Econometrics

Francis X. Diebold
University of Pennsylvania

March 21, 2015
Copyright © 2013 onward, by Francis X. Diebold.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. (Briefly: You can use, copy and distribute non-commercially, so long as you give me attribution and do not modify.) To view a copy of the license, visit http://creativecommons.org/licenses/by-nc-nd/4.0.

The painting is *Enigma*, by Glen Josselsohn, from Wikimedia Commons.
Introduction
Who Uses Econometrics

Statistical analysis of economic data:

- Economics
- Finance
- Business
- Consulting
- Government
What Makes Econometrics Special

Econometrics is not just “statistics using economic data.”

Special issues related to the properties of economic data.
- No experiments; only “observational data”
- Special issues and features that arise routinely in economic data
- Predictive modeling, causal modeling
Types of Recorded Economic Data

- Continuous vs. discrete
- Time series vs. cross section
- Panel

Complement: Explore nominal, ordinal, interval and ratio data
Web Data Resources

- Resources for Economists (AEA)
- FRED (Federal Reserve Economic Data)
- National Bureau of Economic Research
- Quandl
- FRB Phila Real-Time Data Research Center
Software

- R
  - CRAN
  - RStudio
  - R-bloggers

- More: Eviews, Python

- Still more: *Econometrics Journal* software links
Graphics
Let’s have some fun and look at the pictures *first*...
Time Series Plot:
1-Year Government Bond Yield, Levels
Time Series Plot:
1-Year Government Bond Yield, Levels and Changes
Histogram: 1-Year Government Bond Yield
Bivariate Scatterplot
1-Year and 10-Year Government Bond Yields
Scatterplot Matrix:
1-, 10-, 20- and 30-Year Government Bond Yields
Graphics

- Summarize and reveal patterns in univariate time-series data. Time Series plots. Trend, seasonal, cycle, outliers, ...
- Summarize and reveal patterns in univariate cross-section data. Histograms are helpful for learning about distributional shape. Symmetric, skewed, fat-tailed, ...
- Identify relationships and understand their nature, in both multivariate time-series and multivariate cross-section environments. Bivariate scatterplots. Does a relationship exist? Is it linear or nonlinear? Are there outliers?
- Identify relationships and understand their nature in panel data. Cross-sectional histograms across time periods, or time series plots across cross-sectional units.
- Compare different pieces of data via multiple comparisons. Scatterplot matrix.
Univariate and Multivariate Graphics

- Time-series plot
  - levels
  - change

- Density estimate
  - histogram
  - smoothed

- Scatterplot
  - Two-way
  - Multi-way
Principles of Graphical Style

- Know your audience, and know your goals.
- Appeal to the viewer.
- Show the data, and only the data, within the bounds of reason.
  - Avoid distortion. The sizes of effects in graphics should match their size in the data. Use common scales in multiple comparisons.
  - Minimize, within reason, non-data ink. Avoid chartjunk.
  - Third, choose aspect ratios to maximize pattern revelation. Bank to 45 degrees.
  - Maximize graphical data density.
- Revise and edit, again and again (and again). Graphics produced using software defaults are almost never satisfactory.
Distributions of Wages and Log Wages
Probability and Statistics Review
“Sample” EPC: Simple vs. Partial Correlation

(Read them all carefully!)

The set of pairwise scatterplots that comprises a multiway scatterplot provides useful information about the joint distribution of the set of variables, but it’s incomplete information and should be interpreted with care. A pairwise scatterplot summarizes information regarding the simple correlation between, say, \( x \) and \( y \). But \( x \) and \( y \) may appear highly related in a pairwise scatterplot even if they are in fact unrelated, if each depends on a third variable, say \( z \). The crux of the problem is that there’s no way in a pairwise scatterplot to examine the correlation between \( x \) and \( y \) controlling for \( z \), which we call partial correlation. When interpreting a scatterplot matrix, keep in mind that the pairwise scatterplots provide information only on simple correlation.
Moments, Sample Moments and Their Sampling Distributions

- Discrete random variable, $y$
- Discrete probability distribution $p(y)$
- Continuous random variable $y$
- Probability density function $f(y)$
Mean measures location:

\[ \mu = E(y) = \sum_i p_i y_i \quad \text{(discrete case)} \]

\[ \mu = E(y) = \int y f(y) \, dy \quad \text{(continuous case)} \]

Variance, or standard deviation, measures dispersion, or scale:

\[ \sigma^2 = \text{var}(y) = E(y - \mu)^2. \]

- \( \sigma \) easier to interpret than \( \sigma^2 \). Why?
More Population Moments

Skewness measures skewness (!)

\[ S = \frac{E(y - \mu)^3}{\sigma^3}. \]

Kurtosis measures tail fatness relative to a Gaussian distribution.

\[ K = \frac{E(y - \mu)^4}{\sigma^4}. \]
Covariance and Correlation

Multivariate case: Joint, marginal and conditional distributions
\[ f(x, y), f(x), f(y), f(x|y), f(y|x) \]

Covariance measures linear dependence:
\[ \text{cov}(y, x) = E[(y_t - \mu_y)(x_t - \mu_x)]. \]

So does correlation:
\[ \text{corr}(y, x) = \frac{\text{cov}(y, x)}{\sigma_y \sigma_x}. \]

Correlation is often more convenient. Why?
Sampling and Estimation

Sample: \( \{y_i\}_{i=1}^{N} \sim f(y) \)

Sample mean:
\[ \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i \]

Sample variance:
\[ \hat{\sigma}^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N} \]

Unbiased sample variance:
\[ s^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N - 1} \]
Sample skewness:

\[ \hat{S} = \frac{1}{N} \sum_{i=1}^{N} (y_t - \bar{y})^3 \]

Sample kurtosis:

\[ \hat{K} = \frac{1}{N} \sum_{i=1}^{N} (y_t - \bar{y})^4 \]
Still More Sample Moments

Sample covariance:

$$\widehat{\text{cov}}(y, x) = \frac{1}{N} \sum_{i=1}^{N} [(y_i - \bar{y})(x_i - \bar{x})]$$

Sample correlation:

$$\widehat{\text{corr}}(y, x) = \frac{\widehat{\text{cov}}(y, x)}{\hat{\sigma}_y \hat{\sigma}_x}$$
Exact Sampling Distribution of the Sample Mean
(Requires *iid* Normality)

Simple random sampling: \( y_i \sim iid \ N(\mu, \sigma^2), i = 1, \ldots, N \)

\( \bar{y} \) is unbiased, consistent, normally distributed with variance \( \sigma^2/N \),
and minimum variance unbiased (MVUE).

\[
\bar{y} \sim N \left( \mu, \frac{\sigma^2}{N} \right)
\]

\[
\sqrt{N}(\bar{y} - \mu) \sim N(0, \sigma^2)
\]

\[
\mu \in \left[ \bar{y} \pm t_{1-\frac{\alpha}{2}} \left( N - 1 \right) \frac{s}{\sqrt{N}} \right] \text{ w.p. } \alpha
\]

\[
\mu = \mu_0 \implies \frac{\bar{y} - \mu_0}{s/\sqrt{N}} \sim t(N - 1)
\]
Approximate Asymptotic Sampling Distribution
(Does Not Require Normality)

Simple random sampling: \( y_i \sim iid(\mu, \sigma^2), i = 1, \ldots, N \)

\( \bar{y} \) is unbiased, consistent, asymptotically normally distributed with variance \( \sigma^2/N \), and best linear unbiased (BLUE).

\[
\bar{y} \xrightarrow{a} \mathcal{N} \left( \mu, \frac{\sigma^2}{N} \right)
\]

\[ \sqrt{N}(\bar{y} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \]

As \( N \to \infty \), \( \mu \in \left[ \bar{y} \pm z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{N}} \right] \) w.p. \( \alpha \)

As \( N \to \infty \), \( \frac{\bar{y} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{N}}} \sim \mathcal{N}(0, 1) \)
Notational Aside

Standard cross-section notation: \( i = 1, \ldots, N \)

Standard time-series notation: \( t = 1, \ldots, T \)

Much of our discussion will be valid in both cross-section and time-series environments, but still we have to pick a notation.

Without loss of generality, we will use \( t = 1, \ldots, T \).
Regression
Surely the all-time greatest statistical and econometric workhorse...
Distributions of Log Wage, Education and Experience
Scatterplot: Log Wage vs. Education
Regression as Curve Fitting

Fit a line:

\[ y_t = \beta_1 + \beta_2 x_t \]

Solve:

\[ \min_{\beta} \sum_{t=1}^{T} (y_t - \beta_1 - \beta_2 x_t)^2 \]

\( \beta \) is the set of two parameters \( \beta_1 \) and \( \beta_2 \)

\( \hat{\beta} \) is the set of fitted parameters \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \)
Scatterplot: Log Wage vs. Education with Superimposed Regression Line

\[ \hat{LWAGE} = 1.273 + .081\text{EDUC} \]
Actual Values, Fitted Values and Residuals

The fitted values are

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 x_t,$$

$$t = 1, \ldots, T.$$

The residuals are the difference between actual and fitted values,

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

$$t = 1, \ldots, T.$$
Multiple Linear Regression ($K$ RHS Variables))

Solve:

$$\min_{\beta} \sum_{t=1}^{T} (y_t - \beta_1 - \beta_2 x_{2t} - \ldots - \beta_K x_{Kt})^2$$

Fitted line:

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t} + \ldots + \hat{\beta}_K x_{Kt}$$

More compactly:

$$\hat{y}_t = \sum_{i=1}^{K} \hat{\beta}_i x_{it},$$

where $x_{1t} = 1$ for all $t$.

Wage dataset:

$$\text{LWAGE} = .867 + .093EDUC + .013EXPER$$
Regression as a Probability Model

\[ y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_K x_{Kt} + \epsilon_t \]

\[ \epsilon_t \sim iidN(0, \sigma^2), \]

\[ t = 1, \ldots, T. \]

Note:

\[ E(y_t|x_t = x_t^*) = \beta_1 + \beta_2 x_{2t}^* + \ldots + \beta_K x_{Kt}^* \]

Estimation:

\[ \min_\beta \sum_{t=1}^{T} \epsilon_t^2 \]
You already understand matrix ("spreadsheet") notation although you may not know it!

\[ y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{21} & x_{31} & \ldots & x_{K1} \\ 1 & x_{22} & x_{32} & \ldots & x_{K2} \\ \vdots \\ 1 & x_{2T} & x_{3T} & \ldots & x_{KT} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix} \]
Elementary Matrices and Matrix Operations

\[
\begin{align*}
0 &= \begin{pmatrix} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix} \\
I &= \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}
\end{align*}
\]

Transposition: \( A'_{ij} = A_{ji} \)

Addition: For \( A \) and \( B \ \text{n} \times \text{m} \), \( (A + B)_{ij} = A_{ij} + B_{ij} \)

Multiplication: For \( A \ \text{n} \times \text{m} \) and \( B \ \text{m} \times \text{p} \), \( (AB)_{ij} = \sum_{k=1}^{m} A_{ik}B_{kj} \).

Inversion: For non-singular \( A \ \text{n} \times \text{n} \), \( A^{-1} \) satisfies \( A^{-1}A = AA^{-1} = I \). Many algorithms exist for calculation.
We Used to Write This:

\[ y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_K x_{Kt} + \epsilon_t \]
\[ \epsilon_t \sim iidN(0, \sigma^2) \]
\[ t = 1, 2, \ldots, T \]
Now, Equivalently, We Write This:

\[ y = \mathbf{X}\beta + \varepsilon \quad (1) \]
\[ \varepsilon \sim \mathcal{N}(0, \sigma^2 I) \quad (2) \]
The Full Ideal Conditions (FIC)

1. The true data-generating process is:

\[ y = X\beta + \varepsilon \]

\[ \varepsilon \sim N(0, \sigma^2 I), \]

and the fitted model matches it exactly.

1.1 The relationship, if any, is truly linear, with no omitted variables, no measurement error, etc.

1.2 The coefficients, \( \beta \), are fixed.

1.3 \( \varepsilon \sim N \).

1.4 The \( \varepsilon_t \)'s have constant variance \( \sigma^2 \).

1.5 The \( \varepsilon_t \)'s are uncorrelated.

2. \( X \) is of full column rank.

3. \( X \) is fixed in repeated samples.

