Eco 4306 Economic and Business Forecasting
## Chapter 3: Statistics and Time Series

Outline

- in the time series scenario the sample is just *one* observation per unit of time, with no option to obtain more

- question: since we only have one observation per unit of time, how do we construct sample moments estimators for population moments?

- to deal with this issue introduce the concept of stationarity

- next, we develop tools used to estimate forecasting models for time series that are stationary

3.1 Stochastic Process and Time Series

- ▶ **time series sample**: collection of observations ordered by time, $\{y_t\}_{t=1}^T = \{y_1, y_2, \ldots, y_T\}$, where T is the number of periods

- ▶ **time series plot**: graphical representation of the time series sample, with time on the horizontal axis, and values of the variable of interest on the vertical axis

- ▶ example: Dow Jones (DJ) Index monthly data, Closing price at the end of the month, January 1988 to April 2008, and the corresponding monthly return
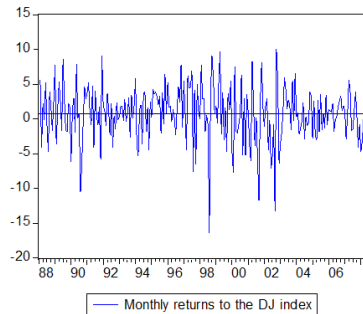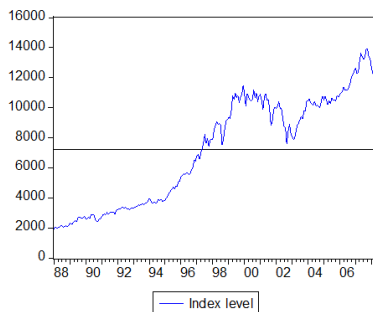
3.1 Stochastic Process and Time Series

recall: a random variable can be characterized in two ways

(1) to fully characterize it, we need to know its probability density function (pdf)
(2) a partial characterization can be obtained using its moments: mean μ, variance σ^2, skewness sk, kurtosis kr, ...

if we know the pdf, we can calculate the moments, but knowing the moments is not enough to find the pdf

3.1 Stochastic Process and Time Series

▶ recall: random variables are in uppercase, possible outcomes of the random variable in lowercase

▶ example: pdf of random variable Y, yearly income - possible outcomes of the variable are on the horizontal axis, for a given y the area under the curve is the probability of such an outcome, e.g. 12% of the U.S. population has yearly income $10,000 or less

▶ cross-sectional sample $\{y_1, y_2, \ldots, y_n\}$ is obtained by drawing individuals from the population and recording their income

▶ cross-sectional sample can then be used to calculate sample moments - e.g. sample mean $\bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i$, sample variance $\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$, and these sample moments are unbiased estimators of the population moments
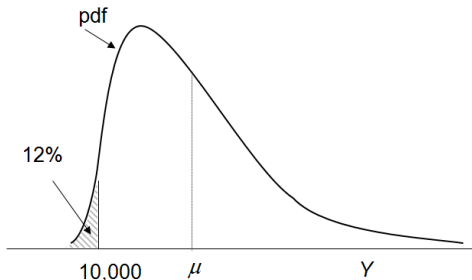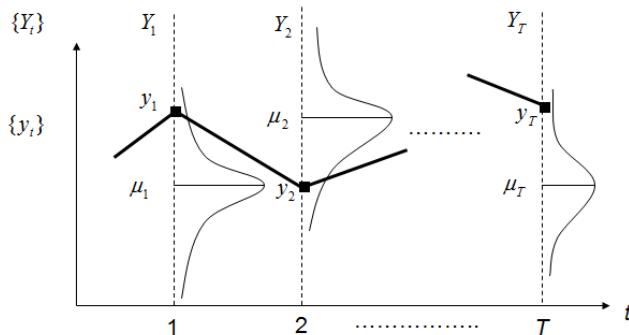
3.1.1 Stochastic Process

▶ **stochastic process** is a collection of random variables indexed by time,
$\{Y_t\} = \{Y_1, Y_2, \ldots, Y_T\}$

▶ note that we use capital letters - all elements are random variables rather than single values

