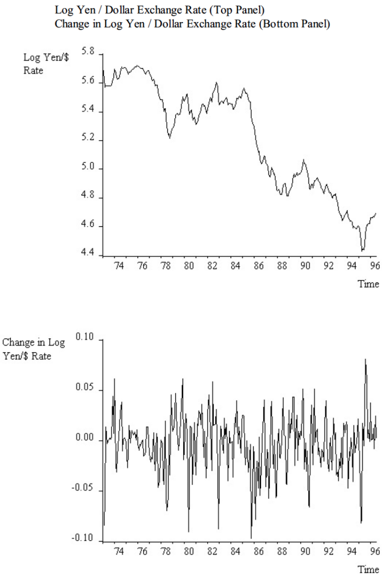
Random Walk, Levels – Sample Autocorrelation Function (Top Panel) and Sample Partial Autocorrelation Function (Bottom Panel)



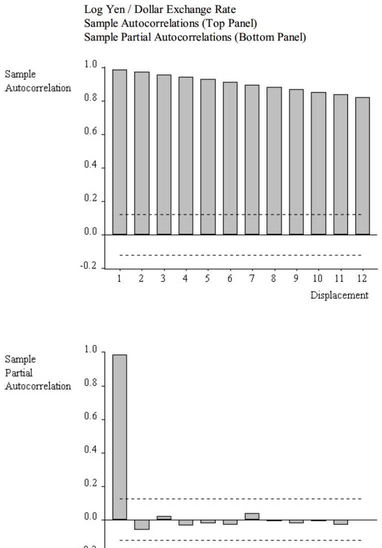
Random Walk, First Differences – Sample Autocorrelation Function (Top Panel) and Sample Partial Autocorrelation Function (Bottom Panel)



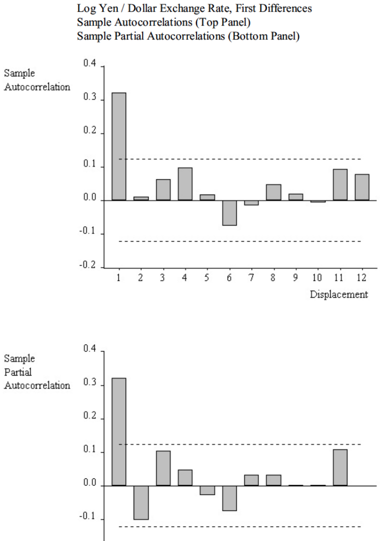
Log Yen / Dollar Exchange Rate (Top Panel) and Change in Log Yen / Dollar Exchange Rate (Bottom Panel)



Log Yen / Dollar Exchange Rate – Sample Autocorrelations (Top Panel) and Sample Partial Autocorrelations (Bottom Panel)



Log Yen / Dollar Exchange Rate, First Differences – Sample Autocorrelations (Top Panel) and Sample Partial Autocorrelations (Bottom Panel)



Log Yen / Dollar Rate, Levels – AIC and SIC Values of Various ARMA Models

Log Yen / Dollar Rate, Levels
AIC Values
Various ARMA Models

			MA Order		
		0	1	2	3
	0		-5.171	-5.953	-6.428
AR Order	1	-7.171	-7.300	-7.293	-7.287
	2	-7.319	-7.314	-7.320	-7.317
	3	-7.322	-7.323	-7.316	-7.308

Log Yen / Dollar Rate, Levels
SIC Values
Various ARMA Models

			MA Order		
		0	1	2	3
	0		-5.130	-5.899	-6.360
AR Order	1	-7.131	-7.211	-7.225	-7.205
	2	-7.265	-7.246	-7.238	-7.221
	3	-7.253	-7.241	-7.220	-7.199

Log Yen / Dollar Exchange Rate – Best-Fitting Deterministic-Trend Model

Log Yen / Dollar Exchange Rate
Best-Fitting Deterministic-Trend Model

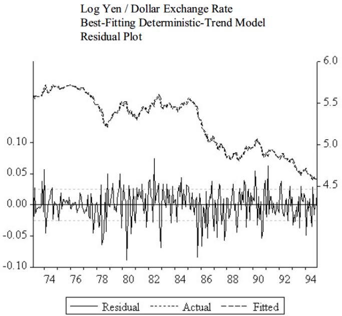
LS // Dependent Variable is LYEN
Sample(adjusted): 1973:03 1994:12
Included observations: 262 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.904705	0.136665	43.20570	0.0000
TIME	-0.004732	0.000781	-6.057722	0.0000
AR(1)	1.305829	0.057587	22.67561	0.0000
AR(2)	-0.334210	0.057656	-5.796676	0.0000

