ECN402—summary

Christian Braathen

November 2018

Contents

Learning outcomes:

- be able to interpret the results of empirical analyses and to be able to choose between competing regression models
- be able to conduct quantitative analysis where several factors can affect an outcome variable simultaneously
- understand what assumptions econometric models are based on
- be able to use STATA for doing econometric analysis, produce do-files and log-files, import data from excel, and produce tables and figures.
- be able to interpret and critically deal with empirical work in applied econometrics
- know the structure and requirements for a master thesis, and be able to develop a research question
- be able to choose and apply an appropriate scientific method for analysing the research question
- understand the ethical issues in collection and interpretation of data
- have a good background for more advanced econometric courses

Perfoming an economic analysis, which this summary has been constructed similar to:

1. Formulate an economic model
2. From the economic model, formulate an econometric model
3. Collect data for the problem at hand
4. Estimate the econometric model
5. Use the estimates for statistical inference

Keys from the summary lecture:

- You have to be able to choose between competing methods and models.
- You need to defend the assumptions.
- You may have to do some calculations from the Chow test.
- Heteroskedasticity is important for standard errors. If heteroskedasticity is present, then you have likely a misspecified model, too.

- Cochrane–Orcutt or Prais–Winsten if you don't know the autocorrelation function.
- Panel data: "Tell me why we should use panel data models." Be very specific on what a_i captures.
- He's not fond of complicated models.
- Use OLS as often as possible.
- He chooses topics he thinks is interesting.
- He has an applied focus.

1 Assumptions

1.1 What the assumptions mean

1. **Linear in parameters (population)**: $\forall i = 1, ..., n.\ y_i = \beta_0 + \sum_{j=1}^{n} \beta_j x_i j + u_i$
2. **Random sampling** (of size n): could be violated if:
   - People opt-in to answer
   - Sample doesn't represent the population, like surveying students but trying to examine the population in the country.
3. **Sample variation in** x: the variable cannot have only one value in all the observations.
   - Additionally, no perfect collinearity (i.e. no exact linear relatinoship among the independent variables, $x_i = \alpha x_j$)
   - Perfect collinearity would give 0 in the denominator of $Var(\beta_{)} = \frac{\hat{\sigma}}{SST_j(1-R_j)}$
4. **Zero conditional mean (population)**: $\mathbb{E}[u|\vec{x}] = 0$.
   - That is, the **average** value of u is independent of the x value.
   - This is key to causality because it can be violated in so many ways.
   - More specifically: if the OLS assumptions hold, and the zero conditional mean is the most crucial one here, then we can interpret the results causally.
   - This assumption is a big one to make.

- If this is violated, then we can't know whether it is changes in x that drives change in y (which we need for the *ceteris paribus*) or whether it's u that does that.
- Implies that $\mathbb{E}[y|\vec{x}] = \beta_0 + \sum_{i=1}^{n} \beta_i x_i$. In other words, that the *population* regression function is a linear function of $\vec{x}$.
- We never know this one for sure, but it's a critical assumption.
- For time series: $\mathbb{E}[u_t|X] = 0 \forall t$. Note that is applies for x values in **all** time periods (strictly exogeneous) and not only within the time period (contemporaneously exogeneous).
- If it breaks for time series, then it's often because of omitted variables or measurement errors. In economics, the strict exogeneity assumption is often violated. For instance, policy variables are often affected by past events.
- For time series, this means that once we have included the lags of x that we did in the model, no other lags of x will affect $mathbbE[u_t|X] = 0 \implies mathbbE[y_t|X] = 0$

5. **Homoskedasticity (population)**: $Var(u|(x_1, ..., x_k)) = \sigma^2 \implies Var[y|x] = \sigma^2$. (if the equality doesn't hold, then we got *heteroskedasticity*
    - That is, constant variance for all $\vec{x}$ values.
    - Or another way: the variance of y given x does not depend on the values of the independent variables.
    - Needed to evaluate efficiency.
    - We don't know σ^2, so we must estimate it with $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{u}_i^2 = \frac{SSR}{n-2}$
    - For time series: $Var(u_t|X) = Var(u_t) = \sigma^2 \implies Var[y|x] = \sigma^2$. Note that is applies for x values in **all** time periods and not only within the time period.
6. **No serial correlation**
    - $Cov[u_t, u_s|X] = 0 \forall t \neq s = \mathbb{E}[u_t, u_s] = 0$
    - This one is often violated in ways that can make usual statistical inference very unreliable.
    - This requirement of unobservables having a constant variance over time could easily be false: policy changes effect the variability of interest rates, for instance.

7. **Independent and identically distributed error term.**
   - Stronger than assumption 4 and 5.
   - With this assumption, the OLS estimators are the *minimum variance unbiased estimators*.
   - This assumption talks about the error term, but the assumption implies that $\hat{\beta}_j$ is normally distributed.
   - Let's us talk about the shape of the distribution. We now assume $\hat{\beta}_j \equiv N(\beta^{H_0}, Var(\hat{\beta}_j))$.
   - So we can invoke the central limit theorem now.
   - We can also standardize the expression above to $\frac{\hat{\beta}_j - \beta_j^{H_0}}{\sigma[\hat{\beta}_j}$.
   - But we don't know σ, so we cannot use that. Instead, we use the standard error $\hat{\sigma} = \sqrt{\frac{SSR}{n-k-1}}$. But then we cannot use the normal distribution. Instead we use the t distribution and say that $t = \frac{\hat{\beta}_j - \beta_j^{H_0}}{\hat{\sigma}[\hat{\beta}_j} \equiv t_{n-k-i}$.
   - For time series: u_t is independent of $\vec{x}_t$ and i.i.d. as $N(0, \sigma^2)$.

1.1.1 Other assumptions

- Related to *measurement errors*: measurement errors in y is statistically independent of every explanatory variable.

1.2 Simple Linear Regression

1. Linear in parameters
2. Random sampling
3. Sample variation in x
4. Zero conditional mean
5. Homoskedasticity

- Assumption 1–3 is needed to derive the estimator formula.
- If any of assumptions 1–4 fails, then we generally get biased estimators.

- If assumptions 1–5 are in place, then $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$
- If assumptions 1–5 are in place, then OLS is BLUE

1.3 Multiple Linear Regression

1. Linear in parameters
2. Random sampling
3. Sample variation in x & no perfect collinearity
4. Zero conditional mean
5. Homoskedasticity
6. Error term is independent and identically distributed

These assumptions only hold when we include the intercept β_0 in the regression.

Assumption 1 to 6 is called the *classic linear model (CLM) assumptions*

1.4 Time Series Regression

1. Linear in parameters

   ~~Random sampling~~ (unnecessary. Now, the sample size equals the number of time periods)
2. Sample variation in x & no perfect collinearity
3. Zero conditional mean (strictly)
4. Homoskedasticity (strictly)
5. **No serial correlation**
6. Error term is independent and identically distributed

Anything that causes the unobservable at time t to be correlated with any of the explanatory variables in any time period causes the assumption about zero conditional mean (TSR4) to fail.

Under $TSR1 - -5$, $\hat{\sigma}$ is an unbiased estimator of σ

Under $TSR1--6$, we can do inference on time series.

1.4.1 Asymptotic time series assumptions

~~Linear in parameters~~ The stochastic process is stationary, weakly dependent, and following a linear model.

~~Random sampling~~ (unnecessary. Now, the sample size equals the number of time periods)

1. Sample variation in x & no perfect collinearity
2. Zero conditional mean (contemporaneously—that is, for only the respective time period—since if it holds in this period, it will hold for all.)
3. Homoskedasticity (contemporaneously)
4. **No serial correlation**
5. Error term is independent and identically distributed

2 Consequences

Beware that we are talking about the estimator, the processing function, being unbiased, efficient, and consistent, not the estimate (the value that the estimator outputs).

Furthermore, beware that each "dart" below is not the observation from a sample, but the estimate $\hat{\beta}$. So the dartboard is the result of drawing tons of samples—one dart per sample—and we make a distribution of all these samples. If the estimator is unbiased, for instance, most samples' estimate hit the population parameter quite well.

- An **unbiased** estimator: the expected value of the estimator's probability distribution equals the parameter we try to estimate. That is, $\forall j = i, ..., k.\ \mathbb{E}[\hat{\beta}_j] = \beta_j$.
  - In other words, if the true parameter is the bulls-eye, then we're able to hit that consistently.
  - Upward bias: $\mathbb{E}[\hat{\beta}] > \beta$.
  - Downward bias: $\mathbb{E}[\hat{\beta}] < \beta$.
  - Biased towards zero: $0 < |\mathbb{E}[\hat{\beta}]| < |\beta|$
- An **efficient** estimator: the minimum variance unbiased estimator.
  - You may not be able to hit bulls-eye all the time, but on average you're doing really well and the spread of the darts are less than for any other attempt.
- A **consistent** estimator: as the sample size increases, the estimates "converge" to the true value of the parameter being estimated—the sampling distribution of the estimator becomes increasingly concentrated at the true parameter value.
  - In other words, you may or may not hit bulls-eye that often in the beginning when you throw only a few darts on the board, but you are consistent in hitting the bulls-eye when you can throw many darts on the board.

Moreover:

- Preferably, we want the estimator to be unbiased and efficient. Then we got $BLUE$: Best Linear Unbiased Estimator. This is true if assumption 1–5 holds for OLS because the OLS estimator is then both unbiased and efficient. Additionally, we want it to be consistent.
  - E: estimator (the function)
  - UE: unbiased estimator.

- LUE: linear unbiased estimator. The estimator can be expressed as a linear function of the data of the dependent variable: $\tilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i$
- $BLUE$: best linear unbiased estimator. **the LUE with the smallest variance.**

- Since we want an unbiased *and* efficient estimator, we can use the mean squared error (MSE) instead because it captures both bias and inefficiency: $MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$
- If we have bias, we may get the wrong significance (significant although it wasn't).
- If we have high, but not perfect, collinearity, we get less efficiency since the denominator in $Var(\beta_) = \frac{\hat{\sigma}}{SST_j(1-R_j)}$ will go towards 0, meaning that the variance, and hence the standard error, will go towards ∞

2.1 Examples

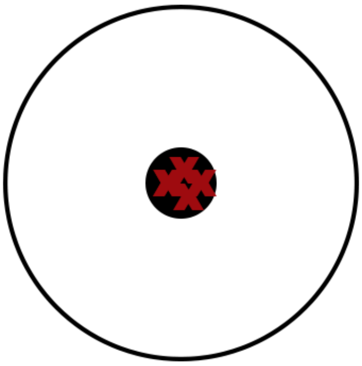


Figure 1: Unbiased

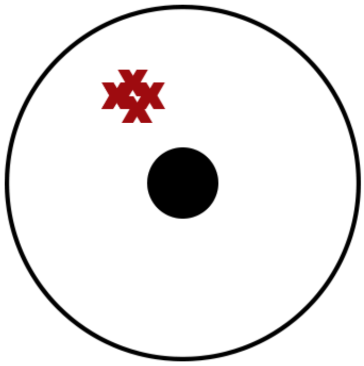


Figure 2: Biased (but efficient)

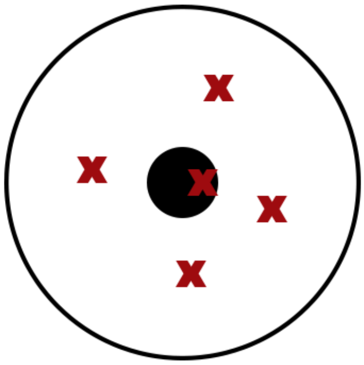


Figure 3: Inefficient (but unbiased—on average, the darts hit the target quite well)

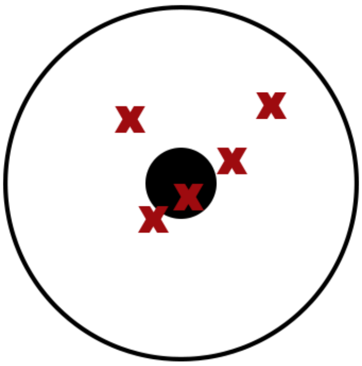


Figure 4: Inconsistent

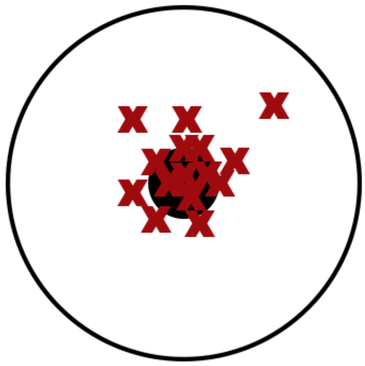


Figure 5: Consistent

Q1.2. If the OLS estimator of β in the regression model Y=α+βX+γZ+u is unbiased, this means:
A) that β is BLUE
B) that $E(\hat{\beta})=0$
C) that $E(\beta)=\hat{\beta}$
D) that $E(\hat{\beta})=\beta$

Figure 6: V17Q1.2. Correct answer is *D*.

Q1.2. If the OLS estimator of β in the regression model Y=α+βX+γZ+u is unbiased, this means:

A) that $E(\beta)=\hat{\beta}$

B) that β is BLUE

C) that $E(\hat{\beta})=\beta$

D) that $E(\hat{\beta})=0$

Figure 7: Correct answer is C.

3 The Input: Data

3.1 Sampling

- Nonrandom samples can lead to biases in OLS.
- When the sample selection is correlated with the error term, OLS is generally biased and inconsistent. This could happen, for instance, if the sample/data availability is selected based on someone's decision or similar, causing the decision to be potentially related to unobserved factors.
- But we don't necessarily get OLS problems if the nonrandom sample selection is "exogeneous".
  - If sample selection is based on the y variable being larger than or smaller to a certain value, then bias always occurs.
    - Example: Only asking people above a certain income level.
  - If sample selection is based on the x variables, this doesn't cause problems for the MLR assumptions except for $MLR2$ as long as we have enough x variables and a large enough n (since the regression function will be the same for any subset of the population).
    - But it does require a fixed rule to include or exclude, though. If we don't have it, it's not so clear-cut though.

3.1.1 Stratified sampling

An alternative sampling method in which the population is split up in stratas (mutually exclusive and collectively exhaustive groups [MECE]) and some groups are asked more often than others relatively to the population size. If the grouping is included as an explanatory variable, then OLS is unbiased and consistent.

3.2 Measurement errors

We're not able to measure a quantitative variable correctly.

- That is, The unobserved variable has a quantitative meaning but we cannot observe it—so we use an imprecise measure instead.

- We just proceed with OLS despite potential measurement errors, but we get more noise because of it—so our estimates becomes less precise relative to the true population.
- Measurement errors in y may give bias in β_j.
- Measurement errors in x is a more serious matter **because a measurement error in *one* x variable causes inconsistency in *all* estimators.**

For time series:

- For time series: $\mathbb{E}[u_t|X] = 0 \forall t$. Note that is applies for x values in **all** time periods (strictly exogeneous) and not only within the time period (contemporaneously exogeneous).
- If it breaks for time series, then it's often because of omitted variables or measurement errors. In economics, the strict exogeneity assumption is often violated. For instance, policy variables are often affected by past events.

Example:

- We want *actual* annual family savings.
- But we get *reported* annual family savings.

Example:

- We want *marginal* tax rate.
- But we get *average* tax rate.

Example:

- We want actual scrap rate.
- But we get reported scrap rate, which may be systematically lower for those who get funds to show that the grant works.

Example:

- We want actual marijuana consumption
- Non-smokers will report correctly, but smokers may report erroneously.

3.3 Missing data

Key: If you are missing a data point (a cell) in an observation (a row), then the entire observation must be skipped (except for panel data).

- Estimators that only use whole observations are called *complete cases estimators.*
- If data are *missing completely at random (MCAR)*, then it causes no statistical problems. Then proceed as follows:
  1. Create $z_{ik} = \begin{cases} x_{ik}, & !is.na(x_{ik}) \\ 0, & is.na(x_{ik} \end{cases}$
  2. Create $m_{ik} = \begin{cases} 1, & !is.na(x_{ik}) \\ 0, & is.na(x_{ik} \end{cases}$
  3. Regress y on $x_1, ..., x_k - 1, z_k, m_k$, **excluding the variabel with missing observations,** x_k. This would give us unbiased and consistent estimators, although they will be less robust.
- But MCAR data are usually unrealistic.
  - If sample selection is based on the y variable being larger than or smaller to a certain value, then bias always occurs.
    - Example: Only asking people above a certain income level.
  - If sample selection is based on the x variables, this doesn't cause problems for the MLR assumptions except for $MLR2$ as long as we have enough x variables and a large enough n (since the regression function will be the same for any subset of the population).
    - But it does require a fixed rule to include or exclude, though. If we don't have it, it's not so clear-cut though.