▶ unit of time can be days, weeks, months, years, . . .
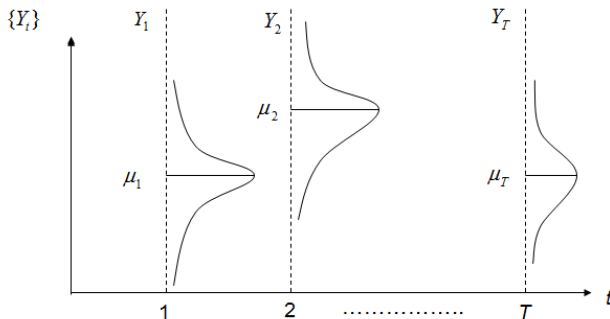
3.1.2 Time Series

▶ a **time series** $\{y_t\} = \{y_1, y_2, \ldots, y_T\}$ is a sample realization of a stochastic process $\{Y_t\} = \{Y_1, Y_2, \ldots, Y_T\}$

▶ for each period $t = 1, 2, \ldots, T$ we observe *a single* realization y_t of the random variable Y_t

3.1.2 Time Series

► we only observe time series, we never observe the stochastic process

► the data that we find in real life are time series, and our task is to infer the characteristics of the stochastic process that generated the time series data in order to build forecasting models

► Dow-Jones Index closing price at the end of month, from January 1988 to April 2008: we observe the time series $\{y_t\} = \{y_1, y_2, \ldots, y_{244}\}$ corresponding to the unobserved underlying stochastic process $\{Y_t\} = \{Y_1, Y_2, \ldots, Y_{244}\}$

3.1.2 Time Series

cross-sectional sample

- ▶ several observations of the random variable
- ▶ we use these observations to compute a cross-sectional sample mean, a cross-sectional sample variance, . . .

time series sample

- ▶ only one observation of the random variable
- ▶ cross-sectional moments can not be constructed, we can only construct time averages

3.1.2 Time Series

▶ time averages:
  ▶ for a time series $\{y_1, y_2, \ldots, y_T\}$ we can compute a time mean

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

  or a time variance

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

▶ but time series may come from stochastic process with potentially different pdfs over time

▶ if the stochastic process has different population means $\{\mu_1, \mu_2, \ldots, \mu_T\}$ and different population variances $\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_T^2\}$, it is not clear which μ and which σ^2 are approximated by time averages

3.1.2 Time Series

- ▶ two conditions need to be imposed on the behavior of the stochastic process, such that time averages are meaningful estimators of population averages

- ▶ **stationarity** focuses on the stochastic process (population information); it requires that the random variables that form the stochastic process have the same population mean and the same population variance

- ▶ **ergodicity** guarantees that under stationarity as the sample becomes larger the time averages converge to the population averages, and thus time averages are good substitutes for cross-sectional averages
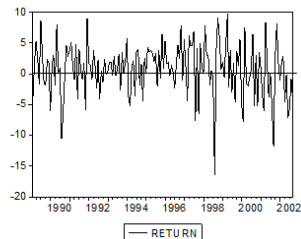
3.2.1 Stationarity

▶ contrast the time series plot of the Dow-Jones Index with the plot of Dow-Jones Index returns
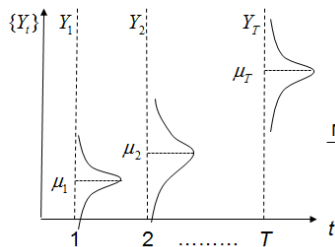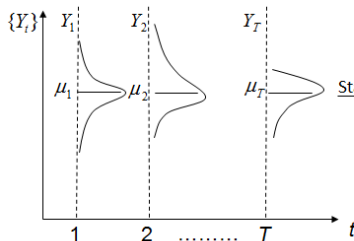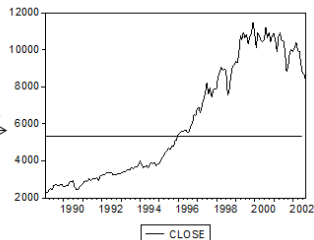
3.2.1 Stationarity

- ▶ mean statistic is a measure of centrality, one should expect the value of the Index to be around the mean value

- ▶ the graph of Dow-Jones however shows that the has been growing over time, without any tendency to revert back over the long run

- ▶ in contrast, the mean return of 0.77% is meaningful because there is a tendency to revert to the mean - the deviations for the mean are not getting larger over time

- ▶ time series of the Dow-Jones Index comes from a nonstationary stochastic process and the time series of Dow-Jones Index returns from a stationary process
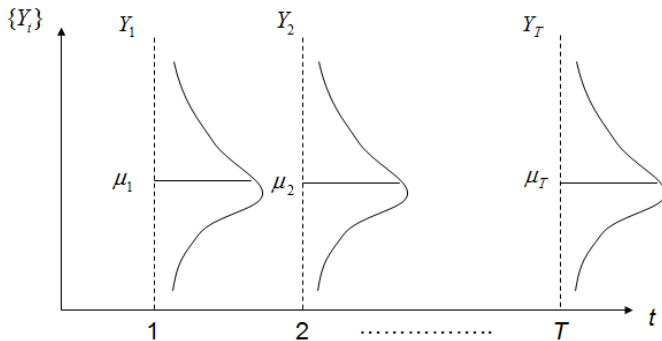
3.2.1 Stationarity

- requiring the process to be stationary, means imposing a degree of homogeneity/similarity in the random variables
- question: how much homogeneity do we need?

