Econometrics

Francis X. Diebold University of Pennsylvania



February 17, 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

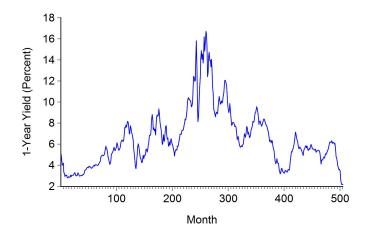


Graphics

Let's have some fun and look at the pictures first ...

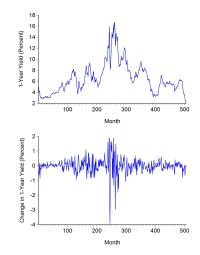


Time Series Plot: 1-Year Goverment Bond Yield, Levels



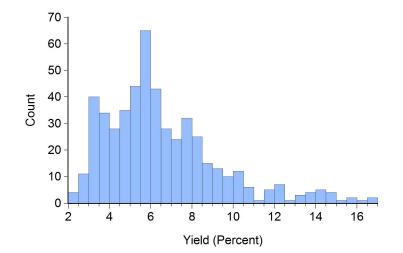


Time Series Plot: 1-Year Goverment Bond Yield, Levels and Changes



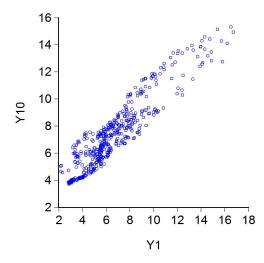


Histogram: 1-Year Government Bond Yield



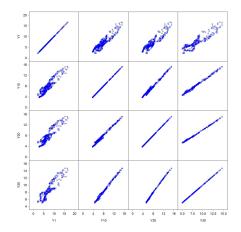


Bivariate Scatterplot 1-Year and 10-Year Government Bond Yields





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

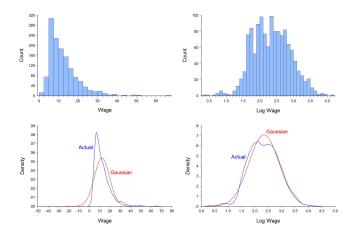


Principles of Graphical Style

- Know your audience, and know your goals.
- Appeal to the viewer.
- Show the data, and only the data, withing 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)



Population Moments: Expectations of Powers of R.V.'s

Mean measures location:

$$\mu = E(y) = \sum_{i} p_{i} y_{i} \text{ (discrete case)}$$
$$\mu = E(y) = \int y f(y) dy \text{ (continuous case)}$$

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

$$\sigma^2 = \operatorname{var}(y) = E(y - \mu)^2.$$

– σ easier to interpret than σ^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:

$$cov(y,x) = E[(y_t - \mu_y)(x_t - \mu_x)].$$

So does correlation:

$$corr(y, x) = \frac{cov(y, x)}{\sigma_y \sigma_x}.$$

Correlation is often more convenient. Why?



Sampling and Estimation

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

Sample mean:

$$ar{y} = rac{1}{N}\sum_{i=1}^N y_i$$

Sample variance:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N}$$

Unbiased sample variance:

$$s^{2} = \frac{\sum_{i=1}^{N}(y_{i} - \bar{y})^{2}}{N-1}$$



More Sample Moments

Sample skewness:

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

Sample kurtosis:

$$\hat{K} = \frac{\frac{1}{N} \sum_{i=1}^{N} (y_t - \bar{y})^4}{\hat{\sigma}^4}$$



Still More Sample Moments

Sample covariance:

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

Sample correlation:

$$\widehat{corr}(y,x) = \frac{\widehat{cov}(y,x)}{\widehat{\sigma}_y \widehat{\sigma}_x}$$



Exact Sampling Distribution of the Sample Mean (Requires *iid* Normality)

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

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

$$ar{y} \sim N\left(\mu, rac{\sigma^2}{N}
ight)$$
 $\sqrt{N}(ar{y} - \mu) \sim N(0, \sigma^2)$

$$\mu \in \left[\bar{y} \pm t_{1-\frac{\alpha}{2}} (N-1) \frac{s}{\sqrt{N}} \right] \text{ w.p. } \alpha$$
$$\mu = \mu_0 \implies \frac{\bar{y} - \mu_0}{\frac{s}{\sqrt{N}}} \sim t(N-1)$$



Approximate Asymptotic Sampling Distribution (Does Not Require Normality)

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

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

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

$$\begin{split} &\sqrt{N}(\bar{y}-\mu) \rightarrow_d N(0,\sigma^2) \\ \text{As } N \rightarrow \infty, \ \mu \in \left[\bar{y} \pm z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{N}} \right] \ w.p. \ \alpha \\ &\text{As } N \rightarrow \infty, \ \frac{\bar{y}-\mu_0}{\frac{\hat{\sigma}}{\sqrt{N}}} \sim N(0,1) \end{split}$$



Standard cross-section notation: i = 1, ..., N

Standard time-series notation: t = 1, ..., 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, ..., T.



Regression

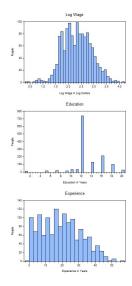




Surely the all-time greatest statistical and econometric workhorse...

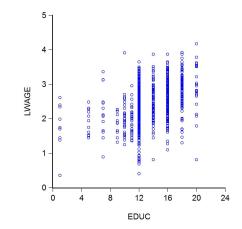


Distributions of Log Wage, Education and Experience





Scatterplot: Log Wage vs. Education





Regression as Curve Fitting

Fit a line:

$$y_t = \beta_1 + \beta_2 x_t$$

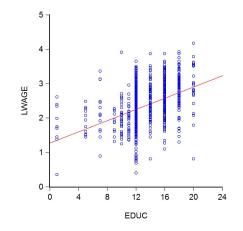
Solve:

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

 β is the set of two parameters β_1 and β_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



 $\widehat{LWAGE} = 1.273 + .081 EDUC$



Actual Values, Fitted Values and Residuals

The fitted values are

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

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

The residuals are the difference between actual and fitted values,

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

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



Multiple Linear Regression (K RHS Variables))

Solve:

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

Fitted line:

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

More compactly:

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

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

Wage dataset:

 $\widehat{LWAGE} = .867 + .093EDUC + .013EXPER$



Regression as a Probability Model

$$y_t = \beta_1 + \beta_2 x_{2t} + \dots + \beta_K x_{Kt} + \varepsilon_t$$
$$\varepsilon_t \sim iidN(0, \sigma^2),$$
$$t = 1, \dots, T.$$

Note:

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

Estimation:





Some Crucial Matrix Notation

You already understand matrix ("spreadsheet") notation although you may not know it!

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{21} & x_{31} & \dots & x_{K1} \\ 1 & x_{22} & x_{32} & \dots & x_{K2} \\ \vdots & & & & \\ 1 & x_{2T} & x_{3T} & \dots & x_{KT} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$



Elementary Matrices and Matrix Operations

$$\underline{0} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \qquad I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Transposition: $A'_{ij} = A_{ji}$ Addition: For A and B $n \times m$, $(A + B)_{ij} = A_{ij} + B_{ij}$ Multiplication: For A $n \times m$ and B $m \times p$, $(AB)_{ij} = \sum_{k=1}^{m} A_{ik}B_{kj}$. Inversion: For non-singular A $n \times n$, A^{-1} satisfies $A^{-1}A = AA^{-1} = I$. Many algorithms exist for calculation.



We Used to Write This:

$$y_t = \beta_1 + \beta_2 x_{2t} + \dots + \beta_K x_{Kt} + \varepsilon_t$$
$$\varepsilon_t \sim iidN(0, \sigma^2)$$
$$t = 1, 2, \dots, T$$



Now, Equivalently, We Write This:

$$y = X\beta + \varepsilon \quad (1)$$

$$\varepsilon \sim N(\underline{0}, \sigma^2 I) \quad (2)$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} 1 & x_{21} & x_{31} & \dots & x_{K1} \\ 1 & x_{22} & x_{32} & \dots & x_{K2} \\ \vdots & & & & \\ 1 & x_{2T} & x_{3T} & \dots & x_{KT} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$
(1)
$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix} \sim N \begin{pmatrix} \begin{pmatrix} 0_1 \\ 0_2 \\ \vdots \\ 0_T \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} \end{pmatrix}$$
(2)



The Full Ideal Conditions (FIC)

1. The true data-generating process is:

$$y = X\beta + \varepsilon$$

$$\varepsilon \sim N(\underline{0}, \sigma^2 I),$$

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, β , are fixed.
- 1.3 $\varepsilon \sim N$.
- 1.4 The ε_t 's have constant variance σ^2 .
- 1.5 The $\varepsilon_t{\rm 's}$ are uncorrelated.
- 2. There is no redundancy among the variables contained in X, so that X'X is non-singular.
- 3. X is a non-stochastic matrix, fixed in repeated samples.

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



Results

The OLS estimator is:

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

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

We write:

$$\hat{\beta}_{LS} \sim N\left(\beta, \ \sigma^2(X'X)^{-1}\right),$$

or equivalently,

$$\sqrt{T}(\hat{\beta} - \beta) \sim N\left(0, \sigma^2\left(\frac{X'X}{T}\right)^{-1}\right)$$



.

Regression Analysis of Wages, Education and Experience

vo\documents\my dropt	Dependent Variable: LWAGE Method: Least Squares Date: 06/27/13 Time: 16:38 Sample (adjusted): 1 1323 Included observations: 1323 after adjustments				
sid	Variable	Coefficient	Std. Error	t-Statistic	Prob.
ole01 pregressions	С	0.867382	0.075331	11,51431	0 0000
ion	EDUC	0.093229	0.005045	18,48002	0.0000
ge gehist	EXPER	0.013104	0.001164	11.26208	0.0000
agehistandstats agekernel	R-squared	0.232224	Mean dependent var		2.341995
	Adjusted R-squared	0.231061			0.561435
	S.E. of regression	0.492318			1.422881
	Sum squared resid	319.9376			1.434644
	Log likelihood	-938.2358			1.427291
	F-statistic	199.6260			1.926045
	Prob(F-statistic)	0.000000			
			EXPE	R	

"Top Matter": Background Information

- Dependent variable
- Method
- Date
- Sample
- Included observations



"Middle Matter": Estimated Regression Function

- Variable
- Coefficient
- Standard error
- t-statistic
- p-value



"Bottom Matter: Statistics"

There are many...



50 / 307

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

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



Regression Statistics: Sum squared resid 319.938

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



53 / 307

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

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

$$SER = \sqrt{s^2} = \sqrt{rac{\sum_{t=1}^T e_t^2}{T-K}}$$



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

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



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

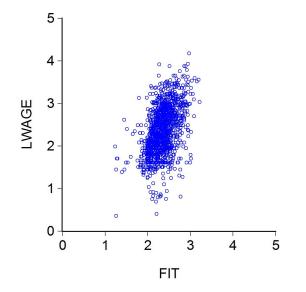
$$\varepsilon_t = \phi \varepsilon_{t-1} + v_t$$

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

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



Residual Scatter





Residual Plot

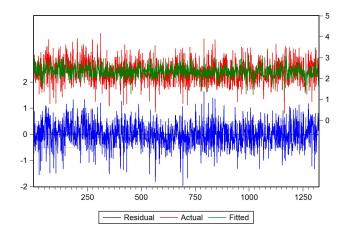


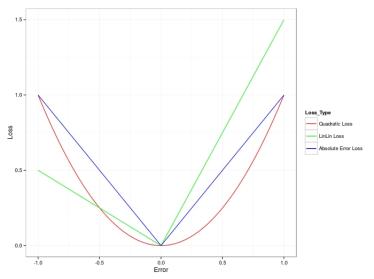
Figure: Wage Regression Residual Plot



63 / 307

Beyond OLS: Non-Quadratic Objectives

Various Loss Functions





Ordinary Least Squares (OLS)

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

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

- 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:

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

 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:

 $mean(y|X) = x\beta$

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

 $median(y|X) = x\beta$

- 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":

 $\min_{\beta} \sum_{t=1}^{T} linlin(\varepsilon_t),$ where: $linlin(e) = \begin{cases} a|e|, & \text{if } e \leq 0 \\ b|e|, & \text{if } e > 0 \end{cases}$ = a|e| I(e < 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: $quantile_d(y|X) = x\beta$ 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?)



Indicator Variables in Cross Sections: Group Effects



Dummy Variables for Group Effects

A dummy variable, or indicator variable, is just a 0-1 variable that indicates something, such as whether a person is female:

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

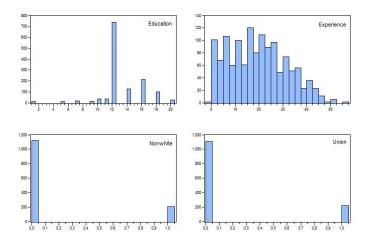
(It really is that simple.)

"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





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:

$LW\!AGE \rightarrow C, EDUC, EXPER$



Wage Regression on Education and Experience

Worldnle GRAPHS - (c/usert/francis x. diebo/documents/my dropt View Proc Objed Plant, Save Detaits -/. Show Fetch Store Detete Range: 11400 1400 obs Sample: 1400 1400 obs Sage Cal wagekernel		View Proc Object Print Name Freez Dependent Variable: LV Method: Least Squares Date: 06/27/13 Time: Sample (adjusted): 1 13 Included observations:	VAGE 16:38 323				
Ø c ⊠ educ ⊠ exper	⊠ nonwhite ⊞ qq ⊠ resid	Variable	Coefficient	Std. Error	t-Statistic	Prob.	
✓ female Image: Final Image: Final Image: Final Image: Final with state	table01 ↓ tworegressions ↓ union ↓ wage	C EDUC	0.867382	0.075331 0.005045	11.51431 18.48002	0.0000	
i graph01 graph02 graph03	교 wage 때 wagehist 때 wagehistandstats 때 wagekernel	EXPER R-squared	0.013104	0.001164 Mean depend		0.0000	
in graph04 in graph05 in histscovariates		J	ŭ	Adjusted R-squared S.E. of regression Sum squared resid	0.231061 0.492318 319.9376	S.D. depende Akaike info cri Schwarz criter	terion
⊠lwage ⊞lwageeduc ⊞lwageeducnoline		Log likelihood F-statistic Prob(F-statistic)	-938.2358 199.6260 0.000000	Hannan-Quinn Durbin-Watso	criter.	1.427291 1.926045	
🖬 lwageexper 🖬 lwagehist 🖬 lwagehistandstats				EXPE	P		
Untitled / New Page /		Path = c:\users\franc	tis x. diebo\documents\d			011\sw3e\eviews ma	



Controlling for Sex, Race and Union Status in the Wage Regression

Now:

$LWAGE \rightarrow C, EDUC, EXPER, FEMALE, NONWHITE, UNION$



Wage Regression on Education, Experience and Group Dummies

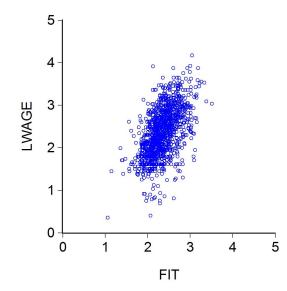
Workfile: GRAPHS - (c\users\francis x. diebo\c View Proc Object Print Save Details -/- Snow Range: 1 1400 - 1400 obs Sample: 1 1400 - 1400 obs Sample: 1 0400 - 1400 obs	View Proc Object Print Name Freez Dependent Variable: LV Method: Least Squares Date: 07/03/13 Time: Sample (adjusted): 1 13 Included observations:	WAGE 13:36 323			
I lwag ⊠educ ⊠nonv	Variable	Coefficient	Std. Error	t-Statistic	Prob.
✓ exper ✓ female ✓ resic ✓ final ■ resic	•	1.000385	0.073180	13.67013	0.0000
an final2 (at table an final2 (at table an finalwithstats an twor So fit So unio an graph01 So wag an graph02 an wag	EDUC EXPER FEMALE NONWHITE UNION	0.090809 0.012707 -0.237535 -0.085286 0.223392	0.004814 0.001119 0.025965 0.035786 0.035307	18.86314 11.35624 -9.148397 -2.383199 6.327126	0.0000 0.0000 0.0000 0.0173 0.0000
ing graph03 ing wag ing graph04 ing wag ing graph05 ing histscovariates i Mayage ing Iwageeduc ing Iwageeduc ing Iwageexper ing Iwageexper ing Iwageinst		0.307856 0.305229 0.467973 288.4212 -869.6356 117.1568 0.000000	Mean depend S.D. depende Akaike info cri Schwarz criter Hannan-Quinn Durbin-Watso	ent var iterion rion n criter.	2.341995 0.561435 1.323712 1.347239 1.332532 1.910120

Graph: GRAPH05 Modifed: 6/27/13 15:31

ĩ

Path = c:\users\francis x. diebo\documents\diebold files\courses\econ104\old\econ104_2011\sw3e\eviews materials

Residual Scatter from Wage Regression on Education, Experience and Group Dummies

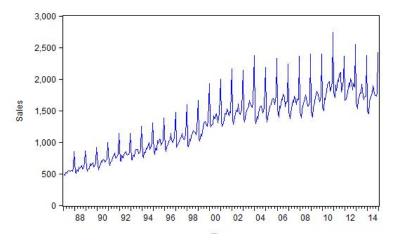




Indicator Variables in Time Series: Trend and Seasonality

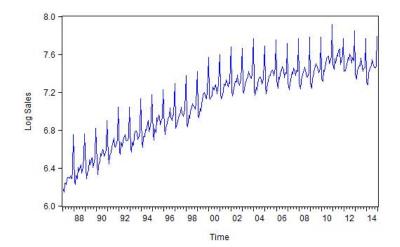


Liquor Sales





Log Liquor Sales





Linear Deterministic Trend

 $Trend_t = \beta_1 + \beta_2 TIME_t$

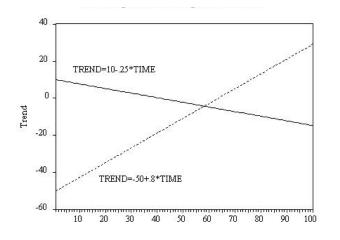
where $TIME_t = t$

Simply run the least squares regression $y \rightarrow c$, *TIME*, where

$$TIME = \begin{pmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ T - 1 \\ T \end{pmatrix}$$