3.4 Scaling

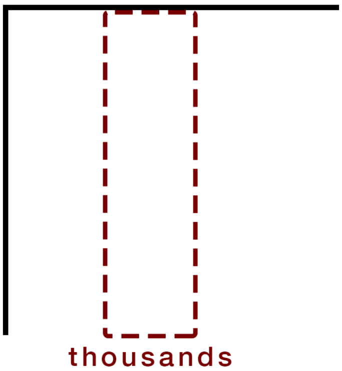


Figure 8: Before

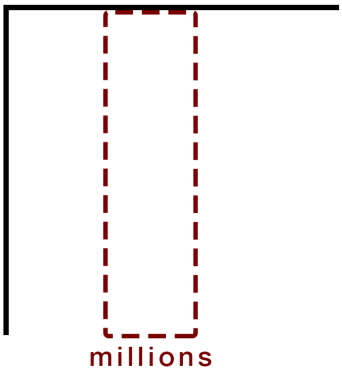


Figure 9: After

Scaling can change the interpretation of the model and it's usually done to make the interpretation easier. Standardizing the coefficients is often interesting. Here we exemplify through scaling by $\cdot\frac{1}{10}$):

- if the x variable is scaled, then its coefficient gets inversely scaled ($\cdot 10$).
- if the y variable is scaled, then the entire right-hand side is scaled similarly.
- if both the x and y variable is scaled by the same amount, then the intercept will be scaled while the x variables coefficient will not since the scale and the inverse scale eliminates each other.

3.4.1 Example 1

Dividing all x_t values by 10 would make x_t's coefficient 10 times larger than the original value. An example with $x_t = 10$ before dividing, giving $y_t = 14.52$:

$$\begin{aligned} 14.52 &= 5.83 + 0.869 \cdot 10 \\ \overset{\text{Dividing } x \text{ by } 10}{\Longleftrightarrow} 14.52 &= 5.83 + 8.69 \cdot \frac{10}{10} \\ \therefore y_t &= 5.83 + 8.69 \cdot x_t^* \end{aligned} \tag{1}$$

The equality must hold, therefore making x_t 10 times smaller must make the coefficient 10 times larger.

Furthermore, dividing all y_t values by 10 would make both the intercept and x_t's coefficient one-tenth of its original value:

$$
\begin{aligned}
14.52 &= 5.83 + 0.869 \cdot 10 \\
\overset{\text{Dividing } y \text{ by } 10}{\Longleftrightarrow} \frac{14.52}{10} &= \frac{5.83}{10} + \frac{0.869}{10} \cdot 10 \\
\therefore y_t^* &= 0.583 + 0.0869 \cdot x_t
\end{aligned}
\tag{2}
$$

Dividing both the x_t and y_t values by 10 will make the intercept one-tenth of its original value but the two effects on x_t's variable will eliminate each other and the coefficient values will be equal to the original value.

$$
\begin{aligned}
14.52 &= 5.83 + 0.869 \cdot 10 \\
\overset{\text{Dividing } x \text{ by } 10}{\Longleftrightarrow} 14.52 &= 5.83 + 8.69 \cdot \frac{10}{10} \\
\overset{\text{Dividing } y \text{ by } 10}{\Longleftrightarrow} \frac{14.52}{10} &= \frac{5.83}{10} + \frac{8.69}{10} \cdot \frac{10}{10} \\
\therefore y_t^* &= 0.583 + 0.869 \cdot x_t^*
\end{aligned}
\tag{3}
$$

3.4.2 Example 2

Q1.8. Changing the units of measurement of the dependent variable by dividing it by 1000 will NOT affect:
A) the estimated intercept
B) R-squared
C) the total sum of squares
D) the estimated standard errors

Figure 10: Correct answer is B. The intercept does change, so it's obviously not correct. SST does change because $(y_i^* - \bar{y}^*)^2 < (y_i - \bar{y})^2$. Furthermore, the standard error changes: Since scaling y will scale the coefficient, and since we obviously don't get any more significance from our data from scaling, the standard errors must change too to keep the t statistic constant. But R^2 does not change. Yes, both SSR and SST will change, but they will change proportionally the same, as is seen in the notes below. We therefore end up with the same R^2, which of course makes sense: if we aim for a high R^2, scaling should obviously not make that score any more beneficial.

3.5 Outliers and influential observations

Key: report results both with and without outliers since it's a hard topic to handle.

- Outliers: the observation diverge substantially from the overall pattern. It can have relatively extreme x or y values, or both.
- Influential observations: outliers that impact the slope estimates. We evaluate these observations by regressing with and without the outlier.
- Common especially in small data sets.
- Try the models both with and without the outliers since they can have large impact on OLS estimates.
- Alternatively, use the *least absolute deviations (LAD)* method instead of OLS since these estimations are less sensitive to outliers. This method is becoming increasingly popular, usually as a supplement to OLS.
- Sometimes, outliers may come from a different population.
- Other times, outliers may occur because a member in the population is very different.
- Outliers provide important information because the variation in the x variables increases.
- It's more useful to report standardized residuals than regular residuals when reporting on outliers.

3.6 Various

- If you could choose, you'd want more variation in the regressors because it allows you to more confidently pin down the relationship between y and $\vec{x}$ in your regression.

4 The Processing: The Model

Population model: $y = \beta_0 + \beta_1 X + u$

Estimated from sample: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{u}$

Nested model: model $i \subset$ model j.

Non-nested model: model $i \not\subset$ model j.

Virtually all applied researchers try different models, estimation techniques, or subsets of data until they get the results they expected. This data mining practice violates the assumptions in our analysis because we're using the same set of data in a model specification search instead of observing a sample and estimating that model once.

For time-series, we study:

- Lagged explanatory variables
- Lagged dependent variables
- Lagged residuals

4.1 OLS

- Note: OLS is an estimation method, not a model.

Some properties:

- $\sum_{i=1}^{n} \hat{u}_i = 0$
- $\sum_{i=1}^{n} x_i \hat{u}_i = 0$
- $(\bar{x}, \bar{y})$ is always on the regression line.

When doing inference, we need $Var[\beta_1] = \frac{\sigma^2/n}{\sigma_x^2}$. So here we see why a lot of variation in x is good and why a lot of variation in the estimator is bad.

4.1.1 Example

Q1.3. The goal of OLS is to find the values of the estimated parameters that:
A) Maximize the sum of the squared stochastic error terms.
B) Minimize the sum of the squared stochastic error terms.
C) Maximize the sum of the squared residuals.
D) Minimize the sum of the squared residuals.

Figure 11: Correct answer is D.

4.2 Transformations: logarithms

- Whenever we use *levels*, we need the units.
- *level–level*: increasing x_1 by 1 unit changes y by β_1 units, everything else equal.
- *log–level*: increasing x_1 by 1 unit changes y by approximately $100 \cdot \beta_1$ percent, everything else equal. Also called semi-elasticity of y with respect to x.
- *level–log*: increasing x_1 by approximately 1 percent changes y by $\frac{1}{100} \cdot \beta_1$ units, everything else equal. **Beware that we divide by 100 here.** Practically never used.
- *log–log*: increasing x_1 by approximately 1 percent changes y by approximately β_1 percent, everything else equal. **Beware that we don't multiply by 100 on *log–log*** Also called the constant elasticity model.
- *percentage points*: if we have percentage points on both sides (i.e. ranging from 0 to 1 (100 %), then we **do** multiply by 100 if the range is 0–1—but beware that a one unit increase means increasing for instance from nothing (0) to everything (1, 100 percentage points), and *don't* multiply anything if range is 0–100.
- In summary:
  - level–level: keep as it is unless for percentage points that goes in the range 0–1, which is multiplied with 100.
  - log–level: multiply β_1 by 100.
  - level–log: divide β_1 by 100.
  - log–log: keep as it is.

Moreover:

- When choosing the functional form, think for instance about how we usually talk about the variable.
- We could also consider whether the transformation makes y and x's relationship linear.
- A common trick with log-transformations: change 0's to 1 before log-transforming because $\ln(1) = 0$.
- Log may yield a distribution that is closer to normal in certain situations.
- Log-transformations has an *approximate* percentage interpretation. The exact interpretation is $100(e^{\beta_1} - 1)$.
- Log is popular when there is a lot of variation and where outliers can influence since the variation is reduced by log-transforming the variable.
- However, when y is close to 0, like they are when they are fractions, then log-transformations would *increase* the variation. Then it's generally not a good idea to log-transform.
- When working with demand, we should talk about percentage changes in prices and percentage changes in quantity (price elasticity of demand). So use log-transformed prices and quantities.
- Note that it's harder to predict with $\ln y$ than y.
- Common scenarios for *log*:
  - Positive money amounts
    * Wages
    * Sales
    * Market value
    * Etc.
  - Number of people
    * Population
    * Number of employees

    * School enrollment
    * Etc.

- Common scenarios for *level*:
  - Years
    * Education
    * Tenure
    * Age
    * Etc.
  - 0 or negative values
- Common scenarios when we could use either:
  - Proportions or percentages (percentage point interpretation)
    * Unemployment rate
    * Etc.

4.2.1 Examples

Q1.4. A fitted regression equation is given by log(CEOsalary)=0.75+0.01firmprofit.
Firmprofit and CEOsalary are both measured in 1 000 NOK.
Which of the following is correct:
A) Increasing firm-profit with 1% increases CEOsalary with 0.01%
B) Increasing firm-profit with 10 000 NOK increases CEOsalary with 0.1%
C) Increasing firm-profit with 1 000 NOK increases CEOsalary with 1%
D) Increasing firm-profit with 10 000 NOK increases CEOsalary with 100 NOK

Figure 12: V17 Q1.4. Correct answer is C

Q1.5. A fitted regression equation is given by log(CEOsalary)=0.7+0.25log(firmprofit). Firmprofit is measured in 100 000 NOK, CEOsalary is measured in 1000 NOK.
Interpret the slope parameter:
A) Increasing firm-profit with 100 000 NOK increases CEOsalary with 25 000 NOK
B) Increasing firm-profit with 100 000 NOK increases CEOsalary with 25%
C) Increasing firm-profit with 1 % increases CEOsalary with 25%
D) Increasing firm-profit with 1% increases CEOsalary with 0.25%

Figure 13: V17 Q1.5. Correct answer is D. Beware that we don't multiply by 100 on *log–log*, only for one *level* and one *log*.

Q1.4 The share of votes (number between 0 and 1) for Donald Trump is regressed on Trump's share (number between 0 and 1) of total election campaign expenditures for Clinton and Trump taken together. The result is: *Trump_voteshare*=0.32+0.28*Trump_Campaignshare*.
The correct interpretation of the coefficient on the variable *Trump_Campaignshare* is:
A) Increasing Trump's share of total campaign expenditures with 10 percentage points increases the votes for Trump with 2.8 percentage points.
B) Increasing Trump's share of total campaign expenditures with 1 percentage point increases the votes for Trump with 2.8 percentage points.
C) Increasing Trump's share of total campaign expenditures with 1 percent increases the votes for Trump with 2.8%.
D) A one-unit increase in Trump's share of total campaign expenditures increases his vote share by 0.028 percent.

Figure 14: Correct answer is A. Beware that when we have percentage points on both sides, we **do** multiply with 100.

4.3 Quadratics

- Often used to capture decreasing or increasing marginal effects, which the sign of the coefficient reveals.
- The downside is that the curve turn around eventually, which may affect a large portion of the population.
- Interpret with partial derivation. But beware that we're not interested in the marginal effect for every observation, but for the average of that explanatory variable.
- You should calculate the turning point and see if it makes sense.
- You should also calculate the slope at various x values and see whether the square of a variable is practically important.

4.4 Interaction terms

- Interpret with partial derivation. But beware that we're not interested in the marginal effect for every observation, but for the average of that explanatory variable.
- We're not that often interested in $\beta_3 x_1 x_2$. Instead, we're often interested in $\beta_3(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)$, which would make the coefficients interpreted as estimated average partial effects.
- Use a simple t test to determine the statistical importance.
- Calculate some values of the variables to see whether the interaction is practically important.

4.5 Control variables

A control variable enters a regression in the same way as an independent variable—the method is the same.

But the interpretation is different.

Control variables are usually variables that you are not particularly interested in, but that are related to the dependent variable. You want to remove their effects from the equation.

Suppose, for example, you were interested in the difference in height of people from different ethnic groups. You could gather a sample of people from different ethinc groups and measure them and compare the heights. But you'd probably want to control for some other variables that are known to relate to height (e.g. sex). It's known that men are taller then women. Even if your samples are random, the sample from each group will not have the same proportion of men and women - so you want to add that as a variable

One kind of control variable is a mediator. An example here is that of the damage caused by a fire. If you are interested in the relationship between the number of firemen and the amount of damage, you want to control for the size of the fire.

4.6 Dummy variables

Interpretation: the change relative to the excluded category.

ILLUSTRATION

- Dummy variables is a good way to capture qualitative information in nominal and ordinal variables. Examples:
  - Race
  - Gender
  - Classification
  - Country names
  - Types of ownership
  - Geographical location
  - Industry classification of firms
  - Political systems
  - Conflict status of countries
  - Etc.
- We can't include all of the categories: we must exclude one (or else we get in the *dummy variable trap*, which gives us multicollinearity issues).
- Interpretation: the change relative to the excluded category.
- The dummy as a stand-alone term ($\delta_1 D_1$) if you want the dummy to affect the intercept. Interpretation: "The difference in y between <what $D_1 = 1$ is> and <what $D_1 = 0$ is> *given* the same $\vec{x}$ values (and u). The differences occurs either because of <what the dummy stands for> or because of factors that are associated with the dummy and that we haven't controlled for."
- Add the dummy as an interaction term ($\delta_1 D_1 x_1$) if you want the dummy to affect the slope.
- Add the dummy as a stand-alone and interaction term ($\delta_1 D_1 + \delta_2 D_1 x_1$) if you want the dummy to affect the slope *and* the intercept.
- Note that if you remove the constant from the regression, the original constant value will be moved over to the dummy.
- When dummies reflect *choices of individuals* or *other economic units*, then causality is an issue again: we don't know what comes first (so a question of reversed causality).

- If you have multiple categories, say five, then make dummies of each one.
- You can easily add interactions among dummies as well, like (single,married) $\times$ (male,female). The interpretation of this interaction would be to test if gender differentials also depends on marital status.
- For time series, dummies can be used to isolate different intervals that may be systematically different form other periods covered by a data set. For example $D^{t \geq 1985}$.
- For time series, dummies can also be used for seasons, like one dummy per quarter of the year.

4.6.1 Examples

Q1.5. In the following regression model, *Exper* is work experience measured in years and U is a dummy variable for union membership

$$Wage_i = \beta_0 + \beta_1 Exper_i + \beta_2 U_i + \beta_3 \, (Exper_i \times U_i) + u_i$$

To test that the model is the same for union members and non-members, you must test
A) the hypothesis that $\beta_3 = 0$.
B) the joint hypothesis that $\beta_0 = 0, \beta_1 = 0$.
C) the joint hypothesis that $\beta_1 = 0, \beta_2 = 0, \beta_3 = 0$
D) the joint hypothesis that $\beta_2 = 0, \beta_3 = 0$

Figure 15: Correct answer is D.

4.7 Instrumental Variables (IV)

Key 1: IV is used as an alternative to OLS in cross-sections, and it's used when we have potential endogeneity issues. The key to successful empirical analysis using instrumental variables is finding valid instruments.

Key 2: Key assumptions: IV 1) uncorrelated with u (exogeneity, $cov(z,u) = 0$, hard to test and requires intuition) and 2) correlated with x (relevance, $cov(z,x) \neq 0$, easy to test), and 3) should have no independent role on y.