3.2.1 Stationarity

▶ stochastic process $\{Y_t\}$ is said to be **strongly stationary** if all random variables have the same probability density function, that is if

$$f_{Y_1}(y) = f_{Y_2}(y) = \ldots = f_{Y_T}(y)$$

3.2.1 Stationarity

▶ stochastic process $\{Y_t\}$ is said to be **first order weakly stationary** if all random variables have the same mean, that is if

$$\mu_1 = \mu_2 = \ldots = \mu_T = \mu$$

▶ stochastic process $\{Y_t\}$ is said to be **second order weakly stationary** if all random variables have the same mean and the same variance, and the covariances do not depend on time, that is if

$$\begin{aligned}
\text{(i)} \quad & \mu_1 = \mu_2 = \ldots = \mu_T = \mu \\
\text{(ii)} \quad & \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_T^2 = \sigma^2 \\
\text{(iii)} \quad & \rho_{Y_t, Y_{t-k}} = \rho_k
\end{aligned}$$

▶ second order weakly stationary processes are also called **covariance-stationary processes**

3.2.1 Stationarity

▶ strong stationarity requirement is very strict - it imposes a very high degree of homogeneity among the random variables since it requires that all pdfs are identical

▶ second order weak stationarity imposes fewer requirements on the random variables than strong stationarity

▶ under first order weak stationarity probability density functions can vary a lot: they only need to have same mean, but can differ in variance, skeweness, kurtosis, or any other higher moment

3.2.1 Stationarity

- ▶ we will first develop forecasting methodology for stationary processes
- ▶ in the second half of the semester we will look at statistical testing procedures to detect nonstationarity in the mean and in the variance and develop forecasting tools for nonstationary processes
- ▶ as a first resource, time series plots are good tools to hint the existence of a nonstationary mean and variance

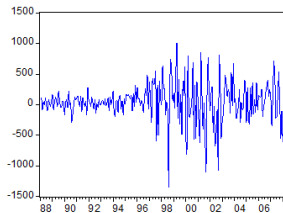# 3.2.2 Useful Transformations of Nonstationary Processes

- **lag operator** L: applying L to Y_t yields Y_{t-1} that is $LY_t = Y_{t-1}$, consequently we also have $L^j Y_t = Y_{t-j}$

- using lag operator the first difference of a time series is calculated as $\Delta y_t = y_t - y_{t-1} = y_t - Ly_t$

- example: Dow Jones (DJ) Index monthly data, Closing price at the end of the month, January 1988 to April 2008, and the corresponding monthly return

- import data from Microsoft Excel file **Figure03_fig01_fig7_DJI.xls** into EViews: click on **File**, select **Import**, then **Import from file**, and **Finish**

- construct first difference $\Delta DJ_t = DJ_t - DJ_{t-1}$ by choosing **Object** then **Generate Series** and entering **d_DJ = d(DJ)**

- construct log transformed data $\log DJ_t$ by chosing **Object** then **Generate Series** and entering **log_DJ = log(DJ)**

- construct first difference of log transformed data $\Delta \log DJ_t = \log DJ_t - \log DJ_{t-1}$ by chosing **Object** then **Generate Series** and entering **d_log_DJ = dlog(DJ)**
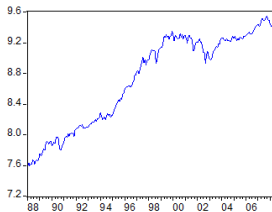
3.2.2 Useful Transformations of Nonstationary Processes

▶ to create a plot with all four time series, first select **Object** then **New Object** and **Group**, enter their names **DJ d_DJ log_DJ d_log_DJ**