Various Linear Trends



Time



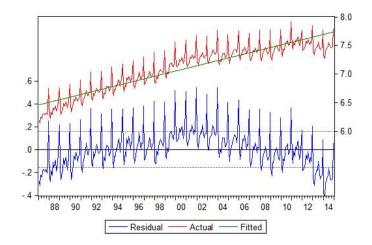
Linear Trend Estimation

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

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C TIME	6.454290 0.003809	0.017468 8.98E-05	369.4834 42.39935	0.0000 0.0000
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood F-statistic Prob(F-statistic)	0.843318 0.842849 0.159743 8.523001 140.5262 1797.705 0.000000	Mean depend S.D. depende Akaike info c Schwarz crite Hannan-Quir Durbin-Watso	ent var riterion erion an criter.	7.096188 0.402962 -0.824561 -0.801840 -0.815504 1.078573



Residual Plot





Deterministic Seasonality

Seasonal_t =
$$\sum_{i=1}^{s} \beta_i SEAS_{it}$$
 (s seasons per year)

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

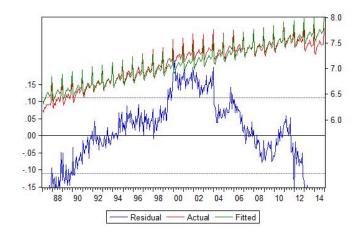
Simply run the least squares regression $y \rightarrow SEAS_1, ..., SEAS_s$ (or blend: $y \rightarrow TIME, SEAS_1, ..., SEAS_s$)

where (e.g., in quarterly data case, assuming Q1 start and Q4 end): $\begin{aligned} SEAS_1 &= (1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, ..., 0)' \\ SEAS_2 &= (0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, ..., 0)' \\ SEAS_3 &= (0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, ..., 0)' \\ SEAS_4 &= (0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, ..., 1)'. \end{aligned}$

Linear Trend with Seasonal Dummies

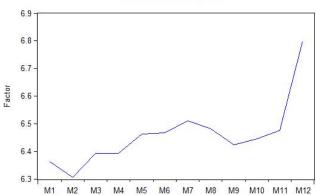
and a stand of the stand	-								
	c Quic	k Options Add-ins Window Help							
1 2014m12 1 2014m12		View Proc Object Print Name Freez	e Estimate Forecast St	ats Resids					
D1 D2 D3 D4 D5 D	D6 D7 D	Dependent Variable: LS	SALES				Â		
a a = a	Graph:	Method: Least Squares							
	w Proc	Date: 09/06/13 Time:	08:01						
Vie Vie Vie Vie	W Proc	Sample: 1987M01 201	4M12						
		Included observations:	336				=		
		Variable	Coefficient	Std. Error	t-Statistic	Prob.			
		TIME	0.003779	6.24E-05	60.57536	0.0000			
=		D1	6.361233	0.023283	273.2148	0.0000	T F		
		D2	6.304412	0.023310	270.4571	0.0000	ŀ	-Statistic	F
=	.6 -	D3	6.391653	0.023338	273.8773	0.0000			
		D4	6.392737	0.023365	273.6004	0.0000		0.57536	0
	.4	D5	6.461768	0.023393	276.2273	0.0000		73.2148	0
=	1	D6	6.466819	0.023421	276.1145	0.0000		70.4571	0
F	.2 -	D7	6.510789	0.023449	277.6602	0.0000		73.8773	0
A		D8	6.482457	0.023477	276.1210	0.0000		73.6004	0
8	.0 +	D9	6.422551	0.023505	273.2406	0.0000		76.2273	0
8		D10	6.444589	0.023533	273.8476	0.0000		76.1145	0
ЦЦ.	2 -	D11	6.476504	0.023562	274.8709	0.0000	- F	77.6602	0
4 W W L F	- 1	D12	6.798519	0.023591	288.1874	0.0000		76.1210	0
F .	4 -	and the second sec					- F	73.2406	0
		R-squared	0.927059	Mean depend		7.096188		73.8476	0
		Adjusted R-squared	0.924350	S.D. depende		0.402962		74.8709	0
		S.E. of regression	0.110833	Akaike info cri		-1.523658	P	88.1874	0
		Sum squared resid	3.967734	Schwarz criter		-1.375972	F		
		Log likelihood	268.9746	Hannan-Quinn	criter.	-1.464786	2	var	7.0
		Durbin-Watson stat	0.100500	for a single state of the same	and dish and first sec		-	4 2011) 2 ->	
			Path = C:\users\	francis x. diebo\docume	nts\alebold files\cou	urses\econ104\old\a	conto	4_2011\SW3E\e	views n
	ĩ	S 🕗 🕺							

Residual Plot





Seasonal Pattern



Estimated Seasonal Factors



Nonlinearity in Cross Sections



Anscombe's Quartet

View Proc Quick Options Add-ins Window Help

	1 .11 1	h Store Delete Genr S	ample					
	1 obs 1 obs			Filter: *				
	⊠ y2							
	⊠ γ3							
	TITLED Workfile: FCST4_4							-
	iject Print Name Freeze		Transpose Edit+/- Smpl+/-					
bs	Y1	X1	Y2	X2	Y3	X3	Y4)
1	8.040000	10.00000	9.140000	10.00000	7.460000	10.00000	6.580000	8.0000
2	6.950000	8.000000	8.140000	8.000000	6.770000	8.000000	5.760000	8.0000
3	7.580000	13.00000	8.740000	13.00000	12.74000	13.00000	7.710000	8.0000
4	8.810000	9.000000	8.770000	9.000000	7.110000	9.000000	8.840000	8.0000
5	8.330000	11.00000	9.260000	11.00000	7.810000	11.00000	8.470000	8.0000
6	9.960000	14.00000	8.100000	14.00000	8.840000	14.00000	7.040000	8.0000
7	7.240000	6.000000	6.130000	6.000000	6.080000	6.000000	5.250000	8.0000
8	4.260000	4.000000	3.100000	4.000000	5.390000	4.000000	12.50000	19.000
9	10.84000	12.00000	9.130000	12.00000	8.150000	12.00000	5.560000	8.0000
10	4.820000	7.000000	7.260000	7.000000	6.420000	7.000000	7.910000	8.0000
11	5.680000	5.000000	4.740000	5.000000	5.730000	5.000000	6.890000	8.0000

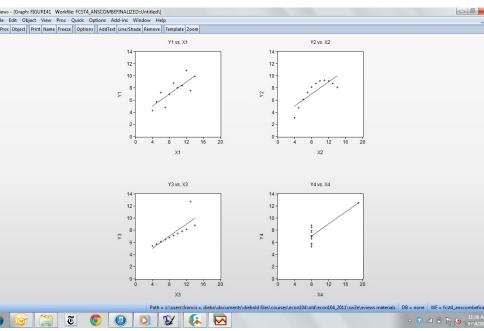
- 0

Anscombe's Quartet: Regressions

LO II DEPERC	dent Variable is	Y1		
Variable	Coefficient	Std Error	T-Statistic	
С	3.00	1.12	2.67	
X1	0.50	0.12	4.24	
R-squared	0.67	S.E.	ofregression	1.24
LS // Depend	dent Variable is	Y2		
Variable	Coefficient	Std Error	T-Statistic	
С	3.00	1.12	2.67	
X2	0.50	0.12	4.24	
R-squared	0.67	S.E.	ofregression	1.24
10.45		70		+
	dent Variable is Coefficient		T-Statistic	
С	3.00	1.12	2.67	
C X3	3.00 0.50	1.12 0.12	2.67 4.24	
Variable C X3 R-squared	3.00	1.12 0.12	2.67	1.24
C X3 R-squared	3.00 0.50 0.67	1.12 0.12 S.E.	2.67 4.24	1.24
C X3 R-squared	3.00 0.50	1.12 0.12 S.E. Y4	2.67 4.24	1.24
C X3 R-squared LS // Depend Variable	3.00 0.50 0.67 dent Variable is	1.12 0.12 S.E. Y4	2.67 4.24 of regression	1.24
C X3 R-squared LS // Depend	3.00 0.50 0.67 dent Variable is Coefficient	1.12 0.12 S.E. Y4 Std. Error	2.67 4.24 of regression T-Statistic	1.24



Anscombe's Quartet: Graphics



Parametric and Nonparametric Nonlinearity...

...and the gray area in between.



94 / 307

Log-Log Regression

$$lny_t = \beta_1 + \beta_2 lnx_t + \varepsilon_t$$

Example: Cobb-Douglas production function

$$y_t = AL_t^{\alpha} K_t^{\beta} exp(\varepsilon_t)$$

$$lny_t = lnA + \alpha lnL_t + \beta lnK_t + \varepsilon_t$$

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

$$lny_t = \beta x_t + \varepsilon$$

Example: Exponential growth

$$y_t = Ae^{rt}$$

$$lny_t = lnA + rt$$

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

Example: LWAGE regression!



Box-Cox Regression

$$B(y_t) = \beta_1 + \beta_2 x_t + \varepsilon_t$$

where

$$B(y_t) = \frac{y_t^{\lambda} - 1}{\lambda}$$

Because

$$\lim_{\lambda\to 0} \left(\frac{y^{\lambda}-1}{\lambda}\right) = \ln(y_t),$$

the Box-Cox model corresponds to the log-lin model in the special case of $\lambda = 0$.



Generalized Linear Model

$$G(y_t) = \beta_1 + \beta_2 x_t + \varepsilon_t,$$

Wide classes of link functions G can be entertained. Log-lin regression, for example, emerges when $G(y_t) = ln(y_t)$, and Box-Cox regression emerges when $G(y_t) = \frac{y_t^{\lambda} - 1}{\lambda}$.



Intrinsically Non-Linear Models

One example is the logistic model,

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

- 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

$$y_t = f(x_t) + \varepsilon_t$$



First 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, $f(x_t, z_t)$:

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

- Equally relevant for dummy variables: "interactions"



 $f(x_t) \approx \beta_1 + \beta_2 sin(x_t) + \beta_3 cos(x_t) + \beta_4 sin(2x_t) + \beta_5 cos(2x_t) + \dots$

 One can also mix Taylor and Fourier approximations by regressing not only on powers and cross products ("Taylor terms"), but also on various sines and cosines ("Fourier terms"). Mixing may facilitate parsimony.



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 I: t and F Tests

Just test for omitted series expansion terms!



104 / 307

Testing for Non-Linearity II: RESET

Run:

$$y_t \to c, X_t$$

and obtain the fitted values \hat{y}_t .

Then run

$$y_t \rightarrow c, X_t, \hat{y}_t^2, ..., \hat{y}_t^m.$$

Note that the powers of \hat{y}_t are linear combinations of powers and cross products of the X variables. No need to include the first power of \hat{y}_t , because that would be redundant with the included X variables. Instead we include powers \hat{y}_t^2 , \hat{y}_t^3 , ... Typically a small *m* is adequate. Significance of the included set of powers of \hat{y}_t can be checked using an *F* test.

Basic Wage Regression

rag rag onv sic ble ror nio ag ag	View Proc Object Print Name Freeze Estimate Forecast Stats Resids Dependent Variable: LWAGE Method: Least Squares Date: 07/03/13 Time: 13:36 Sample (adjusted): 1 1323 Included observations: 1323 after adjustments								
	Variable	Coefficient	Std. Error	t-Statistic	Prob.				
	C EDUC EXPER FEMALE NONWHITE UNION	1.000385 0.090809 0.012707 -0.237535 -0.085286 0.223392	0.073180 0.004814 0.001119 0.025965 0.035786 0.035307	13.67013 18.86314 11.35624 -9.148397 -2.383199 6.327126	0.0000 0.0000 0.0000 0.0000 0.0173 0.0000				
	R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood F-statistic Prob(F-statistic)	0.307856 0.305229 0.467973 288.4212 -869.6356 117.1568 0.000000	Mean dependent var S.D. dependent var Akaike info criterion Schwarz criterion Hannan-Quinn criter. Durbin-Watson stat		2.341995 0.561435 1.323712 1.347239 1.332532 1.910120	•			

Path = c:\users\francis x. diebo\documents\diebold files\courses\econ104\old\econ104_2011\sw3e\eviews materials}

Quadratic Wage Regression

Cut	Vie	Dependent Variable: LV Method: Least Squares Date: 10/02/13 Time: Sample: 1 1323	12:37				* III	_ =	×
c edu	S	Included observations:	1323						
	ß	Variable	Coefficient	Std. Error	t-Statistic	Prob.			
	N N	С	0.473236	0.240586	1.967017	0.0494			
		EDUC	0.109673	0.028918	3.792608	0.0002			
		EXPER	0.064422	0.007652	8.419060	0.0000			
		EDUC2	0.000501	0.000895	0.559994	0.5756			
on hwhite		EXPER2	-0.000705	8.86E-05	-7.962263	0.0000		_	
hite		EDU_EXP	-0.001789	0.000429	-4.173423	0.0000			
EDU		FEMALE	-0.237696	0.025506	-9.319335	0.0000			
		UNION	0.202955	0.034569	5.870998	0.0000			
c edu		NONWHITE	-0.095028	0.034931	-2.720476	0.0066			
		R-squared	0.343072	Mean depend	lent var	2.341995			
		Adjusted R-squared	0.339073	S.D. depende		0.561435			
c edu		S.E. of regression	0.456433	Akaike info cr		1.276028			
		Sum squared resid	273.7465	Schwarz criter	rion	1.311318			
		Log likelihood	-835.0925	Hannan-Quinr	n criter.	1.289257			
od	< >	F-statistic	85.77745	Durbin-Watso	n stat	1.894409	-		-

Path = c:\users\francis x. diebo\documents\diebold files\books\econometrics\dataandsoftware\aaaaaagesmaster\

Dummy Interactions?

	And And Address of the Address of th						
Эbj	ect View Proc Quick Opti	ions Add-ins Window Help					
		View Proc Object Print Name Freeze	e Estimate Forecast St	ats Resids			
ie F	ie View Proc Object Print Nar F Dependent Variab	Dependent Variable: LV Method: Least Squares Date: 10/02/13 Time: Sample: 1 1323 Included observations:	12:48				* III
E S	Method: Least Sq Date: 10/02/13 T	Variable	Coefficient	Std. Error	t-Statistic	Prob.	
0	Sample: 1 1323 Included observat	C EDUC	1.011503 0.090805	0.073797 0.004819	13.70657 18.84231	0.0000	
F	Variable	EXPER FEMALE	0.012674 -0.257621	0.001120 0.030125	11.31689 -8.551814	0.0000 0.0000	
5	С	UNION	0.216911	0.047498	4.566771	0.0000	
F	= C F EDUC T EDUC2 F EXPER2 L EDU_EXP L FEMALE L NONWHITE	NONWHITE FEM_UNI FEM_NON UNI_NON	-0.157606 0.003921 0.125567 0.017743	0.056446 0.072024 0.071781 0.091336	-2.792173 0.054439 1.749313 0.194258	0.0053 0.9566 0.0805 0.8460	
L L		R-squared 0.309487 Mean dependent var Adjusted R-squared 0.305283 S.D. dependent var S.E. of regression 0.467955 Akaike info criterion Sum squared resid 287.7418 Schwarz criterion		ent var iterion	2.341995 0.561435 1.325889 1.361179		
Б	R-squared	Log likelihood	-868.0755	Hannan-Quinr	n criter.	1.339/81/807	7

Everything

íee	Equation: TABLE1FFF Wo	Depe Metho Date:	View Proc Object Print Name Freez Date. 10/02/13 Tille. Sample: 1 1323 Included observations:	12.40	ats Resids			*	
Vie Vie	View Proc Object Print Nar Dependent Variab	Samp	Variable	Coefficient	Std. Error	t-Statistic	Prob.		
NE	Method: Least Sq Date: 10/02/13 T		C EDUC	0.482967	0.240926	2.004623 3.776211	0.0452		
5 C 5 C 1	Sample: 1 1323 Included observat		EXPER EDUC2	0.064269 0.000517	0.007654 0.000900	8.396570 0.573929	0.0000 0.5661		
F == L	Variable		EXPER2 EDU_EXP FEMALE	-0.000701 -0.001796 -0.252921	8.87E-05 0.000429 0.029659	-7.904460 -4.185878 -8.527539	0.0000 0.0000 0.0000		t-
F	C EDUC		UNION NONWHITE	0.200937 -0.161501	0.046575 0.055077	4.314297 -2.932246	0.0000 0.0034	E	18 15
T	EXPER EDUC2 EXPER2		FEM_UNI FEM_NON UNI NON	-0.012956 0.110319 0.033202	0.070740 0.070093 0.089258	-0.183153 1.573909 0.371975	0.8547 0.1157 0.7100		lent v ent va
	EDU_EXP FEMALE UNION	R-squ Adjus	R-squared Adjusted R-squared	0.344357 0.338856	Mean depend S.D. depende	ent var	2.341995	i	iterio rion 1 crite
A L S F		S.E. (Sum Log li	S.E. of regression Sum squared resid	0.456507 273.2109	Akaike info cri Schwarz criter	iterion rion	1.278605 1.325658	•	n sta
	R-squared Adjusted R-square	F-stat Prob(Log likelihood F-statistic	-833.7970 62.59682	Hannan-Quinr Durbin-Watso		1.296244 1.891544	Ŧ	