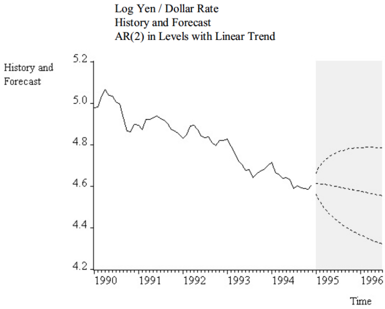
R-squared	0.994468	Mean dependent var	5.253984
Adjusted R-squared	0.994404	S.D. dependent var	0.341563
S.E. of regression	0.025551	Akaike info criterion	-7.319015
Sum squared resid	0.168435	Schwarz criterion	-7.264536
Log likelihood	591.0291	F-statistic	15461.07
Durbin-Watson stat	1.964687	Prob(F-statistic)	0.000000

Inverted AR Roots	.96	.35

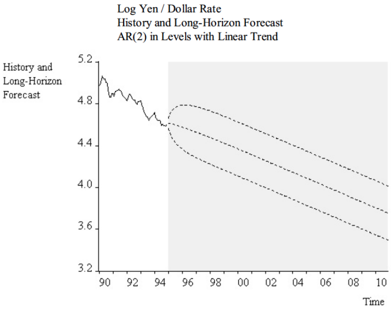
Log Yen / Dollar Exchange Rate – Best-Fitting Deterministic-Trend Model : Residual Plot



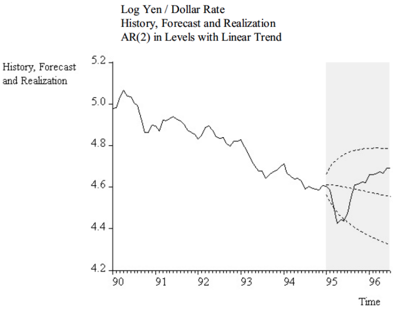
Log Yen / Dollar Rate – History and Forecast : AR(2) in Levels with Linear Trend



Log Yen / Dollar Rate – History and Long-Horizon Forecast : AR(2) in Levels with Linear Trend



Log Yen / Dollar Rate – History, Forecast and Realization : AR(2) in Levels with Linear Trend



Log Yen / Dollar Exchange Rate – Augmented Dickey-Fuller Unit Root Test

Log Yen / Dollar Exchange Rate
Augmented Dickey-Fuller Unit Root Test

Augmented Dickey-Fuller Test Statistic	-2.498863	1% Critical Value	-3.9966
		5% Critical Value	-3.4284
		10% Critical Value	-3.1373

Augmented Dickey-Fuller Test Equation
LS // Dependent Variable is D(LYEN)
Sample(adjusted): 1973:05 1994:12
Included observations: 260 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LYEN(-1)	-0.029423	0.011775	-2.498863	0.0131
D(LYEN(-1))	0.362319	0.061785	5.864226	0.0000
D(LYEN(-2))	-0.114269	0.064897	-1.760781	0.0795
D(LYEN(-3))	0.118386	0.061020	1.940116	0.0535
C	0.170875	0.068474	2.495486	0.0132
@TREND(1973:01)	-0.000139	5.27E-05	-2.639758	0.0088

R-squared	0.142362	Mean dependent var	-0.003749
Adjusted R-squared	0.125479	S.D. dependent var	0.027103
S.E. of regression	0.025345	Akaike info criterion	-7.327517
Sum squared resid	0.163166	Schwarz criterion	-7.245348
Log likelihood	589.6532	F-statistic	8.432417
Durbin-Watson stat	2.010829	Prob(F-statistic)	0.000000

Log Yen / Dollar Rate, Changes – AIC and SIC Values of Various ARMA Models

Log Yen / Dollar Rate, Changes
AIC Values
Various ARMA Models

			MA Order		
		0	1	2	3
	0		-7.298	-7.290	-7.283
AR Order	1	-7.308	-7.307	-7.307	-7.302
	2	-7.312	-7.314	-7.307	-7.299
	3	-7.316	-7.309	-7.340	-7.336

Log Yen / Dollar Rate, Changes
SIC Values
Various ARMA Models

			MA Order		
		0	1	2	3
	0		-7.270	-7.249	-7.228
AR Order	1	-7.281	-7.266	-7.252	-7.234
	2	-7.271	-7.259	-7.238	-7.217
	3	-7.261	-7.241	-7.258	-7.240