Key 3: assume we have $y = \beta_0 + \beta_1 x + u$, $x = \pi_0 + \pi_1 z + v$. v is what causes the endogeneity problem, so the idea behind the 2SLS is to use the problem-free component $x = \pi_0 + \pi_1 z$

Key 4: in addition to adding IVs, you should also discuss whether we need more variables (omitted variable bias) since it reduced the standard errors.

Key 5: if a good proxy exists, an IV approach may not be necessary at all.

Key 6: You must plug in all included *exogeneous* variables in the 2nd step of the 2SLS procedure.

Key 7: we need at least as many excluded exogeneous variables (IVs) as included endogeneous variables.

Key 8: endogeneity problems are more the rule than the exception, so IVs, proxy variables, and panel data are often used.

Key 9: in the covariances, we index on i, but that's a details presumably lacking from this course. So I don't think it's a biggie.

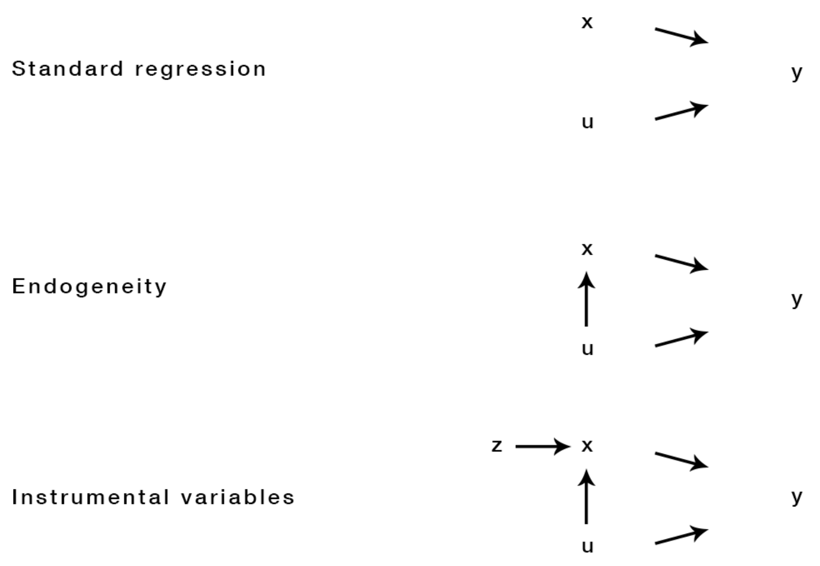


Figure 16

The IV estimator gives the causal effect if we have at least one valid instrumental variable

Q1.10. The estimation of a logarithmic wage regression by the instrumental variable estimator gives the following fitted regression equation: lwage = 10.4 + 0.043 *educ*, where wage is the hourly wage in NOK and *educ* is the years of education. Interpret the slope coefficient.

A) A one year increase in education increases wages by 0.043 percent. Since this is the instrumental variable estimate, we can interpret it as a causal effect of education on wages.
B) A one year increase in education increases wages by (10.4+4.3) NOK per hour. Since this is the instrumental variable estimate, we can interpret it as a causal effect of education on wages.
C) A one year increase in education increases wages by 4.3 percent. The instrumental variable estimator gives the causal effect if we have at least one valid instrumental variable.
D) The coefficients estimated by instrumental variable estimation are the same as by ordinary least squares.

Figure 17

- $\ln(wages) = \beta_0 + \beta_1 edyc0(\beta_2 ability + u)$. You can't say that your parent's education should affect your wage, but it affects your education level. But you cannot use IQ as an IV since it's related to ability, which is in the error term. But you could include IQ as a proxy.
- He stresses that we should write the actual variable names.
- For *wages* as y, let's study some potential IV variables:
  - IQ: relevant for education, but not exogeneous since it's correlated with ability.
  - Mother's and father's education: relevant for education, somewhat exogeneous—could be that this is picking up the effect of whether they're able to affect wages.
  - Distance to nearest university: relevant (could go into the opportunity cost calculations of travel distance?), and maybe exogeneous. But beware that this implies you got no mobility at all.
  - Number of siblings: relevant (employer don't care how many siblings you got) and exogeneous (in the US at least, in which you have to pay for education, so more siblings means less education for you).
- **Most of your time should go to discuss the exogeneity assumption. Ifyou don't have exogeneity, you get biased estimators. So the censor will try to kill most of your exogeneity arguments as much as possible.** We always have that $Var^{OLS} < Var^{IV}$. So OLS is always preferred since this means that OLS is more efficient. IV is used only if OLS assumptions doesn't hold.
- Use as many IVs as possible *as long as they are good.*
- In general, family background variables are often used as IVs for education, but it's

not problem-free. Example: $exam\_score = \beta_0 + \beta_1 skipped\_classes + u$. Skipped is likely correlated with u since it depends on ability, motivation, etc. Potential IVs are distance between living quarters and campus, for instance.

- Remember to talk about whether the IV is positively or negatively related.
- The IV estimator is never entirely unbiased ince x and u are correlated, and the bias may be substantial for small samples.
- Ignore R^2 in IV: it could become negative and goodness-of-fit is not a factor since IV methods are intended to provide better estimates of the *ceteris paribus* effect of x on y when x and u are correlated.
- IV estimates can have large standard errors, especially if $cov(z, x)$ is weak. That is visible because we can rewrite $\beta_1 = \frac{cov(z,y)}{cov(z,x)}$ to $p\lim \hat{\beta}_1 = \beta_1 + \frac{cov(z,u)\sigma_u}{cov(z,x)\sigma_x} \frac{cov(z,y)}{cov(z,x)}$. So if $cov(z, x)$ is low/weak (i.e. more or less insignificant) don't use it because the inconsistency of the IV estimator becomes very large.
- You must plug in all *exogeneous* variables in the 2nd step of the 2SLS procedure. Why? Because exogeneous variables (the exogeneous ones from the OLS plus the additional IVs) are uncorrelated with u, any linear combination of these variables will also be uncorrelated with u. Therefore, we need to include them all in the 2nd step because we can then find the linear combination that is the most highly correlated with y_2.
- Multicollinearity can be a serious problem in 2SLS.
- Rule of thumb: with $F > 10$, we're able to rule out weak instruments.
- If the estimate you get from OLS is significantly different from the estimate you get from 2SLS, then that variable has an endogeneity problem.
- One can also use 2SLS on pooled cross sections and panel data.
- An example: Effect on test scores of class size. Some California schools are forced to close for repairs because of a summer earthquake. Districts closest to the epicenter are most severly affected. A district with some closed schools needs to "double up" its students, temporarily increasing class size. This means that distance from the epicenter satisfies the condition for instrument relevance because it is correlated with class size. But if distance to the epicenter is unrelated to any of the other factors affecting student performance (such as whether the students are still learning English), then it will be exogeneous because it is uncorrelated with the error term.

Thus the IV, distance to the epicenter, could be used to circumvent omitted variables bias and to estimate the effect of class size on test scores.

4.7.1 The two-step least squares method

- Extremely popular, just behind OLS.
- Allows more IVs than endogeneous variables.

1. Run $x = \pi_0 + \pi_1 z + v$ and store the fitted values, $\hat{x}$. This is called the 2SLS estimator.
2. Use $\hat{x}$ instead of x in $y = \beta_0 + \beta_1 x + u$.

4.7.2 Testing if endogeneity is a problem: the Durbin–Wu–Hausman test

Key: we have an endogeneity problem if $cov(x, u) \neq 0$). If it is 0, then use OLS since its estimators are BLUE.

1. Start with original model and regress the variable you suspect is endogeneous on IVs and the included exogeneous variables.
2. Find $\hat{x}$ and $\hat{v}$.
3. Run $y = \beta_0 + \beta_1 x + \delta \hat{v} + e$
4. Test $H_0 : \delta = 0$ (no endogeneity problem, use OLS) against $H_A : \delta \neq 0$ (endogeneity problem, use IV).

4.7.3 Testing if you have multiple suspected endogeneous variables

1. For all suspected endogeneous variables, obtain residuals
2. F-test: $H_0 : \delta_j = 0 \forall j$, $H_A : \exists j.\ \delta_j \neq 0$.

4.7.4 Overidentification

- I.e. we have more instruments than we need to for estimating the parameters consistently. That is, *we have more instruments than endogeneous explanatory variables.* Use the following test whenever we have more instruments than we need.

1. Run regression and obtaint $\hat{u}$.

2. Regress $\hat{u}$ on all exogeneous variables (included and excluded), and get R^2.

3. Under H_0: all IVs are uncorrelated with u, $n \cdot R^2 \equiv \chi^2_q$, where q is the number of excluded IVs minus the number of included endogeneous vairables. H_A: at least some of the IVs are not exogeneous.

I think we'll be asked this on the exam:

- What does overidentification means when we talk about instruments?
  - It means to have more IVs than included endogeneous variables
  - It's an **advantage** because we can test for overidentification, which means we could test for the quality of the instruments and we get better predictions, meaning we also get more narrow standard errors.

4.7.5 2SLS with heteroskedastisticity

1. Regress $\hat{u}$ on $z_1, z_2, ..., z_m$

2. Run F-test

3. H_0: homoskedastic, H_A: heteroskedastic (z_j are jointly significant)

4.7.6 Using multiple IVs

Using multiple IVs is beneficial because we can run some tests relevant for exogeneity after all.

Alternative 1: run test and check for significance for each IV individually.

Alternative 2: if number of IVs are greater than the number of endogeneous explanatory variables:

1. Run 2SLS using all instruments and get residuals $\hat{u}$.

2. Regress $\hat{u}$ on all IVs and all included exogeneous explanatory variables from the original model.

3. If $n \cdot R^2 \geq \chi^{cri}$, go for H_A (i.e. some are endogeneous) (H_0: all IVs are exogeneous, i.e. all the IVs we evaluated are usable.

4.7.7 Multiple endogeneous explanatory variables

- Still need more IVs than endogeneous explanatory variables
- Same procedure as with one endogeneous regressor:
  1. Regress *each of* the endogeneous variables *on all instruments* and *all included exogeneous explanatory variables* in the structure model. So one regression per endogeneous is the fundamental one.
  2. Use the predicted values of the endogeneous variables instead of the actual values in the structural model.

4.7.8 System of equations

We're interested in system of equations like supply and demand because we're particularly interested in price elasticity since we need to know how many customers we lose/get if we increase/reduce prices.

- Demand: $q = \beta_0 + \beta_1 p + u$
- Supply: $p = \alpha_0 + \alpha_1 q + v$
- u can be interpreted as, "I meet this lovely girl who is into French wine, so I start liking it too." You think there are no endogeneity problem, but wait!
- The fact that you start liking it affects the demand q, and that will affect the supply equation as an explanatory variable. This means that higher q will give higher p as an dependent variable.
- But look at this! p is an explanatory variable in the demand equation.
- This means that in the demand equation, we got a relationship between the explanatory variable p and the error term u. Therefore, we have an endogeneity problem after all!
- Another example: a system of equations with supply and demand on butter. Weather can be an IV because below-average rainfall in a dairy region could impair grazing and thus reduce butter production at a given price, so dairy-region rainfall satisfies the condition for instrument relevance. But dairy-region rainfall should not have a direct influence on the demand for butter, so the correlation between dairy-region rainfall and u_i would be zero—that is, dairy-region rainfall satisfies the condition for unstrument exogeneity.

4.8 Trends

Two main types of trend:

- $y_t = \alpha_0 + \alpha_1 t + u_t$, which is a linear time trend (α_1 measures the change in y_t from one period to the next due to the passage of time).
- $\ln(y_t) = \beta_0 + \beta_1 t + u_t$, which is an exponential time trend.
  - Perfect for growth statistics, like GDP who grows by a certain percentage each year.

But you can add other types of trends as well, like quadratics ($y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + u_t$). But don't add too many polynomials: we want to capture *large* movements in y that's not explained by x.

Two ways to handle trend:

- Add `t` in the regression as a variable
- De-trend the dataset before running the regression:
  1. Regress $y_t = \beta_0 + \sum_{j=1}^{n} \beta_j x_{tj} + \delta_1 t + u$ and save $\hat{y}_t$
  2. Run $y_t - \hat{y}_t = \hat{u}$.
  3. Save $\hat{u}$ as y_t^*.
  4. Regress $y_t^* = \textit{no intercept} + \sum_{j=1}^{n} \beta_j x_{tj}^* + \textit{no trend since it's detrended} + u^*$ and obtain the $\hat{\beta}$'s, which is interpreted as coming from a regression without a time trend.

Moreover:

- **Factors in u_t may be correlated with the x variables.** If we don't handle this, then we get **spurious regressions**:
  - we falsely conclude that changes in one variable are statistically significant and causes changes in another variable.
  - But what is really going on is that these two variable are related through a third variable instead. It can often happen in time series with $I(0)$ variables: that we don't have to difference them for the time series to become stationary. More on that on page 85.

- Solution: include the time trend in the regression. Then the model explicitly recognizes that y_t may be growing or shrinking over time for reasons unrelated to the x variables.

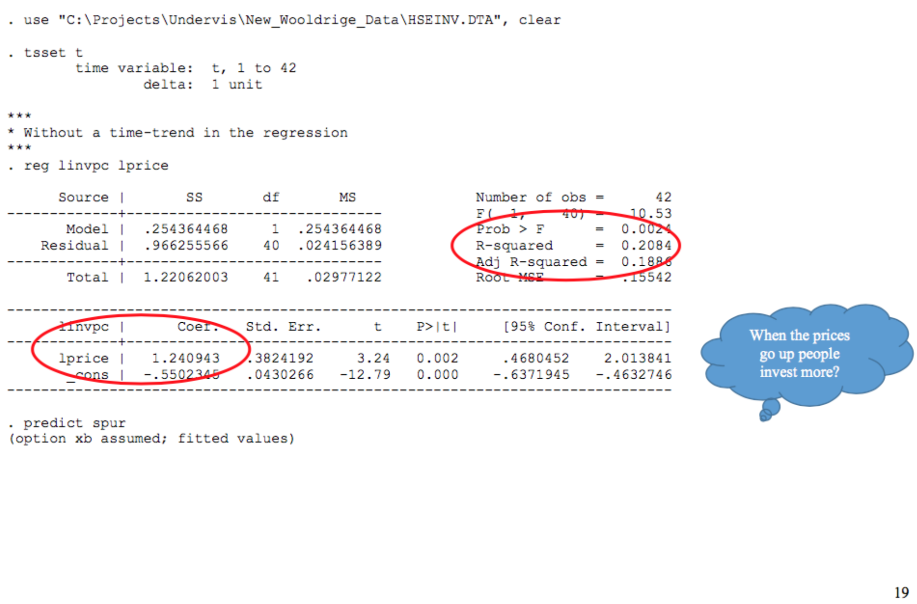
- If you want to know how good the model actually is, then we need to go for the method of de-trending the variables—by taking away the trend, we're using only the variation relative to the trend lines and get a lower but more correct R^2—it's artificially high when the dependent variable is trending.

- Adding a trend can make a key x variable more significant.

- It's a good idea to include a trend in the regression if any independent variable is trending, even if y_t is not.

- We're interested in utilizing the variation in the time series, minus the trend.