Surely these are heroic assumptions in economic environments. Much of econometrics (and this course) is devoted to relaxing them.
The OLS estimator is:

$$\hat{\beta}_{LS} = (X'X)^{-1}X'y$$

Under the full ideal conditions it MVUE and distributed as

$$\hat{\beta}_{LS} \sim N(\beta, \sigma^2(X'X)^{-1})$$.
Regression Analysis of Wages, Education and Experience

Figure: Wage Regression Output

Dependent Variable: LWAGE
Method: Least Squares
Date: 06/27/13   Time: 16:38
Sample (adjusted): 11323
Included observations: 1323 after adjustments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.867382</td>
<td>0.075331</td>
<td>11.51431</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.093229</td>
<td>0.005045</td>
<td>18.48002</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.013104</td>
<td>0.001164</td>
<td>11.26208</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared     0.232224   Mean dependent var  2.341995
Adjusted R-squared 0.231061  S.D. dependent var  0.561435
S.E. of regression  0.492318  Akaike info criterion  1.422881
Sum squared resid   319.9376  Schwarz criterion    1.434644
Log likelihood    -938.2358  Hannan-Quinn criterion  1.427291
F-statistic       199.6260  Durbin-Watson stat    1.926045
Prob(F-statistic) 0.000000

Path = c:\users\francis x. diebo\documents\diebold files\courses\econ104\old\econ104_2011\sw3e\views materials\DB
“Top Matter”: Background Information

- Dependent variable
- Method
- Date
- Sample
- Included observations
“Middle Matter”: Estimated Regression Function

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
</table>

- Variable
- Coefficient
- Standard error
- t-statistic
- p-value
There are many...
Regression Statistics: Mean dependent var 2.342

\[ \bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t \]
Regression Statistics: S.D. dependent var .561

\[ SD = \sqrt{\frac{\sum_{t=1}^{T} (y_t - \bar{y})^2}{T - 1}} \]
Regression Statistics: Sum squared resid 319.938

\[ SSR = \sum_{t=1}^{T} e_t^2 \]
Regression Statistics: Log likelihood -938.236

- Likelihood
- Log likelihood
- Maximum-likelihood estimation
- Hypothesis tests and model selection
Regression Statistics: $F$-statistic 199.626

$$F = \frac{(SSR_{res} - SSR)/(K - 1)}{SSR/(T - K)}$$
Regression Statistics: S.E. of regression \( .492 \)

\[
s^2 = \frac{\sum_{t=1}^{T} e_t^2}{T - K}
\]

\[
SER = \sqrt{s^2} = \sqrt{\frac{\sum_{t=1}^{T} e_t^2}{T - K}}
\]
Regression Statistics: $R$-squared .232

$$R^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$$
Regression Statistics: Adjusted $R$-squared .231

$$\bar{R}^2 = 1 - \frac{\frac{1}{T-K} \sum_{t=1}^{T} e_t^2}{\frac{1}{T-1} \sum_{t=1}^{T} (y_t - \bar{y})^2}$$
Regression Statistics: Schwarz criterion 1.435

\[ SIC = T \left( \frac{k}{T} \right) \frac{\sum_{t=1}^{T} e_t^2}{T} \]
Regression Statistics: Akaike info criterion 1.423

\[ AIC = e^{\left(\frac{2K}{T}\right)} \frac{\sum_{t=1}^{T} e_t^2}{T} \]
Regression Statistics: Durbin-Watson stat. 1.926

\[ \varepsilon_t = \phi \varepsilon_{t-1} + \nu_t \]

\[ \nu_t \sim iidN(0, \sigma^2) \]

\[ DW = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \]
Residual Scatter
Figure: Wage Regression Residual Plot
Beyond OLS: Non-Quadratic Objectives

Various Loss Functions

Loss vs. Error for different loss functions: Quadratic Loss, LinLin Loss, and Absolute Error Loss.
Ordinary Least Squares (OLS)

Recall that the OLS estimator, $\hat{\beta}_{OLS}$, solves:

$$\min_{\beta} \sum_{t=1}^{T} (y_t - \beta_1 - \beta_2 x_{2t} - \ldots - \beta_K x_{Kt})^2 = \min_{\beta} \sum_{t=1}^{T} \varepsilon_t^2$$

- Simple
  (analytic closed-form expression, $(X'X)^{-1}X'y$)

- Wonderful properties under FIC
  (Unbiased, consistent, Gaussian, MVUE)

But other approaches are possible and sometimes useful.
Least Absolute Deviations (LAD)

The LAD estimator, $\hat{\beta}_{LAD}$, solves:

$$\min_\beta \sum_{t=1}^{T} |\varepsilon_t|$$

– Not as simple as OLS, but still simple  
(Solves a linear programming problem)

– Useful properties under some violations of FIC  
(Robust to outliers; more on that later)

– But there’s a much bigger reason to be interested
Conditional Mean and Median Functions

- OLS fits the conditional mean function:

\[
\text{mean}(y|X) = x\beta
\]

- LAD fits the conditional median function (50% quantile):

\[
\text{median}(y|X) = x\beta
\]

- The two are equal under symmetry as with FIC, but not under asymmetry, in which case the median is a better measure of central tendency.
Quantile Regression (QR)

Objective like LAD but unequal slopes on each side of 0.

QR estimator $\hat{\beta}_{QR}$ minimizes “linlin loss,” or “check function loss”:

$$\min_{\beta} \sum_{t=1}^{T} \text{linlin}(\varepsilon_t),$$

where:

$$\text{linlin}(e) = \begin{cases} 
a|e|, \text{ if } e \leq 0 
\phantom{=} 
\phantom{=} 

b|e|, \text{ if } e > 0 
\end{cases}$$

$$= a|e| I(e \leq 0) + b|e| I(e > 0).$$

$I(x) = 1$ if $x$ is true, and $I(x) = 0$ otherwise.

“$I(\cdot)$” stands for “indicator” variable.

“linlin” refers to linearity on each side of the origin.

Not as simple as OLS, but still simple
(solves a linear programming problem)
What Does Quantile Regression Fit?

– QR fits the $d \cdot 100\%$ quantile:

$$\text{quantile}_d(y|X) = x\beta$$

where

$$d = \frac{b}{a + b} = \frac{1}{1 + a/b}$$

– Median regression (LAD) is special case of $d = .5$

– Important generalization of median regression (e.g., How do the wages of people in the far left tail of the wage distribution vary with education and experience, and how does that compare to those in the center of the wage distribution?)
Non-Normality and Outliers
What We’ll Do

– Distributional results under non-normality
  – Detecting non-normality and outliers
  – Dealing with non-normality and outliers (robust estimation)
Recall Sample Mean Under \textit{iid} Normality

\bar{y} \text{ is MVUE, and}

\bar{y} \sim N \left( \mu, \frac{\sigma^2}{T} \right),

and we estimate $\frac{\sigma^2}{T}$ consistently using $\frac{s^2}{T}$

- Small- and large-sample result
Recall Sample Mean Under \textit{iid} (Less Normality)

\[ \bar{y} \text{ is BLUE, and} \]

\[ \bar{y} \overset{a}{\sim} N \left( \mu, \frac{\sigma^2}{T} \right), \]

and we estimate \( \frac{\sigma^2}{T} \) consistently using \( \frac{s^2}{T} \)

– Large-sample result
OLS Under FIC (Including Normality)

$\hat{\beta}_{LS}$ is MVUE, and

\[
\hat{\beta}_{LS} \sim N(\beta, \sigma^2 (X'X)^{-1}),
\]

and we estimate $\sigma^2 (X'X)^{-1}$ consistently using $s^2 (X'X)^{-1}$

– Small- and large-sample result
OLS Under FIC (Less Normality)

\( \hat{\beta}_{LS} \) is BLUE, and

\[
\hat{\beta}_{LS} \sim a \ N ( \beta, \sigma^2 (X'X)^{-1}) ,
\]

and we estimate \( \sigma^2 (X'X)^{-1} \) consistently using \( s^2 (X'X)^{-1} \)

- Large-sample result
Detecting Non-Normality (In Data or in Residuals)

- Sample skewness and kurtosis, $\hat{S}$ and $\hat{K}$
- Jarque-Bera test. Under normality we have:

$$JB = \frac{T}{6} \left( \hat{S}^2 + \frac{1}{4}(\hat{K} - 3)^2 \right) \sim \chi^2_2$$

- Many more
Recall Our “Final” OLS Wage Regression

![EViews output](image)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.360535</td>
<td>0.131792</td>
<td>2.735636</td>
<td>0.0063</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.125028</td>
<td>0.009188</td>
<td>13.60791</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.086130</td>
<td>0.007016</td>
<td>9.424974</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER2</td>
<td>-0.000710</td>
<td>8.82E-05</td>
<td>-8.042035</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDU_EXP</td>
<td>-0.001905</td>
<td>0.000375</td>
<td>-5.078046</td>
<td>0.0000</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.239352</td>
<td>0.025327</td>
<td>-9.450447</td>
<td>0.0000</td>
</tr>
<tr>
<td>UNION</td>
<td>0.202574</td>
<td>0.034553</td>
<td>5.862629</td>
<td>0.0000</td>
</tr>
<tr>
<td>NONWHITE</td>
<td>-0.094903</td>
<td>0.034921</td>
<td>-2.717655</td>
<td>0.0067</td>
</tr>
</tbody>
</table>

R-squared: 0.342915, Mean dependent var: 2.341995
Adjusted R-squared: 0.339418, S.D. dependent var: 0.561435
S.E. of regression: 0.456313, Akaike info criterion: 1.274755
Sum squared resid: 273.8119, Schwarz criterion: 1.306124
Log likelihood: -835.2503, Hannan-Quinn criteria: 1.286514
F-statistic: 98.03775, Durbin-Watson stat: 1.894273
OLS Residual Histogram and Statistics

- **Sample:** 1,323 observations
- **Mean:** -5.82e-16
- **Median:** 0.003600
- **Maximum:** 1.766517
- **Minimum:** -1.888482
- **Std. Dev.:** 0.455104
- **Skewness:** -0.228689
- **Kurtosis:** 3.712251
- **Jarque-Bera:** 39.49685
- **Probability:** 0.000000
QQ Plots

- We introduced histograms earlier...

- ...but if interest centers on the tails of distributions, QQ plots often provide sharper insight as to the agreement or divergence between the actual and reference distributions.

- QQ plot is quantiles of the standardized data against quantiles of a standardized reference distribution (e.g., normal).

- If the distributions match, the QQ plot is the 45 degree line.

- To the extent that the QQ plot does not match the 45 degree line, the nature of the divergence can be very informative, as for example in indicating fat tails.
OLS Wage Regression Residual QQ Plot
Detecting Outliers: OLS Residual Plot
Detecting Outliers: OLS Residual Scatter

![Scatter plot](image.png)
Detecting Outliers: Leverage

Consider:

\[
\left( \hat{\beta}^{(-t)} - \hat{\beta} \right), \ t = 1, \ldots, T
\]

“Leave-one-out plot”

But it can be shown that

\[
\left( \hat{\beta}^{(-t)} - \hat{\beta} \right) = -\frac{1}{1 - h_t} (X'X)^{-1} x_t' e_t ,
\]

where “leverage” \( h_t \) is the \( t \)-th diagonal element of \( X(X'X)^{-1} X' \).

“Leverage plot”
Dealing with Outliers: Robustness Iteration

Fit at “robustness iteration 0”:

\[ \hat{y}^{(0)} = X \hat{\beta}^{(0)} \]

where

\[ \hat{\beta}^{(0)} = \arg \min_\beta \left[ \sum_{t=1}^{T} (y_t - x_t' \beta)^2 \right] \]
Robustness Iteration Continued

Fit at robustness iteration 1:

\[ \hat{y}^{(1)} = X \hat{\beta}^{(1)} \]

where:

\[ \hat{\beta}^{(1)} = \arg\min_\beta \left[ \sum_{t=1}^{T} \rho_t^{(1)} (y_t - x_t' \beta)^2 \right] \]

\[ \rho_t^{(1)} = S \left( \frac{e_t^{(0)}}{6 \text{med} |e_t^{(0)}|} \right) \]

\[ S(z) = 1 \text{ for } z \in [-1, 1] \]

\[ S \text{ downweights outside } [-1, 1] \]

\[ e_t^{(0)} = y_t - \hat{y}_t^{(0)} \]

Continue iterating as desired.

“Weighted least squares”
Dealing with Outliers: Least Absolute Deviations (LAD)

\[ \hat{\beta}_{LAD} = \text{argmin}_\beta \sum_{t=1}^{T} |y_t - x_t' \beta| \]

- You’ve seen it before as a special case of quantile regression
- Fits the conditional median function
- But conditional median is more robust than conditional mean
LAD Wage Regression Estimation

Dependent Variable: LWAGE
Method: Quantile Regression (Median)
Date: 10/21/13 Time: 15:17
Sample: 1 1323
Included observations: 1323
Huber Sandwich Standard Errors & Covariance
Sparsity method: Kernel (Epanechnikov) using residuals
Bandwidth method: Hall-Sheather, bw=0.088501
Estimation successfully identifies unique optimal solution

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.199229</td>
<td>0.127641</td>
<td>1.560855</td>
<td>0.1188</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.073226</td>
<td>0.008260</td>
<td>8.865366</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.137726</td>
<td>0.009575</td>
<td>14.38339</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER2</td>
<td>-0.000720</td>
<td>0.000112</td>
<td>-6.398026</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDU_EXP</td>
<td>-0.002458</td>
<td>0.000471</td>
<td>-5.222012</td>
<td>0.0000</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.227365</td>
<td>0.029209</td>
<td>-7.784041</td>
<td>0.0000</td>
</tr>
<tr>
<td>NONWHITE</td>
<td>-0.129765</td>
<td>0.038623</td>
<td>-3.359806</td>
<td>0.0008</td>
</tr>
<tr>
<td>UNION</td>
<td>0.216760</td>
<td>0.033508</td>
<td>6.468981</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Pseudo R-squared | 0.229337 |
Adjusted R-squared | 0.225235 |
Mean dependent var | 2.341995 |
S.D. dependent var | 0.581435 |
Stochastic Regressors I:

OLS Properties

Omitted Variables, Measurement Error, and Multicollinearity
From FIC to IC:
OLS Without “Fixed X”

– Sometimes logically impossible
– Doesn’t feel right for observational studies

Many earlier results still hold under a key sufficient condition
(plus some regularity conditions on $X$):
– $\hat{\beta}_{OLS}$ still consistent
– $\hat{\beta}_{OLS}$ still asymptotically normally distributed
– $\text{Cov}(\hat{\beta}_{OLS})$ still consistently estimated by $s^2(X'X)^{-1}$

KEY SUFFICIENT CONDITION:
$$E(X'\varepsilon) = 0$$

– Every element of $\varepsilon$ uncorrelated with every element of $X$
– Under the rest of the FIC, every element of $\varepsilon$
  independent of every element of $X$
We can relax normality and “fixed X” simultaneously

Earlier results still hold:
- $\hat{\beta}_{OLS}$ still consistent
- $\hat{\beta}_{OLS}$ still asymptotically normally distributed
- $Cov(\hat{\beta}_{OLS})$ still consistently estimated by $s^2(X'X)^{-1}$.