▶ them, after the spreadsheet with group time series opens select **View** then **Graph** and in **Details** section change **Multiple series** option into **Multiple graphs**
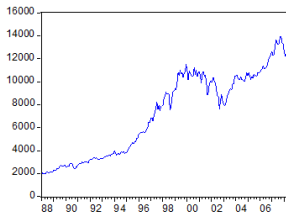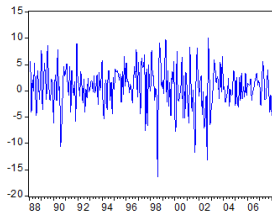

Dow Jones Index level

First difference of Index

Log Index

First difference of the log Index (returns)

3.2.2 Useful Transformations of Nonstationary Processes

first differences of Dow-Jones Index ΔY_t

▶ mean is constant over time so we have obtained a first order weakly stationary process
▶ variance still seems to change with time - series became much more volatile from 1997 to 2008 compared to the beginning of the sample
▶ transformed process ΔY_t thus does not seem to be covariance-stationary

first differences of Dow-Jones Index after the logarithmic transformation $\Delta \log Y_t$

▶ applying natural log to the Dow-Jones Index does not affect its trending behavior
▶ by taking first differences of the log transformed series, we obtain a time series that has a more homogeneous variance than the original series and a mean that is constant over time
▶ transformed process $\Delta \log Y_t$ thus appears to be covariance-stationary

3.2.2 Useful Transformations of Nonstationary Processes

▶ for monthly data, the month-over-month percentage return is calculated as

$$r_t = 100 \times \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

and can be approximated using logarithms because

$$\Delta \log Y_t = \log Y_t - \log Y_{t-1} = \log\left(\frac{Y_t}{Y_{t-1}}\right) = \log\left(1 + \frac{r_t}{100}\right) \approx \frac{r_t}{100}$$

implies that $r_t \approx 100\Delta Y_t$

▶ in some case, we are interested in year-over-year percentage change
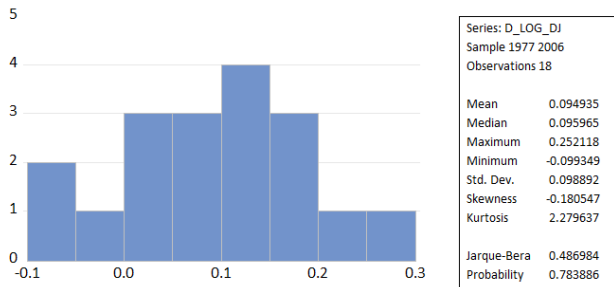
$$r_t^{YoY} = 100 \times \frac{Y_t - Y_{t-12}}{Y_{t-12}}$$

which can be approximated using the following log difference

$$r_t^{YoY} \approx 100\Delta_{12}\log Y_t = 100 \times (\log Y_t - \log Y_{t-12})$$

▶ in EViews we construct $rDJ_t^{YoY} = 100 \times \frac{DJ_t - DJ_{t-12}}{DJ_{t-12}}$ by chosing **Object** then **Generate Series** and entering **r_DJ_YoY = 100 * d(DJ, 0, 12) / DJ(-12)**

▶ in EViews we construct $100 \times (\log DJ_t - \log DJ_{t-12})$ by chosing **Object** then **Generate Series** and entering **d_log_DJ_YoY = 100 * dlog(DJ, 0, 12)**

3.2.2 Useful Transformations of Nonstationary Processes

▶ to obtain histogram and descriptive statistics for Dow Jones returns, double click **D_log_DJ** to open it, select **View** then **Descriptive Statistics & Tests** and then **Histogram and Stats**

Series: D_LOG_DJ	
Sample 1977 2006	
Observations 18	
Mean	0.094935
Median	0.095965
Maximum	0.252118
Minimum	-0.099349
Std. Dev.	0.098892
Skewness	-0.180547
Kurtosis	2.279637
Jarque-Bera	0.486984
Probability	0.783886

3.2.2 Useful Transformations of Nonstationary Processes

▶ we define a measure of **skewness** as

$$sk = E\left[\left(\frac{Y - \mu_Y}{\sigma_Y}\right)^3\right]$$

▶ if the pdf of a random variable is symmetric around mean skewness is zero

▶ positive values of sk: distribution is right skewed, median is smaller than mean

▶ negative values of sk: distribution is left skewed, median is larger than mean

▶ we define **kurtosis** as

$$kr = E\left[\left(\frac{Y - \mu_Y}{\sigma_Y}\right)^4\right]$$

provides information about the "thickness" of the tails of the distribution is

▶ we measure kurtosis in excess to the kurtosis of a normal distribution - for a normal pdf, $kr = 3$