Log Yen / Dollar Exchange Rate – Best-Fitting Stochastic-Trend Model

Log Yen / Dollar Exchange Rate
Best-Fitting Stochastic-Trend Model

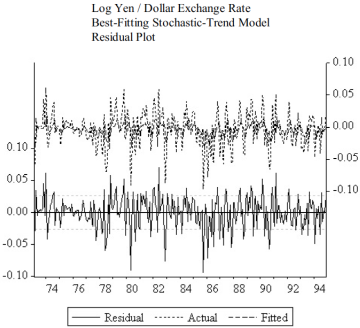
LS // Dependent Variable is DLYEN
Sample(adjusted): 1973:03 1994:12
Included observations: 262 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.003697	0.002350	-1.573440	0.1168
AR(1)	0.321870	0.057767	5.571863	0.0000

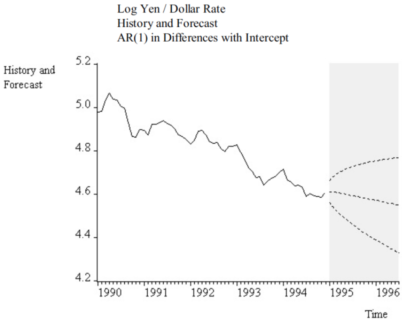
R-squared	0.106669	Mean dependent var	-0.003888
Adjusted R-squared	0.103233	S.D. dependent var	0.027227
S.E. of regression	0.025784	Akaike info criterion	-7.308418
Sum squared resid	0.172848	Schwarz criterion	-7.281179
Log likelihood	587.6409	F-statistic	31.04566
Durbin-Watson stat	1.948933	Prob(F-statistic)	0.000000

Inverted AR Roots .32

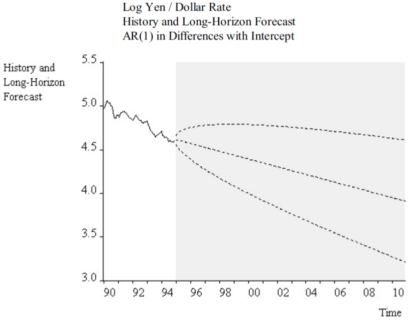
Log Yen / Dollar Exchange Rate – Best-Fitting Stochastic-Trend Model : Residual Plot



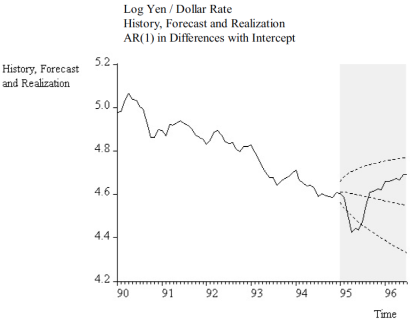
Log Yen / Dollar Rate – History and Forecast : AR(1) in Differences with Intercept



Log Yen / Dollar Rate – History and Long-Horizon Forecast : AR(1) in Differences with Intercept



Log Yen / Dollar Rate – History, Forecast and Realization : AR(1) in Differences with Intercept



Log Yen / Dollar Exchange Rate – Holt-Winters Smoothing

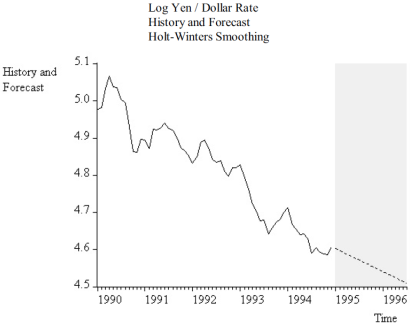
Log Yen / Dollar Exchange Rate
Holt-Winters Smoothing

Sample: 1973:01 1994:12
Included observations: 264
Method: Holt-Winters, No Seasonal
Original Series: LYEN
Forecast Series: LYENSM

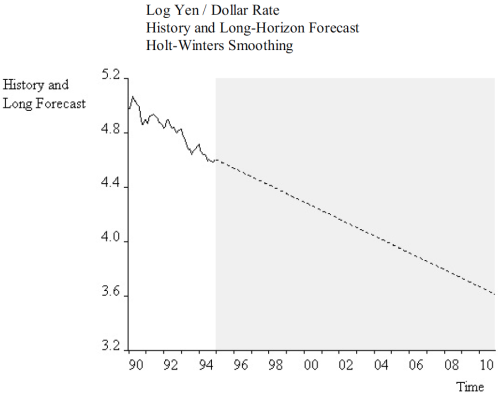
Parameters:	Alpha	1.000000
	Beta	0.090000
Sum of Squared Residuals		0.202421
Root Mean Squared Error		0.027690