**KEY SUFFICIENT CONDITION:**
$X$ and $\varepsilon$ are independent

- Every element of $\varepsilon$
  independent of every element of $X$

Call the weaker conditions the “Ideal Conditions” (IC)
(FIC less normality and with $X$ stochastic but independent of $\varepsilon$)
Even Weaker Sufficient Condition
For Good Asymptotic Properties of OLS
Without “Fixed X” and Normality

\[
E(\varepsilon_t|x_t) = 0
\]

and

\[
\text{var}(\varepsilon_t|x_t) = \sigma^2
\]

(Cross-section case of iid \( x_t \))

\[
E(\varepsilon_t|x_t, x_{t-1}, \ldots) = 0
\]

and

\[
\text{var}(\varepsilon_t|x_t, x_{t-1}, \ldots) = \sigma^2
\]

(Time-series case of serially-correlated \( x_t \))
Omitted Variable Bias

Suppose \( \text{corr}(x, z) > 0 \)

DGP is:

\[
y_t = \beta_1 + \beta_2 x_t + \varepsilon_t
\]

But we incorrectly estimate:

\[
y_t = \delta_1 + \delta_2 z_t + \eta_t
\]

Clearly we’ll estimate a positive effect of \( z \) on \( y \), in large as well as small samples, even though it’s completely spurious and would vanish if \( x \) had been included in the regression.

The problem: \( E(\eta_t|z_t) \neq 0 \)

The bias is positive here because we assumed that \( \text{corr}(x_t, z_t) > 0 \).

– In general the sign of the bias could go either way.
T-Consistency and P-Consistency

T-Consistency: Consistency for a treatment effect
P-Consistency: Consistency for a predictive effect

\[ y_t = \beta_1 + \beta_2 x_t + \varepsilon_t \]

T-consistent under FIC
P-consistent always

This regression is causal (and of course predictive)
(Causality implies correlation, which is always useful for prediction)

\[ y_t = \delta_1 + \delta_2 z_t + \eta_t \]

T-consistent never (FIC fail)
P-consistent always

This regression predictive but not causal.
(Correlation is useful for prediction but does not imply causality)
Measurement Error

DGP:

\[ y_t = \beta_1 + \beta_2 x_t + \varepsilon_t \]

Measurement:

\[ x_t^m = x_t + \nu_t \]

We incorrectly run:

\[ y \rightarrow c, x^m \]

As \( \frac{\sigma_v^2}{\sigma_x^2} \) gets large, the regression is progressively less able to identify the true relationship. In the limit as \( \frac{\sigma_v^2}{\sigma_x^2} \rightarrow \infty \), it is impossible. In any event, \( \hat{\beta}_{LS} \) is biased toward zero, in small as well as large samples.
Collinearity and Multicollinearity

Collinearity: some $x$ equals another

Multicollinearity: some $x$ equals a linear combination of others

Either is a violation of the FIC requirement that: “$X$ is of full column rank”

And collinearity is a special case of multicollinearity, so henceforth we’ll speak only of multicollinearity
Perfect Multicollinearity

Perfect Multicollinearity: $X$ not of full column rank
(“Some $x$ a perfect linear combination of others”)
(“(X’X)^{-1}$ does not exist”)
– Impossible under FIC
– Generally not relevant unless you’ve done something silly.
  – Drop something!

More intuitive now, with stochastic regressors:
(“no $x$ perfectly correlated with a linear combination of others”)
– Generally not relevant unless you’ve done something silly.
  – Drop something!

So we don’t want to drop “no perfect multicollinearity” from FIC
– But we do want to consider the effects of highly-correlated,
  if not perfectly correlated, combinations of regressors
Imperfect Multicollinearity

Some $x$ highly-collinear (but not perfectly correlated) with a linear combination of others

- Large $F$ and $R^2$, yet small $t$’s (large s.e.’s).
  - Hard to parse effects of $x$’s on $y$,
    yet it’s clear that there is an overall relationship.

That’s just the way life is. Not really a “problem.”
That is, it doesn’t necessarily mean that you did something silly.

- OLS is natural: orthogonal projection.

But we might want to know more...
Multicollinearity and Variance Inflation

\[ \text{var}(\hat{\beta}_j) = f \left( \sigma^2, \sigma_{x_j}^2, R_j^2 \right) \]

where \( R_j^2 \) is regression of \( x_j \) on all other regressors

\[ R_j^2 \text{ affects } \text{var}(\hat{\beta}_j) \text{ as } (1 - R_j^2)^{-1} \]

Hence as \( R_j^2 \to 1 \) the variance inflation approaches infinity

(\( x_j \) completely redundant)
Indicator Variables in Cross Sections: Group Effects
A dummy variable, or indicator variable, is just a 0-1 variable that indicates something, such as whether a person is female:

\[ FEMALE_t = \begin{cases} 1 & \text{if person } t \text{ is female} \\ 0 & \text{otherwise} \end{cases} \]

(It really is that simple.)

“Intercept dummies”

Note that the sample mean of a dummy variable is the fraction of the sample with the indicated attribute.
Histograms for Wage Covariates

- Education
- Experience
- Nonwhite
- Union
Important Issues

▶ The intercept corresponds to the “base case” across all dummies (i.e., when all dummies are simultaneously 0), and the dummy coefficients give the extra effects (i.e., when the respective dummies are 1).

▶ Alternatively, use a full set of dummies for each category (e.g., both a union dummy and a non-union dummy) and drop the intercept. (More useful/common for in time-series situations)

▶ Never include a full set of dummies and an intercept. Would be totally redundant: “Perfect Multicollinearity”
Controlling for Sex, Race and Union Status in the Wage Regression

Before:

\[ LWAGE \rightarrow C, EDUC, EXPER \]
Wage Regression on Education and Experience

Dependent Variable: LWAGE
Method: Least Squares
Date: 06/27/13 Time: 16:38
Sample (adjusted): 11323
Included observations: 1323 after adjustments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.867382</td>
<td>0.075331</td>
<td>11.51431</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.093229</td>
<td>0.005045</td>
<td>18.48002</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.013104</td>
<td>0.001164</td>
<td>11.26208</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared | 0.232224 | Mean dependent var | 2.341995 |
Adjusted R-squared | 0.231061 | S.D. dependent var | 0.561435 |
S.E. of regression | 0.492318 | Akaike info criterion | 1.422881 |
Sum squared resid | 319.9376 | Schwarz criterion | 1.434644 |
Log likelihood | -938.2358 | Hannan-Quinn criterion | 1.427291 |
F-statistic | 199.6260 | Durbin-Watson stat | 1.926045 |
Prob(F-statistic) | 0.000000 |  |  |
Controlling for Sex, Race and Union Status in the Wage Regression

Now:

\[ \text{LWAGE} \rightarrow C, \text{EDUC}, \text{EXPER}, \text{FEMALE}, \text{NONWHITE}, \text{UNION} \]
Wage Regression on Education, Experience and Group Dummies

```
Dependent Variable: LWAGE
Method: Least Squares
Date: 07/03/13   Time: 13:36
Sample (adjusted): 1 1323
Included observations: 1323 after adjustments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1.000385</td>
<td>0.073180</td>
<td>13.67013</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.090809</td>
<td>0.004814</td>
<td>18.86314</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.012707</td>
<td>0.001119</td>
<td>11.35624</td>
<td>0.0000</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.237535</td>
<td>0.025965</td>
<td>-9.148397</td>
<td>0.0000</td>
</tr>
<tr>
<td>NONWHITE</td>
<td>-0.085286</td>
<td>0.035766</td>
<td>-2.383199</td>
<td>0.0173</td>
</tr>
<tr>
<td>UNION</td>
<td>0.223392</td>
<td>0.035307</td>
<td>6.327126</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared   | 0.307856    | Mean dependent var | 2.341995 |
Adjusted R-squared | 0.305229 | S.D. dependent var   | 0.561435 |
S.E. of regression | 0.467973 | Akaike info criterion | 1.323712 |
Sum squared resid  | 288.4212 | Schwarz criterion    | 1.347239 |
Log likelihood    | -869.6356 | Hannan-Quinn criter. | 1.332532 |
F-statistic       | 117.1568  | Durbin-Watson stat   | 1.910120 |
Prob(F-statistic) | 0.000000  |                     |          |
```
Residual Scatter from Wage Regression on Education, Experience and Group Dummies
Indicator Variables in Time Series: Trend and Seasonality
Liquor Sales
Log Liquor Sales

![Graph showing log liquor sales over time with a gradual increase and fluctuations. The x-axis represents time from 1988 to 2014, and the y-axis represents log sales ranging from 6 to 8.]
Linear Deterministic Trend

\[ Trend_t = \beta_1 + \beta_2 \text{TIME}_t \]

where \( \text{TIME}_t = t \)

Simply run the least squares regression \( y \rightarrow c, \text{TIME} \), where

\[
\text{TIME} = \begin{pmatrix}
1 \\
2 \\
3 \\
\vdots \\
T - 1 \\
T
\end{pmatrix}
\]
Various Linear Trends

TREND=10-0.25*TIME

TREND=-0.5+0.8*TIME
## Linear Trend Estimation

Method: Least Squares  
Date: 08/08/13  Time: 08:53  
Sample: 1987M01 2014M12  
Included observations: 336

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>6.454290</td>
<td>0.017468</td>
<td>369.4834</td>
<td>0.0000</td>
</tr>
<tr>
<td>TIME</td>
<td>0.003809</td>
<td>8.98E-05</td>
<td>42.39935</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.843318  
Mean dependent var 7.096188  
Adjusted R-squared 0.842849  
S.D. dependent var 0.402962  
Akaike info criterion -0.824561  
Schwarz criterion -0.801840  
Hannan-Quinn criter. -0.815504  
F-statistic 1797.705  
Durbin-Watson stat 1.078573  
Prob(F-statistic) 0.000000
Residual Plot
Deterministic Seasonality

\[ Seasonal_t = \sum_{i=1}^{s} \beta_i SEAS_{it} \quad (s \text{ seasons per year}) \]

where \( SEAS_{it} = \begin{cases} 
1 & \text{if observation } t \text{ falls in season } i \\
0 & \text{otherwise} 
\end{cases} \)

Simply run the least squares regression \( y \rightarrow SEAS_1, \ldots, SEAS_s \)
(or blend: \( y \rightarrow TIME, SEAS_1, \ldots, SEAS_s \))

where (e.g., in quarterly data case, assuming Q1 start and Q4 end):

\[
\begin{align*}
SEAS_1 &= (1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, \ldots, 0)' \\
SEAS_2 &= (0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, \ldots, 0)' \\
SEAS_3 &= (0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, \ldots, 0)' \\
SEAS_4 &= (0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, \ldots, 1)'.
\end{align*}
\]
Linear Trend with Seasonal Dummies

Dependent Variable: LSALES
Method: Least Squares
Date: 09/06/13  Time: 08:01
Sample: 1987M01 2014M12
Included observations: 336

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME</td>
<td>0.003779</td>
<td>6.24E-05</td>
<td>60.57536</td>
<td>0.0000</td>
</tr>
<tr>
<td>D1</td>
<td>6.361233</td>
<td>0.023283</td>
<td>273.2148</td>
<td>0.0000</td>
</tr>
<tr>
<td>D2</td>
<td>6.304412</td>
<td>0.023310</td>
<td>270.4571</td>
<td>0.0000</td>
</tr>
<tr>
<td>D3</td>
<td>6.391653</td>
<td>0.023338</td>
<td>273.8773</td>
<td>0.0000</td>
</tr>
<tr>
<td>D4</td>
<td>6.392737</td>
<td>0.023365</td>
<td>273.6304</td>
<td>0.0000</td>
</tr>
<tr>
<td>D5</td>
<td>6.461768</td>
<td>0.023393</td>
<td>276.2273</td>
<td>0.0000</td>
</tr>
<tr>
<td>D6</td>
<td>6.466819</td>
<td>0.023421</td>
<td>276.1145</td>
<td>0.0000</td>
</tr>
<tr>
<td>D7</td>
<td>6.510789</td>
<td>0.023449</td>
<td>277.6602</td>
<td>0.0000</td>
</tr>
<tr>
<td>D8</td>
<td>6.482457</td>
<td>0.023477</td>
<td>276.1210</td>
<td>0.0000</td>
</tr>
<tr>
<td>D9</td>
<td>6.422551</td>
<td>0.023505</td>
<td>273.2406</td>
<td>0.0000</td>
</tr>
<tr>
<td>D10</td>
<td>6.444589</td>
<td>0.023533</td>
<td>273.8476</td>
<td>0.0000</td>
</tr>
<tr>
<td>D11</td>
<td>6.476504</td>
<td>0.023562</td>
<td>274.8709</td>
<td>0.0000</td>
</tr>
<tr>
<td>D12</td>
<td>6.798519</td>
<td>0.023591</td>
<td>288.1874</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.927059  Mean dependent var 7.096188
Adjusted R-squared 0.924350  S.D. dependent var 0.402962
S.E. of regression 0.110833  Akaike info criterion -1.523658
Sum squared resid 3.967734  Schwarz criterion -1.375972
Log likelihood 268.9746  Hannan-Quinn criter. -1.464786
Durbin-Watson stat 0.100500

Path = c:\users\francis\diebo\documents\diebold\files\courses\econ104\old\econ104_2011\sw3e\views
Residual Plot
Seasonal Pattern
Nonlinearity in Cross Sections
## Anscombe's Quartet

### Data Table

<table>
<thead>
<tr>
<th>obs</th>
<th>Y1</th>
<th>X1</th>
<th>Y2</th>
<th>X2</th>
<th>Y3</th>
<th>X3</th>
<th>Y4</th>
<th>X4</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>5.680000</td>
<td>5.000000</td>
<td>4.740000</td>
<td>5.000000</td>
<td>5.730000</td>
<td>5.000000</td>
<td>6.890000</td>
<td>8.000000</td>
</tr>
</tbody>
</table>
Anscombe's Quartet: Regressions

<table>
<thead>
<tr>
<th>LS // Dependent Variable is Y1</th>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>T-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3.00</td>
<td>1.12</td>
<td>2.67</td>
<td></td>
</tr>
<tr>
<td>X1</td>
<td>0.50</td>
<td>0.12</td>
<td>4.24</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.67</td>
<td>S.E. of regression</td>
<td>1.24</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LS // Dependent Variable is Y2</th>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>T-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3.00</td>
<td>1.12</td>
<td>2.67</td>
<td></td>
</tr>
<tr>
<td>X2</td>
<td>0.50</td>
<td>0.12</td>
<td>4.24</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.67</td>
<td>S.E. of regression</td>
<td>1.24</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LS // Dependent Variable is Y3</th>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>T-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3.00</td>
<td>1.12</td>
<td>2.67</td>
<td></td>
</tr>
<tr>
<td>X3</td>
<td>0.50</td>
<td>0.12</td>
<td>4.24</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.67</td>
<td>S.E. of regression</td>
<td>1.24</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LS // Dependent Variable is Y4</th>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>T-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3.00</td>
<td>1.12</td>
<td>2.67</td>
<td></td>
</tr>
<tr>
<td>X4</td>
<td>0.50</td>
<td>0.12</td>
<td>4.24</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.67</td>
<td>S.E. of regression</td>
<td>1.24</td>
<td></td>
</tr>
</tbody>
</table>
Anscombe’s Quartet: Graphics
Parametric and Nonparametric Nonlinearity...