4.8.1 Example



```
. use "C:\Projects\Undervis\New_Wooldrige_Data\HSEINV.DTA", clear

. tsset t
        time variable:  t, 1 to 42
                delta:  1 unit

***
* Without a time-trend in the regression
***
. reg linvpc lprice

      Source |       SS       df       MS              Number of obs =      42
-------------+------------------------------           F(  1,    40) =   10.53
       Model |  .254364468     1  .254364468           Prob > F      =  0.0024
    Residual |  .966255566    40  .024156389           R-squared     =  0.2084
-------------+------------------------------           Adj R-squared =  0.1886
       Total |  1.22062003    41   .02977122           Root MSE      =  .15542

------------------------------------------------------------------------------
      linvpc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |   1.240943   .3824192     3.24   0.002     .4680452    2.013841
       _cons |  -.5502345   .0430266   -12.79   0.000    -.6371945   -.4632746
------------------------------------------------------------------------------

. predict spur
(option xb assumed; fitted values)
```


Figure 18

```
***
* with a time-trend in the regression
***
. reg linvpc lprice t
      Source |       SS       df       MS              Number of obs =      42
-------------+------------------------------           F(  2,    39) =   10.08
       Model |  .415945108     2  .207972554           Prob > F      =  0.0003
    Residual |  .804674927    39   .02063269           R-squared     =  0.3408
-------------+------------------------------           Adj R-squared =  0.3070
       Total |  1.22062003    41   .02977122           Root MSE      =  .14364
------------------------------------------------------------------------------
      linvpc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |  -.3809612   .6788352    -0.56   0.578    -1.754035    .9921125
           t |   .0098287   .0035122     2.80   0.008     .0027246    .0169328
       _cons |  -.9130595   .1356133    -6.73   0.000    -1.187363   -.6387557
------------------------------------------------------------------------------
```

A more reasonable slope coeff., but insign.

```
***
* detrend ln(price)  [i.e. subtract the predicted values from the actual ones]
***
. reg lprice t
      Source |       SS       df       MS              Number of obs =      42
-------------+------------------------------           F(  1,    40) =  107.57
       Model |  .120404022     1  .120404022           Prob > F      =  0.0000
    Residual |  .044774132    40  .001119353           R-squared     =  0.7289
-------------+------------------------------           Adj R-squared =  0.7222
       Total |  .165178154    41  .004028735           Root MSE      =  .03346
------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   .0044173   .0004259    10.37   0.000     .0035565    .0052781
       _cons |   -.188386   .0105121   -17.92   0.000    -.2096318   -.1671401
------------------------------------------------------------------------------
```

20

Figure 19

```
. predict prd_lpr
(option xb assumed; fitted values)

. gen detrn_lp = lprice - prd_lpr

***
* detrend ln(invest)  [i.e. subtract the predicted values from the actual ones]
***
. reg linvpc t

      Source |       SS       df       MS              Number of obs =      42
-------------+------------------------------           F(  1,    40) =   20.19
       Model |  .409446973     1  .409446973           Prob > F      =  0.0001
    Residual |  .811173061    40  .020279327           R-squared     =  0.3354
-------------+------------------------------           Adj R-squared =  0.3188
       Total |  1.22062003    41   .02977122           Root MSE      =  .14241

------------------------------------------------------------------------------
      linvpc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   .0081459   .0018129     4.49   0.000     .0044819    .0118098
       _cons |  -.8412918    .044744   -18.80   0.000    -.9317228   -.7508608
------------------------------------------------------------------------------

. predict prd_linv
(option xb assumed; fitted values)

. gen detrn_li = linvpc - prd_linv
```

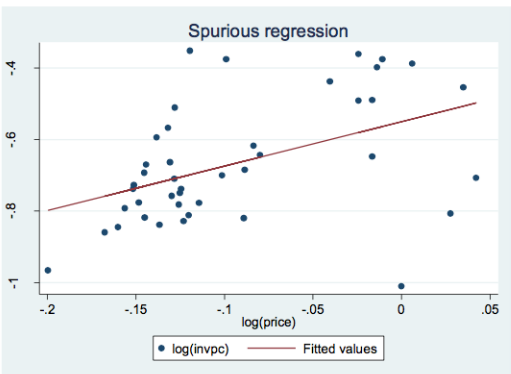
21

Figure 20

```
***
* run a regression on detrended series
***
. reg  detrn_li detrn_lp

      Source |       SS       df       MS              Number of obs =      42
-------------+------------------------------           F(  1,    40) =    0.32
       Model |  .006498137     1  .006498137           Prob > F      =  0.5730
    Residual |  .804674966    40  .020116874           R-squared     =  0.0080
-------------+------------------------------           Adj R-squared = -0.0168
       Total |  .811173104    41   .01978471           Root MSE      =  .14183

------------------------------------------------------------------------------
    detrn_li |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    detrn_lp |  -.3809613    .670296    -0.57   0.573     -1.73568    .9737576
       _cons |  -1.37e-09   .0218855    -0.00   1.000    -.0442322    .0442322
------------------------------------------------------------------------------

. predict prd_detr
(option xb assumed; fitted values)

***
* Note: The R2 of the detrended series reflects how well the explanatory
*       variable(s) explains the dependent variable NET of the effect of the
*       time trend.
***

***
* some graph commands
***
twoway (scatter linvpc lprice)              (line spur lprice)
twoway (scatter linvpc t) (scatter lprice t) (line prd_lpr t) (line prd_linv t)
twoway (scatter detrn_li detrn_lp)          (line prd_detr detrn_lp)
```

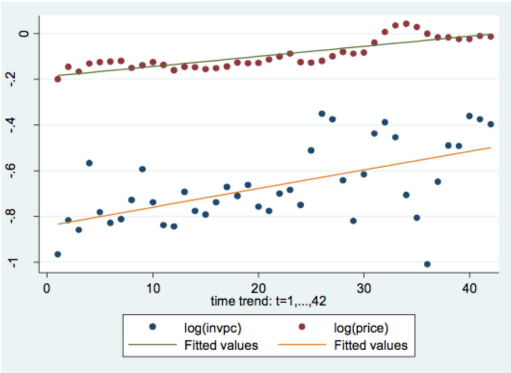
22

Figure 21



23

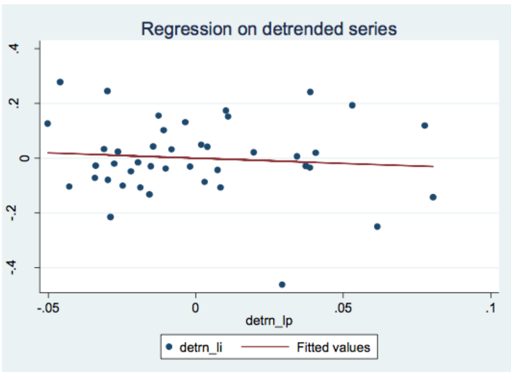
Figure 22



$$\ln(invpc) = \beta_0 + \beta_1 t + u$$
$$\ln(price) = \alpha_0 + \alpha_1 t + w$$

24

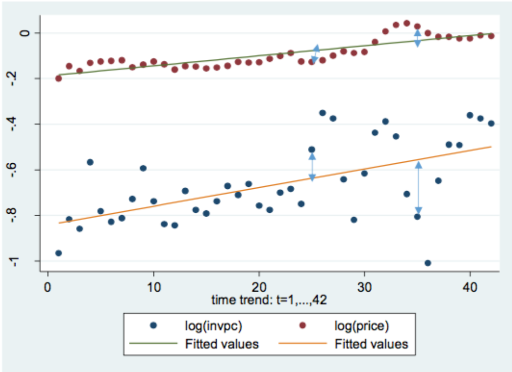
Figure 23



25

Figure 24

By taking away the trend, we are only using variation relative to the trend lines.



26

Figure 25

4.9 Seasonality

Key 1: use dummy variables to remove seasonal factors from the time series.

Key 2: alternatively, de-seasonalize data.

- Use test for joint significance on the dummy variables to see if there exists seasonality.
- De-trending and de-seasonalizing is done before calculating R^2.

How to de-seasonalize data:

1. Run $y_t = \alpha_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \sum_{d=1}^{m} \delta_i D_i$
2. For every x and y variable in the model, withdraw the dummy values and the constant (remember, the last dummy group is the constant) from the variable value. For instance with y_t: $y_t^* = y_t - \hat{\alpha}_0 - \sum_{d=1}^{m} \hat{\delta}_i D_i$.
3. Run $y_t^* = \beta_1 x_{t1}^* + \beta_2 x_{t2}^*$

4.10 Which variables to include

Key 1: the factor that should determine whether a variable belongs in a model is whether that variable has a non-zero partial effect on y *in the population*.

Key 2: excluding variables increases bias and including variables reduces efficiency

Key 3: We should always include x variables that affect y and are uncorrelated with all of the x variables since we reduce the error variance but don't get multicollinearity issues.

Use economic intuition and theory when choosing the specifications.

Patrick prefers less bias—i.e. going for a richer model—but others may disagree.

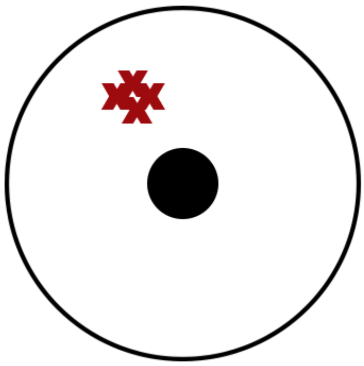


Figure 26: Omitting variables causes bias in the estimators of the variables we do include.

- **Omitted variable bias**: Omitting variables causes bias in the estimators of the variables we do include. And if we have bias, then we can't give a causal interpretation. $\therefore$ **Can't interpret causally when omitted variable bias exists.**
  - This is why it becomes hard to give a causal interpretation with simple linear regression
  - We have positive bias if the **omitted** variable's true β has the same sign as $corr(\text{included } x, \text{excluded } x)$ and negative otherwise.

- Note that omitted variables doesn't lead to biased estimates if the omitted variable i) is uncorrelated with any of the included explanatory variables, or ii) does not affect y. Therefore, we get a bias only when we exclude relevant variables.
- Variables that are not quantitative cannot be included, so such variables that are relevant will always cause omitted variable bias. Examples: ability, patience, preference for risk, management quality, etc.

- Don't remove insignificant variables off the bat: we may want to include them to reduce the bias in another variable. But note that we then generally get more variance in the other estimators' variance.
- The MLR assumptions only hold when we include the intercept β_0 in the regression. Furthermore, we would get biased estimators if we removed it and it's non-zero in the population.
- If we just want to predict, we don't have to be concerned about the omitted variable bias.
- If a variable is statistically significant in only a small fraction of the models estimated, it's likely that the variable has no effect in the population.

For time series:

- For time series: $\mathbb{E}[u_t|X] = 0 \forall t$. Note that is applies for x values in **all** time periods (strictly exogeneous) and not only within the time period (contemporaneously exogeneous).
- If it breaks for time series, then it's often because of omitted variables or measurement errors. In economics, the strict exogeneity assumption is often violated. For instance, policy variables are often affected by past events.

4.10.1 Multicollinearity

- If we have high, but not perfect, collinearity between included variables, we get less efficiency since the denominator in $Var(\beta_) = \frac{\hat{\sigma}}{SST_j(1-R_j)}$ will go towards 0, meaning that the variance, and hence the standard error, will go towards ∞. Note that R_j^2 is the amount of collinearity between x_j and the other x variables—that is, checking how much of the variation in variable x_j that is explained by the other explanatory variables.

- A common way to check for multicollinearity is with $VIF_j = \frac{1}{1-R_j^2}$. A general rule of thumb is that if $VIF_j > 10$, meaning $R_j^2 > 0.9$, then we may have multicollinearity issues.
- Therefore, it's important to think about the relationship between the variables you already included and the variable you consider including. For instance, *mother's education* and *father's education* may be so highly correlated that you would encounter issues. So consider dropping some variables then.
- Collinearity can come from multiples ($x_1 = \alpha x_2$), transformations ($x_1 = \ln(x_2)$)—which by extension is a multiple—or linear dependency between multiple variables. Linear dependency is the fundamental one here.
- If we use robust standard errors and have multicollinearity issues, then the standard errors can become large.
- In time series, multicollinearity makes it hard to get precise estimates for short-run multipliers, but it's not the case for long-run multipliers.

4.10.2 Proxy variables

- When we have unobserved (key) explanatory variables, then we could use proxy variables instead.
- Eliminates or reduced the omitted variable bias, but good proxy variables are hard to find.
- Could potentially use data one a dependent variable from a prior year.

Example:

- We want to regress the model $\ln(wage) = \beta_0 + \beta_1 education + \beta_2 experience + \beta_3 \textbf{ability} + u$
- The problem is that *ability* is unobservable and is suspected to be a key explanatory variable.
- To prevent omitted variable bias, we need to include a **proxy variable: variable that we can observe and is correlated with ability.**
- Let's replace with IQ, which needs a relationship to *ability*: $ability = \delta_0 + \delta_3 IQ + v_3$.
  - We include the constant to allow for different scales

- If $\delta_3 = 0$, then it's not a suitable proxy. Usually, the coefficient is positive.
- And finally, we need the error term since the variable and the proxy are not exactly related. *We also assume that error term is unrelated to education, experience, and IQ.*
- Then regress $\ln(wage) = \boldsymbol{\alpha_0} + \beta_1 education + \beta_2 experience + \boldsymbol{\alpha_3}\mathbf{IQ} + \mathbf{e}$. This four estimators are now unbiased.
  - Note that $\alpha_0 = \beta_0 + \beta_3\delta_0$
  - Note that $\alpha_3 = \beta_3\delta_3$
  - Note that $e = u + \beta_3 v_3$

4.10.3 Lagged dependent variables

Lagged dependent variables can help us get better estimates of the effect of policy variables on various outcomes.

4.11 Functional form misspecifications

Key 1: it's important to think about model specification and don't rely solely on misspecification tests to guide the model choice.

Key 2: midsspecifications changes the conclusions of the model. So always be skeptical and ask if you have a misspecification or not.

Functional form misspecification exists when the model doesn't properly account for the relationship between the y and x variables., meaning it has omitted:

- variables
- certain transformations
- certain interaction terms

Misspecification is a serious problem because it generally leads to biased estimators. But it can be hard to say *why* the model is misspecified.

The problem is not whether we have added irrelevant variables or excluded relevant ones, but rather that we have used the wrong transformation of the variable (square, log, etc) . So

4.11.1 Methods to discover potential misspecifications

Generally: use the F-test for joint exclusion to test for misspecified functional form.

Method 2: RESET (REgression Specification Error Test)

Based on the idea that if $y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + u$ satisfied $MLR4$, then no nonlinear functions of the independent variables should be significant when added to the equation. So to preserve degrees of freedom, it adds polynomials of $\hat{y}$ obtained from OLS. How many polynomials? There's no right answer, but at least a couple is often useful.

1. Regress $y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + u$ and obtain, for instance, $\hat{y}^2$ and $\hat{y}^3$.
2. Then regress $y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \boldsymbol{\delta_1 \hat{y}^2} + \boldsymbol{\delta_2 \hat{y}^3} + u$
3. Run an F-test with with $F_{2,n-k-3}$:
   - $H_0 : \delta_1 = \delta_2 = 0$
   - H_A: At least one δ is $\neq 0$, which means we have some sort of a function form problem.

Problem: if we find a misspecification with RESET, we don't know how to proceed from there.

Please note: The special case White's test is based on a similar idea for heteroskedasticity, but that test uses $\hat{u}$ as the left-hand side variable instead of y

Example:

```
reg cigs lincome lcigpric educ age
predict yhat
gen y2=yhat*yhat
gen y3=y2*yhat
quietly reg cigs lincome lcigpric educ age y2 y3
test y2 y3
```

4.11.2 Methods to test specific models for potential misspecifications

For these tests, we'll use two models for testing to evaluate whether we should use level or log-transformation (transformations of y variables are outside the scope of this course, so we only check for transformations on the x variables):

- model 1: $y = \beta_0 + \sum_{j=1}^{2} \beta_j x_j + u$
- model 2: $y = \beta_0 + \sum_{j=1}^{2} \beta_j \ln(x_j) + u$

The problem with these tests is that a clear winner may not emerge, which ultimately hands the decision over to adjusted R^2 to decide which model to go for. Also, if the effect of key x variables on y are not very different, then it doesn't really matter which model is used.

Method 1: Davidson-MacKinnon

Idea: If model 1 has $\mathbb{E}[u|x_1, x_2] = 0$, then fitted values from model 2 (which has the same y variable and on the same form) will be insignificant when added to model 1. So we can identify if a model is good enough.

1. Regress model 2: $y = \beta_0 + \sum_{j=1}^{2} \beta_j \ln(x_j) + u$ and obtain the fitted values, $\check{y}$
2. Regress model 1 with the fitted values: $y = \beta_0 + \sum_{j=1}^{2} \beta_j x_j + \theta_1 \check{y} + u$
3. Evaluate the significance of θ_1 with the normal t-distribution. If $\theta_1 = 0$, then model 1 is the correct one in terms of respecting $MLR4$. If $\theta_1 \neq 0$, however, then we got specification issues.

Of course, you could test it the other way around as well, starting with model 1 and plugging in the regressed values in model 2.

Note that if one of the models is rejected, it doesn't mean that the other one is correct.

Also note that this was a bad example since adding the logarithm in one model, level in the other, and then join them together would cause multicollinearity issues. So beware that the variables should essentially have been different, not only transformations.