So Drop Dummy Interactions and Tighten the Rest

	Dependent Variable: LW Method: Least Squares Date: 10/03/13 Time: 1 Sample: 1 1323 Included observations: 7 Variable	11:19 1323						×	= x
vie vie CC NN	Date: 10/03/13 Time: 1 Sample: 1 1323 Included observations: 7	1323						×	= x
	Sample: 1 1323 Included observations: 7	1323							
	Included observations: 1								
								-	-
	Variable	0							
1 1		Coefficient	Std. Error	t-Statistic	Prob.	Е			
1 11	С	0.360535	0.131792	2,735636	0.0063				
= =	EDUC	0.125028	0.009188	13.60791	0.0000				-
	EXPER	0.066130	0.007016	9.424974	0.0000				
= =	EXPER2	-0.000710	8.82E-05	-8.042035	0.0000				
	EDU_EXP	-0.001905	0.000375	-5.078046	0.0000				
	FEMALE	-0.239352	0.025327	-9.450447	0.0000		<u> </u>		-
	UNION	0.202574	0.034553	5.862629	0.0000				
	NONWHITE	-0.094903	0.034921	-2.717655	0.0067				
	R-squared	0.342915	Mean depend	ent var	2.341995				
	Adjusted R-squared	0.339418	S.D. depende	nt var	0.561435				
	S.E. of regression	0.456313	Akaike info cri	iterion	1.274755				
	Sum squared resid	273.8119	Schwarz criter	rion	1.306124		<u> </u>		-
	Log likelihood	-835.2503	Hannan-Quinn	criter.	1.286514				
F F	F-statistic	98.03775	Durbin-Watso	n stat	1.894273				
AA	Prob(F-statistic)	0.000000				-		-	-
		EDU_EXP FEMALE UNION NONWHITE R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood F F-statistic Prob(F-statistic)	EDU_EXP -0.001905 FEMALE -0.239352 UNION 0.202574 NONWHITE -0.094903 R-squared 0.342915 Adjusted R-squared 0.339418 S.E. of regression 0.456313 Sum squared resid 273.8119 Log likelihood -835.2503 F. F. Prob(F-statistic) 0.000000	EDU_EXP -0.001905 0.000375 FEMALE -0.239352 0.025327 UNION 0.202574 0.034553 NONWHITE -0.094903 0.034921 R-squared 0.3342915 Mean depend Adjusted R-squared 0.339418 S.D. depende S.E. of regression 0.456313 Akaike info criter Log likelihood -835.2503 Hannan-Quinr F. F. F-statistic 98.03775 Durbin-Watso Prob(F-statistic) 0.000000	EDU_EXP -0.001905 0.000375 -5.078046 FEMALE -0.239352 0.025327 -9.450447 UNION 0.202574 -0.34553 5.862629 NONWHITE -0.094903 0.034921 -2.717655 R-squared 0.342915 Mean dependent var Adjusted R-squared 0.339418 S.D. dependent var S.E. of regression 0.456313 Akaike info criterion Sum squared resid 273.8119 Schwarz criterion Log likelihood -835.2503 Hannan-Quinn criter. F-statistic 98.03775 Durbin-Watson stat Prob(F-statistic) 0.000000	EDU_EXP -0.001905 0.000375 -5.078046 0.0000 FEMALE -0.239352 0.025327 -9.450447 0.0000 UNION 0.202574 0.034553 5.862629 0.0000 NONWHITE -0.094903 0.034921 -2.717655 0.0067 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 criter. 1.286514 F F Sum squared resid 0.000000 1.894273 Prob(F-statistic) 0.000000 Union-Watson stat 1.894273	EDU_EXP -0.001905 0.000375 -5.078046 0.0000 FEMALE -0.239352 0.025327 -9.450447 0.0000 UNION 0.202574 0.034553 5.862629 0.0000 NONWHITE -0.094903 0.034921 -2.717655 0.0067 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 criter. 1.286514 F F-statistic 98.03775 Durbin-Watson stat 1.894273 A Prob(F-statistic) 0.000000 - -	EDU_EXP -0.001905 0.000375 -5.078046 0.0000 FEMALE -0.239352 0.025327 -9.450447 0.0000 UNION 0.202574 -0.34553 5.86229 0.0000 NONWHITE -0.094903 0.034921 -2.717655 0.0067 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 criter. 1.286514 F. F. F-statistic 98.03775 Durbin-Watson stat 1.894273 A A Prob(F-statistic) 0.000000	EDU_EXP -0.001905 0.000375 -5.078046 0.0000 FEMALE -0.239352 0.025327 -9.450447 0.0000 UNION 0.202574 0.034553 5.862629 0.0000 NONWHITE -0.094903 0.034921 -2.717655 0.0067 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 criter. 1.286514 F F-statistic 98.03775 Durbin-Watson stat 1.894273

Nonlinearity in Time Series



Non-Linear Trend: Exponential (Log-Linear)

$$Trend_t = \beta_1 e^{\beta_2 TIME_t}$$

$\ln(Trend_t) = \ln(\beta_1) + \beta_2 TIME_t$



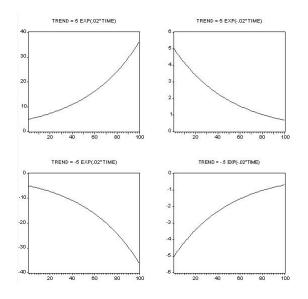


Figure: Various Exponential Trends



Non-Linear Trend: Quadratic

Allow for gentle curvature by including TIME and TIME²:

$$Trend_t = \beta_1 + \beta_2 TIME_t + \beta_3 TIME_t^2$$



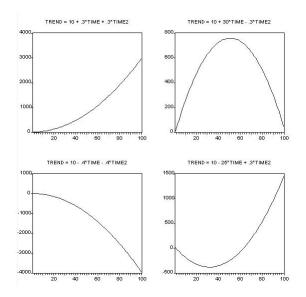


Figure: Various Quadratic Trends



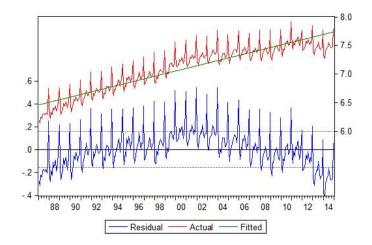
Recall Log-Linear Liquor Sales Trend Estimation

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

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C TIME	6.454290 0.003809	0.017468 8.98E-05	369.4834 42.39935	0.0000 0.0000
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood F-statistic Prob(F-statistic)	0.843318 0.842849 0.159743 8.523001 140.5262 1797.705 0.000000	Mean depend S.D. depende Akaike info c Schwarz crite Hannan-Quir Durbin-Watse	ent var riterion erion an criter.	7.096188 0.402962 -0.824561 -0.801840 -0.815504 1.078573



Residual Plot





117 / 307

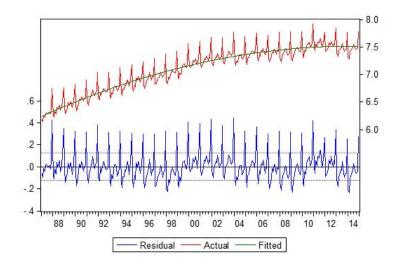
Log-Quadratic Liquor Sales Trend Estimation

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

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C TIME TIME2	6.231269 0.007768 -1.17E-05	0.020653 301.7187 0.000283 27.44987 8.13E-07 -14.44511		0.0000 0.0000 0.0000
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood F-statistic Prob(F-statistic)	0.903676 0.903097 0.125439 5.239733 222.2579 1562.036 0.000000	Mean depen S.D. depend Akaike info c Schwarz crit Hannan-Qui Durbin-Wats	ent var riterion erion nn criter.	7.096188 0.402962 -1.305106 -1.271025 -1.291521 1.754412



Residual Plot





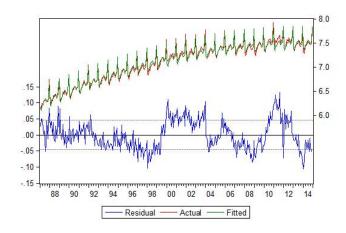
Log-Quadratic Liquor Sales Trend Estimation with Seasonal Dummies

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

Variable	Coefficient	Std. Error	t-Statistic	Prob.
TIME	0.007739	0.000104	74.49828	0.0000
TIME2	-1.18E-05	2.98E-07	-39.36756	0.0000
D1	6.138362	0.011207	547.7315	0.0000
D2	6.081424	0.011218	542.1044	0.0000
D3	6.168571	0.011229	549.3318	0.0000
D4	6.169584	0.011240	548.8944	0.0000
D5	6.238568	0.011251	554.5117	0.0000
D6	6.243596	0.011261	554.4513	0.0000
D7	6.287566	0.011271	557.8584	0.0000
D8	6.259257	0.011281	554.8647	0.0000
D9	6.199399	0.011290	549.0938	0.0000
D10	6.221507	0.011300	550.5987	0.0000
D11	6.253515	0.011309	552.9885	0.0000
D12	6.575648	0.011317	581.0220	0.0000
R-squared	0.987452	Mean depend	dent var	7.096188
Adjusted R-squared	0.986946	S.D. depende	ent var	0.402962
S.E. of regression	0.046041	Akaike info c	riterion	-3.277812
Sum squared resid	0.682555	Schwarz crite	erion	-3.118766
Log likelihood Durbin-Watson stat	564.6725 0.581383	Hannan-Quir	n criter.	-3.214412







Residual Plot

Moving-Average Trend and De-Trending

Two-sided moving average:

$$s_t = \frac{1}{2m+1} \sum_{i=-m}^m y_{t-i}$$

One-sided moving average:

$$s_t = \frac{1}{m+1} \sum_{i=0}^m y_{t-i}$$

One-sided weighted moving average:

$$s_t = \sum_{i=0}^m w_i y_{t-i}$$



Hodrick-Prescott Trend and De-Trending

$$\min_{\{s_t\}_{t=1}^T} \left(\sum_{t=1}^T (y_t - s_t)^2 + \lambda \sum_{t=2}^{T-1} \left((s_{t+1} - s_t) - (s_t - s_{t-1}) \right)^2 \right)$$



More Problems



Measurement Error

DGP:

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

Measurement:

$$x_t^m = x_t + v_t, \quad v_t \sim iid(0, \sigma^2)$$

We incorrectly run:

$$y \to c, x^n$$

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



Omitted Variables

DGP:

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

We incorrectly run:

 $y \rightarrow c, x$

where $corr(x_t, z_t) > 0$.

Clearly we'll estimate a positive effect of x on y, in large as well as small samples, even though it's completely spurious and would vanish if z had been included in the regression. The positive bias arises because in our example we assumed that $corr(z_t, z_t) > 0$; in general the sign of the bias could go either way.

Multicollinearity

Perfect Multicollinearity (e.g., dummy-variable trap):

- Drop a variable!

Imperfect Multicollinearity:

- 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."

- OLS is natural: orthogonal projection.



Multicollinearity and Variance Inflation

$$var(\hat{\beta}_j) = f\left(\underbrace{\sigma^2}_{+}, \underbrace{\sigma^2_{x_j}}_{-}, \underbrace{R_j^2}_{+}\right)$$

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

$$R_j^2$$
 affects $var(\hat{\beta}_j)$ as $(1 - R_j^2)^{-1}$
Hence as $R_j^2 \rightarrow 1$ the variance inflation approaches infinity $(x_j \text{ completely redundant})$



Non-Normality and Outliers

- Distributional results
- Diagnostics
- Outliers
- Robust estimation



Recall Sample Mean Under *iid* Normality

 \bar{y} is unbiased, consistent, normally distributed with variance σ^2/T , and indeed the minimum variance unbiased (MVUE) estimator.

We write:

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

or equivalently

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



Recall Sample Mean Under *iid* (Less Normality)

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

We write:

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

or more precisely, as $T
ightarrow \infty$,

$$\sqrt{T}(\bar{y}-\mu) \rightarrow_d N(0,\sigma^2)$$



OLS Under FIC (Including Normality)

 $\hat{\beta}_{LS}$ is unbiased, consistent, normally distributed with covariance matrix $\sigma^2(X'X)^{-1}$, and indeed MVUE.

We write:

$$\hat{\beta}_{LS} \sim N\left(\beta, \ \sigma^2 (X'X)^{-1}\right)$$

or equivalently

$$\sqrt{T}(\hat{\beta}-\beta) \sim N\left(0,\sigma^2\left(\frac{X'X}{T}\right)^{-1}\right)$$



OLS Under FIC (Less Normality)

 $\hat{\beta}_{LS}$ is consistent, asymptotically normally distributed, and BLUE. We write

$$\hat{\beta}_{LS} \sim N\left(eta, \ \sigma^2(X'X)^{-1}
ight),$$

or more precisely, as $T
ightarrow \infty$,

$$\sqrt{T}(\hat{\beta}_{LS}-\beta) \rightarrow_d N\left(0, \ \sigma^2\left(\frac{X'X}{T}\right)^{-1}\right)$$



Residual Normality Diagnostics

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

$$JB = rac{T}{6} \left(\hat{S}^2 + rac{1}{4} (\hat{K} - 3)^2
ight) \sim \chi_2^2$$

- More exotic: Outlier probabilities, tail indexes
- All can be done on observed data or residuals

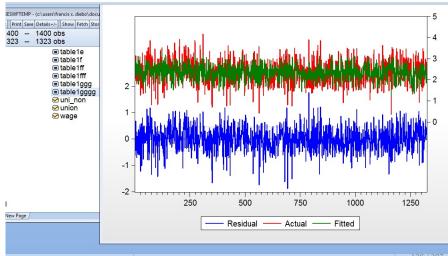


Recall Our "Final" Wage Regression

File Edit Object View Proc Quick O	View Proc Object Print Name Freeze	Estimate Forecast Sta	ats Resids						
	Dependent Variable: LW	AGE				Â			
	Method: Least Squares						- 0	х	ΞX
Rui Vie Vie Vie Vie Vie Vie	Date: 10/03/13 Time: 1	1:19							
	Sample: 1 1323							*	*
	Included observations: 1	323							
	Variable	Coefficient	Std. Error	t-Statistic	Prob.	н			
	С	0.360535	0.131792	2.735636	0.0063				
shd 🗹 = = = = = =	EDUC	0.125028	0.009188	13.60791	0.0000		<u> </u>		- 1
- F	EXPER	0.066130	0.007016	9.424974	0.0000				
^{res} 🖉 = = = L = =	EXPER2	-0.000710	8.82E-05	-8.042035	0.0000				
'Be 🖉	EDU_EXP	-0.001905	0.000375	-5.078046	0.0000				
equ 🖉 🛛 🖡	FEMALE	-0.239352	0.025327	-9.450447	0.0000		<u> </u>		- 1
nor 🖂 🛁	UNION	0.202574	0.034553	5.862629	0.0000				
	NONWHITE	-0.094903	0.034921	-2.717655	0.0067				
RE SAU	R-squared	0.342915	Mean depend	ent var	2.341995				
	Adjusted R-squared	0.339418	S.D. depende	nt var	0.561435				
	S.E. of regression	0.456313	Akaike info cri	iterion	1.274755				
shd 🖾 FLAL	Sum squared resid	273.8119	Schwarz criter	rion	1.306124		<u> </u>		- 1
res FFS = =	Log likelihood	-835.2503	Hannan-Quinn	criter.	1.286514				
	F-statistic	98.03775	Durbin-Watso	n stat	1.894273				
	Prob(F-statistic)	0.000000				-		-	-
-0.001905412242133127	Path	= c:\users\francis x. di	ebo\documents\diebold	d files\books\econom	etrics\dataandsoftw	are\aa	awagesmast	ter\evi	ews DI

Residual Plot

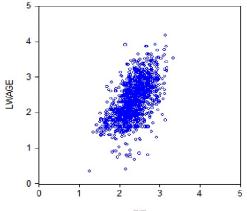
View Proc Ouick Options Add-ins Window Help



Path = c:\users\francis x. diebo\documents\diebold files\books\econometrics\dataandsoftware\aaawagesmaster\eviews | DB = nbab /wBDWages

- 0

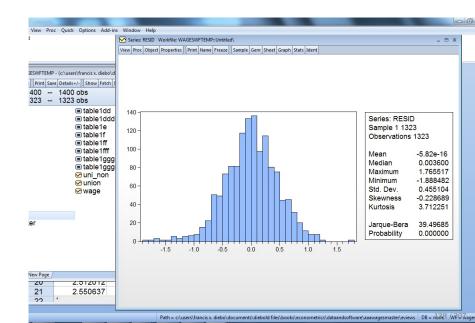
Residual Scatter



FIT



Residual Histogram and Statistics



More Residual Normality Tests

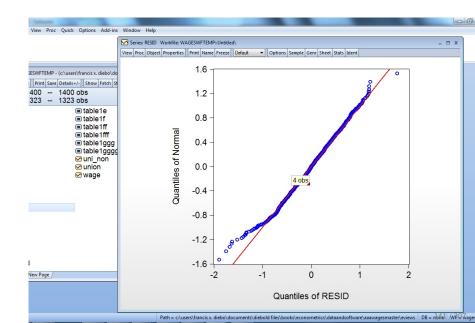
	ick Options Add-ins	Series: RESID Workfile: WAGESWFTEMP	::Untitled\				_ 0
		View Proc Object Properties Print Name	Freeze Sample Genr	Sheet Graph Stats Iden	t]		
rint Save Det	users\francis x. diebo\d ails+/-][Show]Fetch]\$ 00 obs 23 obs	Empirical Distribution Test Hypothesis: Normal Date: 10/21/13 Time: 13: Sample: 1 1323 Included observations: 132	12				
	■ table1dd ■ table1ddd ■ table1e	Method	Value	Adj. Value	Probability		
	■ table1f ■ table1ff	Lilliefors (D) Cramer-von Mises (W2)	0.026247	NA 0.154133	0.0327 0.0210		
	■ table1fff	Watson (U2)	0.138177	0.138229	0.0237		
	 table1ggg table1ggg 	Anderson-Darling (A2)	1.028843	1.029428	0.0104		
	⊠ uni_non ⊠ union ⊠ wage	Method: Maximum Likelihoo Parameter	od - d.f. correct Value	ed (Exact Solut Std. Error	ion) z-Statistic	Prob.	
		MU SIGMA	-6.18E-16 0.455104	0.012512	-4.94E-14 51.41984	1.0000	
		Log likelihood No. of Coefficients	-835.2505 2	Mean depend	dent var.	-5.82E-16 0.455104	
20 21 22	2.512012						

Residual 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 simply a plot of the quantiles of the standardized data against the 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
- Can be done on observed data or residuals



Wage Regression Residual QQ Plot



Outlier Detection and Robust Estimation

Data scatterplots

- Residual plots and scatterplots
- "Leave-one-out" plots:

$$\left(\hat{\beta}_{k}-\hat{\beta}_{k}(-t)\right), \ t=1,...T \ (k=1,...,K)$$

Robust estimation: LAD

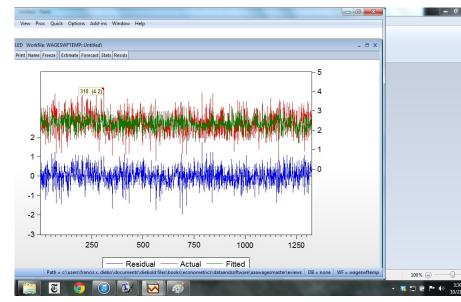
$$\min_{\beta} \sum_{t=1}^{T} |y_t - \beta_1 - \beta_2 x_{2t} - \dots - \beta_K x_{Kt}$$