...and the gray area in between.
Log-Log Regression

\[ \ln y_t = \beta_1 + \beta_2 \ln x_t + \varepsilon_t \]

Example: Cobb-Douglas production function

\[ y_t = AL_t^\alpha K_t^\beta \exp(\varepsilon_t) \]

\[ \ln y_t = \ln A + \alpha \ln L_t + \beta \ln K_t + \varepsilon_t \]

For close \( y_t \) and \( x_t \), \((\ln y_t - \ln x_t)\) is approximately the percent difference between \( y_t \) and \( x_t \). Hence the coefficients in log-log regressions give the expected percent change in \( E(y_t|x_t) \) for a one-percent change in \( x_t \), the elasticity of \( y_t \) with respect to \( x_t \).
**Log-Lin Regression**

\[ \ln y_t = \beta x_t + \varepsilon \]

Example: Exponential growth

\[ y_t = Ae^{rt} \]

\[ \ln y_t = \ln A + rt \]

The growth rate \( r \) gives the approximate percent change in \( E(y_t|t) \) for a one-unit change in time

Example: LWAGE regression!
Intrinsically Non-Linear Models

One example is the “S-curve” model,

\[ y = \frac{1}{a + br^x} \]

\( (0 < r < 1) \)

– No way to transform to linearity

– Use non-linear least squares (NLS)

– Under the remaining FIC (that is, dropping only linearity), \( \hat{\beta}_{NLS} \) has a sampling distribution similar to that of \( \hat{\beta}_{LS} \) under the FIC
Really no such thing as an intrinsically non-linear model...

In the bivariate case we can think of the relationship as

\[ y_t = g(x_t, \varepsilon_t) \]

or slightly less generally as

\[ y_t = f(x_t) + \varepsilon_t \]
Consider Taylor series expansions of $f(x_t)$. The linear (first-order) approximation is

$$f(x_t) \approx \beta_1 + \beta_2 x,$$

and the quadratic (second-order) approximation is

$$f(x_t) \approx \beta_1 + \beta_2 x_t + \beta_3 x_t^2.$$

In the multiple regression case, Taylor approximations also involve interaction terms. Consider, for example, $f(x_t, z_t)$:

$$f(x_t, z_t) \approx \beta_1 + \beta_2 x_t + \beta_3 z_t + \beta_4 x_t^2 + \beta_5 z_t^2 + \beta_6 x_t z_t + \ldots$$

– Equally relevant for dummy variables
A Key Insight

The ultimate point is that so-called “intrinsically non-linear” models are themselves linear when viewed from the series-expansion perspective. In principle, of course, an infinite number of series terms are required, but in practice nonlinearity is often quite gentle (e.g., quadratic) so that only a few series terms are required.

– So non-linearity is in some sense really an omitted-variables problem
Testing for Non-Linearity: \( t \) and \( F \) Tests, \( AIC \) and \( SIC \)

Just test for omitted series expansion terms using \( t \)'s and \( F \).

Use \( AIC \) and \( SIC \) as always.
Basic Wage Regression

Dependent Variable: LWAGE
Method: Least Squares
Date: 07/03/13  Time: 13:36
Sample (adjusted): 11323
Included observations: 1323 after adjustments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1.000385</td>
<td>0.073180</td>
<td>13.67013</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.090809</td>
<td>0.004814</td>
<td>18.86314</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.012707</td>
<td>0.001119</td>
<td>11.35624</td>
<td>0.0000</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.237535</td>
<td>0.025965</td>
<td>-9.148397</td>
<td>0.0000</td>
</tr>
<tr>
<td>NONWHITE</td>
<td>-0.085286</td>
<td>0.035786</td>
<td>-2.383199</td>
<td>0.0173</td>
</tr>
<tr>
<td>UNION</td>
<td>0.223392</td>
<td>0.035307</td>
<td>6.327126</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared          | 0.307856 | Mean dependent var | 2.341995 |
Adjusted R-squared | 0.305229 | S.D. dependent var | 0.561435 |
S.E. of regression | 0.467973 | Akaike info criterion | 1.323712 |
Sum squared resid  | 288.4212 | Schwarz criterion   | 1.347239 |
Log likelihood     | -869.6356| Hannan-Quinn criter. | 1.332532 |
F-statistic        | 117.1568 | Durbin-Watson stat  | 1.910120 |
Quadratic Wage Regression

Dependent Variable: LWAGE
Method: Least Squares
Date: 10/02/13   Time: 12:37
Sample: 1 1323
Included observations: 1323

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.473236</td>
<td>0.240586</td>
<td>1.967017</td>
<td>0.0494</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.109673</td>
<td>0.028918</td>
<td>3.792608</td>
<td>0.0002</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.064422</td>
<td>0.007652</td>
<td>8.419060</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDUC2</td>
<td>0.000501</td>
<td>0.000895</td>
<td>0.559994</td>
<td>0.5756</td>
</tr>
<tr>
<td>EXPER2</td>
<td>-0.000705</td>
<td>8.86E-05</td>
<td>-7.962263</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDU_EXP</td>
<td>-0.001789</td>
<td>0.000429</td>
<td>-4.173423</td>
<td>0.0000</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.237696</td>
<td>0.025506</td>
<td>-9.319335</td>
<td>0.0000</td>
</tr>
<tr>
<td>UNION</td>
<td>0.202955</td>
<td>0.034569</td>
<td>5.870998</td>
<td>0.0000</td>
</tr>
<tr>
<td>NONWHITE</td>
<td>-0.095028</td>
<td>0.034931</td>
<td>-2.720476</td>
<td>0.0066</td>
</tr>
</tbody>
</table>

R-squared    | 0.343072    | Mean dependent var | 2.341995 |
Adjusted R-squared | 0.339073 | S.D. dependent var   | 0.561435 |
S.E. of regression    | 0.456433    | Akaike info criterion   | 1.276028 |
Sum squared resid     | 273.7465    | Schwarz criterion      | 1.311318 |
Log likelihood        | -835.0925   | Hannan-Quinn criter.   | 1.289257 |
F-statistic           | 85.77745    | Durbin-Watson stat     | 1.894409 |
Dummy Interactions?

Figure: Wage Regression on Education, Experience, Group Dummies, and Interactions

Dependent Variable: LWAGE
Method: Least Squares
Date: 10/02/13  Time: 12:48
Sample: 1 1323
Included observations: 1323

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1.011503</td>
<td>0.073797</td>
<td>13.70657</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.090805</td>
<td>0.004819</td>
<td>18.84231</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.012674</td>
<td>0.001120</td>
<td>11.31689</td>
<td>0.0000</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.257621</td>
<td>0.030125</td>
<td>-8.551814</td>
<td>0.0000</td>
</tr>
<tr>
<td>UNION</td>
<td>0.216911</td>
<td>0.047498</td>
<td>4.566771</td>
<td>0.0000</td>
</tr>
<tr>
<td>NONWHITE</td>
<td>-0.157606</td>
<td>0.056446</td>
<td>-2.792173</td>
<td>0.0053</td>
</tr>
<tr>
<td>FEM_UNI</td>
<td>0.003921</td>
<td>0.072024</td>
<td>0.054439</td>
<td>0.9566</td>
</tr>
<tr>
<td>FEM_NON</td>
<td>0.125567</td>
<td>0.071781</td>
<td>1.749313</td>
<td>0.0805</td>
</tr>
<tr>
<td>UNI_NON</td>
<td>0.017743</td>
<td>0.091336</td>
<td>0.194258</td>
<td>0.8460</td>
</tr>
</tbody>
</table>

R-squared  0.309487  Mean dependent var  2.341995
Adjusted R-squared  0.305283  S.D. dependent var  0.561435
S.E. of regression  0.467955  Akaike info criterion  1.325889
Sum squared resid  287.7418  Schwarz criterion  1.361179
Log likelihood  -868.0755  Hannan-Quinn criterion  1.339118
Figure: Wage Regression with Continuous Non-Linearities and Interactions, and Discrete Interactions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.482967</td>
<td>0.240926</td>
<td>2.004623</td>
<td>0.0452</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.109522</td>
<td>0.029003</td>
<td>3.776211</td>
<td>0.0002</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.064269</td>
<td>0.007654</td>
<td>8.396570</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDUC2</td>
<td>0.000517</td>
<td>0.000900</td>
<td>0.573929</td>
<td>0.5661</td>
</tr>
<tr>
<td>EXPER2</td>
<td>-0.000701</td>
<td>8.87E-05</td>
<td>-7.904460</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDU_EXP</td>
<td>-0.001796</td>
<td>0.000429</td>
<td>-4.186578</td>
<td>0.0000</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.252921</td>
<td>0.029659</td>
<td>-8.527539</td>
<td>0.0000</td>
</tr>
<tr>
<td>UNION</td>
<td>0.200937</td>
<td>0.046575</td>
<td>4.314297</td>
<td>0.0000</td>
</tr>
<tr>
<td>NONWHITE</td>
<td>-0.161501</td>
<td>0.050777</td>
<td>-3.232246</td>
<td>0.0034</td>
</tr>
<tr>
<td>FEM_UNI</td>
<td>-0.012956</td>
<td>0.070740</td>
<td>-0.183153</td>
<td>0.8547</td>
</tr>
<tr>
<td>FEM_NON</td>
<td>0.110319</td>
<td>0.070939</td>
<td>1.573909</td>
<td>0.1157</td>
</tr>
<tr>
<td>UNI_NON</td>
<td>0.033202</td>
<td>0.069258</td>
<td>0.371975</td>
<td>0.7100</td>
</tr>
</tbody>
</table>

R-squared: 0.344357
Adjusted R-squared: 0.338856
S.E. of regression: 0.456507
Sum squared resid: 274.2109
Schwarz criterion: 1.325658
Hannan-Quinn criterion: 1.295244
Durbin-Watson stat: 1.891544

Dependent Variable: WAGE
So Drop Dummy Interactions and Tighten the Rest

Figure: "Final" Wage Regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.360535</td>
<td>0.131792</td>
<td>2.735636</td>
<td>0.0063</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.125028</td>
<td>0.009188</td>
<td>13.60791</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.086130</td>
<td>0.007016</td>
<td>9.424974</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER2</td>
<td>-0.000710</td>
<td>8.82E-05</td>
<td>-8.042035</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDU_EXP</td>
<td>-0.001905</td>
<td>0.000375</td>
<td>-5.078046</td>
<td>0.0000</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.239352</td>
<td>0.025327</td>
<td>-9.450447</td>
<td>0.0000</td>
</tr>
<tr>
<td>UNION</td>
<td>0.202574</td>
<td>0.034553</td>
<td>5.862629</td>
<td>0.0000</td>
</tr>
<tr>
<td>NONWHITE</td>
<td>-0.094903</td>
<td>0.034921</td>
<td>-2.717655</td>
<td>0.0067</td>
</tr>
</tbody>
</table>

R-squared | 0.342915 | Mean dependent var | 2.341995
Adjusted R-squared | 0.339418 | S.D. dependent var | 0.561435
S.E. of regression | 0.456313 | Akaike info criterion | 1.274755
Sum squared resid | 273.8119 | Schwarz criterion | 1.306124
Log likelihood | -835.2503 | Hannan-Quinn criterion | 1.286514
F-statistic | 98.03775 | Durbin-Watson stat | 1.894273
Prob(F-statistic) | 0.000000 |
Nonlinearity in Time Series
Non-Linear Trend: Exponential (Log-Linear)

\[ Trend_t = \beta_1 e^{\beta_2 \text{TIME}_t} \]

\[ \ln(Trend_t) = \ln(\beta_1) + \beta_2 \text{TIME}_t \]
Figure: Various Exponential Trends
Non-Linear Trend: Quadratic

Allow for gentle curvature by including $TIME$ and $TIME^2$:

$$Trend_t = \beta_1 + \beta_2 TIME_t + \beta_3 TIME^2_t$$
**Figure:** Various Quadratic Trends
Recall Log-Linear Liquor Sales Trend Estimation

Dependent Variable: LSALES
Method: Least Squares
Date: 08/08/13   Time: 08:53
Sample: 1987M01 2014M12
Included observations: 336