Method 4: an alternative method for non-nested alternatives

1. Merge the two models together: $y = \beta_0 + \sum_{j=1}^{2} \beta_j x_j + \sum_{j=1}^{2} \beta_{j+2} \ln(x_j) + u$
2. Run the hypothesis test:
   - H_0: $\beta_1 = \beta_2 = 0$ (model 1 has no effect, so model 2 should be used instead)
   - H_1: $\beta_3 = \beta_4 = 0$ (model 2 has no effect, so model 1 should be used instead)

4.12 Pooled OLS and Panel Data

Enormously useful for policy analysis and program evaluation.

They key difference between the two is the "units" we follow. I am defining units as households, countries, or whatever we are collecting data on. In pooled cross section, we will take random samples in different time periods, of different units, i.e. each sample we take, will be populated by different individuals. This is often used to see the impact of policy or programmes. For example we will take household income data on households X, Y and Z, in 1990. And then we will take the same income data on households G, F and A in 1995. Although we are interested in the same data, we are taking different samples (using different households) in different time periods.

In pure panel data, we are following the same units i.e. the same households or individuals over time. For example we will follow the same set of households X, Y and Z, for each time period we collect data i.e. in 1990 and we will also interview the same households in 1995.

Therefore the fundamental difference, is simply the units we observe the data for.

It's important to have several observations for the same unit because you can control for more things and therefore get different conclusions.

Key 1: We need to handle the individual-specific effect that is unobserved and hence captured in the residual: $v_{it} = a_i + u_{it}$. As you can see from the index, this is constant over time.

Key 2: It's important to handle this because if there is correlation between a_i and x_{it}, then OLS will be biased and inconsistent since the zero conditional mean assumption will be violated and we'll have omitted variable bias. Additionally, we will likely also have autocorrelation in the residuals when using OLS.

Key 3: If there is correlation, then we have three options to choose among when dealing with a_i. If there is no correlation, then we can utilize some of the variation in a_i through the random effects method as well.

Key 4: Typically the first question on the exam questions about panel data is, "Why don't we use OLS?". And the answer is that if we don't control for a_i, then we get omitted variable bias.

Key 5: The benefit of random effects over fixed effects is that FE uses degrees

of freedom like crazy. However, we must be sure a_i and the explanatory variables are uncorrelated if we're using RE.

- *Balanced sample*: T observations per unit.
- *Unbalanced sample*: we don't have T observations for all units. *This is the most common one.*
- *Individual unobserved effect/**fixed effect**/unobserved heterogeneity*: a_i
- *idiosyncracit error term*: *uit*.
- *composite error*: $v_i t = a_i + u_{it}$
- A pooled sample ignored the individual specific effects a_i since we don't follow the same units over time. Since the samples are independent across time, autocorrelation is not a problem.

You need institutional knowledge to know what a_i consists of, but here are some examples:

- Individual:
  - Ability
  - Motivation
  - Family background
  - Etc
- Firm:
  - Efficiency
  - Etc
- Country:
  - Political system
  - Citizen's attitude
  - Etc

4.12.1 Handling panel data

We'll exemplify the methods with a two-period model—i.e. that we have two periods worth of data for every individual. The general model is:

$$\forall t = 1, 2.\ y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{itj} + \delta_2 D^{t=2} + \mathbf{a_i} + \mathbf{u_{it}}$$

Method 1: First differencing (DF)

$$\Delta y_{i2} = \beta_1 \sum_{j=1}^{k} \Delta x_{i2j} + \delta_2 \cdot 1 + \mathbf{\Delta u_{i2}}$$

- By taking the difference between two time periods, we remove the constant terms like a_i.
- But note that we also remove the intercept, so it's just the dummies that determines the intercept and the base group is forced through the origin.
- The interpretation of β is the same as the one we have with levels.
- FD'ing greatly reduces the variation in the explanatory variables since many variables don't change that much over time, so it's hard then to get a precise estimator.
- Potential pitfalls:
  - Key variable doesn't vary over time
  - We must have strict exogeneity of the regressors
  - It can be worse than pooled OLS if one or more variables has measurement error.
- Δu_i is uncorrelated with Δx_i if *both* u_{i1} and u_{i2} is uncorrelated with both x_{i2} and x_{i1}. This is called *strict exogeneity.*

Method 2: The least squares dummy variable approach

$$y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{itj} + \sum_{t'=2}^{T} \delta_{t'} D^{t=t'} + \sum_{\mathbf{i'=2}}^{\mathbf{N}} \mathbf{a_{i'} D^{i=i'}} + u_{it}$$

- We take a_i out of the error term and hence won't violate the zero conditional mean assumption.
- The problem with this method is that many units (individuals, firms, etc) will give many variables. For 1000ish units or less, this works fine.

Method 3: Within group/fixed effects

1. $y^*_{it} = y_{it} - \bar{y}_i$
2. $\forall j = 1, ..., k.\ x^*_{itj} = x_{itj} - \bar{x}_i$
3. $y^*_{it} = \sum_{j=1}^{k} x^*_{itj} + u_{it}$

- Likely the most used one.
- Every term is transformed by subtracting the unit's average over time.
- Constant terms like β_0 and a_i has an average of β_0 and a_i and are therefore removed.
- So note that we regress *without* the intercept, so this must be specified in STATA.
- The interpretation of β is the same as the one we have with levels.
- Interpretation: imagine that each unit has its own coordinate system (variation between groups). By removing the average, we move all individuals in the vector space to the same coordinate system (i.e. removing the variation between the groups). That way, we only utilize the variation *within* the groups.
- Can be used on unbalanced panel data since we subtract the average, but then you must assume that the missing data is not systematically related to u_{it}, or else we get biased FE estimators.
- This estimation method ignores important information on how the variables change over time.
- We cannot include time-fixed variables like gender because it disappears.
- Degrees of freedom $= N \cdot T - N - k$

- R^2 is interpreted as the amount of time variation in y_{it} that is explained by the time variation in the explanatory variables.
- The FE is widely thought to be a more convincing tool for estimating ceteris paribus effects than the RE model.
- When it comes to unbalaneced data sets, the FE estimator is more robust than the RE estimator.

Method 4: Random effects

$$y_{it} - \lambda\bar{y}_i = \beta_0(1-\lambda_1) + \beta_1(x_i - \bar{x}_i) + (\mathbf{v_{it}} - \lambda\mathbf{\bar{v}_i}),\ \lambda \in [0;1]$$

$$\Longleftrightarrow$$

$$y_{it} - \lambda\bar{y}_i = \beta_0(1-\lambda_1) + \beta_1(x_i - \bar{x}_i) + (\mathbf{1}-\lambda)\mathbf{a_i} + (\mathbf{u_{it}} - \lambda\mathbf{\bar{u}_{it}}),\ \lambda \in [0;1]$$

- What is different here is that we draw the errors from a distribution such that the correlation between the errors and the explanatory variables is 0: $u_{it} \equiv N(0, \sigma_u^2), a_i \equiv N(0, \sigma_a^2$.
- Then we don't have to exclude a_i completely, and can thus utilize some of the variation between the individuals.
  - We can explain this through the coordinate system interpretation again. in FE, we moved all individuals into the same coordinate system. Now, we move each units coordinate system closer towards a united coordinate system but not completely.
- Note that we don't subtract with $\hat{a}_i$ since a_i is time-independent and hence $\hat{a}_i = a_i$.
- It's crucial that there is no correlation between a_i and the explanatory variables if we are going to utilize some of the variation, or else we will have an endogeneity problem. So be really sure about this.
- The benefit of random effects over fixed effects is that FE uses degrees of freedom like crazy. So if there is no correlation, then RE is better.
- Include the intercept so that we can test whether $a_i = 0$. The intercept then captures something that applies to all individuals over all time periods.
- The RE model is just *generalized least squares* (GLS)

4.12.2 The Hausman test: testing whether we can use RE or whether we need to go for FE

.

- H_0 : uncorrelated—use RE.
- H_A : correlated—use FE.
- Test statistic: $m = \frac{(\hat{\beta}^{FE} - \hat{\beta}^{RE})^2}{VC(\hat{q}_1)} \equiv \chi^2_{df=k}$

4.12.3 Other notes

- In a production function, for instance, time dummies is measured in percentage points.
- Robust standard errors in panel data: use *cluster* robust standard errors.
- If there exists correlation between the explanatory variables and a_i, whether you should choose FD or FE depends on *autocorrelation* and *measurement errors*. So use both in their own models and compare them (the results will be identical for $T = 2$).
- Large N, small T: if $\vec{x}$ is uncorrelated with u_{it}, then FE is more efficient than FD. Else, FD is better.
- Large T, especially when N is not large: go for FD since FE inference can be sensitive. FD has the advantage of turning an integrated time series process into a weakly dependent time series.
- Unabalanced panels: observations can be included in the average calculations *if and only if* no data points (cells) are missing whatsoever from that particular observation (row).
- It's important to understand why the panel data set is unbalanced.
- FE and RE methods can be applied to a cluster sample: with cluster sampling, the researcher divides the population into separate groups, called clusters. Then, a simple random sample of clusters is selected from the population. The researcher conducts his analysis on data from the sampled clusters.
- Advantages of panel data:
  - Increase the sample size

  - Able to reduce multicollearity problems because of
    * Variation between cross-sections
    * Variation over time
  - Are able to build dynamic model
  - Are able to control for unobserved effects better than in cross-sections or time-series
- Ignoring the existence of a_i when estimating the model, we could think of this as an omitted variable bias problem.

4.12.4 Pooled OLS

- Use time dummies so that we have different intercepts over time.
- Since a_i disappears in some of the methods above, then we can estimate with pooled OLS.
- There may be heteroskedasticity in the error term underlying the estimated equation, and the error variance may change over time even if it doesn't change the cross-sectional values.
- Interact time dummies with key explanatory variables to see if the effect of that variable has changed over time.
- Beware: we need to work with real amounts, not nominal amounts. So convert.
- The Chow test can be used to determine whether a regression function differs across two groups and it's applicable across two different time periods as well.

4.12.5 Risky Business: The Market for Unprotected Commercial Sex

- Sex workers are willing to risk infection by not using condoms with clients if they are adequately compensated.
- A condom will not be used when the client's maximum willingsness to pay not to use a condom is greater than the minimum the sez worker is willing to accept to take the risk.

- The model also predicts that when the client is worried about the risk of infection from unprotected sex, he may be charged more for using a condom than for unprotected sex.
- Similarly, when the sex worker prefers not to use a condom, the client is given a discount for not using a condom.
- Mexican sex workers received a 23 percent premium for unprotected sex from clients who requested not to use a condom, and this premium jumped to 46 percent if the worker was considered very attractive.
- A common approach is to use IVs. However, in principle, there are no omitted variables that could be used as instruments for condom use in the price equation.
- Instead, we shall take advantage of the fact that we have transaction data and multiple transactions for each sex worker by including a sex worker fixed effect. The sex worker FE controls also for the value of the sex worker's outside option and the FE from the client's value function. The FE controls for bias from both unobserved sex worker heterogeneity and client selection based on unobserved sex worker characteristics.
- Conclusion: sex workers in Mexico are responding rationally to financial incentives given their risk preferences.
- The most effective interventions for reducing HIV/STI transmission through commercial sex will be those that target both the supply side and the demand side of the market.

5 The Output: The Inference

5.1 Degrees of freedom

$$df = n - k - 1$$

Key: k is *not* unique variables. It's instead the number of terms on the right-hand side including variables. So if there's 50 observations, 5 unique variables, and 1 variable that is included once more but transformed, then we have $df = 50 - (5 + 1) - 1 = 43$.

5.1.1 Examples

Q1.4. When estimating the following regression model, with a dataset of 44 observations:

$$Wage_i = \beta_0 + \beta_1 Exper_i + \beta_2 \ (Exper_i)^2 + \beta_3 \ Male_i + \beta_4 \ Married_i + \beta_5 \ Educ_i + u_i \ ,$$

the degrees of freedom for this model are:

A) 44
B) 38
C) 39
D) 37

Figure 27: V18 Q1.4. Correct answer is B.

Q1.7. When estimating regression model below, you have a dataset with 30 observations.

$$Wage_i = \beta_0 + \beta_1 Exper_i + \beta_2 \ Educ_i + \beta_3 \ Male_i + \beta_4 \ Married_i + u_i$$

The degrees of freedom for this model is:

A) 26
B) 25
C) 24
D) 23

Figure 28: V17 Q1.7. Correct answer is B.

5.2 The Standard Error

The standard error gives us an idea of how precise the estimate is.

Key: the more total sample variation we have in x_j, the less variance we will have in the estimator, giving us more precise confidence intervals. Therefore, more sample variation is good.

- Independent and identically distributed error term.
  - Stronger than assumption 4 and 5.
  - Let's us talk about the shape of the distribution. We now assume $\hat{\beta}_j \equiv N(\beta^{H_0}, Var(\hat{\beta}_j))$.
  - We can also standardize the expression above to $\frac{\hat{\beta}_j - \beta_j^{H_0}}{\sigma[\hat{\beta}_j}$.
  - But we don't know σ, so we cannot use that. Instead, we use the standard error $\hat{\sigma} = \sqrt{\frac{SSR}{n-k-1}}$. But then we cannot use the normal distribution. Instead we use the t distribution and say that $t = \frac{\hat{\beta}_j - \beta_j^{H_0}}{\hat{\sigma}[\hat{\beta}_j} \equiv t_{n-k-i}$.
  - **Therefore, with assumption 6 in place, we can start hypothesizing about the unknown population value of β_j.**

5.3 The Error Variance

Under A1–A5 for simple OLS, we have:

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \hat{x})^2} = \frac{SSR}{SST_x} \tag{4}$$

Under A1–A5 for multiple OLS, we change it to:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2} = \frac{SSR}{SST_x} \tag{5}$$

So we see that a high R_j^2—which is R^2 when excluding variable j—gives high $Var(\hat{\beta}_j)$, which will affect our efficiency. And the smaller and more precise this is, the smaller the confidence intervals will be and the more accurate will our hypothesis tests be.

Including an extra variable will always increase R^2, so therefore we encounter the key question: **Should we add another variable to reduce bias but face the cost of more variance (i.e. less efficiency)? Or should we skip adding another variable to have more efficiency but at the cost of more bias?**

An unbiased estimator of $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2 = \frac{SSR}{n-2}$. Two degrees of freedoms are used in estimating $\hat{\beta}_0$ and $\hat{\beta}_1$.

Replacing σ^2 in the formula for the OLS estimator's variance with $\hat{\sigma}^2$ gives us an estimate of the variance of the OLS estimates. The smaller the estimated variance, the more precise/efficient the OLS estimates.

$\hat{\sigma}$ is not an unbiased estimator of σ but it's consistent at least.

5.4 The Error Term

u_i: error of disturbance. **The error**.

$\hat{u}_i$: unobservables for observation i that affect y_i. **The residual**.

Reasons why the error term is not zero:

- Relevant excluded variables affect the dependent variable
- Wrong functional form
- Errors in measuring the dependent variable
- Randomness in behavior
- Etc.

5.4.1 Examples

Q1.1. A fitted regression equation is given by $\hat{Y}$=20+X, what is the residual for the observation where X=10 and Y=40?
A) 0
B) 10
C) −20
D) −10

Figure 29: Correct answer is B.

Q1.1. A fitted regression equation is given by $\hat{Y} = -20 + 0.1X$, what is the residual for the observation where X=100 and Y=10?
A) 0
B) -10
C) 20
D) 10

Figure 30: Correct answer is C.

5.5 Interpreting Coefficients

$$\beta_1 = \frac{\text{sample covariance}(x_i, y_i)}{\text{sample variance}(x_i)}$$

$$\beta_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

You can only interpret parameters for the observed interval—we know there's a linear relationship if we can read it for the model, but we don't know if this linear relationship expands outside of the interval.