End of Period Levels:	Mean	4.606969
	Trend	-0.005193

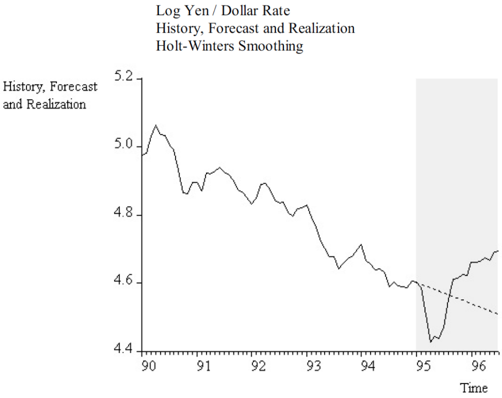
Log Yen / Dollar Rate – History and Forecast : Holt-Winters Smoothing



Log Yen / Dollar Rate – History and Long-Horizon Forecast : Holt-Winters Smoothing



Log Yen / Dollar Rate – History, Forecast and Realization : Holt-Winters Smoothing



Volatility Measurement, Modeling and Forecasting

The main idea:

$$\varepsilon_t \mid \Omega_{t-1} \;\sim\; (0,\; \sigma_t^2)$$

$$\Omega_{t-1} \;=\; \{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$$

We'll look at:

- Basic Structure and properties
- Time variation in volatility and prediction-error variance
- ARMA representation in squares
- GARCH(1,1) and exponential smoothing
- Unconditional symmetry and leptokurtosis
- Convergence to normality under temporal aggregation
- Estimation and testing

Basic Structure and Properties

Standard models (e.g., ARMA):

- Unconditional mean: constant
- Unconditional variance: constant
- Conditional mean: varies
- Conditional variance: constant (unfortunately)
- k-step-ahead forecast error variance: depends only on k, not on Ω_t (again unfortunately)

The Basic ARCH Process

$$y_t = B(L)\varepsilon_t$$

$$B(L) = \sum_{i=0}^{\infty} b_i L^i \qquad \sum_{i=0}^{\infty} b_i^2 < \infty \qquad b_0 = 1$$

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \gamma(L)\varepsilon_t^2$$

$$\omega > 0 \qquad \gamma(L) = \sum_{i=1}^{p} \gamma_i L^i \qquad \gamma_i \geq 0 \text{ for all } i \qquad \sum \gamma_i < 1.$$

The Basic ARCH Process cont.

ARCH(1) process:

$$r_t \mid \Omega_{t-1} \sim (0,\ \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2$$

- Unconditional mean: $E(r_t) = 0$
- Unconditional variance: $E(r_t - E(r_t))^2 = \frac{\omega}{1-\alpha}$
- Conditional mean: $E(r_t \mid \Omega_{t-1}) = 0$
- Conditional variance:
  $E([r_t - E(r_t \mid \Omega_{t-1})]^2 \mid \Omega_{t-1}) = \omega + \alpha r_{t-1}^2$

The GARCH Process

$$y_t = \varepsilon_t$$

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2$$

$$\alpha(L) = \sum_{i=1}^{p} \alpha_i L^i, \qquad \beta(L) = \sum_{i=1}^{q} \beta_i L^i$$

$$\omega > 0, \qquad \alpha_i \geq 0, \qquad \beta_i \geq 0, \qquad \sum \alpha_i + \sum \beta_i < 1.$$

Time Variation in Volatility and Prediction Error Variance

Prediction error variance depends on Ω_{t-1}

- e.g. 1-step-ahead prediction error variance is now

$$\sigma_{\mathrm{t}}^2 \;=\; \omega + \alpha \mathrm{r}_{\mathrm{t}-1}^2 + \beta \sigma_{\mathrm{t}-1}^2$$

Conditional variance is a serially correlated RV

- Again, follows immediately from

$$\sigma_{\mathrm{t}}^2 \;=\; \omega + \alpha \mathrm{r}_{\mathrm{t}-1}^2 + \beta \sigma_{\mathrm{t}-1}^2$$