Wage Regression LAD Estimation

					- 0 X	
View Proc Quick Options Add-in						Type a question for hel
	Equation: UNTITLED Workfile: WA					×
	View Proc Object Print Name Freez	e Estimate Forecast St	ats Resids			
GESWFTEMP - (c\userb/francis x. diebo); d Pimt Same Details -/. Snow retch 1400 1400 obs 1323 - 1323 obs II table 1 II tabl	Dependent Variable: LV Method: Quantile Regre Date: 10/21/13 Time: Sample: 1 1323 Included observations: Huber Sandwich Standa Sparsity method: Kerne Bandwidth method: Hall Estimation successfully Variable C C EXPER EDUC EXPER EDUC EXPER2 EDU_EXP FEMALE NONWHITE UNION Pseudo R-squared	VAGE ssion (Median) 15:17 1323 ard Errors & Cov I (Epanechnikov -Sheather, bw=I	Variance (*) using residua (*) 0.88501 e optimal solut Std. Error 0.127641 0.008260 0.009575 0.000112 0.000471 0.029209 0.038623 0.033508 Mean depend	t-Statistic 1.560855 8.865366 14.38339 -6.398026 -5.222012 -7.784041 -3.359806 6.468981 dent var	Prob. 0.1188 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 2.341995 0.561435	everzon.net he Atlas Cubs when he was a Atlas Cubs from 1923 and my o
	Adjusted R-squared	0.225235	S.D. depende	entvar	0.561435	
Path = c:\users\francis x. di	ebo\documents\diebold files\books\ecc	nometrics\dataandsoftw	are\aaawagesmaster\e	views DB = none	WF = wageswftemp	
👸 🙆 🌍 🧕) 😰 🔽					 ▲ ▲ ▲ 第 ● 3:2 10/2

Residual Plot



Generalized Least Squares (GLS)

Consider the FIC except that we now let:

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

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 it is consistent, normally distributed with covariance matrix $\sigma^2(X'\Omega^{-1}X)^{-1}$, and MVUE:

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



Heteroskedasticity in Cross-Section Regression

Homoskedasticity: variance of ε_i is constant across i

Heteroskedasticity: variance of ε_i is not constant across i

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

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

- Can arise for many reasons

- Engel curve (e.g., food expenditure vs. income) is classic example

OLS 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

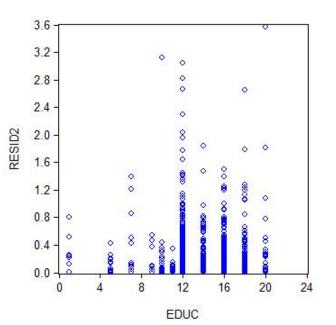
Problem: Purely pairwise



Recall Our "Final" Wage Regression

File Edit Object	View	Proc	Qu	ick	Options Add-ins Window Help	eeze Estimate Enrecast St	ats Resids		-				
					Dependent Variable:					ŕ			
	= =		E	eí	Method: Least Square							x	ΞX
	Vie Vie	_	Vie	Vie	Date: 10/03/13 Time	e: 11:19						_	
nor R			-		Sample: 1 1323	100000						*	*
she S	9 5	15	Ŋ	5	Included observations	s: 1323							
		E S	2 1 2		Variable	Coefficient	Std. Error	t-Statistic	Prob.	E			
equ 🗹 👌			3	3	c	0.360535	0.131792	2,735636	0.0063				
she 🗹 💳			-		EDUC	0.125028	0.009188	13.60791	0.0000				- 1
		F			EXPER	0.066130	0.007016	9,424974	0.0000				
		4 4	=	=	EXPER2	-0.000710	8.82E-05	-8.042035	0.0000				
'Be 🗹		-			EDU EXP	-0.001905	0.000375	-5.078046	0.0000				
equ 🔀		F			FEMALE	-0.239352	0.025327	-9.450447	0.0000				- 1
nor 🗹 💻					UNION	0.202574	0.034553	5.862629	0.0000				
sho 🗹 F	_	T			NONWHITE	-0.094903	0.034921	-2.717655	0.0067				
RE S	A	17			R-squared	0.342915	Mean depend	ent var	2.341995				
equ 🔳 S	S =	4 L			Adjusted R-squared	0.339418	S.D. depende	ent var	0.561435				
	SF	1 -			S.E. of regression	0.456313	Akaike info cr	iterion	1.274755				
she 🗳 F	L A	4 4			Sum squared resid	273.8119	Schwarz crite	rion	1.306124		<u> </u>		-
res 🔳 F	F S		=	=	Log likelihood	-835.2503	Hannan-Quinr	criter.	1.286514				
	FS	F	F	F	F-statistic	98.03775	Durbin-Watso	n stat	1.894273				
< Þ	- !	L	A	A	Prob(F-statistic)	0.000000				-		-	-
-0.0019054122421	33127		_			Path = c:\users\francis x. di	ebo\documents\diebol	d files\books\econom	etrics\dataandsoftv	vare\aa	awagesmas	ter\evi	ews DB
			2	Ĕ	👩 😰 📖				-				😽 🖞

Squared Residual vs. EDUC





151/307

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

EViews				And the second			
	uick Options Add-ins Window He	lp					
Is Iwage c educ exper exper2 edu_ex	View Proc Object Print Name Freez	e Estimate Forecast St	ats Resids				
	Heteroskedasticity Test	Breusch-Paga	n-Godfrey		*		
	Heteroskedasticity Test: Breusch-Pagan-Godfrey						
Workfile: WAGESWFTEMP - (c:	F-statistic	5.414870	Prob. F(7,131	(5)	0.0000		
View Proc Object Print Save De	Obs*R-squared	37.06628	Prob. Chi-Squ		0.0000		
Range: 1 1400 14		49.66045	Prob. Chi-Squ		0.0000		
Sample: 1 1323 13		40.00040	TIOD. OIL OQ		0.0000		
⊠ age							
D D	Test Equation:						
🗠 edu_exp	Dependent Variable: RI						
🗠 educ	Method: Least Squares						
educ2	Date: 10/30/13 Time:						
⊠ exper	Sample: 1 1323	10.54					
⊠ exper2 ⊠ fem_non		1000					
iem_non ⊠ fem_uni	Included observations: 1323						
M female		0 15 1 1	0.1 5				
M Iwage	Variable	Coefficient	Std. Error	t-Statistic	Prob.		
M nonwhite							
🗹 resid	C	-0.170309	0.097349	-1.749473	0.0804		
■ table1	EDUC	0.024074	0.006787	3.547204	0.0004		
table1a	EXPER	0.011701	0.005183	2.257616	0.0241		
table1b table1c	EXPER2	-5.53E-05	6.52E-05	-0.849150	0.3960		
E table1d	EDU_EXP	-0.000478	0.000277	-1.725513	0.0847		
■ table1dd	FEMALE	-0.009757	0.018708	-0.521530	0.6021		
■ table1ddd	UNION	-0.079648	0.025523	-3.120623	0.0018		
	NONWHITE	0.000486	0.025794	0.018829	0.9850 🛫		
Untitled New Page							

Path = c:\users\francis x. diebo\documents\diebold files\books\econometrics\dataandsoftware\aaawagesmaster\eviews DB



💥 🛍

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)



White Test

Views

A Suffrage and Suffrage

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

Is Iwage c educ exper exper2 edu_exp female union nonwhite

Workfile: WAGESWFTEMP - (c:\u	users\francis x. diebo\documents\diebold f	iles\books\econometri	cs\da 🗖 🗙			
View Proc Object Print Save Deta	ails+/- Show Fetch Store Delete Genr	Sample				
Range: 1 1400 140	00 obs		Filter: *			
Sample: 1 1323 132						
⊠ age	≡ table1e					
ອ ເ	■ table1f					
🗠 edu_exp	Equation: UNTITLED Workfile: WAG	ESWFTEMP::Untitled			_ = ×	
🖂 educ	View Proc Object Print Name Freeze	Estimate Forecast St	ats Resids			
⊠ educ2			1 1			
⊠ exper ⊠ exper2	Heteroskedasticity Test:	White				
i fem non						
⊠ fem_uni	F-statistic	2.431488	Prob. F(29		0.0000	
M female	Obs*R-squared	68.41804			0.0000	
✓ Iwage	Scaled explained SS	91.66473	Prob. Chi-S	Square(29)	0.0000	
M nonwhite						
🗠 resid					-	
table1 table1a						
E table1b						
Table1c						
■ table1d						
table1dd						
table1ddd						
Untitled New Page						
-2.682949803289767		Path = c:\users\franci	s x. diebo\documen	, ts\diebold files\books\e	conometrics\dataandso	ftware\aaawagesmaster\eviews D
📀 💽 🚞 🔇	ĕ 📀 🐖 🕸	💌 🛷	and the	100 V		- 🎕 i

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 $1/\sigma_i$:

Уi	=	$x'_i\beta$	+	ε_i
σ_i		σ_i	'	σ_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 σ_i^2



Remark on Weighted Least Squares

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

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



Feasible Weighted Least Squares

Intuition: Replace the unknown σ_i^2 values with estimates

Some good ideas:

▶ Use w_i = 1/e_i², where e_i² are from the BGP test regression
 ▶ Use w_i = 1/e_i², where e_i² 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 $e^2 \rightarrow X$



Regression Weighted by Fit From White Test Regression

nr resid2=resid*2 resid2 c educ exper exper2 edu_		View Proc Object Print Name Freez	e Estimate Forecast St	ats Resids			
nr resid2fit=resid2 - resid		Dependent Variable: LV					-
Workfile: WAGESWFTEMP - (c:\users\francis x. diebo\d	Method: Least Squares Date: 10/30/13 Time:					
View Proc Object Print Save E	Details+/- Show Fetch	Sample: 1 1323	12.32				C
Range: 1 1400 1 Sample: 1 1323 1		Included observations: Weighting series: RESI					
⊠ age Ø c	■ table1dd ■ table1ddd	Weight type: Variance ()			
✓ edu_exp ✓ educ ✓ educ2	 table1e table1f table1ff 	Variable	Coefficient	Std. Error	t-Statistic	Prob.	
✓ eque2	■ table1fff	С	0.406917	0.116809	3,483612	0.0005	
✓ exper2	table1ggg	EDUC	0.122743	0.008715	14.08407	0.0000	
M fem_non	table1ggg	EXPER	0.062997	0.006428	9,799839	0.0000	
✓ fem_uni ✓ female	⊠ uni_non ⊠ union	EXPER2	-0.000659	8.42E-05	-7.829347	0.0000	
✓ remaie ✓ lwage	M union M wage	EDU EXP	-0.001870	0.000374	-5.001801	0.0000	
M nonwhite	waye	FEMALE	-0.229424	0.024288	-9.445895	0.0000	
⊠ resid		UNION	0.234806	0.029741	7.895118	0.0000	
iresid2 iresid2fit		NONWHITE	-0.100705	0.032644	-3.084963	0.0021	
■ table1 ■ table1a			Weighted	Statistics			
table1b table1c		R-squared	0.388984	Mean depend	ent var	2.274204	
■ table1d		Adjusted R-squared	0.385731	S.D. depende	ent var	0.562167	-
Untitled New Page /	L				1	m	_

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

3

le Edit Object View Proc	Quick Options Add-ins						
lwage c educ exper exper2 edu	_exp female union nonwhite	Equation: UNTITLED Workfile: WA				- 🗆	×
nr resid2=resid^2		View Proc Object Print Name Freez	e Estimate Forecast St	ats Resids			
resid2 c educ exper exper2 edu nr resid2=resid^2	u_expiremale union nonwhite	Dependent Variable: LV	MACE				1
Workfile: WAGESWFTEMP							
View Proc Object Print Save			12:42				-
Range: 1 1400		Sample: 1 1323					
Sample: 1 1323	1323 obs	Included observations:					
🗠 age	table1dd	White heteroskedasticit	y-consistent sta	ndard errors &	covariance		
₿ C	table1ddd						
🗹 edu_exp	table1e	Variable	Coefficient	Std. Error	t-Statistic	Prob.	
✓ educ	■ table1f						
educ2	table1ff table1ff table1ff	С	0.360535	0.130391	2.765026	0.0058	
✓ exper ✓ exper2	i≣ table1ggg	EDUC	0.125028	0.009890	12.64118	0.0000	
i fem non	■ table1ggg	EXPER	0.066130	0.006967	9,491284	0.0000	
⊠ fem_uni	⊠ uni non	EXPER2	-0.000710	8.86E-05	-8.004870	0.0000	
⊠ female	✓ union	EDU EXP	-0.001905	0.000412	-4.623006	0.0000	
⊠ lwage	🖾 wage	FEMALE	-0.239352	0.025499	-9.386559	0.0000	
nonwhite	°	UNION	0.202574	0.031386	6.454196	0.0000	
🗠 resid		NONWHITE	-0.094903	0.034074	-2.785164	0.0054	
✓ resid2			-0.034303	0.034074	-2.700104	0.0004	
✓ resid2fit		R-squared	0.342915	Mean depend	lantwor	2.341995	
E table1							
■ table1a ■ table1b		Adjusted R-squared	0.339418	S.D. depende		0.561435	
■ table1c		S.E. of regression	0.456313	Akaike info cr		1.274755	
■ table1d		Sum squared resid	273.8119	Schwarz crite		1.306124	
		Log likelihood	-835.2503			1.286514	
Untitled New Page		F-statistic	98.03775	Durbin-Watso	n stat	1.894273	-

- 9

Path = c:\users\francis x. diebo\documents\diebold files\books\econometrics\dataandsoftware\aaawagesmaster\eviews

A Tiny Bit of Time-Series Theory: White Noise and AR(1) Processes

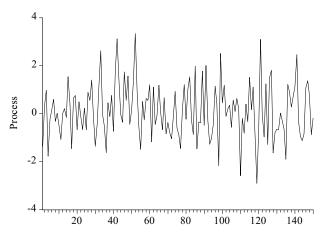
White noise: $y_t \sim WN(\mu, \sigma^2)$ (serially uncorrelated)

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

iidIndependent (strong) white noise: $y_t \sim (0, \sigma^2)$

Gaussian white noise: $y_t \sim N(0, \sigma^2)$





Realization of White Noise Process

Time



Autocovariance, Autocorrelation and Partial Autocorrelation Functions

Population autocovariances: $\gamma_y(\tau) = cov(y_t, y_{t-\tau}), \ \tau = 0, 1, 2, ...$

Population autocorrelations:

$$\rho_y(\tau) = \frac{\gamma_y(\tau)}{\gamma_y(0)} = corr(y_t, y_{t-\tau}), \ \tau = 0, 1, 2, ...$$

Population partial autocorrelations: $p_y(\tau)$ is the coefficient on $y_{t-\tau}$ in the projection $y_t \rightarrow c, y_{t-1}, ..., y_{t-(\tau-1)}, y_{t-\tau}, \ \tau = 0, 1, 2, ...$



Moment Structure of Strong White Noise

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

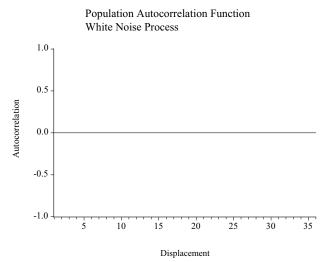
$$var(y_t | \Omega_{t-1}) = E[(y_t - E(y_t | \Omega_{t-1}))^2 | \Omega_{t-1}] = \sigma^2$$
where $\Omega_{t-1} = \{y_{t-1}, \ y_{t-2}, \ ...\}$

$$\gamma(\tau) = \begin{cases} \sigma^2, \ \tau = 0 \\ 0, \ \tau \ge 1 \end{cases}$$

$$\rho(\tau) = \begin{cases} 1, \ \tau = 0 \\ 0, \ \tau \ge 1 \end{cases}$$

$$p(\tau) = \begin{cases} 1, \ \tau = 0 \\ 0, \ \tau \ge 1 \end{cases}$$







Zero-Mean AR(1)

$$egin{aligned} &y_t = \phi y_{t-1} + arepsilon_t \ &arepsilon_t \sim \mathit{iidN}(0, \ \sigma^2), \ |\phi| < 1 \end{aligned}$$

- Regression on just a lagged dependent variable

- "Autoregression"

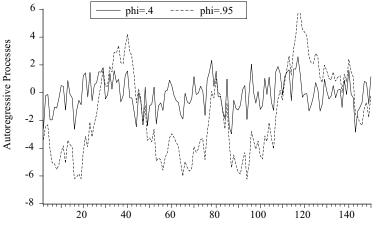
Back-substitution reveals that:

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

$$\implies E(y_t) = 0$$



Realizations of Zero-Mean Two AR(1) Processes



Time



Moment Structure of the Zero-Mean AR(1) Process

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

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

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

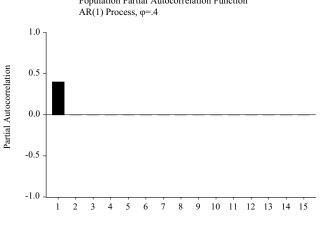
$$E(y_t | \Omega_{t-1}) = \phi y_{t-1} \text{ (obvious)}$$

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

$$\rho(\tau) = \begin{cases} 1, \ \tau = 0 \\ \phi^{\tau}, \ \tau \ge 1 \end{cases} \text{ (hmmm...)}$$

$$p(\tau) = \begin{cases} 1, \ \tau = 0 \\ \phi, \ \tau = 1 \\ 0, \ \tau \ge 2 \end{cases} \text{ (obvious)}$$

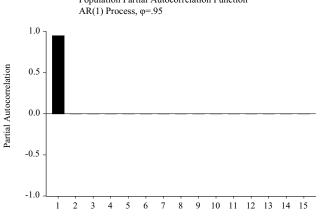




Population Partial Autocorrelation Function

Displacement





Population Partial Autocorrelation Function

Displacement



AR(1) Autocorrelation Function

$$y_{t} = \phi y_{t-1} + \varepsilon_{t}$$
$$\implies \quad y_{t} y_{t-\tau} = \phi y_{t-1} y_{t-\tau} + \varepsilon_{t} y_{t-\tau} \quad (1)$$

First consider $\tau = 0$. Immediately:

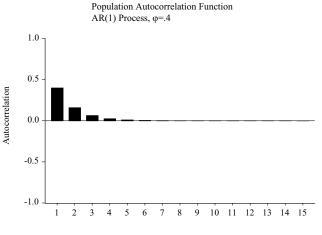
$$\gamma(0) = var(y_t) = \frac{\sigma^2}{1 - \phi^2}$$

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

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

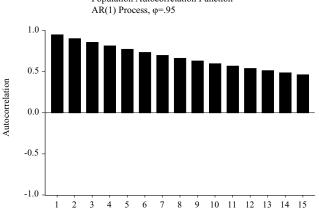
Hence $\gamma(\tau) = \phi^{ au} \frac{\sigma^2}{1-\phi^2}$, so $\rho(\tau) = \phi^{ au}, \ \tau = 0, 1, 2, ...$





Displacement





Population Autocorrelation Function

Displacement



 $\gamma(\tau)$, $\rho(\tau)$, and $p(\tau)$ for Generic AR(p)

AR(p)Process:

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

 $\gamma(au)
ightarrow 0$ as $au
ightarrow \infty$, gradually ho(au)
ightarrow 0 as $au
ightarrow \infty$, gradually p(au)
ightarrow 0 at au = p, 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

$$y_t = \mu + \varepsilon_t$$

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

$$v_t \sim \textit{iidN}(0, \sigma^2), \ |\phi| < 1$$



The Sample Autocorrelation Function

Autocorrelations:

$$\rho_{y}(\tau) = corr(y_{t}, y_{t-\tau}) = \frac{cov(y_{t}, y_{t-\tau})}{\sqrt{var(y_{t})}\sqrt{var(y_{t-\tau})}} = \frac{cov(y_{t}, y_{t-\tau})}{var(y_{t})}$$

Sample autocorrelations:

$$\hat{\rho}_{y}(\tau) = \frac{\widehat{cov}(y_{t}, y_{t-\tau})}{\widehat{var}(y_{t})} = \frac{\frac{1}{T}\sum_{t} y_{t}y_{t-\tau}}{\frac{1}{T}\sum_{t} y_{t}^{2}}$$

We view $\hat{\rho}_{y}(\tau)$ as a function of τ and examine its shape.