▶ if $kr > 3$, we say that there is **leptokurtosis** or **fat tails**; this means that there is more probability in the tails than in the density of normal distribution

▶ if $kr < 3$, we say that the distribution has **thin tails**, less probability in tails than normal distribution

3.3 A New Tool of Analysis: The Autocorrelation Functions

▶ two useful tools to detect time patterns: **autocorrelation function** (ACF) and **partial autocorrelation function** (PACF)

▶ ACF and PACF will be useful to construct time series models for forecasting

▶ consider stochastic process $\{Y_t\} = \{Y_1, Y_2, \ldots, Y_T\}$ and two of its random variables k periods apart, Y_t and Y_{t-k}

▶ calculate **autocorrelation coefficient or order** k, the measure of linear association of Y_t with Y_{t-k}

$$\rho_{Y_t, Y_{t-k}} = \frac{cov(Y_t, Y_{t-k})}{\sqrt{var(Y_t)} \sqrt{var(Y_{t-k})}}$$

▶ **autocorrelation function** (ACF) is a function that assigns to any two random variables that are k periods apart their correlation coefficient, $\rho : k \to \rho_{Y_t, Y_{t-k}}$

3.3 A New Tool of Analysis: The Autocorrelation Functions

▶ a covariance stationary process, the autocorrelation function can be simplified since the first and second moments are time invariant, and thus

$$\rho_k = \frac{cov(Y_t, Y_{t-k})}{\sqrt{var(Y_t)}\sqrt{var(Y_{t-k})}} = \frac{\gamma_k}{\gamma_0}$$

where γ_k is the autocovariance of order k, and γ_0 is the variance of the process, which is the autocovariance of order zero

▶ **sample autocorrelation function** for a covariance stationary process is constructed using

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\frac{1}{T-k}\sum_{t=k+1}^{T}(y_t - \bar{\bar{y}})(y_{t-k} - \bar{\bar{y}})}{\frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{\bar{y}})^2}$$

where $\bar{\bar{y}} = \frac{1}{T}\sum_{t=1}^{T} y_t$ and $\bar{\bar{x}} = \frac{1}{T}\sum_{t=1}^{T} x_t$

3.3 A New Tool of Analysis: The Autocorrelation Functions

▶ the autocorrelation coefficient ρ_k is affected by all the random variables between t and $t + k$, that is $Y_{t+1}, Y_{t+2}, \ldots, Y_{t+k-1}$

▶ if we want to know the autocorrelation between Y_t and Y_{t+k} after we had removed the information in between we need to control for the effects of $Y_{t+1}, Y_{t+2}, \ldots, Y_{t+k-1}$

▶ this type of correlation is given by the **partial autocorrelation coefficient**, to obtain it consider set of OLS regressions

$$Y_{t+k} = \beta_{0,1} + \beta_{1,1}Y_{t+k-1} + \varepsilon_{t+k}$$
$$Y_{t+k} = \beta_{0,2} + \beta_{1,2}Y_{t+k-1} + \beta_{2,2}Y_{t+k-2} + \varepsilon_{t+k}$$
$$Y_{t+k} = \beta_{0,3} + \beta_{1,3}Y_{t+k-1} + \beta_{2,3}Y_{t+k-2} + \beta_{3,3}Y_{t+k-3} + \varepsilon_{t+k}$$
$$\vdots$$
$$Y_{t+k} = \beta_{0,k} + \beta_{1,k}Y_{t+k-1} + \beta_{2,k}Y_{t+k-2} + \beta_{3,k}Y_{t+k-3} + \ldots + \beta_{k,k}Y_t + \varepsilon_{t+k}$$

▶ estimated coefficients $\hat{\beta}_{1,1}, \hat{\beta}_{2,2}, \ldots, \hat{\beta}_{k,k}$ form the **sample partial autocorrelation function** (PACF)

3.3 A New Tool of Analysis: The Autocorrelation Functions

▶ ACF and PACF are commonly reported in econometric software by plotting the **correlogram**

▶ example: Annual Working Hours per Employee in the United States, 1977 to 2006

▶ import data from Microsoft Excel file **Figure03_10_HoursWorkedUSA.XLS** into EViews: click on **File**, select **Import**, then **Import from file**, and **Finish**

▶ obtain correlogram opening the series **usa** and choosing **View → Correlogram**