The *ceteris paribus* is used on simple linear regression, too. But it cannot be used for variables that is included multiple times in a regression, like *inc* in $y = \beta_0 + \beta_1 inc + \beta_2 inc^2 + \beta_3 inc \cdot exp$. *So the fundamental interpretation of the ceteris paribus meaning is the partial derivative of y with respect to the specific explanatory variable (like inc).*

5.5.1 Examples: interpreting coefficients

Q1.10. The estimation of a logarithmic wage regression by the instrumental variable estimator gives the following fitted regression equation: lwage = 10.4 + 0.043 *educ*, where wage is the hourly wage in NOK and *educ* is the years of education. Interpret the slope coefficient.
A) A one year increase in education increases wages by 0.043 percent. Since this is the instrumental variable estimate, we can interpret it as a causal effect of education on wages.
B) A one year increase in education increases wages by (10.4+4.3) NOK per hour. Since this is the instrumental variable estimate, we can interpret it as a causal effect of education on wages.
C) A one year increase in education increases wages by 4.3 percent. The instrumental variable estimator gives the causal effect if we have at least one valid instrumental variable.
D) The coefficients estimated by instrumental variable estimation are the same as by ordinary least squares.

Figure 31: V18 Q1.2. Correct answer is C

Q1.3. A fitted regression equation is given by wage=5.5+2education.
Education is measured in years of education, wage is measured in 10 000 NOK.
Interpret the slope parameter:
A) Increasing educ. with 1 year increases wage with 2
B) Increasing educ. with 1 year increases wage with 20 000 NOK
C) Increasing educ. with 1 year increases wage with 2 000 NOK
D) Increasing educ. with 1 year increases wage with 5.5+2=7.5

Figure 32: V17 Q1.3. Correct answer is B

Q1.4. A fitted regression equation is given by log(CEOsalary)=0.75+0.01firmprofit.
Firmprofit and CEOsalary are both measured in 1 000 NOK.
Which of the following is correct:
A) Increasing firm-profit with 1% increases CEOsalary with 0.01%
B) Increasing firm-profit with 10 000 NOK increases CEOsalary with 0.1%
C) Increasing firm-profit with 1 000 NOK increases CEOsalary with 1%
D) Increasing firm-profit with 10 000 NOK increases CEOsalary with 100 NOK

Figure 33: V17 Q1.4. Correct answer is C

Q1.5. A fitted regression equation is given by log(CEOsalary)=0.7+0.25log(firmprofit).
Firmprofit is measured in 100 000 NOK, CEOsalary is measured in 1000 NOK.
Interpret the slope parameter:
A) Increasing firm-profit with 100 000 NOK increases CEOsalary with 25 000 NOK
B) Increasing firm-profit with 100 000 NOK increases CEOsalary with 25%
C) Increasing firm-profit with 1 % increases CEOsalary with 25%
D) Increasing firm-profit with 1% increases CEOsalary with 0.25%

Figure 34: V17 Q1.5. Correct answer is D. Beware that we don't multiply by 100 on *log–log*, only for one *level* and one *log*.

Q1.6. A fitted regression equation is given by log(CEOsalary)=0.4+0.005firmprofit.
Firmprofit and CEOsalary are both measured in 100 000 NOK. Which of the following is correct:
A) Increasing firm-profit with 10 000 NOK increases CEOsalary with 0.05%
B) Increasing firm-profit with 100 000 NOK increases CEOsalary with 500 NOK
C) Increasing firm-profit with 1% increases CEOsalary with 0.5%
D) Increasing firm-profit with 1 million NOK increases CEOsalary with 0.5%

Figure 35: Correct answer is A.

Q1.7. A fitted regression equation is given by log(CEOsalary)=0.3+0.07log(firmprofit).
Firmprofit is measured in 100 000 NOK, CEOsalary is measured in 1000 NOK.
Interpret the slope parameter:
A) Increasing firm-profit with 100 000 NOK increases CEOsalary with 0.07%
B) Increasing firm-profit with 100 000 NOK increases CEOsalary with 7000 NOK
C) Increasing firm-profit with 1% increases CEOsalary with 0.07%
D) Increasing firm-profit with 1 % increases CEOsalary with 7%

Figure 36: Correct answer is C.

5.6 Hypotheses

Key 1: we're testing hypotheses about the *population* parameters. Key 2: the *p-value* is interpreted as the probability of observing something more extreme than we did given that the null hypothesis is true. Key 3: significance level is interpreted as being willing to mistankely reject H_0 when it's true (Type I error) a maximum of $100 \cdot \alpha\%$ of the time. Key 4: If, for instance, we reject H_0: then it is the *estimate* that is statistically significant. That is, for instance, "$\hat{\beta}_j$ is statistically greater than $+$ at the 5 % significance level." Key 5: It's important to think about the practical/economic significance as well. Too much focus on the statistical significance could make you erroneously conclude that a variable is important for explaining y. Key 6: they are not overly focused on the p-values because the conclusions can be misleading if we have misspecified the model. So say at which levels they are significant.

It's often normal to test:

$H_0 : \beta_j = 0$

$H_A : \beta_j \neq 0$

$H_A : \beta_j > 0$

$H_A : \beta_j < 0$

- Keeping H_0: "We don't have sufficient evidence to discard H_0 in favor of H_A".
- Going for H_A: there is evidence against the null hypothesis at a $100 \cdot \alpha\%$ significance level.
- $H_A : \beta_j > 0$ essentially means that $H_0 : \beta_j \leq 0$. But since $H_0 : \beta_j = 0$ is harder to reject than $H_0 : \beta_j < 0$, we would reject both < 0 and $= 0$ if we reject the latter. Therefore, it's enough to have $H_0 : \beta_j = 0$.

Recipe:

1. Specify the hypotheses
2. Determine the significance level. Take a look at n, the number of observations, when deciding.
3. Find the rejection rule
4. Check for statistical significance. If it is significant, and if it's a key variable, discuss economical significance.
5. See the p-value. For small samples, we can accept p-values down to 20 % but we're at thin ice so far out. So don't put too much weight on the estimate being slightly insignificant.
6. Give conclusion, don't stop at the p-value.

Moreover:

- You need to choose a hypothesis before looking at the data.
- With the t statistic, we can hypothesize over one coefficient parameter at a time. We can only use the t test to test one-sided alternatives, which the F test cannot.
- With the F statistic, we can hypotehesize over several coefficient parameters at a time. Note that $F = \sqrt{t}$. This is useful to check if the variables we're testing over should actually be in the model. If one is significant, then we don't know which of the variables are. But on the other hand, we get the opportunity to test for joint significance, which the t test won't pick up. Additionally, multicollinearity isn't that much of a problem for F tests.
- If we're testing nested models with the F test, then $F = \frac{(R_{ur}^2 - R_r^2)/q}{(1-R_{ur}^2/(n-k-1)}$. *This is the most important use of the F test since we can test exclusion restrictions*
- The conclusion we make out of the hypothesis test may not always be the one we should make:
  - Type I error: reject H_0 given that it is true (you are put in prison although you are innocent—the worst error to make)
  - Type II error: accept H_0 given that it is false (you are not imprisoned although you are guilty).
- It's hard to find significance with small sample sizes, so many researchers are willing

to use larger significance values for such samples (like 10% instead of 5% for $n = 29$, for instance).

Examples of some other alternatives to test:

- $H_0 : \beta_j = 0$
- $H_0 : \beta_j = 1$
- $H_0 : \beta_1 + \beta_2 = 1$ (linear combination of parameters. But note that $se[\beta_1 + \beta_2] = \sqrt{se[\hat{\beta}_1]^2 + se[\hat{\beta}_1]^2 + Cov(b\hat{e}ta_1, b\hat{e}ta_2)}$)
- $H_0 : \beta_1 = 1 - \beta_2$ (transforming a variable—solve by substituting for β_1 in the regression)
- $H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$
- $H_0 : \beta_1 = 0, \beta_2 = 1, \beta_3 = 1$

Hypotheses intervals

A *confidence interval* is a range of likely values for the *population* parameter.

Note that the confidence interval is only as good as the underlying assumptoins used to construct it.

$$H_0 : \beta^{H_0} \in \hat{\beta}_j \pm t^c \cdot se[\hat{\beta}_j]$$

$$H_A : \beta^{H_0} \notin \hat{\beta}_j \pm t^c \cdot se[\hat{\beta}_j]$$

5.6.1 Examples: confidence intervals

Q1.1. You have estimated by ordinary least squares the following relationship between the logarithm of *production* and the variable *number_employees*. The dataset contains 28 observations of firms. Production is measured in 10 000 NOK, and *number_employees* is how many employees the firm employs. The results are as follows (standard errors in parentheses).

$$\underset{}{log(production)} = \underset{(0.53)}{0.87} + \underset{(0.04)}{0.12}\ number\_employees, \qquad R2 = 0.43$$

Assuming that the estimator has a normal distribution, the 95% confidence interval for the slope coefficient in front of the variable *number_employees*, (β_1) is ***approximately*** the interval

A) [0.042, 0.198]
B) [0.080, 0.160]
C) [0.038, 0.202]

Figure 37: V18 Q1.1. Correct answer is C.

Q1.8. You have estimated the following relationship between y and x . The data set used contains 1500 observations and the results are as follows (standard errors in parentheses).

$$y = \underset{(0.620)}{5.25} + \underset{(0.075)}{0.341}\ x, \qquad R2 = 0.62$$

Assuming that the estimator has a normal distribution, the 95% confidence interval for the slope coefficient in front of the variable *x*, (β_x) is ***approximately*** the interval

A) [0.19, 0.49]
B) [0.27, 0.42]
C) [0, 2]
D) [-0.5, 0.5]

Figure 38: V17 Q1.8. Correct answer is A.

Q1.10. You have estimated the following relationship between y and the variables x and z. The data set used contains 30 observations and the results are as follows (standard errors in parentheses).

$$y = \underset{(0.101)}{2.814} - \underset{(0.053)}{0.233}\ x + \underset{(0.077)}{0.330}\ z, \qquad R2 = 0.36$$

Assuming that the estimator has a normal distribution, the 95% confidence interval for the slope coefficient in front of the variable *x*, (β_x) is ***approximately*** the interval

A) [2.61, 3.01]
B) [-0.33, 0.13]
C) [-0.34, -0.13]
D) [-0.36, 0.36]

Figure 39: Correct answer is C.

5.6.2 Examples: t-values, p-values, and conclusions

Q1.2. Consider the estimated coefficient of some variable $\beta_x = 2.2$, with a standard error of 0.9. You have 25000 degrees of freedom. The t-statistic and corresponding conclusion for testing H0: β_x=1 against H1: β_x>1 at the 10% level is:
A) t=2.44, do not reject H0
B) t=2.44, reject H0
C) t=1.33, do not reject H0
D) t=1.33, reject H0

Figure 40: V18 Q1.2. Correct answer is D

Q1.6. Finding a large *p*-value (e.g. higher than 10%)
A) implies that the *t*-statistic is bigger than 1.96.
B) indicates evidence against the null hypothesis.
C) will only happen roughly one in twenty samples.
D) indicates evidence in favor of the null hypothesis.

Figure 41: V17 Q1.6. Correct answer is D.

Q1.9. Consider the estimation in Q1.8. The t-statistic and corresponding conclusion for testing H0: β_x=0.2 against H1: β_x>0.2 at the 5% level is:
A) t=1.88, do not reject H0
B) t=1.88, reject H0
C) t=4.55, do not reject H0
D) t=4.55, reject H0

Figure 42: V17 Q1.9. Correct answer is B.

Q1.9. Finding a small *p*-value (e.g. less than 5%)
A) indicates evidence in favor of the null hypothesis.
B) implies that the *t*-statistic is less than 1.96.
C) indicates evidence against the null hypothesis.
D) will only happen roughly one in twenty samples.

Figure 43: Correct answer is C.

Q1.11. Consider the estimation in Q1.10. The t-statistic and corresponding conclusion for testing H0: β_z=0.2 against H1: β_z>0.2 at the 5% level is:
A) t=1.688, reject H0
B) t=1.688, do not reject H0
C) t=5.116, reject H0
D) t=5.116, do not reject H0

Figure 44: Correct answer is B.

5.7 Predictions

- While inference about the x variables are statements about the population parameters—what most of the section is about—inference about the y variable are *predictions*.
- Predictions: an estimate of the expected value of y given the particular values of the x variables.
- The variance of the prediction is smallest at the mean values of the x_j.
- The prediction interval is not the same as the confidence interval because we need in addition to account for the variance in the ubonbserved error.

5.8 Goodness-of-fit

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$$

$SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ (a measure of how spread out $\hat{y}$ is)

$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$ (a measure of how spread out y is)

$SSR = \sum_{i=1}^{n}\hat{u}_i^2$

- Don't use as the only metric
- Don't use across models with different dependent variables.

- Don't use across models with different samples.
- In other words, don't use for model selection.
- Too much R^2 may indicate there's some underlying issue with the model, too.
- High R^2 has nothing to do with causality.
- Interpretation: the fraction of sample variance in y that is explained by $\vec{x}$.
- Often low in cross-sections for social sciences.
- Be careful interpreting the typically reported R^2 when the constant is not included: it doesn't remove $\bar{y}$ when calculating SST.
- Including additional variables will always increase R^2 except for when we have missing observations in the data set (then we don't know what happens).
- A low R^2 indicates it's hard to predict individual outcomes on y with much accuracy.

5.8.1 Adjusted R^2

Key: use when comparing different functional forms of explanatory variables.

- This is an adjustment of the R-squared that penalizes the addition of extraneous predictors to the model (but rewards good predictors, i.e. those who have $|t| > 1$.
- Adjusted R-squared is computed using the formula $1 - (1 - R^2)\frac{n-1}{n-k-i}$, where k is the number of x variables.
- It's especially useful for non-nested models since the F-test only work on nested.

5.8.2 Comparing models in which one is level and the other is log-transformed

We cannot use R^2 to compare different y's, so we need to get the log-transformed variable to level.

1. Estimate $\ln(y)$ and obtain $\widehat{\ln(y)}$ (fitted values)
2. Generate $\hat{m}_i = \exp[\widehat{\ln(y_i)}$
3. Regress on y on $\hat{m}_i$ *without* an intercept.
4. Find the sample correlation between $\tilde{y}$ and y in the sample, $corr(\tilde{y}, y$.

5. Square this and compare with R^2 of the model with y in levels.

5.9 Zero Conditional Mean

$$\mathbb{E}[u|\vec{x}] = 0$$

If this is not in place:

- OLS is not unbiased
- We don't get a causal effect.

5.9.1 Examples

We have:

$$wages = \beta_0 + \beta_1 education + u$$

We got *ability* in u, which is likely correlated with *education* since it's easier to take more education when it comes easier to you. So we cannot say, for instance, that $\mathbb{E}[ability|education = 10] = \mathbb{E}[ability|education = 23]$. Therefore, we cannot interpret education's effect on wages causally since the zero conditional mean assumption doesn't hold.

5.10 Heteroscedasticity

Key: if we suspect heteroscedasticity, the usual OLS standard errors are invalid and some corrective action should be taken. So heteroskedasticity only affect the efficiency of the estimators.

Testing for heteroskedasticity: look at the graph or go for the special case of the White's test. Handling heteroskedasticity: the easiest one is to use robust standard errors

On the exam, they expect a definition of heteroscedasticity and a short statement of its consequences.

Heteroskedasticity means that the error term ***does not*** *have constant variance, which in turn will make the* ***variance*** *in y vary for different x values/different segments. Heteroskedasticity doesn't affect the slope, but it does affect the standard errors and makes the tests invalid. In short, we cannot rely on t-tests and the corresponding confidence intervals or hypothesis tests because the heteroskedasticity could lead us to conclude erroneously. Consequently, OLS is not BLUE in the presence of heteroskedasticity.*

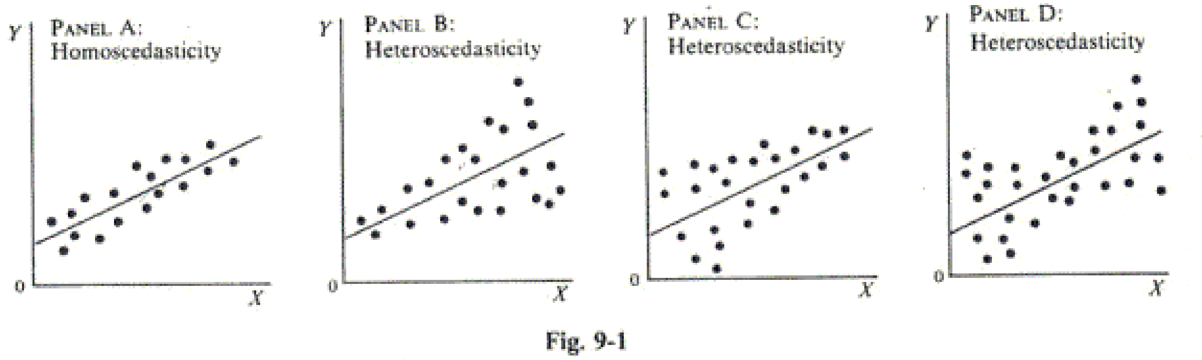


Figure 45

- So we don't get bias in $\hat{\beta}_j$, but we do get bias in $Var[\hat{\beta}_j]$, which invalidates the standard errors.
- If we use robust standard errors and have multicollinearity issues, then the standard errors can become large.
- Robust standard errors can be both larger than or smaller than normal standard errors, but they are usually larger.
- Reasons for robust standard errors: 1) the standard errors are more often valid, and 2) they are useful with large sample sizes because one is then getting closer to the t-distribution
- Reasons against robust standard errors: 1) the t-statistic is exactly t-distributed, regardless of sample size.
- If we get heteroskedasticity, we should ask ourselves why we have such problems.