The Sample Partial Autocorrelation Function

Partial autocorrelations:

 $\hat{p}_y(\tau)$ is the coefficient on $y_{t-\tau}$ in the projection $y_t \rightarrow c, y_{t-1}, ..., y_{t-(\tau-1)}, y_{t-\tau}, \ \tau = 0, 1, 2, ...$

Sample partial autocorrelations:

 $\hat{p}_y(\tau)$ is the coefficient on $y_{t-\tau}$ in the regression $y_t \rightarrow c, y_{t-1}, ..., y_{t-(\tau-1)}, y_{t-\tau}, t = 1, ..., T, \tau = 0, 1, 2, ...$

We view $\hat{\rho}_{y}(\tau)$ as a function of τ and examine its shape.



Bartlett Standard Errors

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

(1)
$$\hat{\rho}_{y}(\tau) \stackrel{a}{\sim} N\left(0, \frac{1}{T}\right), \ \forall \tau$$

(used for inference on individual autocorrelations) 95% "Bartlett bands" under the *iid* null: $0 \pm \frac{2}{\sqrt{T}}$

(2)
$$cov(\hat{\rho}_y(\tau),\hat{\rho}_y(\tau+v))=0, \ \forall \tau, \ v$$

(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 $T \to \infty$):

$$Q_{BP} = T \sum_{ au=1}^m \hat{
ho}^2(au) \sim \chi_m^2$$

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

(We test an *implication* of *iid*, $\rho(1) = \rho(2) = ... = \rho(m) = 0$)



(Part of a) Correlogram

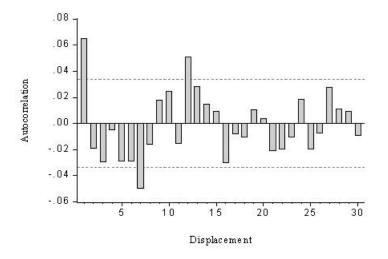


Figure: Sample Acorr Fn, Daily Stock Market Returns



Serial Correlation in Time-Series Regression

Consider:

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

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

We've already seen heteroskedasticity.

Now we consider "serial correlation" or "autocorrelation."

 \rightarrow ε_t 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 Ω with Heteroskedasticity vs. Serial Correlation

With heteroskedasticity, ε_i is independent across *i* but not identically distributed across *i* (variance of ε_i varies with *i*):

$$\sigma^{2}\Omega = \begin{pmatrix} \sigma_{1}^{2} & 0 & \dots & 0 \\ 0 & \sigma_{2}^{2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{N}^{2} \end{pmatrix}$$

With serial correlation, ε_t is correlated across t but unconditionally identically distributed across t:

$$\sigma^{2}\Omega = \begin{pmatrix} \sigma^{2} & \gamma(1) & \dots & \gamma(T-1) \\ \gamma(1) & \sigma^{2} & \dots & \gamma(T-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(T-1) & \gamma(T-2) & \dots & \sigma^{2} \end{pmatrix}$$



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

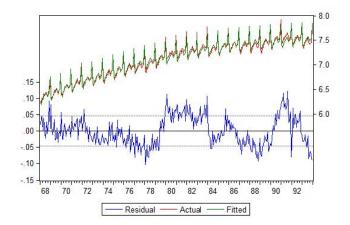
Dependent Variable: LSALES Method: Least Squares Date: 10/13/12 Time: 12:32 Sample: 1968M01 1993M12 Included observations: 312

Variable	Coef	Std. Err	t-Statistic	Prob.
TIME	0.007656	0.000123	62.35882	0.0000
TIME2	-1.14E-05	3.56E-07	-32.06823	0.0000
D1	6.147456	0.012340	498.1699	0.0000
D2	6.088653	0.012353	492.8890	0.0000
D3	6.174127	0.012366	499.3008	0.0000
D4	6.175220	0.012378	498.8970	0.0000
D5	6.246086	0.012390	504.1398	0.0000
D6	6.250387	0.012401	504.0194	0.0000
D7	6.295979	0.012412	507.2402	0.0000
D8	6.268043	0.012423	504.5509	0.0000
D9	6.203832	0.012433	498,9630	0.0000
D10	6.229197	0.012444	500.5968	0.0000
D11	6.259770	0.012453	502,6602	0.0000
D12	6.580068	0.012463	527.9819	0.0000
P. couprad		0.086111	Maan dan	andant var

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		



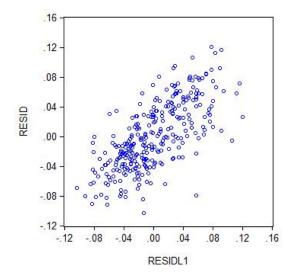
Graphical Diagnostics - Residual Plot





189 / 307

Graphical Diagnostics - Scatterplot of e_t against e_{t-1}





Formal Tests and Analyses: Durbin-Watson (0.59!)

Simple paradigm (AR(1)):

$$y_t = x'_t \beta + \varepsilon_t$$

$$arepsilon_t = \phi arepsilon_{t-1} + v_t$$

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

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

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

Hence as $T \to \infty$: $DW \approx \frac{\sigma^2 + \sigma^2 - 2cov(e_t, e_{t-1})}{\sigma^2} = 2(1 - \underbrace{corr(e_t, e_{t-1})}_{\rho_e(1)})$

 $\implies DW \in [0,4], \ DW \rightarrow 2 \ \text{as} \ \phi \rightarrow 0, \ \text{and} \ DW \rightarrow 0 \ \text{as} \ \phi \rightarrow 1$



Formal Tests and Analyses: Breusch-Godfrey

General AR(p) environment:

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

$$\begin{aligned} \varepsilon_t &= \phi_1 \varepsilon_{t-1} + \ldots + \phi_p \varepsilon_{t-p} + v_t \\ v_t &\sim \textit{iidN}(0, \sigma^2) \end{aligned}$$

We want to test H_0 : $(\phi_1, ..., \phi_p) = \underline{0}$ against H_1 : $(\phi_1, ..., \phi_p) \neq \underline{0}$

• Regress $y_t \rightarrow x_t$ and obtain the residuals e_t

• Regress
$$e_t \rightarrow x_t, e_{t-1}, ..., e_{t-p}$$

• Examine TR^2 . In large samples $TR^2 \sim \chi_p^2$ under the null.

Does this sound familiar?



BG for AR(1) Disturbances ($TR^2 = 168.5$, p = 0.0000)

2 6

ile Edit View Tools Hel	View Proc Object Print Name	Freeze Estimate	Forecast Stats Resi	ds			- 0 X	5	m
	Test Equation:						-		1
Organize 👻 🔽 Open 👻	Dependent Variable: RI	ESID							
🔶 Favorites	Method: Least Squares								
	Date: 11/15/13 Time:								
Downloads	Sample: 1987M01 201								
E Desktop	Included observations: 3								
Recent Places	Presample missing value	lagged resid	uals set to zero						
😂 Dropbox		00							
	Variable	Coefficient	Std. Error	t-Statistic	Prob.				
Documents									
	TIME	8.05E-06	7.35E-05	0.109640	0.9128				
J Music	TIME2	-3.01E-08 -0.001578	2.11E-07 0.007925	-0.142439 -0.199158	0.8868				
E Pictures	D1 D2	-0.001578	0.007925	-0.028949	0.8423				
Videos	D2 D3	-0.000230	0.007932	-0.028949	0.9769		E		
	D3	-0.000228	0.007948	-0.028089	0.9773		_		
E Computer	D4 D5	-0.000220	0.007955	-0.028423	0.9776				
🏭 OS (C:)	D6	-0.000222	0.007962	-0.027871	0.9778				
🙀 FNCEFICweb\$ (\\acadws	D7	-0.000220	0.007969	-0.027585	0.9780				
	D8	-0.000218	0.007976	-0.027293	0.9782				
📭 Network	D9	-0.000215	0.007983	-0.026995	0.9785				
	D10	-0.000213	0.007990	-0.026690	0.9787				
	D11	-0.000211	0.007996	-0.026378	0.9790				
	D12	-0.000209	0.008002	-0.026060	0.9792				
	RESID(-1)	0.709791	0.039491	17.97369	0.0000				
	R-squared	0.501594	Mean depend	lont var	5.87E-17				
	Adjusted R-squared	0.479857	S.D. depende		0.045138		-		
liquor Date		0.479007	S.D. depende	ni vai	0.043130	DB = none WF = liquor	uftomo		

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

2

11

M EViews	5		-				T is named	-		
File Edit	t Obj	ject Vie	w Proc Quick	Options Add-ins Window	Help					
			ſ	Equation: TABLE103 Work	file: LIOUORWFTEMP::	Untitled\			x	
83	Wor	Hile: 110	UORWFTEMP - (c	View Proc Object Print Nam					_	
	\overline{c}			Test Louation.	e_rreeze) [cstimate_r	precase stats Resi	asj			
		Gra	ph: FIGURE103 W	Dependent Variable: F	ESID				^	
Vie Vie	Vie	View P	roc Object Print	Method: Least Square	S					
				Date: 11/15/13 Time:	12:37					
	9			Sample: 1987M01 201	4M12					
	M			Included observations:						
	DS			Presample missing val	ue lagged residua	als set to zero.				
	In									
				Variable	Coefficient	Std. Error	t-Statistic	Prob.		
			1 Martin	TIME	1.74E-05	6.23E-05	0.279045	0.7804		
		.6 -	Ant	TIME2	-6.48E-08	1.79E-07	-0.361811	0.7177		
		4 -	1.	D1	-0.002187	0.006721	-0.325305	0.7452		
		.4 -		D2	-0.001492	0.006729	-0.221721	0.8247	E	
		.2 -		D3	-0.001016	0.006736	-0.150781	0.8802		
	R			D4 D5	-0.000180 -0.000488	0.006742	-0.026771	0.9787		
		.0 -	MMMMM	D5 D6	-0.000484	0.006747	-0.072302	0.9424		
	A S S		Y Y Y Y	D6	-0.000484	0.006753	-0.071596	0.9430		
-	S	2 -		D8	-0.000479	0.006765	-0.070136	0.9435		
	L			D9	-0.000474	0.006771	-0.069382	0.9447		
-	F	4 -	here here here here here h	D10	-0.000465	0.006776	-0.068611	0.9453		
	P		88 90	D10	-0.000460	0.006782	-0.067823	0.9460		
				D12	-0.000455	0.006787	-0.067020	0.9466		
				RESID(-1)	0.356563	0.055390	6.437279	0.0000		
				RESID(-2)	0.255694	0.053824	4.750556	0.0000		
•	U	ntitled 🖌	New Page /	RESID(-3)	0.425333	0.053844	7.899419	0.0000		
				RESID(-4)	-0.164177	0.055567	-2.954595	0.0034		
									Ψ.	

.

Path = c:\users\francis x. diebo\documents\diebold files\books\econometrics\dataandsoftware\aaaliquorsalesmaster\eviews DB

BG for AR(8) Disturbances ($TR^2 = 219.0, p = 0.0000$)

EViews

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

3

11

×	x			::Untitled\		Equation: TABLE200 Workfil	aí =	a l	Grap		m	ш
-	~		.]	Forecast Stats Resi				_	View Pr	Vie	Vie	(ie
		0.9111	-0.103472	Forecast Stats Resi	-0.000097	w Proc Object Print Name	Vie	Vie	View	Vie	Vie	le
^	^	0.9392	-0.076323	0.006738	-0.000514	D7				D		
		0.9187	-0.102103	0.006744	-0.000689	D8			Г	M		
		0.9354	-0.081066	0.006748	-0.000547	D9				D		
		0.9361	-0.080187	0.006753	-0.000542	D10				S		
		0.9369	-0.079289	0.006759	-0.000536	D11				In		
		0.9376	-0.078372	0.006764	-0.000530	D12				=		
		0.0000	6.560221	0.056379	0.369859	RESID(-1)						
		0.0001	4.110231	0.059906	0.246228	RESID(-2)			.6 -	=		
		0.0000	5.887687	0.061281	0.360805	RESID(-3)			.0			
		0.0112	-2.551620	0.064757	-0.165236	RESID(-4)			.4 -			
		0.9938	0.007815	0.064799	0.000506	RESID(-5)						
		0.1123	1.592575	0.061505	0.097952	RESID(-6)			.2 -	=		
		0.1406	-1.477162	0.060271	-0.089030	RESID(-7)				R		
		0.2066	1.265623	0.056707	0.071770	RESID(-8)		-	.0 -	A		-
		5 87E-17	lent var	Mean depend	0.651708	-squared	R		2 -	S		-
		0.045138		S.D. depende	0.628414	djusted R-squared				Ē		
		-4.284907		Akaike info cr	0.027515	E. of regression		-	4 -	E		1
		-4.034977	rion	Schwarz criter	0.237729	um squared resid				P		
		-4.185278	criter.	Hannan-Quinr	741.8643	og likelihood				=		
					2.000282	urbin-Watson stat						
• E				301	1.978	Durbin-Watson stat	ige /	Jew Pa	titled 🔏	U		

Path = c:\users\francis x. diebo\documents\diebold files\books\econometrics\dataandsoftware\aaaliquorsalesmaster\eviews DB

196

Formal Tests and Analyses: Residual Correlogram

$$\hat{\rho}_{e}(\tau) = \frac{\widehat{cov}(e_{t}, e_{t-\tau})}{\widehat{var}(e_{t})} = \frac{\frac{1}{T}\sum_{t} e_{t}e_{t-\tau}}{\frac{1}{T}\sum_{t} e_{t}^{2}}$$

 $\hat{p}_e(au)$ is the coefficient on $e_{t- au}$ in the regression $e_t o c, e_{t-1}, ..., e_{t-(au-1)}, e_{t- au}$

Approximate 95% "Bartlett bands" under the *iid* N null: $0 \pm \frac{2}{\sqrt{T}}$

$$Q_{BP} = T \sum_{\tau=1}^{m} \hat{\rho}_{e}^{2}(\tau) \sim \chi_{m-K}^{2} \text{ under } \textit{iid N}$$

$$\frac{m}{2} \left(-\frac{1}{2} \right) = 0$$

$$Q_{LB} = T(T+2)\sum_{ au=1} \left(rac{1}{T- au}
ight) \hat{
ho}_e^2(au) \sim \chi^2_{m-K}$$



Residual Correlogram for Trend + Seasonal Model

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

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.700	0.700	154.34	0.000
	· 🗖	2	0.686	0.383	302.86	0.000
		3	0.725	0.369	469.36	0.000
		4	0.569	-0.141	572.36	0.000
·	i)i	5	0.569	0.017	675.58	0.000
	ip	6	0.577	0.093	782.19	0.000
·	1	7	0.460	-0.078	850.06	0.000
	լոր	8	0.480	0.043	924.38	0.000
·	լոր	9	0.466	0.030	994.46	0.000
· 🗖		10	0.327	-0.188	1029.1	0.000
	լոր	11	0.364	0.019	1072.1	0.000
· •	ן ו	12	0.355	0.089	1113.3	0.000
· 🗖	 	13	0.225	-0.119	1129.9	0.000
· 🗖	լոր	14	0.291	0.065	1157.8	0.000
· 🗖 ·	 	15	0.211	-0.119	1172.4	0.000
· 🗖 ·	10	16	0.138	-0.031	1178.7	0.000
· 🗖 ·	ון ו	17	0.195	0.053	1191.4	0.000
· 🗖	10	18	0.114	-0.027	1195.7	0.000
i þi	101	19	0.055	-0.063	1196.7	0.000
1	ip	20	0.134	0.089	1202.7	0.000
ւի	10	21	0.062	0.018	1204.0	0.000
1 1		22	-0.006	-0.115	1204.0	0.000
i þ		23	0.084	0.086	1206.4	0.000
i di i		24	-0 020	_0 194	1206.0	0 000

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:

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

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:

$$\hat{eta}_{GLS} \sim N\left(eta, \ \sigma^2(X'\Omega^{-1}X)^{-1}
ight)$$



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:

$$\phi y_{t-1} = \phi x'_{t-1}\beta + \phi \varepsilon_{t-1} \quad (1a*)$$

$$\implies (y_t - \phi y_{t-1}) = (x'_t - \phi x'_{t-1})\beta + (\varepsilon_t - \phi \varepsilon_{t-1}) \text{ (just (1a) - (1a*))}$$

$$\implies y_t = \phi y_{t-1} + x'_t \beta - x'_{t-1}(\phi \beta) + v_t$$

- 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.)



201 / 307

Feasible GLS

(1) Replace the unknown ϕ 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: $\hat{\beta}_1, \hat{\phi}_1, \hat{\beta}_2, \hat{\phi}_2, \dots$

(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): $y_t \rightarrow y_{t-1}, ..., y_{t-p}, x_t, x_{t-1}, ..., x_{t-p}$ (Select p using the usual AIC, SIC, etc.)