ARMA Representation in Squares

r_t^2 has the ARMA(1,1) representation:

$$r_t^2 = \omega + (\alpha + \beta) r_{t-1}^2 - \beta \nu_{t-1} + \nu_t$$

where $\nu_t = r_t^2 - \sigma_t^2$

Important result:
The above equation is simply

$$\begin{aligned} r_t^2 &= (\omega + (\alpha + \beta) r_{t-1}^2 - \beta \nu_{t-1}) + \nu_t \\ &= \sigma_t^2 + \nu_t \end{aligned}$$

Thus r_t^2 is a *noisy* indicator of σ_t^2

GARCH(1,1) and Exponential Smoothing

Exponential smoothing recursion:

$$\bar{r}_t^2 \;=\; \gamma r_t^2 + (1-\gamma)\bar{r}_{t-1}^2$$

Back substitution yields:

$$\bar{r}_t^2 \;=\; \sum w_j r_{t-j}^2$$

where

$$w_j \;=\; \gamma(1-\gamma)^j$$

GARCH(1,1)

$$\sigma_t^2 \;=\; \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$$

Back substitution yields:

$$\sigma_t^2 \;=\; \frac{\omega}{1-\beta} + \alpha \sum \beta^{j-1} r_{t-j}^2$$

Unconditional Symmetry and Leptokurtosis

- Volatility clustering produces unconditional leptokurtosis
- Conditional symmetry translates into unconditional symmetry
  Unexpected agreement with the facts!

  Convergence to Normality under Temporal Aggregation

- Temporal aggregation of covariance stationary GARCH processes produces convergence to normality.
  Again, unexpected agreement with the facts!

Estimation and Testing

Estimation: easy!
Maximum Likelihood Estimation

$$L(\theta; r_1, ..., r_T) \;=\; f(r_T \mid \Omega_{T-1}; \theta)\; f(r_{T-1} \mid \Omega_{T-2}; \theta)...$$

If the conditional densities are Gaussian,

$$f(r_t \mid \Omega_{t-1}; \theta) \;=\; \frac{1}{\sqrt{2\pi}} \sigma_t^2(\theta)^{-1/2} \exp\left(-\frac{1}{2}\frac{r_t^2}{\sigma_t^2(\theta)}\right).$$

We can ignore the $f(r_p, ..., r_1; \theta)$ term, yielding the likelihood:

$$-\frac{T-p}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=p+1}^{T} \ln \sigma_t^2(\theta) - \frac{1}{2}\sum_{t=p+1}^{T} \frac{r_t^2}{\sigma_t^2(\theta)}.$$

Testing: likelihood ratio tests
Graphical diagnostics: Correlogram of squares, correlogram of squared standardized residuals

Variations on Volatility Models

We will look at:

- Asymmetric response and the leverage effect
- Exogenous variables
- GARCH-M and time-varying risk premia

Asymmetric Response and the Leverage Effect:

TGARCH and EGARCH

Asymmetric response I: TARCH

Standard GARCH:

$$\sigma_t^2 \;=\; \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$$

TARCH:

$$\sigma_t^2 \;=\; \omega + \alpha r_{t-1}^2 + \gamma r_{t-1}^2 D_{t-1} + \beta \sigma_{t-1}^2$$

where

positive return (good news): α effect on volatility

negative return (bad news): $\alpha + \gamma$ effect on volatility

$\gamma \neq 0$: Asymmetric news response

$\gamma > 0$: "Leverage effect"

Asymmetric Response and the Leverage Effect Cont.

Asymmetric Response II: E-GARCH

$$\ln(\sigma_t^2) \;=\; \omega + \alpha \left| \frac{r_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{r_{t-1}}{\sigma_{t-1}} + \beta \ln(\sigma_{t-1}^2)$$

- Log specification ensures that the conditional variance is positive.
- Volatility driven by both size and sign of shocks
- Leverage effect when $\gamma < 0$

Introducing Exogenous Variables

$$r_t \mid \Omega_{t-1} \sim N(0,\ \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma X_t$$

where:

γ is a parameter vector

X is a set of positive exogenous variables.

Component GARCH

Standard GARCH:

$$(\sigma_t^2 - \bar{\omega}) \;=\; \alpha(r_{t-1}^2 - \bar{\omega}) + \beta(\sigma_{t-1}^2 - \bar{\omega}),$$

for constant long-run volatility $\bar{\omega}$.
Component GARCH:

$$(\sigma_t^2 - q_t) \;=\; \alpha(r_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1}),$$

for time-varying long-run volatility q_t, where

$$q_t \;=\; \omega + \rho(q_{t-1} - \omega) + \phi(r_{t-1}^2 - \sigma_{t-1}^2)$$

Component GARCH Cont.