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>6.454290</td>
<td>0.017468</td>
<td>369.4834</td>
<td>0.0000</td>
</tr>
<tr>
<td>TIME</td>
<td>0.003809</td>
<td>8.98E-05</td>
<td>42.39935</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.843318     Mean dependent var 7.096188
Adjusted R-squared 0.842849     S.D. dependent var 0.402962
S.E. of regression 0.159743     Akaike info criterion -0.824561
Sum squared resid 8.523001     Schwarz criterion -0.801840
Log likelihood 140.5262     Hannan-Quinn criter. -0.815504
F-statistic 1797.705     Durbin-Watson stat 1.078573
Prob(F-statistic) 0.000000
Residual Plot
### Log-Quadratic Liquor Sales Trend Estimation

- **Dependent Variable:** LSALES
- **Method:** Least Squares
- **Date:** 08/08/13  **Time:** 08:53
- **Sample:** 1987M01 2014M12
- **Included observations:** 336

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>6.231269</td>
<td>0.020653</td>
<td>301.7187</td>
<td>0.0000</td>
</tr>
<tr>
<td>TIME</td>
<td>0.007768</td>
<td>0.000283</td>
<td>27.44987</td>
<td>0.0000</td>
</tr>
<tr>
<td>TIME2</td>
<td>-1.17E-05</td>
<td>8.13E-07</td>
<td>-14.44511</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

- **R-squared:** 0.903676
- **Mean dependent var:** 7.096188
- **Adjusted R-squared:** 0.903097
- **S.D. dependent var:** 0.402962
- **Akaike info criterion:** -1.305106
- **Schwarz criterion:** -1.271025
- **Log likelihood:** 222.2579
- **Durbin-Watson stat:** 1.754412
- **Prob(F-statistic):** 0.000000
Residual Plot
Log-Quadratic Liquor Sales Trend Estimation with Seasonal Dummies

Dependent Variable: LSALES
Method: Least Squares
Date: 08/08/13   Time: 08:53
Sample: 1987M01 2014M12
Included observations: 336

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME</td>
<td>0.007739</td>
<td>0.000104</td>
<td>74.49828</td>
<td>0.0000</td>
</tr>
<tr>
<td>TIME2</td>
<td>-1.18E-05</td>
<td>2.98E-07</td>
<td>-39.36756</td>
<td>0.0000</td>
</tr>
<tr>
<td>D1</td>
<td>6.138362</td>
<td>0.011207</td>
<td>547.7315</td>
<td>0.0000</td>
</tr>
<tr>
<td>D2</td>
<td>6.081424</td>
<td>0.011218</td>
<td>542.1044</td>
<td>0.0000</td>
</tr>
<tr>
<td>D3</td>
<td>6.168571</td>
<td>0.011229</td>
<td>549.3318</td>
<td>0.0000</td>
</tr>
<tr>
<td>D4</td>
<td>6.169584</td>
<td>0.011240</td>
<td>548.9444</td>
<td>0.0000</td>
</tr>
<tr>
<td>D5</td>
<td>6.238568</td>
<td>0.011251</td>
<td>554.5117</td>
<td>0.0000</td>
</tr>
<tr>
<td>D6</td>
<td>6.243596</td>
<td>0.011261</td>
<td>554.4513</td>
<td>0.0000</td>
</tr>
<tr>
<td>D7</td>
<td>6.287566</td>
<td>0.011271</td>
<td>557.8584</td>
<td>0.0000</td>
</tr>
<tr>
<td>D8</td>
<td>6.259257</td>
<td>0.011281</td>
<td>554.8647</td>
<td>0.0000</td>
</tr>
<tr>
<td>D9</td>
<td>6.199399</td>
<td>0.011290</td>
<td>549.0938</td>
<td>0.0000</td>
</tr>
<tr>
<td>D10</td>
<td>6.221507</td>
<td>0.011300</td>
<td>550.5987</td>
<td>0.0000</td>
</tr>
<tr>
<td>D11</td>
<td>6.253515</td>
<td>0.011309</td>
<td>552.9885</td>
<td>0.0000</td>
</tr>
<tr>
<td>D12</td>
<td>6.575648</td>
<td>0.011317</td>
<td>581.0220</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared          0.987452  Mean dependent var  7.096188
Adjusted R-squared 0.986946  S.D. dependent var  0.402962
S.E. of regression  0.046041  Akaike info criterion -3.277812
Sum squared resid   0.682555  Schwarz criterion    -3.118766
Log likelihood      564.6725  Hannan-Quinn criter. -3.214412
Durbin-Watson stat  0.581383
Residual Plot
Structural Change: Gradual

\[ y_t = \beta_1 + \beta_2 t x_t + \varepsilon_t \]

where

\[ \beta_1 t = \gamma_1 + \gamma_2 TIME_t \]

\[ \beta_2 t = \delta_1 + \delta_2 TIME_t \]

Then we have:

\[ y_t = (\gamma_1 + \gamma_2 TIME_t) + (\delta_1 + \delta_2 TIME_t)x_t + \varepsilon_t \]

We simply run:

\[ y_t \rightarrow c, , Time_t, x_t, TIME_t \ast x_t \]

Use regression to test for structural change (F test of \( \gamma_2 = \delta_2 = 0 \)).

Use regression to accommodate structural change if present.
Structural Change: Sharp

\[ y_t = \begin{cases} 
\beta_1^1 + \beta_1^2 x_t + \varepsilon_t, & t = 1, \ldots, T^* \\
\beta_2^1 + \beta_2^2 x_t + \varepsilon_t, & t = T^* + 1, \ldots, T 
\end{cases} \]

Let

\[ D_t = \begin{cases} 
0, & t = 1, \ldots, T^* \\
D_t = 1, & t = T^* + 1, \ldots, T 
\end{cases} \]

Then we can write the model as:

\[ y_t = (\beta_1^1 + (\beta_1^2 - \beta_1^1)D_t) + (\beta_2^1 + (\beta_2^2 - \beta_2^1)D_t)x_t + \varepsilon_t \]

We run:

\[ y_t \rightarrow c, \ D_t, \ x_t, \ D_t \cdot x_t \]

Use regression to test for structural change \((F \ test)\)

Use regression to accommodate structural change if present.
The “Chow test” is what we’re really calculating:

$$Chow = \frac{(e' e - (e'_1 e_1 + e'_2 e_2))/K}{(e'_1 e_1 + e'_2 e_2)/(T - 2K)}$$

Distributed $F$ under the no-break null (and the rest of the FIC)
Structural Change: Sharp Endogenous

\[ \text{MaxChow} = \max_{\tau_{\text{min}} \leq \tau \leq \tau_{\text{max}}} \text{Chow}(\tau), \]

where \( \tau \) denotes potential break location as fraction of sample

(Typically we take \( \tau_{\text{min}} = .15 \) and \( \tau_{\text{max}} = .85 \))

The null distribution of \( \text{MaxChow} \) has been tabulated.
Heteroskedasticity in Cross-Section Regression
Generalized Least Squares (GLS)

Consider the FIC except that we now let:

$$\varepsilon \sim N(0, \sigma^2 \Omega)$$

The old case is $\Omega = I$, but things are very different when $\Omega \neq I$:

– OLS parameter estimates consistent but inefficient (no longer MVUE or BLUE)

– OLS standard errors biased and inconsistent. Hence $t$ ratios do not have the $t$ distribution in finite samples and do not have the $N(0, 1)$ distribution asymptotically

The GLS estimator is:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

Under the remaining full ideal conditions GLS is MVUE with:

$$\hat{\beta}_{GLS} \sim N(\beta, \sigma^2(X'\Omega^{-1}X)^{-1})$$.
Heteroskedasticity

Homoskedasticity: variance of $\varepsilon_i$ is constant across $i$

Heteroskedasticity: variance of $\varepsilon_i$ is not constant across $i$

Relevant cross-sectional heteroskedasticity situation (on which we focus for now):
$\varepsilon_i$ independent across $i$ but not identically distributed across $i$

$$\Omega = \begin{pmatrix}
\sigma_1^2 & 0 & \ldots & 0 \\
0 & \sigma_2^2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \sigma_N^2
\end{pmatrix}$$

- Can arise for many reasons
- Engel curve (e.g., food expenditure vs. income) is classic example
Consequences

Same as for any violation of “$\sigma^2 I$”:

- OLS consistent but inefficient (no longer MVUE or BLUE), in finite samples and asymptotically
- Standard errors biased and inconsistent. Hence $t$ ratios do not have the $t$ distribution in finite samples and do not have the $N(0, 1)$ distribution asymptotically
Detection

- Graphical heteroskedasticity diagnostics
- Formal heteroskedasticity tests
Graphical Diagnostics

Graph $e_i^2$ against $x_i$, for various regressors

Problem: Purely pairwise
Recall Our “Final” Wage Regression
Squared Residual vs. EDUC
The Breusch-Godfrey-Pagan Test (BGP)

- Estimate the OLS regression, and obtain the squared residuals
- Regress the squared residuals on all regressors
- To test the null hypothesis of no relationship, examine $NR^2$ from this regression. In large samples $NR^2 \sim \chi^2$ under the null.
BPG Test

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic | 5.414870 | Prob. F(7, 1315) | 0.0000
Obs*R-squared | 37.06628 | Prob. Chi-Square(7) | 0.0000
Scaled explained SS | 49.66045 | Prob. Chi-Square(7) | 0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 10/30/13  Time: 10:54
Sample: 1 1323
Included observations: 1323

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-0.170309</td>
<td>0.097349</td>
<td>-1.749473</td>
<td>0.0804</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.024074</td>
<td>0.006787</td>
<td>3.547204</td>
<td>0.0004</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.011701</td>
<td>0.005183</td>
<td>2.257616</td>
<td>0.0241</td>
</tr>
<tr>
<td>EXPER2</td>
<td>-5.53E-05</td>
<td>6.52E-05</td>
<td>-0.849150</td>
<td>0.3960</td>
</tr>
<tr>
<td>EDU_EXP</td>
<td>-0.000478</td>
<td>0.0000277</td>
<td>-1.725513</td>
<td>0.0847</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.009757</td>
<td>0.018708</td>
<td>-0.521530</td>
<td>0.6021</td>
</tr>
<tr>
<td>UNION</td>
<td>-0.079648</td>
<td>0.025523</td>
<td>-3.120623</td>
<td>0.0018</td>
</tr>
<tr>
<td>NONWHITE</td>
<td>0.000486</td>
<td>0.025794</td>
<td>0.018829</td>
<td>0.9850</td>
</tr>
</tbody>
</table>
White’s Test

- Estimate the OLS regression, and obtain the squared residuals

- Regress the squared residuals on all regressors, squared regressors, and pairwise regressor cross products

- To test the null hypothesis of no relationship, examine $NR^2$ from this regression. In large samples $NR^2 \sim \chi^2$ under the null.

(White’s test is a natural and flexible generalization of the Breusch-Pagan-Godfrey test)
Is l wage c educ exper exper2 edu_exp female union nonwhite

Heteroskedasticity Test: White

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-statistic</td>
<td>2.431488</td>
<td>0.0000</td>
</tr>
<tr>
<td>Obs*R-squared</td>
<td>68.41804</td>
<td>0.0000</td>
</tr>
<tr>
<td>Scaled explained SS</td>
<td>91.66473</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
GLS for Heteroskedasticity

▶ “Weighted least squares” (WLS)

– Take a stand on the DGP. Get consistent standard errors and efficient parameter estimates.
(Infeasible) Weighted Least Squares

DGP:

\[ y_i = x_i' \beta + \varepsilon_i \]

\[ \varepsilon_i \sim idN(0, \sigma_i^2) \]

Weight the data \((y_i, x_i)\) by \(1/\sigma_i\):

\[ \frac{y_i}{\sigma_i} = \frac{x_i' \beta}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i} \]

The DGP is now:

\[ y_i^* = x_i^* ' \beta + \varepsilon_i^* \]

\[ \varepsilon_i^* \sim iidN(0, 1) \]

- OLS is MVUE!
- Problem: We don't know \(\sigma_i^2\)
Remark on Weighted Least Squares

Weighting the data by $1/\sigma_i$ is the same as weighting the residuals by $1/\sigma_i^2$:

$$
\min_{\beta} \sum_{i=1}^{N} \left( \frac{y_i - x_i' \beta}{\sigma_i} \right)^2 = \min_{\beta} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (y_i - x_i' \beta)^2
$$
Feasible Weighted Least Squares

Intuition: Replace the unknown $\sigma_i^2$ values with estimates

Some good ideas:

- Use $w_i = 1/e_i^2$, where $e_i^2$ are from the BGP test regression
- Use $w_i = 1/\hat{e}_i^2$, where $\hat{e}_i^2$ are from the White test regression

What about WLS directly using $w_i = 1/e_i^2$?

- Not such a good idea
- $e_i^2$ too noisy; we’d like to use not $e_i^2$ but rather $E(e_i^2|x_i)$. So we use an estimate of $E(e_i^2|x_i)$, namely $\hat{e}_i^2$ from $e^2 \to X$
Regression Weighted by Fit From White Test Regression

Dependent Variable: LWAGE
Method: Least Squares
Date: 10/30/13  Time: 12:32
Sample: 1 1323
Included observations: 1323
Weighting series: RESID2FIT
Weight type: Variance (average scaling)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.406917</td>
<td>0.116809</td>
<td>3.483612</td>
<td>0.0005</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.122743</td>
<td>0.008715</td>
<td>14.08407</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.062997</td>
<td>0.006428</td>
<td>9.799839</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER2</td>
<td>-0.000659</td>
<td>8.42E-05</td>
<td>-7.829347</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDU_EXP</td>
<td>-0.001870</td>
<td>0.000374</td>
<td>-5.001801</td>
<td>0.0000</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.229424</td>
<td>0.024288</td>
<td>-9.445895</td>
<td>0.0000</td>
</tr>
<tr>
<td>UNION</td>
<td>0.234806</td>
<td>0.029741</td>
<td>7.895118</td>
<td>0.0000</td>
</tr>
<tr>
<td>NONWHITE</td>
<td>-0.100705</td>
<td>0.032644</td>
<td>-3.084963</td>
<td>0.0021</td>
</tr>
</tbody>
</table>

Weighted Statistics

R-squared          | 0.388984 | Mean dependent var | 2.274204
Adjusted R-squared | 0.385731 | S.D. dependent var | 0.562167
A Different Approach
(Advanced but Very Important)
White’s Heteroskedasticity-Consistent Standard Errors

Perhaps surprisingly, we make direct use of $e_i^2$

Don’t take a stand on the DGP
Give up on efficient parameter estimates, but get consistent s.e.’s.