Trend + Seasonal Model with AR(4) Disturbances

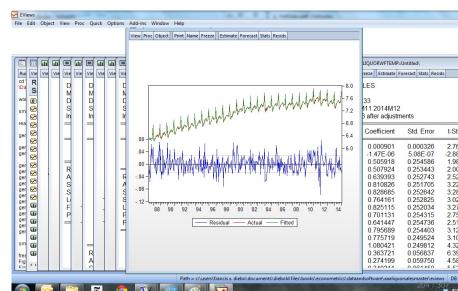
3

🧿

								ons	Add-ins Win	iect	rint Name	Freeze Esti	mate I	Forecast Stats R	esids		_			
									included o	DServ	auons.	after 5 itera	ijusir	nents			^			
зÍ		_		J		III	u	Ξ	Va	ariabl	э	Coeffici	ent	Std. Error	t-Statistic	Prob.	-	LIQUORWFTEN	1P::Untitled\	
d 1	Vie V	ie V	e Vie	Vie	Vie	Vie	Vie	Vie		IME		0.0081	29	0.000560	14.51705	0.0000)	11	Forecast Stats Re	sids
Da	R		D		D			d	Т	IME2		-1.28E	-05	1.49E-06	-8.592611	0.0000		LES		
1	S		M		M			Ν		D1		6.1077	40	0.045124	135.3540	0.0000				
0	B		D		D			C		D2		6.0501	93	0.045154	133.9901	0.0000		33		
m			S		S			S		D3		6.1375		0.045195	135.8017			111 2014M		
			In		In			Ir		D4		6.1397	90	0.045252	135.6806			6 after adjus	tments	
a			=		=			=		D5		6.2085	530	0.045138	137.5442	0.0000)			-
er										D6		6.2139		0.045137	137.6693			Coefficient	Std. Error	
			=		=			=		D7		6.2585		0.045130	138.6763					_
er										D8		6.2302		0.045075	138.2193			0.000901	0.000326	
er										D9		6.1707		0.045078	136.8913			-1.47E-06		
er			=							D10		6.1931		0.045066	137.4237			0.505918		
			R					=		D11		6.2252		0.045045	138.2010			0.507924		
er		1	A					F		D12		6.5476		0.045050	145.3420			0.639393		
			S					A		R(1)		0.3481		0.055751	6.243965			0.810826		
			S					9		R(2)		0.2574		0.053823	4.783041			0.828685		
er								9		R(3)		0.4292		0.053804	7.977784			0.764161	0.252825	
			F			-		4	/	AR(4)		-0.1616	533	0.055771	-2.898162	0.0040		0.825115		
en	90 10		Ρ					1	Desma			0.0055	205	Mana da		7 40700		0.701131	0.254315	
er	an An		=					Ч	R-squared			0.9953		Mean deper		7.10702		0.641447		
CI	an An							-	Adjusted I			0.9950		S.D. depend		0.392974		0.795689		
	an An								S.E. of reg			0.0275		Akaike info		-4.292292		1.080421	0.249524	
					R				Sum squa		sia	0.2384		Schwarz crit		-4.085990		0.363721	0.056837	
sd.	a				A				Log likelih			730.52		Hannan-Qui	nn criter.	-4.210019	,	0.303721		
9	< >				A				Durbin-Wa	uson	stat	1.9829	921					0.274199	0.059750	

63)

Trend + Seasonal Model with AR(4) Disturbances **Residual Plot**





Trend + Seasonal Model with AR(4) Disturbances Residual Correlogram

2

EViews	140	-			-			a. 8	8	Lan		- Salarah				
File Edit Obje	ect Vie	ew P	roc Q	luick	Opti	ons	Add-ins Window Help	ame Freeze Estimate Fore Correlogram of			s					
Rui Vie Vie cd R VDa Si			JU (E) Vie (Vie D	Vie			Date: 11/15/13 Time Sample: 1987M05 20 Included observations Q-statistic probabilitie	e: 15:39 014M12 s: 332 es adjusted for 4 ARN		rm(s)				LIQUORWFTEM reeze Estimate	P::Untitled\ [Forecast Stats Res	ilds
wol 🔊		D S In	D S In					Partial Correlation	1 2	AC 0.005 0.003	PAC 0.005 0.003	Q-Stat 0.0077 0.0100	Prob	:33 111 2014M1 5 after adjus		
rea 🖌 ger 🖌 ger 🖌			=			=			3 4 5 6	-0.057 -0.009 -0.002 0.017	-0.057 -0.008 -0.002 0.014	1.0926 1.1176 1.1190 1.2161	0.290	Coefficient 0.000901	Std. Error 0.000326	
ger 🗹 ger 🗹		R					i ĝi Li pi Li pi	1011 1011 1011	7 8 9	-0.025 0.068 0.059	-0.026 0.068 0.061	1.4316 2.9954 4.1934	0.698 0.559 0.522	-1.47E-06 0.505918 0.507924 0.639393	5.08E-07 0.254586 0.253443 0.252743	-
ger 🗹 ger 🗹 ger 🖉 ger		SSL							10 11 12 13	-0.123 0.015 0.174 -0.023	-0.128 0.024 0.189 -0.043	9.4373 9.5145 19.988 20.170		0.839393 0.810826 0.828685 0.764161	0.252743 0.251705 0.252642 0.252825	
jer 111 jer 111 jer 111		F	-	-		L F			14 15 16	0.149 -0.108 -0.066	0.149 -0.092 -0.071	27.928 31.981 33.495	0.002 0.001 0.001	0.825115 0.701131 0.641447	0.252034 0.254315 0.254736	
sm CD						-			17 18 19	0.043	0.059 -0.074 -0.055	34.146 35.510 37.142		0.795689 0.775719 1.080421 0.363721	0.254403 0.249524 0.249812 0.056837	
Fig 🕡			A				· []		20 21	0.074 0.060	0.044 0.047	39.077 40.351	0.001	0.274199	0.059750	

.

Path = c:\users\francis x. diebo\documents\diebold files\books\econometrics\dataandsoftware\aaaliguorsalesmaster\eviews

Trend + Seasonal Model with Four Lags of Dep. Var.

12/ 18

ĩ

								View Proc Object Print Nam Date: 111-3/13 Time.	Freeze Estimate	Forecast Stats Resi	ds						
								Sample (adjusted): 198					1				
Í	í m	E	m	e	ш	m		Included observations:	332 after adjust	ments					- 0	х	٦
	/ie Vie	 Vie	Vie	Vie			Vie [Variable	Coefficient	Std. Error	t-Statistic	Prob.					
	R Si	D		D			D	TIME	0.000993	0.000303	3.274924	0.0012				ń	
		M		M			M	TIME2	-1.63E-06	4.82E-07	-3.379250	0.0008				=	
11.7	8	D		D			D	D1	0.577182	0.240084	2.404080	0.0168				-	
8		s		s			S	D2	0.579618	0.239820	2.416883	0.0162					1
8		In		In			In	D3	0.667059	0.238627	2.795401	0.0055					1
8		=		=			=	D4	0.894665	0.237447	3.767847	0.0002			_		1
8								D5	0.893728	0.232717	3.840401	0.0001		Statistic	Prob.		1
8		-		=			-	D6	0.827871	0.233806	3.540838	0.0005					1
8								D7	0.865982	0.235247	3.681158	0.0003		.865169	0.0044		1
18								D8	0.791626	0.236419	3.348398	0.0009	Ξ	.036983	0.0026		1
8	2	=						D9	0.739295	0.237199	3.116777	0.0020		.084902	0.0379		
	2	R					-	D10	0.771468	0.236858	3.257093	0.0012		.027856	0.0434		1
8		AS					R	D11	0.830449	0.236573	3.510331	0.0005		.366898	0.0185		
	2	S					A	D12	1.156867	0.236231	4.897183	0.0000		.116444	0.0020		
	7	S					S	LSALES(-1)	0.348107	0.055751	6.243965	0.0000		.419130	0.0007		1
	J)	4			1		S	LSALES(-2)	0.257435	0.053823	4.783041	0.0000		.183296	0.0016		1
	an an	F	1		-		4	LSALES(-3)	0.429234	0.053804	7.977784	0.0000		.285394	0.0011		
	an An	Ρ					F	LSALES(-4)	-0.161633	0.055771	-2.898162	0.0040		.974953	0.0032		1
		=	1		-		Р							.703747	0.0072		1
	10						=	R-squared	0.995335	Mean depend		7.107025		.830854	0.0049		l
	10							Adjusted R-squared	0.995082	S.D. depende	ent var	0.392974		.052611	0.0025		1
	JJ			=				S.E. of regression	0.027559	Akaike info cr	iterion	-4.292292		462497	0.0000		1
	10			R				Sum squared resid	0.238480	Schwarz crite		-4.085990		.491207	0.0000		1
2	•			A				Log likelihood	730.5205	Hannan-Quinr	n criter.	-4.210019	-	.180509	0.0000	-	1

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) = -3.797AR(2) = -3.941AR(3) = -4.080AR(4) = -4.086AR(5) = -4.071AR(6) = -4.058AR(7) = -4.057AR(8) = -4.040



Heteroskedasticity-and-Autocorrelation Consistent (HAC) Standard Errors

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

M EViews	
----------	--

other Tablecont

.....

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

e

19 📖 🛷

2	1	J		Ш) (í m		View Proc Object Print Na		orecast Stats Resi	ds		RWFTE	MP::Untitled\	
u N	/ie Vie	Vie	Vie	vie Vi	e Vi	e Vie	Vie Vie	Dependent Variable:					Estima	te Forecast Stats R	esids
1	R	-		_				Method: Least Squar							
)a	s		D		2		D	Date: 11/15/13 Time							
			M		M		M	Sample: 1987M01 20							
14			D		2		D	Included observations						110	
			S		3		S	HAC standard errors		tlett kernel, Ne	wey-West fixed	1	2014N		
	2		In		1		In	bandwidth = 6.00	000)				er adju	istments	
er 6	2							Variable	Coefficient	Std. Error	t-Statistic	Prob.	efficier	nt Std. Error	
	2				1			TIME	0.007739	0.000197	39.26392	0.0000	00090	1 0.000326	
	7							TIME2	-1.18E-05	6.34E-07	-18.52563	0.0000	47E-0	6 5.08E-07	-
	7		-					D1	6.138362	0.013502	454.6369	0.0000	50591	8 0.254586	
			R				-	D2	6.081424	0.012767	476.3219	0.0000	50792	4 0.253443	
			A				R	D3	6.168571	0.012301	501.4707	0.0000	\$3939	3 0.252743	
1	2		S				A	D4	6.169584	0.010921	564.9371	0.0000	\$1082	6 0.251705	
			S				S	D5	6.238568	0.011627	536.5598	0.0000	\$2868	5 0.252642	
r C			L			-	S	D6	6.243596	0.011226	556.1570	0.0000	76416		
14	ш		F	-		-	L	D7	6.287566	0.010984	572.4238	0.0000	82511	5 0.252034	
	Ш		P				F	D8	6.259257	0.012412	504.2842	0.0000	70113		
	Ш		=	-		-	P	D9	6.199399	0.011517	538.2667	0.0000	64144		
	ш							D10	6.221507	0.011771	528.5468	0.0000	79568		
	Ш							D11	6.253515	0.013011	480.6421	0.0000	7571		
16	ш			=				D12	6.575648	0.013389	491.1133	0.0000	08042		
6 6	ш				2								+ 86372		
9 4					1								27419	9 0.059750	

A Partner of Columb

Structural Change



Structural Change: Gradual

 $\mathbf{v}_t = \beta_1 + \beta_{2t} \mathbf{x}_t + \varepsilon_t$ where $\beta_{1t} = \gamma_1 + \gamma_2 TIME_t$ $\beta_{2t} = \delta_1 + \delta_2 TIME_t$ Then we have: $y_t = (\gamma_1 + \gamma_2 TIME_t) + (\delta_1 + \delta_2 TIME_t)x_t + \varepsilon_t$ We simply run: $v_t \rightarrow c$, Time_t, x_t , TIME_t * x_t

This is yet another important use of dummies. The regression can be used both to test for structural change (*F* test of $\gamma_2 = \delta_2 = 0$), and to accommodate it if present.

Structural Change: Sharp Exogenous

$$y_{t} = \begin{cases} \beta_{1}^{1} + \beta_{2}^{1} x_{t} + \varepsilon_{t}, \ t = 1, ..., T^{*} \\ \beta_{1}^{2} + \beta_{2}^{2} x_{t} + \varepsilon_{t}, \ t = T^{*} + 1, ..., T \end{cases}$$

Let
$$D_{t} = \begin{cases} 0, \ t = 1, ..., T^{*} \\ D_{t} = 1, \ t = T^{*} + 1, ..., T \end{cases}$$

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 simply run:

$$y_t \rightarrow c, D_t, x_t, D_t \times x_t$$



Structural Change: Sharp Exogenous, Continued

The regression can be used both to test for structural change, and to accommodate it if present. It represents yet another use of dummies. The no-break null corresponds to the joint hypothesis of zero coefficients on D_t and $D_t \times x_t$, for which the "F" statistic is distributed χ^2 asymptotically (and F in finite samples under normality).

In the general case, under the no-break null the so-called Chow breakpoint test statistic,

$$Chow = \frac{(e'e - (e'_1e_1 + e'_2e_2))/K}{(e'_1e_1 + e'_2e_2)/(T - 2K)},$$

is distributed F in finite samples (under normality) and χ^2 asymptotically.



Structural Change: Sharp Endogenous

$$MaxChow = \max_{\tau_1 \leq \tau \leq \tau_2} Chow(\tau),$$

where τ denotes sample fraction (typically we take $\tau_1 = .15$ and $\tau_2 = .85$).

The distribution of MaxChow has been tabulated.



Recursive Estimation

$$y_t = \sum_{k=1}^{n} \beta_k x_{kt} + \varepsilon_t$$
$$\varepsilon_t \sim iidN(0, \sigma^2),$$
$$t = 1, ..., T.$$
OLS estimation uses the full sample, $t = 1, ..., T$.

V

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



Recursive Residuals

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

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

The corresponding forecast errors, or recursive residuals, are

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

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

where $r_t > 1$ for all t



Standardized Recursive Residuals and CUSUM

$$w_{t+1,t} \equiv \frac{\hat{e}_{t+1,t}}{\sigma \sqrt{r_t}},$$
$$t = K, ..., T - 1.$$

Under the maintained assumptions,

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

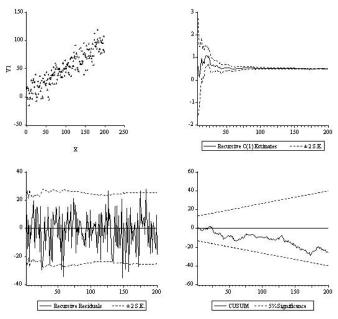
Then

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

is just a sum of *iid* N(0,1)'s (i.e. a Gaussian random walk).

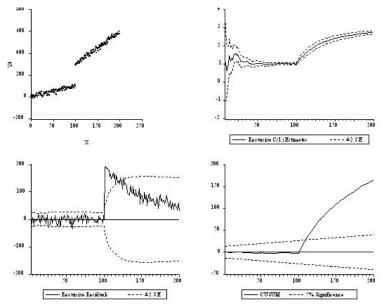


Recursive Analysis, Constant Parameter





Recursive Analysis, Breaking Parameter





Regime Switching I: Observed-Regime Threshold Model

$$y_{t} = \begin{cases} c^{(u)} + \phi^{(u)}y_{t-1} + \varepsilon_{t}^{(u)}, & \theta^{(u)} < y_{t-d} \\ c^{(m)} + \phi^{(m)}y_{t-1} + \varepsilon_{t}^{(m)}, & \theta^{(l)} < y_{t-d} < \theta^{(u)} \\ c^{(l)} + \phi^{(l)}y_{t-1} + \varepsilon_{t}^{(l)}, & \theta^{(l)} > y_{t-d} \end{cases}$$



Regime Switching II: Markov-Switching Model

Regime governed by latent 2-state Markov process:

$$M = egin{pmatrix} p_{00} & 1-p_{00} \ 1-p_{11} & p_{11} \end{pmatrix}$$

Switching mean:

$$f(y_t|s_t) = rac{1}{\sqrt{2\pi\sigma}} \exp\left(rac{-(y_t - \mu_{s_t})^2}{2\sigma^2}
ight).$$

Switching regression:

$$f(y_t|s_t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(y_t - x'_t\beta_{s_t})^2}{2\sigma^2}\right).$$



Rolling Regression for Generic Structural Change



222 / 307

Heteroskedasticity in Time Series

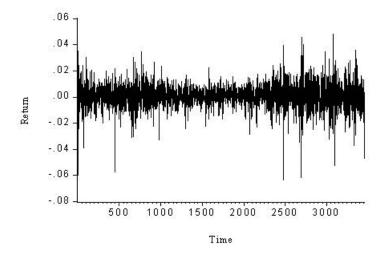


Figure: Time Series of Daily NYSE Returns.



Key Fact 1: Stock Returns are Approximately Serially Uncorrelated

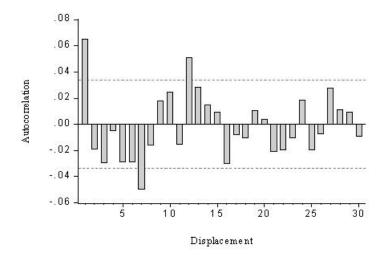


Figure: Correlogram of Daily Stock Market Returns.



Key Fact 2: Returns are Unconditionally Non-Gaussian

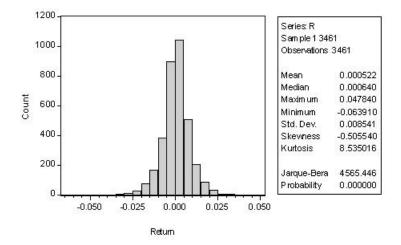


Figure: Histogram and Statistics for Daily NYSE Returns.



Unconditional Volatility Measures

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

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: γ 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 I

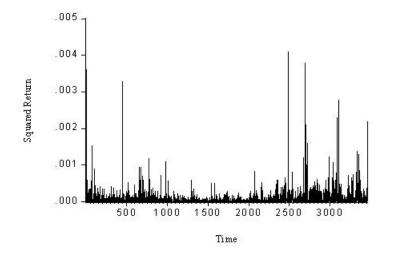


Figure: Time Series of Daily Squared NYSE Returns



Key Fact 3: Returns are Conditionally Heteroskedastic II

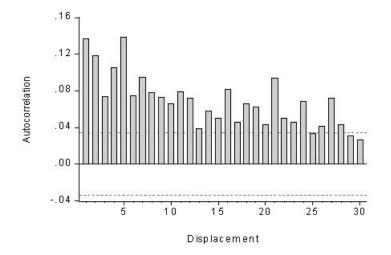


Figure: Correlogram of Daily Squared NYSE Returns.