- Transitory dynamics governed by $\alpha + \beta$
- Persistent dynamics governed by ρ
- Equivalent to nonlinearly restricted GARCH(2,2)
- Exogenous variables and asymmetry can be allowed:

$$(\sigma_t^2 - q_t) \ = \ \alpha(r_{t-1}^2 - q_{t-1}) + \gamma(r_{t-1}^2 - q_{t-1})D_{t-1} + \beta(\sigma_{t-1}^2 - q_{t-1}) +$$

Regression with GARCH Disturbances

$$y_t = x_t'\beta + \varepsilon_t$$

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0, \ \sigma_t^2)$$

GARCH-M and Time-Varying Risk Premia

Standard GARCH regression model:

$$y_t = x_t' \beta + \varepsilon_t$$

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0,\ \sigma_t^2)$$

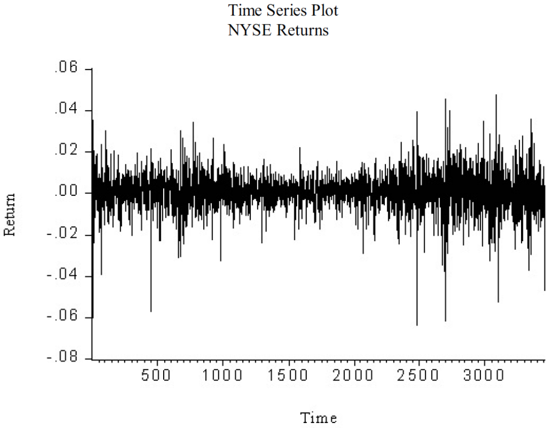
GARCH-M model is a special case:

$$y_t = x_t' \beta + \gamma \sigma_t^2 + \varepsilon_t$$

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0,\ \sigma_t^2)$$

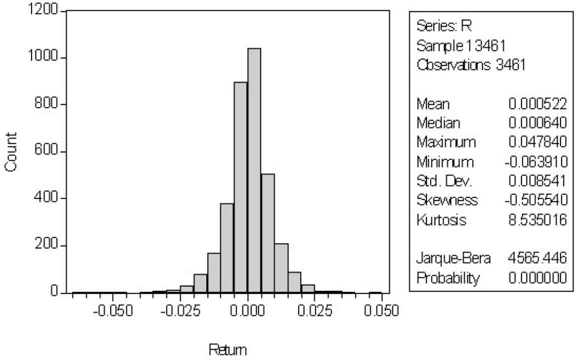
- Time-varying risk premia in excess returns

Time Series Plot – NYSE Returns

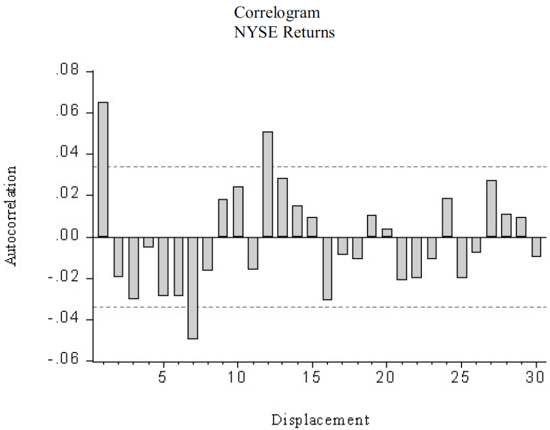


Histogram and Related Diagnostic Statistics – NYSE Returns

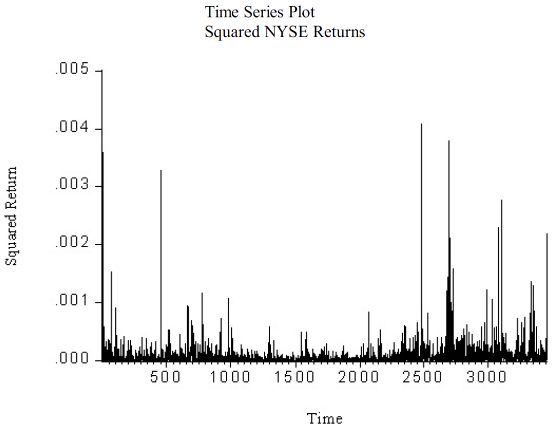
Histogram and Related Diagnostic Statistics
NYSE Returns