Using advanced methods, one can obtain consistent s.e.’s (if not an efficient $\hat{\beta}$) using only $e_i^2$

- Standard errors are rendered consistent.
- $\hat{\beta}$ remains unchanged at its OLS value. (Is that a problem?)

“Robustness to heteroskedasticity of unknown form”
Regression with White’s Heteroskedasticity-Consistent Standard Errors

Dependent Variable: LWAGE
Method: Least Squares
Date: 10/30/13  Time: 12:42
Sample: 1 1323
Included observations: 1323

White heteroskedasticity-consistent standard errors & covariance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.360535</td>
<td>0.130391</td>
<td>2.765026</td>
<td>0.0058</td>
</tr>
<tr>
<td>EDUC</td>
<td>0.125028</td>
<td>0.009890</td>
<td>12.64118</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER</td>
<td>0.066130</td>
<td>0.006967</td>
<td>9.491284</td>
<td>0.0000</td>
</tr>
<tr>
<td>EXPER2</td>
<td>-0.000710</td>
<td>8.86E-05</td>
<td>-8.004870</td>
<td>0.0000</td>
</tr>
<tr>
<td>EDU_EXP</td>
<td>-0.001905</td>
<td>0.000412</td>
<td>-4.623006</td>
<td>0.0000</td>
</tr>
<tr>
<td>FEMALE</td>
<td>-0.239352</td>
<td>0.025499</td>
<td>-9.386559</td>
<td>0.0000</td>
</tr>
<tr>
<td>UNION</td>
<td>0.202574</td>
<td>0.031386</td>
<td>6.454196</td>
<td>0.0000</td>
</tr>
<tr>
<td>NONWHITE</td>
<td>-0.094903</td>
<td>0.034074</td>
<td>-2.785164</td>
<td>0.0054</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.342915</td>
<td>Mean dep. var</td>
<td>2.341995</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.339418</td>
<td>S.D. dep. var</td>
<td>0.561435</td>
<td></td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>0.456313</td>
<td>Akaike info criterion</td>
<td>1.274755</td>
<td></td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>273.8119</td>
<td>Schwarz criterion</td>
<td>1.306124</td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-835.2503</td>
<td>Hannan-Quinn criterion</td>
<td>1.288514</td>
<td></td>
</tr>
<tr>
<td>F-statistic</td>
<td>98.03775</td>
<td>Durbin-Watson stat</td>
<td>1.894273</td>
<td></td>
</tr>
</tbody>
</table>
Serial Correlation in Time-Series Regression
A Tiny Bit of Time-Series Theory:
White Noise and \( AR(1) \) Processes

White noise: \( y_t \sim WN(\mu, \sigma^2) \) (serially uncorrelated)

Zero-mean white noise: \( y_t \sim WN(0, \sigma^2) \)

Independent (strong) white noise: \( y_t \sim iid (0, \sigma^2) \)

Gaussian white noise: \( y_t \sim iid N(0, \sigma^2) \)
Autocovariance, Autocorrelation and Partial Autocorrelation Functions

Population autocovariances:
\[ \gamma_y(\tau) = \text{cov}(y_t, y_{t-\tau}), \ \tau = 0, 1, 2, \ldots \]

Population autocorrelations:
\[ \rho_y(\tau) = \frac{\gamma_y(\tau)}{\gamma_y(0)} = \text{corr}(y_t, y_{t-\tau}), \ \tau = 0, 1, 2, \ldots \]

Population partial autocorrelations:
\[ p_y(\tau) \text{ is the coefficient on } y_{t-\tau} \text{ in the projection } \]
\[ y_t \rightarrow c, y_{t-1}, \ldots, y_{t-(\tau-1)}, y_{t-\tau}, \ \tau = 0, 1, 2, \ldots \]
Moment Structure of Strong White Noise

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

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

where \( \Omega_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\} \)

\[
\gamma(\tau) = \begin{cases} 
\sigma^2, & \tau = 0 \\
0, & \tau \geq 1 
\end{cases}
\]

\[
\rho(\tau) = \begin{cases} 
1, & \tau = 0 \\
0, & \tau \geq 1 
\end{cases}
\]

\[
p(\tau) = \begin{cases} 
1, & \tau = 0 \\
0, & \tau \geq 1 
\end{cases}
\]
Population Autocorrelation Function
White Noise Process
Zero-Mean AR(1)

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

\[ \varepsilon_t \sim iidN(0, \sigma^2), \ |\phi| < 1 \]

– Regression on just a lagged dependent variable

– "Autoregression"

Back-substitution reveals that:

\[ y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \]

\[ \implies E(y_t) = 0 \]
Realizations of Zero-Mean Two AR(1) Processes

\[ \phi = 0.4 \quad \text{---} \quad \phi = 0.95 \]
Moment Structure of the Zero-Mean AR(1) Process

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

\[ E(y_t) = 0 \text{ (of course)} \]

\[ \text{var}(y_t) = \frac{\sigma^2}{1 - \phi^2} \text{ (hmmm...)} \]

\[ E(y_t|\Omega_{t-1}) = \phi y_{t-1} \text{ (obvious)} \]

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

\[ \rho(\tau) = \begin{cases} 
1, & \tau = 0 \\
\phi^\tau, & \tau \geq 1 
\end{cases} \text{ (hmmm...)} \]

\[ p(\tau) = \begin{cases} 
1, & \tau = 0 \\
\phi, & \tau = 1 \\
0, & \tau \geq 2 
\end{cases} \text{ (obvious)} \]
Population Partial Autocorrelation Function
AR(1) Process, $\phi=.4$
Population Partial Autocorrelation Function
AR(1) Process, $\varphi=.95$

Displacement
AR(1) Autocorrelation Function

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

\[ \implies y_t y_{t-\tau} = \phi y_{t-1} y_{t-\tau} + \varepsilon_t y_{t-\tau} \quad (1) \]

First consider \( \tau = 0 \). Immediately:

\[ \gamma(0) = \text{var}(y_t) = \frac{\sigma^2}{1 - \phi^2} \]

Now consider \( \tau > 0 \). Taking expectations of (1) produces:

\[ \gamma(\tau) = \phi \gamma(\tau - 1) \]

Hence \( \gamma(\tau) = \phi^\tau \frac{\sigma^2}{1 - \phi^2} \), so \( \rho(\tau) = \phi^\tau \), \( \tau = 0, 1, 2, ... \)
Population Autocorrelation Function
AR(1) Process, $\varphi=.4$
Population Autocorrelation Function
AR(1) Process, $\phi = .95$
\(\gamma(\tau), \rho(\tau), \text{ and } p(\tau)\) for Generic AR\((p)\)

\(\text{AR}(p)\text{Process :}\)

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

\(\gamma(\tau) \to 0 \text{ as } \tau \to \infty, \text{ gradually}\)

\(\rho(\tau) \to 0 \text{ as } \tau \to \infty, \text{ gradually}\)

\(p(\tau) \to 0 \text{ at } \tau = p, \text{ sharply}\)
Non-Zero Mean I ($AR(1)$ Example): Regression on an Intercept and $y_{t-1}$, With White Noise Disturbances

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

$$\varepsilon_t \sim iidN(0, \sigma^2), \ |\phi| < 1$$

$$\implies y_t = c + \phi y_{t-1} + \varepsilon_t, \text{ where } c = \mu(1 - \phi)$$

Back-substitution reveals that:

$$y_t = \mu + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

$$\implies E(y_t) = \mu$$
Non-Zero Mean II (AR(1) Example, Cont’d):
Regression on an Intercept Alone, with AR(1) Disturbances

\[ y_t = \mu + \varepsilon_t \]

\[ \varepsilon_t = \phi \varepsilon_{t-1} + \nu_t \]

\[ \nu_t \sim iidN(0, \sigma^2), \ |\phi| < 1 \]
The Sample Autocorrelation Function

Autocorrelations:

\[ \rho_y(\tau) = \text{corr}(y_t, y_{t-\tau}) = \frac{\text{cov}(y_t, y_{t-\tau})}{\sqrt{\text{var}(y_t)} \sqrt{\text{var}(y_{t-\tau})}} = \frac{\text{cov}(y_t, y_{t-\tau})}{\text{var}(y_t)} \]

Sample autocorrelations:

\[ \hat{\rho}_y(\tau) = \frac{\hat{\text{cov}}(y_t, y_{t-\tau})}{\hat{\text{var}}(y_t)} = \frac{1}{T} \sum_t y_t y_{t-\tau} = \frac{1}{T} \sum_t y_t^2 \]

We view \( \hat{\rho}_y(\tau) \) as a function of \( \tau \) and examine its shape.
The Sample Partial Autocorrelation Function

Partial autocorrelations:

\( \hat{\rho}_y(\tau) \) is the coefficient on \( y_{t-\tau} \) in the projection

\[ y_t \rightarrow c, y_{t-1}, \ldots, y_{t-(\tau-1)}, y_{t-\tau}, \quad \tau = 0, 1, 2, \ldots \]

Sample partial autocorrelations:

\( \hat{\rho}_y(\tau) \) is the coefficient on \( y_{t-\tau} \) in the regression

\[ y_t \rightarrow c, y_{t-1}, \ldots, y_{t-(\tau-1)}, y_{t-\tau}, \quad t = 1, \ldots, T, \quad \tau = 0, 1, 2, \ldots \]

We view \( \hat{\rho}_y(\tau) \) as a function of \( \tau \) and examine its shape.
Bartlett Standard Errors

Under $H_0: y_t \sim iidN(0, \sigma^2)$, we have (as $T \to \infty$):

\[(1) \hat{\rho}_y(\tau) \sim \mathcal{N}(0, \frac{1}{T}), \forall \tau\]

(used for inference on individual autocorrelations)

95% “Bartlett bands” under the iid null: $0 \pm \frac{2}{\sqrt{T}}$

\[(2) \text{cov} (\hat{\rho}_y(\tau), \hat{\rho}_y(\tau + \nu)) = 0, \forall \tau, \nu\]

(used to derive distributions of Box-Pierce and Ljung-Box stats)
Box-Pierce and Ljung-Box $Q$ Statistics

Under $H_0: y_t \sim iidN(0, \sigma^2)$, we have (as $T \to \infty$):

$$Q_{BP} = T \sum_{\tau=1}^{m} \hat{\rho}^2(\tau) \sim \chi^2_m$$

$$Q_{LB} = T(T + 2) \sum_{\tau=1}^{m} \left( \frac{1}{T - \tau} \right) \hat{\rho}^2(\tau) \sim \chi^2_m$$

(We test an *implication* of iid, $\rho(1) = \rho(2) = \ldots = \rho(m) = 0$)
Figure: Sample Acorr Fn, Daily Stock Market Returns
Serial Correlation in Time-Series Regression

Consider:

$$\varepsilon \sim N(0, \sigma^2 \Omega)$$

The FIC case is $$\Omega = I$$. When is $$\Omega \neq I$$?

We’ve already seen heteroskedasticity.

Now we consider “serial correlation” or “autocorrelation.”

$$\rightarrow \quad \varepsilon_t \text{ is correlated with } \varepsilon_{t-\tau} \leftarrow$$

Can arise for many reasons, but they all boil down to:

The included $$X$$ variables fail to capture all the dynamics in $$y$$.