Background: Financial Economics Changes Fundmentally 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:

$$(r_{it} - r_{ft}) = \alpha + \beta(r_{mt} - r_{ft})$$
$$\beta = \frac{cov((r_{it} - r_{ft}), (r_{mt} - r_{ft}))}{var(r_{mt} - r_{ft})}$$

Conditional CAPM:

$$\beta_t = \frac{cov_t((r_{it} - r_{ft}), (r_{mt} - r_{ft}))}{var_t(r_{mt} - r_{ft})}$$



Asset Pricing III: Derivatives

Black-Scholes:

$$C = N(d_1)S - N(d_2)Ke^{-r\tau}$$
$$d_1 = \frac{ln(S/K) + (r + \sigma^2/2)\tau}{\sigma\sqrt{\tau}}$$
$$d_2 = \frac{ln(S/K) + (r - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}$$

$$P_C = BS(\sigma, ...)$$

(Standard Black-Scholes options pricing)

Completely different when σ varies!



Conditional Return Distributions

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

Are returns conditional variance independent? No way!

Squared returns serially correlated, often with very slow decay.



Linear Models (e.g., AR(1))

$$egin{aligned} & r_t = \phi r_{t-1} + arepsilon_t \ arepsilon_t & au_t \ arepsilon_t \sim \mathit{iid}(0, \ \sigma^2), \ ert \phi ert < 1 \end{aligned}$$

Uncond. mean: $E(r_t) = 0$ (constant) Uncond. variance: $E(r_t^2) = \sigma^2/(1 - \phi^2)$ (constant) Cond. mean: $E(r_t \mid \Omega_{t-1}) = \phi r_{t-1}$ (varies) Cond. variance: $E([r_t - E(r_t \mid \Omega_{t-1})]^2 \mid \Omega_{t-1}) = \sigma^2$ (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")

$$r_t \mid \Omega_{t-1} \sim N(0, h_t)$$
$$h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}$$

$$E(r_t) = 0$$

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

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

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

Well-defined and covariance stationary if $0 < \alpha < 1$, $0 < \beta < 1$, $\alpha + \beta < 1$



GARCH(1,1) and Exponential Smoothing

Exponential smoothing recursion:

$$\hat{\sigma}_t^2 = \lambda \hat{\sigma}_{t-1}^2 + (1 - \lambda) r_t^2$$
$$\implies \hat{\sigma}_t^2 = (1 - \lambda) \sum_j \lambda^j r_{t-j}^2$$

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

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

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

$$\ln L = const - \frac{1}{2} \sum_{t} \ln h_t(\theta) - \frac{1}{2} \sum_{t} \frac{r_t^2}{h_t(\theta)}$$



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



242 / 307

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,

 $rac{r_t}{\sqrt{h_t}} \sim leptokurtic$

So take:

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

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): α effect on volatility

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

 $\gamma \neq$ 0: Asymetric news response $\gamma >$ 0: "Leverage effect"



Regression with GARCH Disturbances

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

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



245 / 307

Time-Varying Risk Premia: GARCH-M

Standard GARCH regression model:

 $y_t = x'_t eta + arepsilon_t$ $arepsilon_t | \Omega_{t-1} \sim N(0, h_t)$

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)

Range: 1 4000 4 Sample: 1 3500 3	3500 obs	Specification Options - Equation specification - Dependent variable followed by list of regressors including ARMA and PDL terms, OR an explicit equation like Y=c(1)+c(2)*X.
☑ ecsdrgarch11 ☑ ecvrgarch11 ☑ fcst ☑ fig1410 ☑ fig1411 湎 figure1410 쿄 figure1411 쿄 figure1412 茴 figure142 茴 figure143 쿄 figure143 쿄 figure144 茴 figure145 茴 figure147 쿄 figure148 쿄 figure149 ➢ history ➢ r2	 ✓ r2smooth ✓ r2sqsmooth ☑ rarch5 ✓ resid ☑ rgarch11 ✓ se I table141 II table142 III table143 ✓ vfcst ✓ yhat 	Estimation settings Method: LS - Least Squares (NLS and ARMA) Sample: LS - Least Squares (NLS and ARMA) (TSLS - Two-Stage Least Squares (TSLS and ARMA) (MM - Generalized Method of Moments LIML - Limited Information Maximum Likelihood and K-Class (COMMR Generalized Method of Moments LIML - Limited Information Maximum Likelihood and K-Class (COMMR Generalized Method of Ordered Choice COMMR - Generalized Method of Data (Refs - Quartile Regression (Including LD) (QM - Generalized Linear Models STEPLS - Stepwise Least Squares
Section_4 New Page /		

Range: 1 4000 Sample: 1 3461		Specification Options	X
2 c ✓ ecsdrgarch11 ✓ fcst © fig1410 © fig1411 ↓ figure141 ↓ figure141 ↓ figure142 ↓ figure142 ↓ figure144 ↓ figure144 ↓ figure144 ↓ figure145 ↓ figure148 ↓ figure149 ↓ figure148 ↓ figure149 ↓ figure149 ↓ figure148 ↓ figure149 ↓ figure148 ↓ figure148 ↓ figure149 ↓ figure148 ↓ figure148 ↓ figure149 ↓ figure148 ↓ figure149 ↓ figure148 ↓ figure149 ↓ figure148 ↓ figure148 ↓ figure148 ↓ figure188 ↓ figure188 ↓ figure188 ↓ figure188 ↓ fig	 ■ r2ar5 ✓ r2smooth ✓ r2sqsmooth ■ rarch5 ✓ resid ■ rgarch11 ✓ se I table141 II table142 II table143 ✓ vfcst ✓ yhat 	Mean equation Dependent followed by regressors & ARMA terms OR explicit e r c Variance and distribution specification Model: GARCH/TARCH Variance regressors Order: ARCH: 1 Threshold order: 0 GARCH: 1 Threshold order: 0 Estimation settings Method: ARCH - Autoregressive Conditional Heteroskedastic Sample: 13461	ARCH-M: None

Range: 1 4000	Details+/- Show Fetch Store De 4000 obs 3461 obs	View Proc Object Prin Dependent Variable: R Method: ML - ARCH (M Date: 12/03/12 Time:	arquardt) - Norr		ecast Stats F	Resids	
 B c M ecsdrgarch11 M ecvrgarch11 M fcst 	 ■ r2ar5 ✓ r2smooth ✓ r2sqsmooth ■ rarch5 	Sample: 1 3461 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)					
G fig1410	resid	Variable	Coefficient	Std. Error	z-Statistic	Prob.	
© fig1411 I figure141	Image: Image	С	0.000641	0.000127	5.039437	0.0000	
III figure1410	table141	Variance Equation					
m figure1411 m figure142 m figure143	⊞ table142 ⊞ table143 ⊠ vfcst	C RESID(-1)^2 GARCH(-1)	1.06E-06 0.067408 0.919717	1.49E-07 0.004959 0.006128	7.127979 13.59218 150.0893	0.0000	
교 figure144 া figure145 i figure146 i figure147 i figure147 i figure148 i figure149 i fistory i r r r r r r r	ifigure145 ifigure146 ifigure147 ifigure148 ifigure149 ifigure149 istory fr	R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood Durbin-Watson stat	-0.000193 -0.000193 0.008542 0.252471 11889.09 1.861386	S.D. dependent var Akaike info criterion Schwarz criterion Hannan-Quinn criter.		0.000522 0.008541 -6.868008 -6.860901 -6.865470	

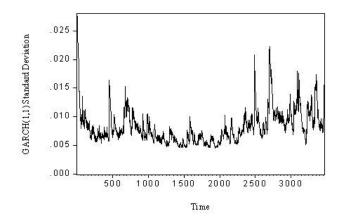


Figure: Estimated Conditional Standard Deviation, Daily NYSE Returns.



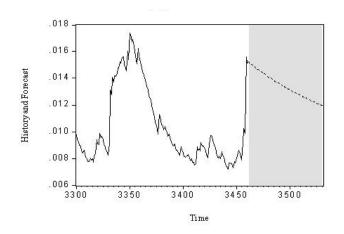


Figure: Conditional Standard Deviation, History and Forecast, Daily NYSE Returns.



A Useful Specification Diagnostic

$$\begin{split} r_t |\Omega_{t-1} &\sim N(0, h_t) \\ r_t &= \sqrt{h_t} \varepsilon_t, \ \varepsilon_t &\sim \textit{iid} N(0, 1) \\ \frac{r_t}{\sqrt{h_t}} &= \varepsilon_t, \ \varepsilon_t &\sim \textit{iid} N(0, 1) \end{split}$$

Infeasible: examine ε_t . iid? Gaussian?

Feasible: examine $\hat{\varepsilon}_t = r_t / \sqrt{\hat{h}_t}$. iid? Gaussian?

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



GARCH(1,1)

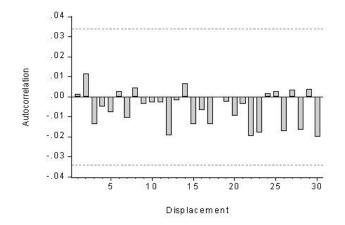


Figure: Correlogram of Squared Standardized GARCH(1,1) Residuals, Daily NYSE Returns.



"Fancy" GARCH(1,1)

Sample: 1 3461 3461 obs	Fourtion Estimation
∎ r2ar5	Equation Estimation
✓ ecsdrgarch11 ✓ r2smooth	Specification Options
✓ ecvrgarch11 ✓ r2sqsmooth	- Mean equation
✓ fcst	Dependent followed by regressors & ARMA terms OR explicit equation:
In fig1410	r cr(-1)
	▼ Std. Dev. ▼
Im figure 141	~Variance and distribution specification
in figure 1410 in table 141	Variance regressors:
Im figure1411 Im table142 Im figure142 Im table143	Model: GARCH/TARCH
m figure143 ⊠ vfcst	Order: ARCH: 1 Threshold order: 1
ingure144	
m figure 145	Error distribution:
m figure 146	Restrictions: None Student's t
m figure 147	-Estimation settings
III figure 148	
🖩 figure149	Method: ARCH - Autoregressive Conditional Heteroskedasticity
M history	Sample: 13461
🖾 r	
⊠ r2	
Section_4 New Page	OK Cancel

"Fancy" GARCH(1,1)

Dependent Variable: R Method: ML - ARCH (Marquardt) - Student's t distribution Date: 04/10/12 Time: 13:48 Sample (adjusted): 2 3461 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)

Variable	Coefficient	Std. Error	z-Statistic	Prob.	
@SQRT(GARCH) C R(-1)	0.083360 1.28E-05 0.073763	0.053138 0.000372 0.017611	1.568753 0.034443 4.188535	0.1167 0.9725 0.0000	
Variance Equation					
C RESID(-1)^2 RESID(-1)^2*(RESID(- 1)<0) GARCH(-1)	1.03E-06 0.014945 0.094014 0.922745	2.23E-07 0.009765 0.014945 0.009129	4.628790 1.530473 6.290700 101.0741	0.0000 0.1259 0.0000 0.0000	
T-DIST. DOF	5.531579	0.478432	11.56188	0.0000	



Causal Predictive Modeling

Consider a standard linear regression setting with K regressors and sample size N.



T-Consistency

We will say that an estimator $\hat{\beta}$ is consistent for a treatment effect ("T-consistent") if $plim\hat{\beta}_k = \partial E(y|x)/\partial x_k, \ \forall k = 1, ..., K$; that is, if $\left(\hat{\beta}_k - \frac{\partial E(y|x)}{\partial x_k}\right) \rightarrow_p 0, \ \forall k = 1, ..., K.$

Hence in large samples $\hat{\beta}_k$ provides a good estimate of the effect on y of a one-unit "treatment" or "intervention" performed on x_k . T-consistency is the standard econometric notion of consistency. OLS is T-consistent under the FIC. OLS is generally **not** T-consistent without the FIC.



And Remember How Stringent the FIC Are!

1. The fitted model is:

$$y = X\beta + \varepsilon$$
$$\varepsilon \sim N(\underline{0}, \sigma^2 I),$$

and it matches the true data-generating process.

- 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 ε 's have constant variance $\sigma^2.$
- 1.5 The $\varepsilon{\rm 's}$ are uncorrelated.
- 2. There is no redundancy among the variables contained in X, so that X'X is non-singular.
- 3. X is a non-stochastic matrix, fixed in repeated samples (old style), or X is a stochastic matrix such that $E(\varepsilon|X) = 0$ (new style).

Non-Causal Predictive Modeling

Again consider a standard linear regression setting with K regressors and sample size N.



P-Consistency

Assuming quadratic loss, the predictive risk of a parameter configuration β is

$$R(\beta) = E(y - x'\beta)^2.$$

Let B be a set of β 's and let $\beta^* \in B$ minimize $R(\beta)$.

We will say that $\hat{\beta}$ is consistent for a predictive effect ("P-consistent") if $plimR(\hat{\beta}) = R(\beta^*)$; that is, if

$$\left(R(\hat{\beta})-R(\beta^*)\right)
ightarrow_p 0.$$

Hence in large samples $\hat{\beta}$ provides a good way to predict y for any hypothetical x: simply use $x'\hat{\beta}$. OLS is effectively **always** P-consistent; we require almost no conditions of any kind!



Correlation vs. Causality, and P-Consistency vs. T-consistency

The distinction between P-consistency and T-consistency is related to the distinction between correlation and causality. As is well known, correlation does not imply causality! As long as x and yare correlated, we can exploit the correlation (as captured in $\hat{\beta}$) very generally to predict y given knowledge of x. That is, there will be a nonzero "predictive effect" of x knowledge on y. But nonzero correlation doesn't necessarily tell us anything about the causal "treatment effect" of x treatments on y. That requires the full ideal conditions. Even if there is a non-zero predictive effect of x on y (as captured by $\hat{\beta}_{LS}$), there may or may not be a nonzero treatment effect of x on y, and even if nonzero it will generally not equal the predictive effect.



Correlation vs. Causality, and P-Consistency vs. T-consistency, Continued

So, assembling things:

P-consistency is consistency for a non-causal predictive effect. (Almost trivially easy to obtain.)

T-consistency is consistency for a causal predictive effect. (Notoriously difficult to obtain.)



An Example of Correlation Without Causality

To take a simple example, suppose that y and x are in fact causally *un*related, so that the true treatment effect of x on y is 0 by construction. But suppose that x is also highly correlated with an unobserved variable z that *does* cause y. Then y and x will be correlated due to their joint dependence on z, and that correlation can be used to predict y given x, despite the fact that, by construction, x treatments (interventions) will have no effect on y.



A Thought Experiment

True DGP:

$$y_i = z_i + \varepsilon_i$$

Suppose also that there exists a variable x such that corr(x, z) > 0.

Fitted OLS Regression Model:

 $y \rightarrow x$

Is $\hat{\beta}_{OLS}$ P-consistent? Is $\hat{\beta}_{OLS}$ T-consistent?



Nonstationarity



265 / 307

Nonstationarity and Random Walks

Random walk:

 $y_t = y_{t-1} + \varepsilon_t$ $\varepsilon_t \sim iid(0, \sigma^2)$

Just a simple special case of $AR(1) \ \phi = 1$



Random Walk with Drift

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

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

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

$$E(y_t) = y_0 + t\delta$$
$$var(y_t) = t\sigma^2$$
$$\lim_{t \to \infty} var(y_t) = \infty$$



267 / 307

Recall Properties of AR(1) with $|\phi| < 1$

– Shocks ε_t have persistent but not permanent effects

$$y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \pmod{\phi_j \to 0}$$

– Series y_t varies but not too extremely

$$\operatorname{var}(y_t) = rac{\sigma^2}{1-\phi^2} \ (ext{note } \operatorname{var}(y_t) < \infty)$$

– Autocorrelations $\rho(\tau)$ nonzero but decay to zero

$$\rho(\tau) = \phi^{\tau} \quad (\text{note } \phi^{\tau} \to \mathbf{0})$$



Properties of the Random Walk (AR(1) With $|\phi| = 1$)

- Shocks have permanent effects

$$y_t = y_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j}$$

- Series is infinitely variable

 $E(y_t) = y_0$ $var(y_t) = t\sigma^2$ $\lim_{t \to \infty} var(y_t) = \infty$

- Autocorrelations ho(au) do not decay ho(au) pprox 1 (formally not defined)



A Key Insight Regarding the Random Walk

- Level series y_t is non-stationary (of course)

- Differenced series y_t is stationary (indeed white noise)!

 $\Delta y_t = \varepsilon_t$

A series is called I(d) if it is non-stationary in levels but is appropriately made stationary by differencing d times.

> Random walk is the key I(1) process. Other I(1) processes are similar. Why?



The Beveridge-Nelson Decomposition

$$y_t \sim I(1) \implies y_t = x_t + z_t$$

 $x_t = random walk$
 $z_t = covariance stationary$

Hence the random walk is the key ingredient for all I(1) processes.

The Beveridge-Nelson decomposition implies that shocks to any I(1) process have some permanent effect, as with a random walk. But the effects are not *completely* permanent, unless the process is a pure random walk.