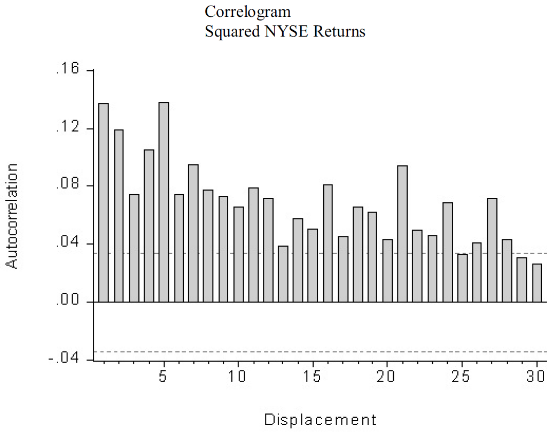
Correlogram – NYSE Returns



Time Series Plot – Squared NYSE Returns



Correlogram – Squared NYSE Returns



AR(5) Model – Squared NYSE Returns

AR(5) Model
Squared NYSE Returns

Dependent Variable: R2
Method: Least Squares

Sample(adjusted): 6 3461
Included observations: 3456 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.40E-05	3.78E-06	11.62473	0.0000
R2(-1)	0.107900	0.016137	6.686547	0.0000
R2(-2)	0.091840	0.016186	5.674167	0.0000
R2(-3)	0.028981	0.016250	1.783389	0.0746
R2(-4)	0.039312	0.016481	2.385241	0.0171
R2(-5)	0.116436	0.016338	7.126828	0.0000

R-squared	0.052268	Mean dependent var	7.19E-05
Adjusted R-squared	0.050894	S.D. dependent var	0.000189
S.E. of regression	0.000184	Akaike info criterion	-14.36434
Sum squared resid	0.000116	Schwarz criterion	-14.35366
Log likelihood	24827.58	F-statistic	38.05372
Durbin-Watson stat	1.975672	Prob(F-statistic)	0.000000

ARCH(5) Model – NYSE Returns

ARCH(5) Model
NYSE Returns

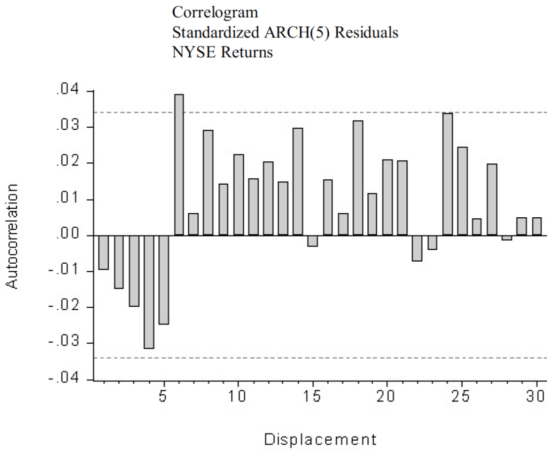
Dependent Variable: R
Method: ML - ARCH (Marquardt)

Sample: 1 3461
Included observations: 3461
Convergence achieved after 13 iterations
Variance backcast: ON

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000689	0.000127	5.437097	0.0000
Variance Equation				
C	3.16E-05	1.08E-06	29.28536	0.0000
ARCH(1)	0.128948	0.013847	9.312344	0.0000
ARCH(2)	0.166852	0.015055	11.08281	0.0000
ARCH(3)	0.072551	0.014345	5.057526	0.0000
ARCH(4)	0.143778	0.015363	9.358870	0.0000
ARCH(5)	0.089254	0.018480	4.829789	0.0000

R-squared	-0.000381	Mean dependent var	0.000522
Adjusted R-squared	-0.002118	S.D. dependent var	0.008541
S.E. of regression	0.008550	Akaike info criterion	-6.821461
Sum squared resid	0.252519	Schwarz criterion	-6.809024
Log likelihood	11811.54	Durbin-Watson stat	1.861036