Sample: 1977 2006
Included observations: 30

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.737	0.737	17.987	0.000
		2	0.364	-0.392	22.537	0.000
		3	0.062	-0.058	22.676	0.000
		4	-0.086	0.039	22.951	0.000
		5	-0.162	-0.126	23.957	0.000
		6	-0.288	-0.295	27.270	0.000
		7	-0.352	0.052	32.432	0.000
		8	-0.253	0.185	35.229	0.000
		9	-0.064	0.001	35.416	0.000
		10	0.114	0.034	36.035	0.000

3.3.2 Statistical Tests for Autocorrelation Coefficients

▶ when we compute the sample ACF we obtain an estimate $\hat{\rho}_k$ of the population parameter ρ_k

▶ to test whether some autocorrelation coefficient is zero ... following null hypothesis $H_0 : \rho_k = 0$

▶ in a large sample $\hat{\rho}_k$ is a normal random variable with mean 0 and variance $1/T$ when the null hypothesis is true, $\hat{\rho}_k \sim N(0, 1/T)$

▶ thus using the t-ratio test, at the customary 5% significance level, we reject the null hypothesis whenever

$$|\hat{\rho}_k / \sqrt{1/T}| > 1.96$$

▶ equivalently, the 95% confidence interval is

$$(\hat{\rho}_k - 1.96\sqrt{1/T}, \hat{\rho}_k + 1.96\sqrt{1/T})$$

▶ in EViews correlogram, the dashed lines running vertically in both correlograms are the bands corresponding to a 95% confidence interval centered at zero

▶ the autocorrelation coefficients that fall within the bands are statistically zero

3.3.2 Statistical Tests for Autocorrelation Coefficients

► the following joint hypothesis test

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_k = 0 \quad \text{against} \quad H_1 : \rho_j \neq 0 \text{ for some } j \in \{1, 2, \ldots, k\}$$

is evaluated using the **Ljung-Box Q-statistic**

$$Q_k = T(T+2) \sum_{j=1}^{k} \frac{\hat{\rho}_j^2}{T-j}$$

which under null hypothesis is chi-square distributed with degrees of freedom k, the number of autocorrelations tested in the null
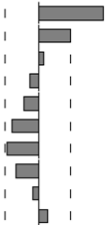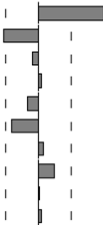
► rejecting H_0 means that some or all autocorrelations up to order k are different from zero

3.3.2 Statistical Tests for Autocorrelation Coefficients

▶ for example, for Annual Working Hours per Employee in the United States from 1977 to 2006, for $H_0 : \rho_1 = \rho_2 = \rho_3 = 0$ we have Q-statistic is $Q_3 = 22.676$ and the associated p-value is 0 so we strongly reject the null hypothesis

Sample: 1977 2006
Included observations: 30

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.737	0.737	17.987	0.000
		2	0.364	-0.392	22.537	0.000
		3	0.062	-0.058	22.676	0.000
		4	-0.086	0.039	22.951	0.000
		5	-0.162	-0.126	23.957	0.000
		6	-0.288	-0.295	27.270	0.000
		7	-0.352	0.052	32.432	0.000
		8	-0.253	0.185	35.229	0.000
		9	-0.064	0.001	35.416	0.000
		10	0.114	0.034	36.035	0.000

3.4 Conditional Moments and Time Series: What Lies Ahead

▶ in regression analysis, the goal is to find $E(Y|X)$ the expected value of dependent variable Y for a given value of explanatory variable X

▶ this yields the conditional mean of Y as a function of X, that is, $E(Y|X) = g(X)$

▶ in forecasting, regression and conditional moments play an important role - **forecast is a function of conditional moments**

▶ for a stochastic process $\{Y_t\}$ and an information set $I_t = \{y_1, y_2, \ldots y_t\}$, we construct a forecast for Y_{t+h}, as a function of the information set $f_{t,h} = g(I_t)$

▶ function g is a function of conditional moments - often the forecast is just the conditional mean (but sometimes it also depends on the conditional variance)

▶ thus for instance we can have

$$f_{t,h} = g_1(I_t) = E(Y_{t+h}|I_t) = E(Y_{t+h}|y_t, y_{t-1}, \ldots, y_1)$$

▶ then, if we specify the regression model as

$$E(Y_{t+h}|Y_t, Y_{t-1}, \ldots, Y_1) = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}$$

and estimate it using OLS to obtain $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, we can construct the 1-period-ahead forecast of Y_{t+1} as

$$f_{t,1} = \hat{\beta}_0 + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2}$$