5.10.1 Identify heteroskedasticity

Method 1: study graphically

- Run OLS, plot residual against $\hat{y}$
- Run OLS, plot residual against the suspected $\hat{x}$ variable.

Method 2: Breusch-Pagan

We can run the following test on the entire model or only on specific variables—the latter is useful if we suspect only specific variables causes heteroskedasticity.

1. Estimate $y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + u$ and obtain $\hat{u_i}^2$
2. Run $\hat{u_i}^2 = \delta_0 + \sum\limits_{j=1}^{k} \delta_j x_j + error$
3. Choose one of the following to test whether all δ's except δ_0 is 0, and then compute the p-value:
   - $F \equiv F_{k,n-k-1}$
   - $LM \equiv \chi_k^2$
4. H_0: homoskedastic. H_1: heteroskedastic.

Method 3: White's test

The main difference between White's test and Breusch-Pagan is that White includes all kinds of interactions between the explanatory variables as well.

1. Estimate $y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + u$ and obtain $\hat{u_i}^2$
2. Run $\hat{u_i}^2 = \delta_0 + \sum\limits_{j=1}^{k} \delta_j x_j + \sum\limits_{\mathbf{j=1}}^{\mathbf{k}} \sum\limits_{\mathbf{j'\in\{j'=1,\ldots,k:j\geq j'\}}}^{\mathbf{k}} \delta_{\mathbf{k+(j-1)*k+j'}} \mathbf{x_i x_j} + error$
3. Choose one of the following to test whether all δ's except δ_0 is 0, and then compute the p-value:
   - $F \equiv F_{k,n-k-1}$
   - $LM \equiv \chi_k^2$
4. H_0: homoskedastic. H_1: heteroskedastic.

The weakness of this test that with all these interactions, we're using degrees of freedom like crazy. The solution is the *special case White's test*

Method 4: Special case White's test

This method is a great test because we can test for the interactions but at the same time keep many degrees of freedom.

1. Estimate $y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + u$ and obtain $\hat{u_i}^2$ and as many polynomials of the fitted values as you want (for example $\hat{y}$ and $\hat{y}^2$.

2. Run $\hat{u_i}^2 = \delta_0 + \delta_{\mathbf{1}}\hat{\mathbf{y}}$ and $\delta_{\mathbf{2}}\hat{\mathbf{y}}^{\mathbf{2}} + error$
3. Choose one of the following to test whether all δ's except δ_0 is 0, and then compute the p-value:
   - $F \equiv F_{k,n-k-1}$
   - $LM \equiv \chi^2_k$
4. H_0: homoskedastic. H_1: heteroskedastic.

Please note: The RESET test is based on a similar idea for misspecifications, but that test uses y as the left-hand side variable instead of $\hat{u}$

5.10.2 Handle heteroskedasticity

If one of the tests above indicate heteroskedasticity, then we should do corrective measures like one of the methods below. The new model we generate will then be the one we infer from.

A key concern is to evaluate whether we know the form of the heteroskedasticity or not.

Also, if we use a weight function, then a key concern is whether we have misspecified the weight function or not. But in cases of strong heteroskedasticity, it's often better to use a wrong form heteroskedasticity and apply WLS than to ignore heterskedasticity altogether in estimation and use OLS.

Method 1, which is the easiest one: Robust standard errors

Use the option `robust` to deal with heteroskedasticity without changing the coefficients (as it should be).

Method 2: feasible generalized least squares (FGLS)

Anti-Nilsen recommends this one because one could have misspecifications of the model, which this method deals with better.

Not unbiased, but consistent and asymptotically efficient.

We have heteroskedastic error if the error variance is on the form $Var[u_i] = \sigma^2 z_i^2$, in which $z_i = z_i(x)$.

- If we know the form of the heteroskedastic error, then we can go for something called *the weighted least squares (WLS) method.*

- The key here is to divide every term in $y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + u$ by $\sqrt{h_i}$, which is chosen in such a way that $MLR5$ is respected: $\mathbb{E}[\frac{u_i}{\sqrt{h_i}}|\vec{x}_i] = \mathbf{0}$ and $Var[\frac{u_i}{\sqrt{h_i}}|\vec{x}_i] = \sigma$. The key is that it gives less weight to observations with a higher error variance. If heteroskedasticity was the only problem in the model, $MLR1 - -6$ will then be respected.

- However, the key here is that we know the form, and then WLS would be more efficient than OLS. But the problem is that we rarely know the form. Therefore, a slight variation of this is the following FGLS method.

Furthermore, we should go for this if the form of the heteroskedasticity is not known (which is common), which means that we need to estimate the weighting function.

1. Run $y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + u$ and obtain $\hat{u}$.
2. Generate $\ln(\hat{u}^2)$
3. Run $\ln(\hat{\mathbf{u}}^{\mathbf{2}}) = \delta_0 + \sum_{j=1}^{k} \delta_j x_j + u$ and obtain $\hat{u}$ and obtain the fitted values, $\hat{g} = \widehat{\ln(\hat{u}^2)}$
4. Generate $\hat{h} = \exp[\hat{g}]$ because exponentiating gives us positive estimated variances, which is required.
5. Estimate $y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + u$ with WGS using $\frac{1}{\hat{h}}$ as weights.

5.11 Endogeneity

- Variables in the error term is correlated with the included x variables.
- The opposite, *exogeneity*, means that all factors in the error term is uncorrelated with the included x variables.
  - This means that we have correctly accounted for the functional relationships between the explained and explanatory variables.
- We have this issue if assumption 4 doesn't hold.

5.11.1 Examples

5.12 Chow

The Chow statistic is an F statistic to test whether the *true* coefficents in two regressions on different data sets are equal. Used for different:

- groups (men and women, for example); or
- time periods (before and after a policy change, for example).

Moreover:

- So this can be considered a bit of an extreme variant of dummies. With dummies, we allow for different slopes and/or intercept. If we want to know whether there is *any* difference, then you would have to create dummy interactions with all variables. With a large model, this would be tedious. Moreover, we could get large standard errors from this, and then it would be hard to tell how the groups differ through the interactions. **It would therefore be easier to use the Chow statistic to compare groups.**
- Note that there are two variations: allowing the intercept to be different and not allowing the intercept to be different.

1. Split the sample up in the different groups (let's say two groups, g_1, g_2.
2. Then $\forall g = 1, 2.\ y = \beta_g 0 + \sum_{j=1}^{k} \beta_{gj} x_j + u$. Consider these as restricted models.
3. Obtain the SSR from each, SSR_1 and SSR_2.
4. $H_0 : \forall j = 0, ..., k.\ \beta_1 j = \beta_2 j$ for variant 1: that even the intercept is equal.
5. $H_0 : \forall j = 1, ..., k.\ \beta_1 j = \beta_2 j$ for variant 2: that we allow the intercept to be unequal. Often more useful since the intercept will mostly always be different.
6. Imagine an unrestricted model that includes 1) the intercept, 2) the variables, 3) the group dummy, and 4) the interactions between the dummies and the variables. Degrees of freedom: $n - 2(k + 1)$.
7. Obtain the SSR from the unrestricted model, SSR_P.
8. Then calculate the Chow statistic below. Critical assumption: homoskedasticity since F tests (which the Chow test essentially is) are only valid then.

$$F = \frac{[SSR_P - (SSR_1 + SSR_2)]}{SSR_1 + SSR_2} \frac{n - 2(k+1)}{k+1}$$

5.12.1 If the model is so small that you can just add all interaction terms instead of dividing the sample in groups

Key: you just set the H_0 that the interaction terms with the dummy are all equal to zero.

This variation does *not* require homoscedasticity.

What I want to test: Are all coefficients equal across groups?

Restricted model:

$$price = \beta_0 + \beta_1 sqft + \beta_2 pool + \beta_3 age + u$$

Unrestricted model:

$$\begin{aligned} price = \beta_0 + \beta_1 sqft + \beta_2 pool + \beta_3 age + \delta_0 Utown + \delta_1 sqft \cdot Utown + \\ \delta_2 pool \cdot Utown + \delta_3 age \cdot Utown + u \end{aligned}$$

Listing 1: Testing all coefficients including the intercept

```
* First, generate interaction terms and run regression
gen sqftU = sqft*Utown
gen ageU = age*Utown
gen poolU = pool*Utown
reg price sqft age pool Utown sqftU ageU poolU

* Then run the F-test:
test Utown sqftU ageU poolU
```

Listing 2: Testing all coefficients excluding the intercept.

```
* The only difference is that we remove Utown from the test.
* First, generate interaction terms and run regression
gen sqftU = sqft*Utown
gen ageU = age*Utown
```

```
gen poolU = pool*Utown
reg price sqft age pool Utown sqftU ageU poolU

* Then run the F-test:
test sqftU ageU poolU
```

5.12.2 If the model is so big that you rather go for splitting the sample

```
reg price sqft pool age
reg price sqft pool age if utown==1
reg price sqft pool age if utown==0
```

5.13 Short-Run and Long-Run Multipliers

- A shock may have both immediate and long-run consequences.
- Test the significance of the long-run multiplier, like this: `test(_b[inf] + _b[L1.inf]`
- If both the y and the x variables are log-transformed, then the multipliers are interpreted as *short-run elasticity* and *long-run elasticity*.

5.13.1 Finding the short-run multiplier

- The short-run multiplier is just the coefficient to x_t—the non-lagged explanatory variable.
- This is therefore the effect we see from a *temporary* increase in x.
- Hard to get precice short-run multiplier estimates because of multicollinearity.

5.13.2 Finding the long-run multiplier

- The long-run multiplier is the effect we see from a *permanent* increase in y—so the cummulative effect after we reached the steady state.

- The long-run multiplier doesn't suffer from multicollinearity issues so we usually get good estimates from this one.

Let's study a 1^{st} order autoreggresive distributive T-finite lag model—**ARDL(1,T)**—and identify the multipliers:

$$y_t = \alpha + \sum_{lag=0}^{T} \delta_{lag} z_{t-lag} + \gamma y_{t-1} + u_t$$

Then the crucial step: assume we reached a steady state, meaning that all x and y values are the same for the last $\max(1, T)$ periods:

$$y = \alpha + \sum_{lag=0}^{T} \delta_{lag} z + \gamma y + u_t$$

Moving over and factorizing:

$$(1-\gamma)y = \alpha + \sum_{lag=0}^{T} \delta_{lag} z + u_t$$

Dividing:

$$y = \frac{\alpha}{(1-\gamma)} + \frac{\boldsymbol{\sum_{lag=0}^{T} \delta_{lag}}}{\boldsymbol{(1-\gamma)}} z + \frac{u_t}{(1-\gamma)}$$

The parameter in front of z is our long-term multiplier.

5.13.3 Evaluating the statistical significance

Let's simiplify a bit and study a model that doesn't include an autoregressive component.

For short-run multipliers, we just take $\hat{\delta_0}$, $se[\hat{\delta_0}]$, and calculate the t-value and p-value.

However, this is not as straightforward with the long-run multipliers since it involves many terms. Here's the process:

1. Use $\theta_0 = \delta_0 + \sum_{lag=1}^{T} \delta_{lag}$ and rearrange to $\delta_0 = \theta_0 - \sum_{lag=1}^{T} \delta_{lag}$
2. Substitute for δ_0 in the basic model, $y_t = \alpha + \sum_{lag=0}^{T} \delta_{lag} z_{t-lag} + u_t$ so that we get:
   - $y_t = \alpha + \boldsymbol{\theta_0 x_t} + \sum_{lag=0}^{T} \delta_{lag}(z_{t-lag} - \mathbf{z_t}) + u_t$
3. Then regress, obtain $\hat{\theta_0}$ and $se[\hat{\theta_0}]$, and calculate its t-value and p-value to see whether the long-run multiplier is statistically significant.

5.14 Autocorrelation and Stationarity

Autocorrelation is about the function that estimates ρ, $\hat{\rho} = \frac{Cov[y_t, y_{t+s}]}{Var[y_t]}$.

The stationarity issues deals with models/processes, and looks at specific forms of processes that may or may not contain autocorrelation.

Examples: Unit root processes, trend stationary processes, autoregressive processes, and moving average processes are specific forms of processes with autocorrelation.

5.15 Autocorrelation

Key 1: autocorrelation is lags in the *error term*–that is, error correlation over time: $Corr[u_t, u_s|\vec{x}_t] \neq 0 \forall t \neq s$.

Key 2: autocorrelation may give you the wrong standard errors, which makes the least squares estimates *inefficient* (but they are still unbiased except for small samples—then it can even change the slopes). So everything regarding inference will be a problem when you have autocorrelation.

Key 3: autocorrelation is a sign of dynamic misspecification: that is, misspecification of a model that spans multiple time periods through lags.

1. Positive autocorrelation: the next period's error term usually has the *same* sign as this period's.
    - For instance: negative news seems to follow negative news (Venezuela, Greece, bankruptcy).
2. Negative autocorrelation: the next period's error term usually has the *opposite* sign as this period's.

5.15.1 Example with $AR(1)$: note that $AR(1)$ does not include any x variables while $ARDL(1,0)$ includes one x variable for this time period.

- We have $\forall t = 1, 2, ..., T$. $y_t = \beta_0 + \beta_1 x_t + u_t$
- A potential problem is that we could have correlation over time, making the residual behave badly if $\rho \neq 0$ in $u_t = \rho u_{t-1} + e_t$. Note that we removed the constant from that model because is has to be 0 by definition of OLS. So we therefore assume it's 0 and take it away.

- However, that e_t is well-behaved, being $e_t \equiv N(0, \sigma_e^2), \mathbb{E}[e_t \cdot e_s] \forall t \neq s$
- So the key is that we should first evaluate whether we do have a autocorrelation problem ($\rho \neq 0$), and if so, apply a method to deal with it.

5.15.2 How to detect autocorrelation

1^{st} **alternative: look at the graphs**

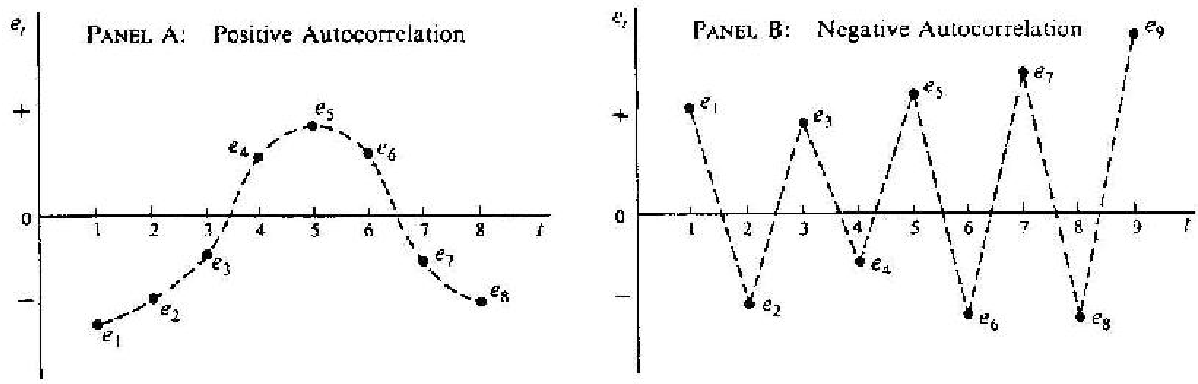


Fig. 9-2

Figure 46

$2n^{nd}$ **alternative Only for AR(1): Durbin–Watson**

The *Durbin–Watson statistic*: $d = 2(1 - \hat{\rho})$

- Positive autocorrelation if $d \in \langle 0; d_L \rangle$
- Negative autocorrelation if $d \in \langle 4 - d_L; 4 \rangle$

Downside: often inconclusive

If $R^2 > d$, then one should suspect a spurious regression.