Correlogram – Standardized ARCH(5) Residuals : NYSE Returns



GARCH(1,1) Model – NYSE Returns

GARCH(1,1) Model
NYSE Returns

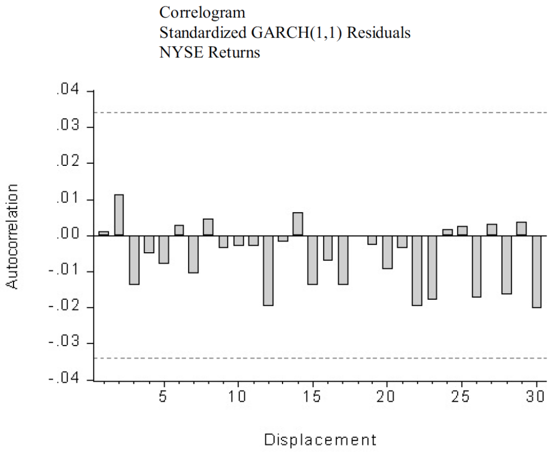
Dependent Variable: R
Method: ML - ARCH (Marquardt)

Sample: 1 3461
Included observations: 3461
Convergence achieved after 19 iterations
Variance backcast: ON

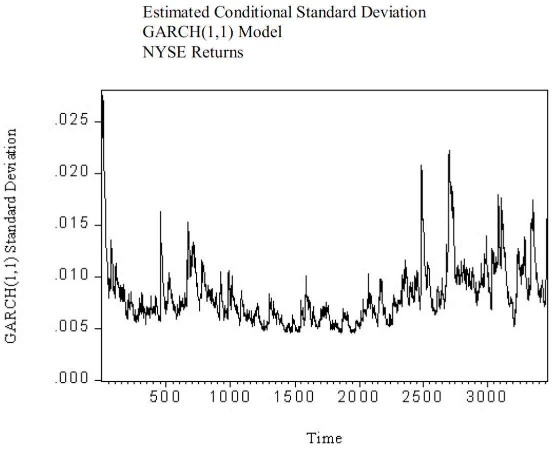
Coefficient	Std. Error	z-Statistic	Prob.	
C	0.000640	0.000127	5.036942	0.0000
Variance Equation				
C	1.06E-06	1.49E-07	7.136840	0.0000
ARCH(1)	0.067410	0.004955	13.60315	0.0000
GARCH(1)	0.919714	0.006122	150.2195	0.0000

R-squared	-0.000191	Mean dependent var	0.000522
Adjusted R-squared	-0.001059	S.D. dependent var	0.008541
S.E. of regression	0.008546	Akaike info criterion	-6.868008
Sum squared resid	0.252471	Schwarz criterion	-6.860901
Log likelihood	11889.09	Durbin-Watson stat	1.861389

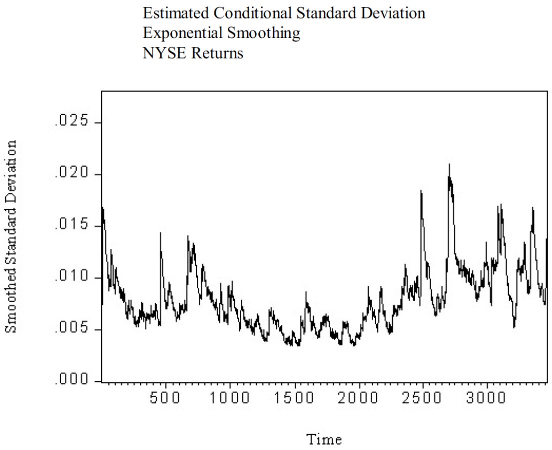
Correlogram – Standardized GARCH(1,1) Residuals : NYSE Returns



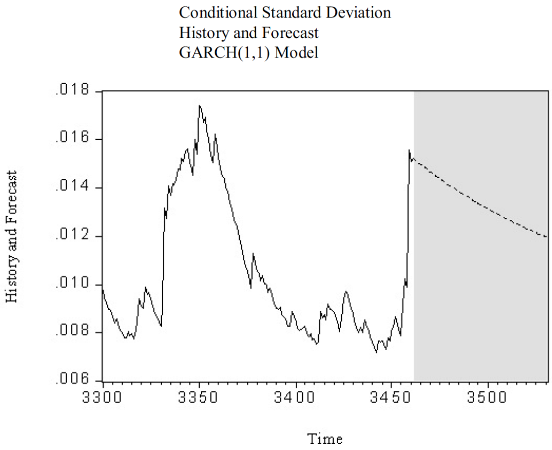
Estimated Conditional Standard Deviation – GARCH(1,1) Model : NYSE Returns



Estimated Conditional Standard Deviation – Exponential Smoothing : NYSE Returns



Conditional Standard Deviation – History and Forecast : GARCH(1,1) Model



Conditional Standard Deviation – Extended History and Extended Forecast : GARCH(1,1) Model