– No additional explanation needed!
On $\Omega$ with Heteroskedasticity vs. Serial Correlation

With heteroskedasticity, $\varepsilon_i$ is independent across $i$ but not identically distributed across $i$ (variance of $\varepsilon_i$ varies with $i$):

$$
\sigma^2 \Omega = \begin{pmatrix}
\sigma_1^2 & 0 & \ldots & 0 \\
0 & \sigma_2^2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \sigma_N^2
\end{pmatrix}
$$

With serial correlation, $\varepsilon_t$ is correlated across $t$ but unconditionally identically distributed across $t$:

$$
\sigma^2 \Omega = \begin{pmatrix}
\sigma^2 & \gamma(1) & \ldots & \gamma(T-1) \\
\gamma(1) & \sigma^2 & \ldots & \gamma(T-2) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(T-1) & \gamma(T-2) & \ldots & \sigma^2
\end{pmatrix}
$$
Consequences of Serial Correlation

OLS inefficient (no longer BLUE),
in finite samples and asymptotically

Standard errors biased and inconsistent. Hence $t$ ratios do not have the $t$ distribution in finite samples and do not have the $N(0, 1)$ distribution asymptotically

Does this sound familiar?
Detection

- Graphical autocorrelation diagnostics
  - Residual plot
  - Scatterplot of $e_t$ against $e_{t-\tau}$

- Formal autocorrelation tests and analyses
  - Durbin-Watson
  - Breusch-Godfrey
  - Residual correlogram
# Liquor Sales Regression on Trend and Seasonals

Dependent Variable: LSALES  
Method: Least Squares  
Date: 10/13/12  Time: 12:32  
Sample: 1968M01 1993M12  
Included observations: 312

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef</th>
<th>Std. Err</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME</td>
<td>0.007656</td>
<td>0.000123</td>
<td>62.35882</td>
<td>0.0000</td>
</tr>
<tr>
<td>TIME2</td>
<td>-1.14E-05</td>
<td>3.56E-07</td>
<td>-32.06823</td>
<td>0.0000</td>
</tr>
<tr>
<td>D1</td>
<td>6.147456</td>
<td>0.012340</td>
<td>498.1699</td>
<td>0.0000</td>
</tr>
<tr>
<td>D2</td>
<td>6.08653</td>
<td>0.012353</td>
<td>492.8890</td>
<td>0.0000</td>
</tr>
<tr>
<td>D3</td>
<td>6.174127</td>
<td>0.012366</td>
<td>499.3008</td>
<td>0.0000</td>
</tr>
<tr>
<td>D4</td>
<td>6.175220</td>
<td>0.012378</td>
<td>498.8970</td>
<td>0.0000</td>
</tr>
<tr>
<td>D5</td>
<td>6.246086</td>
<td>0.012390</td>
<td>504.1398</td>
<td>0.0000</td>
</tr>
<tr>
<td>D6</td>
<td>6.250387</td>
<td>0.012401</td>
<td>504.0194</td>
<td>0.0000</td>
</tr>
<tr>
<td>D7</td>
<td>6.295979</td>
<td>0.012412</td>
<td>507.2402</td>
<td>0.0000</td>
</tr>
<tr>
<td>D8</td>
<td>6.268043</td>
<td>0.012423</td>
<td>504.5509</td>
<td>0.0000</td>
</tr>
<tr>
<td>D9</td>
<td>6.203832</td>
<td>0.012433</td>
<td>498.9630</td>
<td>0.0000</td>
</tr>
<tr>
<td>D10</td>
<td>6.229197</td>
<td>0.012444</td>
<td>500.5968</td>
<td>0.0000</td>
</tr>
<tr>
<td>D11</td>
<td>6.259770</td>
<td>0.012453</td>
<td>502.6602</td>
<td>0.0000</td>
</tr>
<tr>
<td>D12</td>
<td>6.580068</td>
<td>0.012463</td>
<td>527.9819</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

- 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: -3.291086  
- Sum squared resid: 0.621448  
- Schwarz criterion: -3.123131  
- Log likelihood: 527.4094  
- Hannan-Quinn criter: -3.223959  
- Durbin-Watson stat: 0.586187
Graphical Diagnostics - Residual Plot
Graphical Diagnostics - Scatterplot of $e_t$ against $e_{t-1}$
Formal Tests and Analyses: Durbin-Watson (0.59!)

Simple paradigm (AR(1)):

\[ y_t = x_t' \beta + \varepsilon_t \]

\[ \varepsilon_t = \phi \varepsilon_{t-1} + \nu_t \]

\[ \nu_t \sim iid \ N(0, \sigma^2) \]

We want to test \( H_0 : \phi = 0 \) against \( H_1 : \phi \neq 0 \)

Regress \( y \rightarrow X \) and obtain the residuals \( e_t \)

\[ DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \]
Understanding the Durbin-Watson Statistic

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

\[ = \frac{1}{T} \sum_{t=2}^{T} e_t^2 + \frac{1}{T} \sum_{t=2}^{T} e_{t-1}^2 - 2 \frac{1}{T} \sum_{t=2}^{T} e_t e_{t-1}}{\frac{1}{T} \sum_{t=1}^{T} e_t^2} \]

Hence as \( T \to \infty \):

\[ DW \approx \frac{\sigma^2 + \sigma^2 - 2\text{cov}(e_t, e_{t-1})}{\sigma^2} = 2\left(1 - \text{corr}(e_t, e_{t-1})\right) \]

\[ = 2\left(1 - \rho_e(1)\right) \]

\[ \implies DW \in [0, 4], \ DW \to 2 \text{ as } \phi \to 0, \text{ and } DW \to 0 \text{ as } \phi \to 1 \]
Formal Tests and Analyses: Breusch-Godfrey

General AR($p$) environment:

\[ y_t = x_t' \beta + \varepsilon_t \]

\[ \varepsilon_t = \phi_1 \varepsilon_{t-1} + \ldots + \phi_p \varepsilon_{t-p} + \nu_t \]

\[ \nu_t \sim iidN(0, \sigma^2) \]

We want to test \( H_0 : (\phi_1, \ldots, \phi_p) = 0 \) against \( H_1 : (\phi_1, \ldots, \phi_p) \neq 0 \)

- Regress \( y_t \rightarrow x_t \) and obtain the residuals \( e_t \)
- Regress \( e_t \rightarrow x_t, e_{t-1}, \ldots, e_{t-p} \)
- Examine \( TR^2 \). In large samples \( TR^2 \sim \chi^2_p \) under the null.

Does this sound familiar?
BG for AR(1) Disturbances

\( TR^2 = 168.5, \ p = 0.0000 \)
BG for AR(4) Disturbances
($TR^2 = 216.7, \ p = 0.0000$)
BG for AR(8) Disturbances

\( TR^2 = 219.0, \ p = 0.0000 \)
Formal Tests and Analyses: Residual Correlogram

\[ \hat{\rho}_e(\tau) = \frac{\hat{\text{cov}}(e_t, e_{t-\tau})}{\hat{\text{var}}(e_t)} = \frac{1}{T} \sum_t e_t e_{t-\tau} \]

\[ \hat{\rho}_e(\tau) \text{ is the coefficient on } e_{t-\tau} \text{ in the regression} \]
\[ e_t \to c, e_{t-1}, \ldots, e_{t-(\tau-1)}, e_{t-\tau} \]

Approximate 95\% “Bartlett bands” under the iid N null: \( 0 \pm \frac{2}{\sqrt{T}} \)

\[ Q_{BP} = T \sum_{\tau=1}^{m} \hat{\rho}_e^2(\tau) \sim \chi^2_{m-K} \text{ under iid N} \]

\[ Q_{LB} = T(T+2) \sum_{\tau=1}^{m} \left( \frac{1}{T-\tau} \right) \hat{\rho}_e^2(\tau) \sim \chi^2_{m-K} \]
Residual Correlogram for Trend + Seasonal Model

Date: 10/14/12  Time: 18:32  
Sample: 1968M01 1993M12  
Included observations: 312

<table>
<thead>
<tr>
<th>Autocorrelation</th>
<th>Partial Correlation</th>
<th>AC</th>
<th>PAC</th>
<th>Q-Stat</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.700</td>
<td>0.700</td>
<td>154.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.686</td>
<td>0.383</td>
<td>302.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.725</td>
<td>0.369</td>
<td>469.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0.569</td>
<td>-0.141</td>
<td>572.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>0.569</td>
<td>0.017</td>
<td>675.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>0.577</td>
<td>0.093</td>
<td>782.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>0.460</td>
<td>-0.078</td>
<td>850.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>0.480</td>
<td>0.043</td>
<td>924.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>0.466</td>
<td>0.030</td>
<td>994.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>0.327</td>
<td>-0.188</td>
<td>1029.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>0.364</td>
<td>0.019</td>
<td>1072.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>0.355</td>
<td>0.089</td>
<td>1113.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
<td>0.225</td>
<td>-0.119</td>
<td>1129.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>0.291</td>
<td>0.065</td>
<td>1157.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>0.211</td>
<td>-0.119</td>
<td>1172.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16</td>
<td>0.138</td>
<td>-0.031</td>
<td>1178.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17</td>
<td>0.195</td>
<td>0.053</td>
<td>1191.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>0.114</td>
<td>-0.027</td>
<td>1195.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19</td>
<td>0.055</td>
<td>-0.063</td>
<td>1196.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>0.134</td>
<td>0.089</td>
<td>1202.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>21</td>
<td>0.062</td>
<td>0.018</td>
<td>1204.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>22</td>
<td>-0.006</td>
<td>-0.115</td>
<td>1204.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23</td>
<td>0.084</td>
<td>0.086</td>
<td>1206.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24</td>
<td>0.038</td>
<td>-0.124</td>
<td>1206.0</td>
</tr>
</tbody>
</table>
Correcting for Autocorrelation

- Generalized least squares
  - Transform the data such that the classical conditions hold

- Heteroskedasticity and autocorrelation consistent (HAC) s.e.’s
  - Use OLS, but calculate standard errors robustly
Recall Generalized Least Squares (GLS)

Consider the FIC except that we now let:

\[ \varepsilon \sim N(0, \sigma^2 \Omega) \]

The GLS estimator is:

\[ \hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \]

Under the remaining full ideal conditions it is consistent, normally distributed with covariance matrix \( \sigma^2(X'\Omega^{-1}X)^{-1} \), and MVUE:

\[ \hat{\beta}_{GLS} \sim N(\beta, \sigma^2(X'\Omega^{-1}X)^{-1}) \]
Infeasible GLS
(Illustrated in the Durbin-Watson AR(1) Environment)

\[ y_t = x_t' \beta + \varepsilon_t \quad (1a) \]
\[ \varepsilon_t = \phi \varepsilon_{t-1} + v_t \quad (1b) \]
\[ v_t \sim iid \ N(0, \sigma^2) \quad (1c) \]

Suppose that you know \( \phi \). Then you could form:

\[ \phi y_{t-1} = \phi x_{t-1}' \beta + \phi \varepsilon_{t-1} \quad (1a*) \]

\[ \Rightarrow (y_t - \phi y_{t-1}) = (x_t' - \phi x_{t-1}') \beta + (\varepsilon_t - \phi \varepsilon_{t-1}) \] (just \( (1a) - (1a*) \))

\[ \Rightarrow y_t = \phi y_{t-1} + x_t' \beta - x_{t-1}' (\phi \beta) + v_t \]

– Satisfies the classical conditions! Note the restriction.

So, two key closely-related regressions:

\[ y_t \rightarrow x_t \] (with AR(1) disturbances)
\[ y_t \rightarrow y_{t-1}, x_t, x_{t-1} \] (with WN disturbances and a coef. restr.)
Feasible \textit{GLS}

(1) Replace the unknown $\phi$ value with an estimate and run the OLS regression:

$$(y_t - \hat{\phi} y_{t-1}) \rightarrow (x'_t - \hat{\phi} x'_{t-1})$$

- Iterate if desired: $\hat{\beta}_1, \hat{\phi}_1, \hat{\beta}_2, \hat{\phi}_2, \ldots$

(2) Run the OLS Regression

$$y_t \rightarrow y_{t-1}, x_t, x_{t-1}$$

subject to the constraint noted earlier (or not)

- Generalizes trivially to $AR(p)$:

  $$y_t \rightarrow y_{t-1}, \ldots, y_{t-p}, x_t, x_{t-1}, \ldots, x_{t-p}$$

  (Select $p$ using the usual \textit{AIC}, \textit{SIC}, etc.)
Trend + Seasonal Model with $AR(4)$ Disturbances

![Figure: Trend + Seasonal Model with AR(4) Disturbances](image)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME</td>
<td>0.008129</td>
<td>0.000560</td>
<td>14.51705</td>
<td>0.0000</td>
</tr>
<tr>
<td>TIME2</td>
<td>-1.28E-05</td>
<td>1.49E-06</td>
<td>-8.592611</td>
<td>0.0000</td>
</tr>
<tr>
<td>D1</td>
<td>6.107740</td>
<td>0.045124</td>
<td>133.5340</td>
<td>0.0000</td>
</tr>
<tr>
<td>D2</td>
<td>6.050193</td>
<td>0.045154</td>
<td>133.9901</td>
<td>0.0000</td>
</tr>
<tr>
<td>D3</td>
<td>6.137542</td>
<td>0.045195</td>
<td>133.8017</td>
<td>0.0000</td>
</tr>
<tr>
<td>D4</td>
<td>6.139790</td>
<td>0.045252</td>
<td>133.6806</td>
<td>0.0000</td>
</tr>
<tr>
<td>D5</td>
<td>6.208530</td>
<td>0.045138</td>
<td>137.5442</td>
<td>0.0000</td>
</tr>
<tr>
<td>D6</td>
<td>6.213975</td>
<td>0.045137</td>
<td>137.6693</td>
<td>0.0000</td>
</tr>
<tr>
<td>D7</td>
<td>6.258524</td>
<td>0.045130</td>
<td>138.6763</td>
<td>0.0000</td>
</tr>
<tr>
<td>D8</td>
<td>6.230220</td>
<td>0.045075</td>
<td>138.2193</td>
<td>0.0000</td>
</tr>
<tr>
<td>D9</td>
<td>6.170731</td>
<td>0.045078</td>
<td>136.8913</td>
<td>0.0000</td>
</tr>
<tr>
<td>D10</td>
<td>6.193151</td>
<td>0.045066</td>
<td>137.4237</td>
<td>0.0000</td>
</tr>
<tr>
<td>D11</td>
<td>6.225272</td>
<td>0.045045</td>
<td>138.2010</td>
<td>0.0000</td>
</tr>
<tr>
<td>D12</td>
<td>6.547684</td>
<td>0.045050</td>
<td>145.3420</td>
<td>0.0000</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.348107</td>
<td>0.055751</td>
<td>6.243965</td>
<td>0.0000</td>
</tr>
<tr>
<td>AR(2)</td>
<td>0.257425</td>
<td>0.053823</td>
<td>4.783041</td>
<td>0.0000</td>
</tr>
<tr>
<td>AR(3)</td>
<td>0.429234</td>
<td>0.053804</td>
<td>7.977784</td>
<td>0.0000</td>
</tr>
<tr>
<td>AR(4)</td>
<td>-0.181633</td>
<td>0.055771</td>
<td>-2.898162</td>
<td>0.0040</td>
</tr>
</tbody>
</table>

R-squared: 0.995335
Adjusted R-squared: 0.995082
S.E. of regression: 0.027559
Sum squared resid: 0.238480
Log likelihood: 730.5205
Durbin-Watson stat: 1.982921

Path: c:\users\francis x. diebo\documents\diebold files\books\econometrics\data and software\aaliquo sales master\eviews DB 8
Trend + Seasonal Model with AR(4) Disturbances
Residual Plot
Trend + Seasonal Model with $AR(4)$ Disturbances

Residual Correlogram
Trend + Seasonal Model with Four Lags of Dep. Var.
How Did we Arrive at $AR(4)$ Dynamics?