I(1) Processes and "Unit Roots"

Random walk is an I(1) AR(1) process: $y_t = y_{t-1} + \varepsilon_t$ $\underbrace{(1-L)}_{t} y_t = \varepsilon_t$ deg 1 One (unit) root, L = 1 Δy_t is standard covariance-stationary WN More general I(1) AR(p) process: $\underbrace{\Phi(L)}_{t} y_t = \varepsilon_t$ deg p $[\underbrace{\Phi'(L)}_{t}\underbrace{(1-L)}_{t}]y_t = \varepsilon_t$ (deg p-1)(deg 1) p-1 stationary roots, one unit root Δy_t is standard covariance stationary AR(p-1)



Unit Root Distribution for the AR(1) Process

Key issue (hypothesis) in economics:

I(1) vs. I(0), unit root vs. stationary process

 $\begin{array}{ll} \mbox{When } |\phi| < 1, \\ & d \\ \sqrt{T}(\hat{\phi}_{LS} - \phi) \ \ \rightarrow \ \ N \end{array}$

When
$$\phi=1$$
, d
 $T(\hat{\phi}_{LS}-1) \
ightarrow DF$

Superconsistent Nonstandard limiting distribution Downward finite-sample bias ("Dickey-Fuller bias")



Studentized Statistic

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

Not t in finite samples Not N(0, 1) asymptotically

Trick: Don't run $y_t o y_{t-1}$ Instead run $\Delta y_t o y_{t-1}$



AR(1) With Nonzero Mean Under the Alternative

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

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

Random walk null vs. mean-reverting alternative

Studentized statistic $\hat{\tau}_{\mu}$



AR(1) With Trend Under the Alternative

$$(y_t - a - bt) = \phi(y_{t-1} - a - b(t-1)) + \varepsilon_t$$
$$y_t = \alpha + \beta t + \phi y_{t-1} + \varepsilon_t$$
where $\alpha = a(1 - \phi) + b\phi$ and $\beta = b(1 - \phi)$
$$H_0: \phi = 1 \text{ (unit root)}$$
$$H_1: \phi < 1 \text{ (stationary root)}$$

Studentized statistic $\hat{\tau}_{\tau}$

"Random walk with drift" vs. "stat. AR(1) around linear trend" "Stochastic trend" vs. "deterministic trend"



Stochastic Trend vs. Deterministic Trend

NYSE Returns .08 .06 .04 Autocorrelation .02 .00 -.02 -.04 --.06 10 15 20 25 30 5

Correlogram

Displacement



AR(p)

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

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

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

Studentized statistic $\hat{\tau}$ is still relevant



AR(p) With Nonzero Mean Under the Alternative

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

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

where $\alpha = \mu (1 + \sum_{j=1}^{p} \phi_j)$

Studentized statistic $\hat{\tau}_{\mu}$ is still relevant



AR(p) With Trend Under the Alternative

-

$$(y_t - a - bt) + \sum_{j=1}^p \phi_j (y_{t-j} - a - b(t-j)) = \varepsilon_t$$
$$y_t = k_1 + k_2 \ t + \rho_1 y_{t-1} + \sum_{j=2}^p \rho_j (y_{t-j+1} - y_{t-j}) + \varepsilon_t$$
$$k_1 = a \left(1 + \sum_{i=1}^p \phi_i \right) - b \sum_{i=1}^p i \phi_i$$
$$k_2 = b \left(1 + \sum_{i=1}^p \phi_i \right)$$

Under the null hypothesis, $k_1 = -b \sum_{i=1}^p i \phi_i$ and $k_2 = 0$

Studentized statistic $\hat{\tau}_{\tau}$ is still relevant



"Trick Form" of ADF in the General AR(p) Case

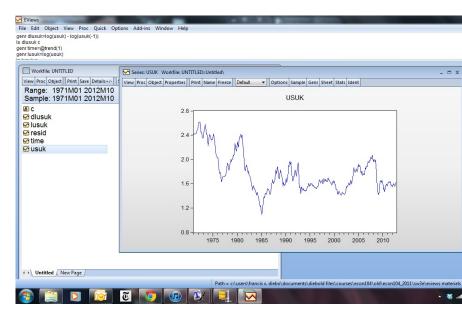
$$(y_t - y_{t-1}) = (\rho_1 - 1)y_{t-1} + \sum_{j=2}^{k-1} \rho_j (y_{t-j+1} - y_{t-j}) + \varepsilon_t$$

– Unit root corresponds to $(
ho_1-1)=0$

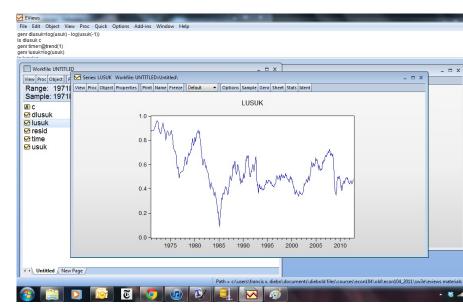
 Use standard automatically-computed *t*-statistic (which of course does not have the *t*-distribution)



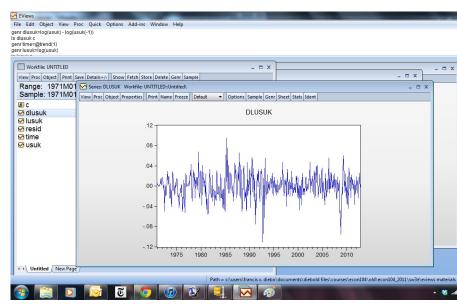
USD/GBP Exchange Rate, 1971.01-2012.10



Log USD/GBP Exchange Rate, 1971.01-2012.10



Change in USD/GBP Exchange Rate, 1971.01-2012.10



Trend-Stationary Model

File Edit Object View Proc Ouick Ontions Adduins Window Help

U

EViews

					-	= ×	
Dependent Variable: L Method: Least Squares Date: 11/12/12 Time: Sample: 1971M01 201	JSUK 16:30 2M10	Estimate For	ecast Stats R	esids			
Variable	Coefficient	Std. Error	t-Statistic	Prob.			
C TIME	0.495082 -0.000579	0.008211 4.52E-05	60.29146 -12.80648	0.0000 0.0000			
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood F-statistic Prob(F-statistic)	0.146788 10.77337 251.9151	S.D. depende Akaike info cr Schwarz crite Hannan-Quin	nt var terion rion n criter.	-0.978870			
	View Proc Object Print Dependent Variable L Method Least Squares Date: 11/12/12 Time: Sample: 1971M01201 Included observations: Variable C TitleE R-squared Adjusted R-squared SE of repression Sum squared resid Log likelihood F-statistic	View Proc Object Print Name Freeze Dependent Variable USUK Method Less Squares Name Freeze Data Variable Cefficient Sample 197 VM01 2012M0 Included doservations. 502 Variable Coefficient Coefficient Cefficient Adjusted R-squared 0.245498 S. Gregression 0.146748 Sum oguared resident def 10.177337 Log Kleinhood 251.9151 Pratatistic 164.0000	View Proc Object Print Name Freeze Estimate From Dependent Variable LUSUK Method Lession Lession Sample 1712 Time Food Sample 1710140 2012/101 Included observations 502 Variable Coefficient SId Error C 0.495082 0.008211 Included observations R-squared 0.244995 SD depending 4.52E-503 Sumsel or forgression 0.44788 SD depending Sc forgression Sum squared reid 0.7373 Schwarz of the observations Log likelihood 251.9151 Hanna-Ouin -statistic 164.0000 Durinh-Watter 164.0000 Durinh-Watter Log likelihood Durinh-Watter	Dependent Variable. LUSUK Method. Lass Squares Sample 197 (2004) Method. Lass Squares Sample 197 (2004) Included observations: 502 Variable Coefficient Stage 197 (2004) Included observations: 502 Variable Coefficient Quarter 0.09521 60.29146 0.24595 R-squared 0.24595 August of regression 0.14798 Sum squared evid 10 0.14737 Sch fregression 0.14798 Jum squared evid 10 27373 Log likelihood 251.9151 Hanna-Quino riterion 147408	View Proc Object Print Name Freeze Estimate Forecast Status Relids Dependent Variable: LUSUK Method: Less Squares Data Status Relids Relids Data Display Status Relids Relids Relids Variable Coefficient Std. Error I-Statistic Prob. C 0.495082 0.008211 60.29146 0.0000 R-spuared -0.40595 4.525-05 -12.80648 0.0000 R-spuared 0.244959 Buan dependent var 0.584170 -0.995870 Sum squared reid 0.14788 Analue inforderinor -0.995870 -0.998870 Sum squared reid 0.14788 Analue inforderinor -0.978870 -0.988004 Sum squared reid 0.14778 Analue inforderino -0.978870 -0.988004 Sum squared reid 0.14778 Analue inforderino -0.978870 -0.988004 Sum square reid 0.147780 Analue inforderino -0.978870 -0.988004 0.026995 Sum square reid	View Proc Object Print Name Freeze Estimate Forecast Stats Resids Dependent Variable: LUSUK Method: Less Siguraes Data Stats Resids Stats Resids Data Induct Siguraes Data Stats Resids Stats Stats Resids Data Inducted observations: 502 Stats Stats Prob. Stats Stats<	View Proc Object Prim Name Preeze Estimate Forecast Stats Resids Dependent Variable: LUSUK Method Least Stguares Date: 1/12/12 Time: 16:30 Sample: 197/10/10 2012M10 Induded observations: 502 Variable Coefficient Std Error IsStatistic Prob. Coefficient Std Error IsStatistic Prob. Coefficient Std Error IsStatistic Prob. Auguste Rouge Coefficient Std Error IsStatistic Prob. Rougerd Coefficient Std Error IsStatistic Prob. Sc orgenession Octoor Std Action Coefficient Statistic Prob. Sum squared Coefficient Octoor Octoor Octoor Octoor Octoor Statistic Prob. Sum squared roid 10 77737 Schwarz creation - 0.995678 Sum squared roid 10 7737 Schwarz creation - 0.995677 Log likelihood 251 9151 Hannan-Ouino criterio - 0.989084 F-statistic 164 0000 Outroh-Viston Station 0.026995 <

Trend-Stationary Model



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

Is lusuk c

Is lusuk c



Difference-Stationary Model (Random Walk With Drift)

EViews

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

0

õ 🧔 🖸

Is lusuk c Is lusuk c time

Is dlusuk c

L = Nort	Hypothesis: LUSUK has a unit root		
Workfile: UNTITLED	Enonthesis: LUSUK has a unit root Equation: UNTITLED Workfile: UN View [Proc] object] [Print] Name [Prezz Dependent Variable: DLUSUK Method: Least Squares Date: 11/12/12 Time: 17.11 Sample (adjusce): 1971/1002 2012/M Included observations: 501 after adju	e) [Estimate Forecast Stats Resids	
⊠ resid	Variable Coefficient	Std. Error t-Statistic Prob.	
⊠ time ⊠ usuk	C -0.000804	0.001077 -0.746262 0.4559	
	R-squared 0.00000 Adjutet R-squared 0.00000 SE of regression 0.024116 Sum squared reside 0.20000 Log likelihood 1155771 Durbin-Watson stat 1.281071	S.D. dependent var 0.024116 Akaike info criterion -4.609863 Schwarz criterion -4.601447 Hannan-Quinn criter4.606561	
Untitled / New Page /			
		Path = c:\users\francis x. diebo\documents\	diebold files\courses\econ104\old\econ104_2011\sw3e\eviews materials

🔺 💸 all

B

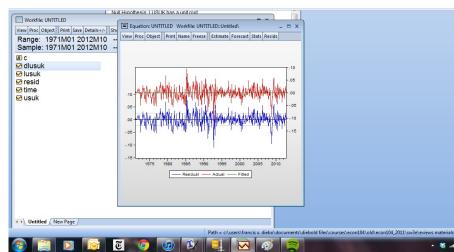
M 🔊

Difference-Stationary Model (Random Walk With Drift)

EViews

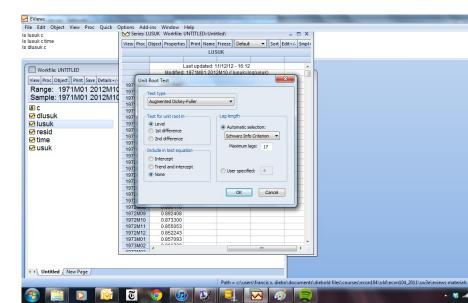
File Edit Object View Proc Oujck Options Add-ins Window Help Is lusuk c Is lusuk c time le dlusuk c

-



- 22

DF Tests - Option Screen



ADF Test, Allowing for Trend Under the Alternative

usuk c usuk c time	Series: LUSUK Workfile: UNTITLED::Untitled\							
flusuk c	View Proc Object Properties Print Name Freeze Sample Genr Sheet Graph S Augmented Dickey-Fuller Unit Root Test on LUSUK					Stats I		
Workfile: UNTITLE	Null Hypothesis: LUSUK has a unit root Exogenous: Constant, Linear Trend						x	1
View Proc Object Pr	Lag Length: 1 (Automatic - based on SIC, maxlag=17)							
Range: 1971N				t-Statistic	Prob.*		r: *	
Sample: 1971N	Augmented Dickey-Fulle	r test statistic		-2.815425	0.1923	Ξ		
₿ C	Test critical values:	1% level		-3.976517	0.1020			
M dlusuk		5% level 10% level		-3.418834 -3.131954				
M lusuk		10% level		-3.131954				
M resid	*MacKinnon (1996) one-sided p-values.							
⊠ time ⊠ usuk								
	Augmented Dickey-Fuller Test Equation Dependent Vrastable: DL(USUK) Method: Least Squares Date: 111/212: Time: 17.23 Sample (adjusted): 197/1403 2012/1410 Included observations: 500 after adjustments							
	Variable	Coefficient	Std. Error	t-Statistic	Prob.			
	LUSUK(-1)	-0.019346	0.006871	-2.815425	0.0051			
	D(LUSUK(-1))	0.362419 0.012131	0.041782	8.674026 2.314242	0.0000			
	@TREND(1971M01)	-7.40E-06	8.02E-06	-0.923110	0.3564			
	R-squared	0.134675	Mean depende S.D. depende Akaike info cr	ent var iterion	-0.000816 0.024139 -4.746653			
	Adjusted R-squared S.E. of regression Sum squared resid Log likelihood F-statistic	0.250093 1190.663	Schwarz crite Hannan-Quin Durbin-Watso	n criter.	-4.712936 -4.733422 1.930399			



The Lag Operator

$$Ly_t = y_{t-1}$$

$$AR(1)$$
 illustration:

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

$$y_t = \phi L y_t + \varepsilon_t$$

 $y_t - \phi L y_t = \varepsilon_t$

$$(1-\phi L)y_t = \varepsilon_t$$

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

 $\Phi(L)$ is a polynomial of degree 1 in the L



The Lag Operator, Continued (AR(p))

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t$$
$$y_t = \phi_1 L y_t + \dots + \phi_p L^p y_t + \varepsilon_t$$
$$y_t - \phi_1 L y_t - \dots - \phi_p L^p y_t = \varepsilon_t$$
$$(1 - \phi_1 L - \dots - \phi_p L^p) y_t = \varepsilon_t$$
$$\Phi(L) y_t = \varepsilon_t$$

 $\Phi(L)$ is a polynomial of degree p in the lag operator

Roots of $\Phi(L)$ are important for nature and stability of dynamics



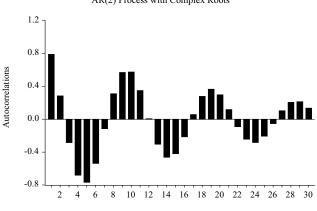
Covariance Stationarity in AR(p)

AR(p) is

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t$$
$$(1 - \phi_1 L - \dots - \phi_p L^p) y_t = \varepsilon_t$$
$$\Phi(L) y_t = \varepsilon_t$$

Stable if the p roots of $\Phi(L)$ are outside the unit circle





Population Autocorrelation Function AR(2) Process with Complex Roots

Displacement



Big Data



Selection, Shrinkage and Derived Inputs

"Data-rich" environments

"Wide data"

Dimensionality reduction is key: Selection, shrinkage, more.



Selection Methods

• All Subsets

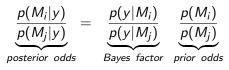
Quickly gets hard as there are 2^{K} subsets of K regressors!

- Greedy Forward Selection
- Start with intercept only and add the new regressor that minimizes RSS, then take the one variable model and add the new regressor that minimizes RSS, etc.

• Greedy Backward Selection

Start with K-variable model and remove the "least significant" variable, then take that K-1-variable model and remove the "least significant" variable, etc.

• Posterior odds and marginal likelihood:



• Information criteria:



Model Selection by MSE (or R^2)

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

$$R^{2} = 1 - \frac{\frac{1}{T} \sum_{t=1}^{T} e_{t}^{2}}{\frac{1}{T} \sum_{t=1}^{T} (y_{t} - \bar{y})^{2}} = 1 - \frac{MSE}{\frac{1}{T} \sum_{t=1}^{T} (y_{t} - \bar{y})^{2}}$$

Selection by MSE (or R^2) produces in-sample over-fitting



Model Selection by s^2 (or \bar{R}^2)

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

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

Selection by s^2 (or \overline{R}^2) still produces in-sample over-fitting



Information Criteria for Model Selection

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

"Oracle property" No over-fitting (asmptotically)!

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



Information Criteria for Model Selection

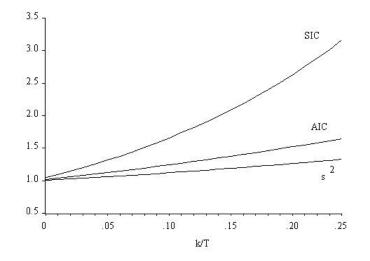


Figure: Degrees-of-Freedom Penalties



Shrinkage Methods

• Bayesian regression:

$$\hat{\beta}_{bayes} = \omega_1 \hat{\beta}_{MLE} + \omega_2 \beta_0$$

- Ridge Regression:
- $\hat{\beta}_{ridge} = (X'X + \lambda I)^{-1}X'y$
 - Penalized regression:

$$\tilde{\beta} = \operatorname{argmin}_{\beta_1 \dots \beta_K} \left(\sum_{t=1}^T \left(y_t - \sum_{i=1}^K \beta_i x_{it} \right)^2 + \lambda \sum_{i=1}^K |\beta_i|^q \right)$$

- penalties smooth at the origin produce shrinkage - penalties non-differentiable at the origin produce selection - q = 2 is ridge; q = 1 is lasso; $q \rightarrow 0$ is selection.



Aside: Review of Principal Components Analysis (PCA) for ata (X Matrix) Description

Think of a wide X matrix and how to "reduce" it.

X'X eigendecomposition:

 $X'X = VD^2V'$

The j^{th} column of V, v_j , is the j^{th} eigenvector of X'XDiagonal matrix D^2 contains the descending eigenvalues of X'X

First principal component:

$$z_1=Xv_1$$
 $var(z_1)=d_1^2/7$

(maximal sample var among all possible l.c.'s of columns of X)

In general:

$$z_j = Xv_j \perp z_{j'}, j' \neq j$$

 $var(z_j) \leq d_j^2/T$



Derived Input Variable Methods I: PC Regression (PCR) and its First Problem

> "Factor-Augmented Regression" "Distill, then select then proceed"

Ridge and PCR are both shrinkage procedures.

BUT:

Ridge effectively includes all PC's and shrinks according to sizes of eigenvalues associated with the PC's.

PCR effectively shrinks some PCs completely to zero (those not included) and doesn't shrink others at all (those included).

Awkward



Derived Input Variable Methods I (Continued): PC Regression (PCR) and its Second Problem

No "supervision"



Derived Input Variable Methods II: Partial Least Squares Regression (PLS)