Everything points there:
- Supported by original trend + seasonal residual correlogram
  - Supported by $DW$
  - Supported by $BG$
  - Supported by $SIC$ pattern:
    $AR(1) = -3.797$
    $AR(2) = -3.941$
    $AR(3) = -4.080$
    $AR(4) = -4.086$
    $AR(5) = -4.071$
    $AR(6) = -4.058$
    $AR(7) = -4.057$
    $AR(8) = -4.040$
Using advanced methods, one can obtain consistent standard errors (if not an efficient $\hat{\beta}$), under minimal assumptions

- “HAC standard errors”
- “Robust standard errors”
- “Newey-West standard errors”
- $\hat{\beta}$ remains unchanged at its OLS value. Is that a problem?
Trend + Seasonal Model with HAC Standard Errors
Heteroskedasticity in Time Series
A Typical Financial Asset Return Series

Figure: Time Series of Daily NYSE Returns.
Key Fact 1: Stock Returns are Approximately Serially Uncorrelated

Figure: Correlogram of Daily Stock Market Returns.
Key Fact 2: Returns are Unconditionally Non-Gaussian

Figure: Histogram and Statistics for Daily NYSE Returns.
Unconditional Volatility Measures

Variance: \( \sigma^2 = E(r_t - \mu)^2 \) (or standard deviation: \( \sigma \))

Mean Absolute Deviation: \( MAD = E|r_t - \mu| \)

Interquartile Range: \( IQR = 75\% - 25\% \)

Outlier probability: \( P|r_t - \mu| > 5\sigma \) (for example)

Tail index: \( \gamma \) s.t. \( P(r_t > r) = k \, r^{-\gamma} \)

Kurtosis: \( K = E(r - \mu)^4 / \sigma^4 \)

\( p\% \) Value at Risk (\( VaR^p \)): \( x \) s.t. \( P(r_t < x) = p \)
Key Fact 3: Returns are Conditionally Heteroskedastic

Figure: Time Series of Daily Squared NYSE Returns
Key Fact 3: Returns are Conditionally Heteroskedastic II

Figure: Correlogram of Daily Squared NYSE Returns.
Background: Financial Economics Changes Fundamentally When Volatility is Dynamic

- Risk management
- Portfolio allocation
- Asset pricing
- Hedging
- Trading
Asset Pricing I: Sharpe Ratios

Standard Sharpe:
\[
\frac{E(r_{it} - r_{ft})}{\sigma}
\]

Conditional Sharpe:
\[
\frac{E(r_{it} - r_{ft})}{\sigma_t}
\]
Asset Pricing II: CAPM

Standard CAPM:

\[(r_{it} - r_{ft}) = \alpha + \beta (r_{mt} - r_{ft})\]

\[\beta = \frac{\text{cov}((r_{it} - r_{ft}), (r_{mt} - r_{ft}))}{\text{var}(r_{mt} - r_{ft})}\]

Conditional CAPM:

\[\beta_t = \frac{\text{cov}_t((r_{it} - r_{ft}), (r_{mt} - r_{ft}))}{\text{var}_t(r_{mt} - r_{ft})}\]
Asset Pricing III: Derivatives

Black-Scholes:

\[ C = N(d_1)S - N(d_2)Ke^{-r\tau} \]

\[
\begin{align*}
    d_1 &= \frac{\ln(S/K) + (r + \sigma^2/2)\tau}{\sigma\sqrt{\tau}} \\
    d_2 &= \frac{\ln(S/K) + (r - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}
\end{align*}
\]

\[ P_C = BS(\sigma, \ldots) \]

(Standard Black-Scholes options pricing)

Completely different when \( \sigma \) varies!
Conditional Return Distributions

\[ f(r_t) \text{ vs. } f(r_t|\Omega_{t-1}) \]

Key 1: \( E(r_t|\Omega_{t-1}) \)

Are returns conditional mean independent? Arguably yes.

Returns are (arguably) approximately serially uncorrelated, and (arguably) approximately free of additional non-linear conditional mean dependence.
Conditional Return Distributions, Continued

Key 2: \( \text{var}(r_t|\Omega_{t-1}) = E((r_t - \mu)^2|\Omega_{t-1}) \)

Are returns conditional variance independent? No way!

Squared returns serially correlated, often with very slow decay.
Linear Models (e.g., AR(1))

\[ r_t = \phi r_{t-1} + \varepsilon_t \]

\[ \varepsilon_t \sim iid(0, \sigma^2), \quad |\phi| < 1 \]

Uncond. mean: \( E(r_t) = 0 \) (constant)

Uncond. variance: \( E(r_t^2) = \sigma^2/(1 - \phi^2) \) (constant)

Cond. mean: \( E(r_t | \Omega_{t-1}) = \phi r_{t-1} \) (varies)

Cond. variance: \( E([r_t - E(r_t | \Omega_{t-1})]^2 | \Omega_{t-1}) = \sigma^2 \) (constant)

– Conditional mean adapts, but conditional variance does not
ARCH(1) Process

\[ r_t | \Omega_{t-1} \sim N(0, h_t) \]

\[ h_t = \omega + \alpha r_{t-1}^2 \]

\[ E(r_t) = 0 \]

\[ E(r_t^2) = \frac{\omega}{(1 - \alpha)} \]

\[ E(r_t | \Omega_{t-1}) = 0 \]

\[ E([r_t - E(r_t | \Omega_{t-1})]^2 | \Omega_{t-1}) = \omega + \alpha r_{t-1}^2 \]
GARCH(1,1) Process ("Generalized ARCH")

\[ r_t \mid \Omega_{t-1} \sim N(0, h_t) \]

\[ h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1} \]

\[ E(r_t) = 0 \]

\[ E(r_t^2) = \frac{\omega}{(1 - \alpha - \beta)} \]

\[ E(r_t \mid \Omega_{t-1}) = 0 \]

\[ E([r_t - E(r_t \mid \Omega_{t-1})]^2 \mid \Omega_{t-1}) = \omega + \alpha r_{t-1}^2 + \beta h_{t-1} \]

Well-defined and covariance stationary if

\[ 0 < \alpha < 1, \, 0 < \beta < 1, \, \alpha + \beta < 1 \]
Exponential smoothing recursion:

\[
\hat{\sigma}^2_t = \lambda \hat{\sigma}^2_{t-1} + (1 - \lambda) r_t^2
\]

\[
\implies \hat{\sigma}^2_t = (1 - \lambda) \sum_j \lambda^j r_{t-j}^2
\]

But in GARCH(1,1) we have:

\[
h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}
\]

\[
h_t = \frac{\omega}{1 - \beta} + \alpha \sum \beta^{j-1} r_{t-j}^2
\]
Unified Theoretical Framework

- Volatility dynamics (of course, by construction)
- Volatility clustering produces unconditional leptokurtosis
- Temporal aggregation reduces the leptokurtosis
Tractable Empirical Framework

\[ L(\theta; r_1, \ldots, r_T) = f(r_T|\Omega_{T-1}; \theta)f((r_{T-1}|\Omega_{T-2}; \theta)\ldots, \]

where \( \theta = (\omega, \alpha, \beta)' \)

If the conditional densities are Gaussian,

\[ f(r_t|\Omega_{t-1}; \theta) = \frac{1}{\sqrt{2\pi}} h_t(\theta)^{-1/2} \exp \left( -\frac{1}{2} \frac{r_t^2}{h_t(\theta)} \right), \]

so

\[ \ln L = \text{const} - \frac{1}{2} \sum_t \ln h_t(\theta) - \frac{1}{2} \sum_t \frac{r_t^2}{h_t(\theta)} \]
Variations on the GARCH Theme

- Explanatory variables in the variance equation: GARCH-X
- Fat-tailed conditional densities: t-GARCH
- Asymmetric response and the leverage effect: T-GARCH
- Regression with GARCH disturbances
- Time-varying risk premia: GARCH-M
Explanatory variables in the Variance Equation: GARCH-X

\[ h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1} + \gamma z_t \]

where \( z \) is a positive explanatory variable
Fat-Tailed Conditional Densities: t-GARCH

If $r$ is conditionally Gaussian, then

$$r_t = \sqrt{h_t} \ N(0, 1)$$

But often with high-frequency data,

$$\frac{r_t}{\sqrt{h_t}} \sim \text{leptokurtic}$$

So take:

$$r_t = \sqrt{h_t} \ \frac{t_d}{\text{std}(t_d)}$$

and treat $d$ as another parameter to be estimated.
Asymmetric Response and the Leverage Effect: T-GARCH

Standard GARCH: \( h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1} \)

T-GARCH: \( h_t = \omega + \alpha r_{t-1}^2 + \gamma r_{t-1}^2 D_{t-1} + \beta h_{t-1} \)

\[ D_t = \begin{cases} 
1 & \text{if } r_t < 0 \\
0 & \text{otherwise} 
\end{cases} \]

positive return (good news): \( \alpha \) effect on volatility

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

\( \gamma \neq 0 \): Asymmetric news response

\( \gamma > 0 \): “Leverage effect”
Regression with GARCH Disturbances

\[ y_t = x_t' \beta + \varepsilon_t \]

\[ \varepsilon_t | \Omega_{t-1} \sim N(0, h_t) \]
Time-Varying Risk Premia: GARCH-M

Standard GARCH regression model:

\[ y_t = x_t' \beta + \varepsilon_t \]

\[ \varepsilon_t | \Omega_{t-1} \sim N(0, h_t) \]

GARCH-M model is a special case:

\[ y_t = x_t' \beta + \gamma h_t + \varepsilon_t \]

\[ \varepsilon_t | \Omega_{t-1} \sim N(0, h_t) \]
Back to Empirical Work – “Standard” GARCH(1,1)
GARCH(1,1)
GARCH(1,1)

![GARCH(1,1) Estimation, Daily NYSE Returns.](image)

```
Dependent Variable: R
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 12/03/12   Time: 07:58
Sample: 13461   Included observations: 3461
Convergence achieved after 16 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>z-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.000641</td>
<td>0.000127</td>
<td>5.039437</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
</tr>
<tr>
<td>RESID(-1)^2</td>
</tr>
<tr>
<td>GARCH(-1)</td>
</tr>
</tbody>
</table>

R-squared: 0.000193
Adjusted R-squared: 0.000193
S.E. of regression: 0.008542
S.D. dependent var: 0.008541
Akaike info criterion: -6.868008
Schwarz criterion: -6.860901
Hannan-Quinn criter.: -6.865470
Durbin-Watson stat: 1.861386
```
GARCH(1,1)

**Figure:** Estimated Conditional Standard Deviation, Daily NYSE Returns.
GARCH(1,1)

**Figure:** Conditional Standard Deviation, History and Forecast, Daily NYSE Returns.
A Useful Specification Diagnostic

\[ r_t \mid \Omega_{t-1} \sim N(0, h_t) \]

\[ r_t = \sqrt{h_t} \varepsilon_t, \quad \varepsilon_t \sim iidN(0, 1) \]

\[ \frac{r_t}{\sqrt{h_t}} = \varepsilon_t, \quad \varepsilon_t \sim iidN(0, 1) \]

Infeasible: examine \( \varepsilon_t \). iid? Gaussian?

Feasible: examine \( \hat{\varepsilon}_t = \frac{r_t}{\sqrt{\hat{h}_t}} \). iid? Gaussian?

Key deviation from iid is volatility dynamics. So examine correlogram of squared standardized returns, \( \hat{\varepsilon}_t^2 \)
GARCH(1,1)

Figure: Correlogram of Squared Standardized GARCH(1,1) Residuals, Daily NYSE Returns.
“Fancy” GARCH(1,1)
Dependent Variable: R
Method: ML - ARCH (Marquardt) - Student's t distribution
Date: 04/10/12   Time: 13:48
Sample (adjusted): 23461
Included observations: 3460 after adjustments
Convergence achieved after 19 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*RESID(-1)^2*(RESID(-1)<0)
+ C(7)*GARCH(-1)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>z-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>@SQRT(GARCH)</td>
<td>0.083360</td>
<td>0.053138</td>
<td>1.568753</td>
<td>0.1167</td>
</tr>
<tr>
<td>C</td>
<td>1.28E-05</td>
<td>0.000372</td>
<td>0.034443</td>
<td>0.9725</td>
</tr>
<tr>
<td>R(-1)</td>
<td>0.073763</td>
<td>0.017611</td>
<td>4.188535</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance Equation</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1.03E-06</td>
<td>2.23E-07</td>
<td>4.628790</td>
<td>0.0000</td>
</tr>
<tr>
<td>RESID(-1)^2</td>
<td>0.014945</td>
<td>0.009765</td>
<td>1.530473</td>
<td>0.1259</td>
</tr>
<tr>
<td>RESID(-1)^2*(RESID(-1)&lt;0)</td>
<td>0.094014</td>
<td>0.014945</td>
<td>6.290700</td>
<td>0.0000</td>
</tr>
<tr>
<td>GARCH(-1)</td>
<td>0.922745</td>
<td>0.009129</td>
<td>101.0741</td>
<td>0.0000</td>
</tr>
<tr>
<td>T-DIST. DOF</td>
<td>5.531579</td>
<td>0.478432</td>
<td>11.56188</td>
<td>0.0000</td>
</tr>
</tbody>
</table>