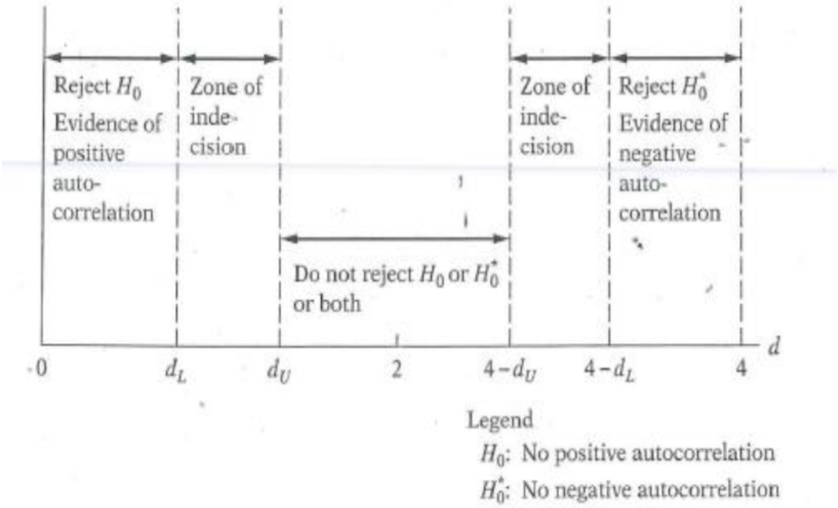


Figure 47

3^{rd} alternative Only for AR(1): study the error term

1. Run $y = \beta_0 + \sum_{i=1}^{k} \beta_j x_{it} + u_t$
2. Obtain $\hat{u}_t$
3. Run $\forall t = 2, ..., n$. $\hat{u}_t = \delta + \rho \hat{u}_{t-1} + \sum\limits_{i=1}^{k} \theta_i x_{it} + v_t$
4. t-test ρ, with $H_0 : \rho = 0$ (no autocorrelation)

Please note that the lag doesn't have to be one period behind. It could also be, for instance, four periods behind so that you can test for quarter-season autocorrelation.

4^{th} alternative for all orders, so a more fundamental one than the alternative above: study the error term

1. Run $y = \beta_0 + \sum_{i=1}^{k} \beta_j x_{it} + u_t$
2. Obtain $\hat{u}_t$
3. Run $\forall t = q+1, ..., n$. $\hat{u}_t = \delta + \sum\limits_{j=1}^{q} \rho_j \hat{u}_{t-j} + \sum\limits_{i=1}^{k} \theta_i x_{it} + v_t$
4. F-test with H_0 : all $\rho = 0$ (no autocorrelation)

5^{th} alternative: Correlogram

See page 88.

6^{th} alternative: Study autocorrelation in the shocks with the restricted form

1. Start with an $y_t = \beta_0 + \beta_1 x_t + u_t$ and $y_{t-1} = \beta_0 + \beta_1 x_{t-1} + u_{t-1}$.
2. Multiply the latter one with ρ, then subtract go "today's regression" $-$ "yesterday's regression".
3. Then we'll get $(y_t - \rho y_{t-1}) = \beta_0(1-\rho) + \beta_1(x_t - \rho x_{t-1}) + (u_t - \rho u_{t-1})$, of which $(u_t - \rho u_{t-1})$.
4. Move ρy_{t-1} over and split up the x parenthesis on the right-hand side, giving us $y_t = \beta_0(1-\rho) + \beta_1 x_t - \beta_1 \rho x_{t-1} + \rho y_{t-1} + (u_t - \rho u_{t-1})$
5. Regress the model in the line above and test whether $|\rho| < 1$ (easiest it to read of the confidence interval whether 1 is in there. If 1 is *not* in the confidence interval, then $|\rho| < 1$ and hence $y = \beta_0 + \sum_{i=1}^{k} \beta_j x_{it} + u_t, u_t = \rho u_{t-1} + e_t$ is the true model.

7^{th} alternative: The common factor approach (the unrestricted form)

1. Perform step 1–4 from the restricted form method above.
2. Rewrite to $y_t = \pi_0 + \pi_1 x_t - \pi_2 x_{t-1}) + \pi_3 y_{t-1} + (u_t - \rho u_{t-1})$
3. Regress the model in the line above and test $H_0 : \pi_2 = -\pi_1\pi_3$. If we don't reject H_0, then $y = \beta_0 + \sum_{i=1}^{k} \beta_j x_{it} + u_t, u_t = \rho u_{t-1} + e_t$ is the true model.

5.15.3 How to correct autocorrelation

Correcting autocorrelation depends on one key question first: do you know ρ? I.e. do you know the structure of the autocorrelation?

Method 1: The quasi-difference data method—used only if *we know* ρ:

1. Perform step 1–3 from the restricted form method above.
2. In STATA, generate two new variables, `new_y` for $(y_t - \rho y_{t-1})$ and `new_x` for $(x_t - \rho x_{t-1})$.
3. Then regress `new_y` on `new_x`. The residual is e_t, which is well-behaved.

Method 3: The Cochrane–Orcutt procedure—an FGLS method used if *we do not know* ρ

Key: we estimate a *static* model but allow the shocks, u_t to be serially correlated.

1. Run $y_t = \beta_0 + \beta_1 x_t + u_t$, with $u_t = \rho u_{t-1} + e_t$ and save the residuals $\hat{u}_t$

2. Run $\forall t \geq 2$. $\hat{u}_t = \delta_+ \rho \hat{u}_{t-1} + v_t$ and save the estimated ρ
3. Transform the y and x variable from the first regression:
   - $y^* = y_t - \hat{\rho} y_{t-1}$
   - $x^* = x_t - \hat{\rho} x_{t-1}$
4. Run $y^* = \beta_0^* + \beta_1^* x_t^* + (u_t - \hat{\rho} u_{t-1})$. Repeat step 2.–4. until $\hat{\rho}$ converges.

Method 4: The Prais–Winsten method

A slight variation of the Cochrane–Orcutt procedure. Cochrane–Orcutt omits the first observation, while Prais-Winsten uses it.

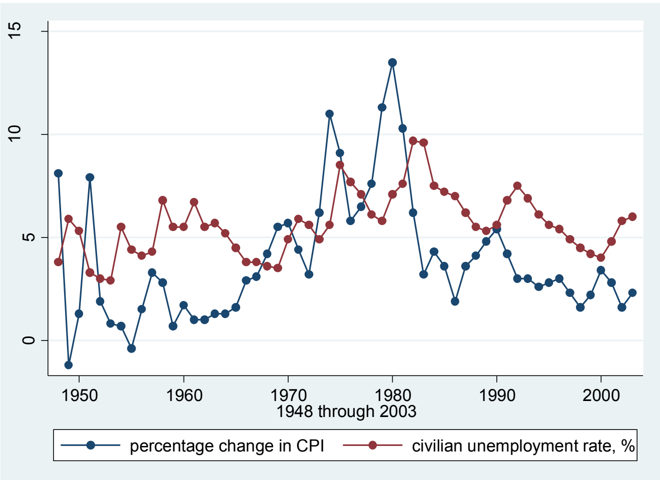
5.15.4 Example 1

Example: Phillips curve (inflation as a function of unemployment)

$$\pi_t = \beta_0 + \beta_1 unempl_t + u_t$$

What sign should the slope have?

```
. use phillips.dta

. tsset year
        time variable:  year, 1948 to 2003
                delta:  1 unit

***
* plot the two series
***
. twoway (connected inf year) (connected unem year)
```

```
***
* Initial regression
***
. reg inf unem
      Source |       SS       df       MS              Number of obs =      56
-------------+------------------------------           F(  1,    54) =    3.58
       Model |   31.599858     1   31.599858           Prob > F      =  0.0639
    Residual |  476.815691    54   8.8299202           R-squared     =  0.0622
-------------+------------------------------           Adj R-squared =  0.0448
       Total |  508.415549    55  9.24391907           Root MSE      =  2.9715
------------------------------------------------------------------------------
         inf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |   .5023782   .2655624     1.89   0.064    -.0300424    1.034799
       _cons |   1.053566   1.547957     0.68   0.499    -2.049901    4.157033
------------------------------------------------------------------------------
```

Reasonable slope coeff.?

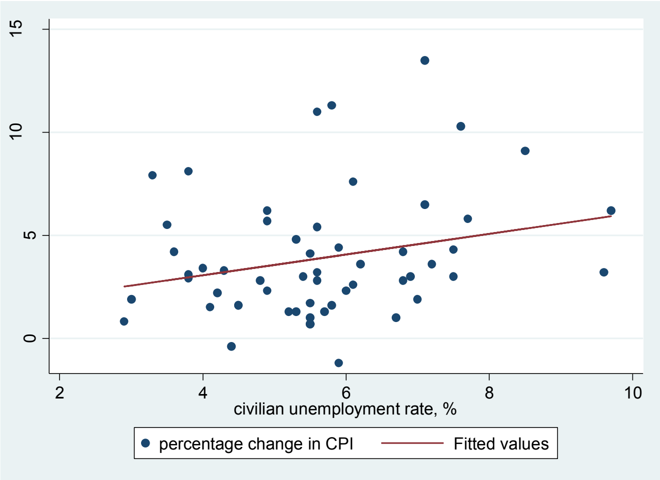
```
***
* Could the time-trend be a point?
***
. reg inf unem year
      Source |       SS       df       MS              Number of obs =      56
-------------+------------------------------           F(  2,    53) =    1.83
       Model |   32.800234     2   16.400117           Prob > F      =  0.1708
    Residual |  475.615315    53  8.97387386           R-squared     =  0.0645
-------------+------------------------------           Adj R-squared =  0.0292
       Total |  508.415549    55  9.24391907           Root MSE      =  2.9956
------------------------------------------------------------------------------
         inf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |    .468465   .2833218     1.65   0.104    -.0998065    1.036736
        year |    .009586   .0262102     0.37   0.716    -.0429849    .0621569
       _cons |  -17.69255   51.27949    -0.35   0.731    -120.5462    85.16109
------------------------------------------------------------------------------
```



```
***
* re-run initial regression
***
. reg inf unem

      Source |       SS       df       MS              Number of obs =      56
-------------+------------------------------           F(  1,    54) =    3.58
       Model |   31.599858     1   31.599858           Prob > F      =  0.0639
    Residual |  476.815691    54   8.8299202           R-squared     =  0.0622
-------------+------------------------------           Adj R-squared =  0.0448
       Total |  508.415549    55  9.24391907           Root MSE      =  2.9715

------------------------------------------------------------------------------
         inf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |   .5023782   .2655624     1.89   0.064    -.0300424    1.034799
       _cons |   1.053566   1.547957     0.68   0.499    -2.049901    4.157033
------------------------------------------------------------------------------

. predict u_hat, resid
. predict inf_hat, xb

***
* durbin watson
***
. estat dwatson
Durbin-Watson d-statistic(  2,    56) =  .8014823
```

We will see this DW d-stat. later on too

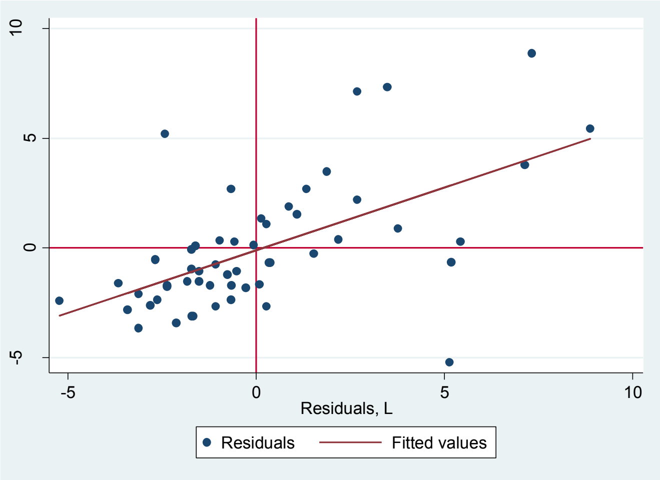
```
***
* plot the data points and the predicted regressions line
***
. twoway (scatter inf unem) (line inf_hat unem)
```



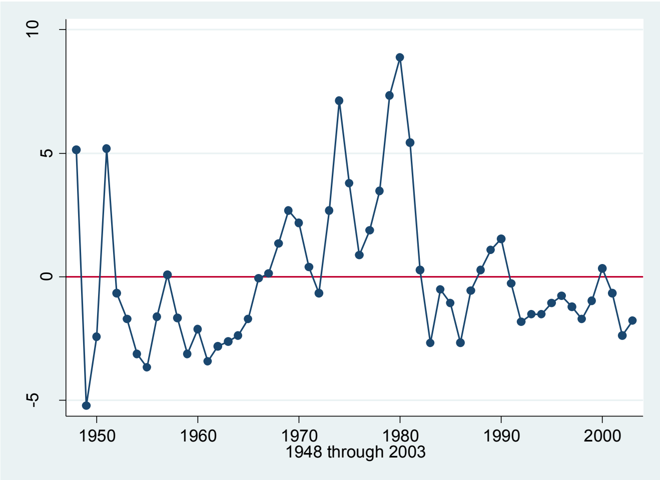

```
***
* testing for autocorrelation
***
. reg u_hat l.u_hat

      Source |       SS       df       MS              Number of obs =      55
-------------+------------------------------           F(  1,    53) =   27.91
       Model |  155.221317     1  155.221317           Prob > F      =  0.0000
    Residual |  294.721637    53   5.5607856           R-squared     =  0.3450
-------------+------------------------------           Adj R-squared =  0.3326
       Total |  449.942953    54  8.33227692           Root MSE      =  2.3581

------------------------------------------------------------------------------
       u_hat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       u_hat |
         L1. |   .5724722   .1083545     5.28   0.000     .3551407    .7898038
             |
       _cons |  -.1118079   .3179895    -0.35   0.727    -.7496141    .5259983
------------------------------------------------------------------------------

. predict u_hat_hat, xb
```

Keep an eye on this coefficient

```
***
* plot the residual as a function of lagged residual
***
. twoway (scatter u_hat l1.u_hat) (line u_hat_hat l1.u_hat) , xline(0) yline(0)
```




```
***
* plot the residual as a function of time
***
. twoway (connected u_hat year)  , yline(0)
```



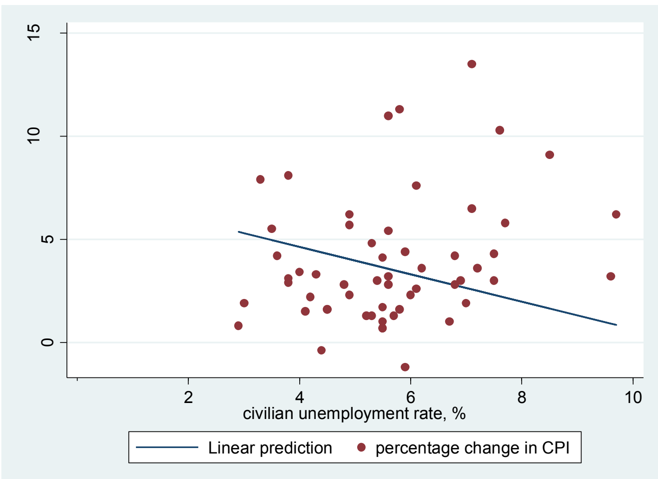
```
***
* Cochrane-Orcutt
***
. prais  inf unem, corc
Iteration 0:  rho = 0.0000
Iteration 1:  rho = 0.5721
Iteration 2:  rho = 0.7204
Iteration 3:  rho = 0.7683
Iteration 4:  rho = 0.7793
Iteration 5:  rho = 0.7815
Iteration 6:  rho = 0.7819
Iteration 7:  rho = 0.7820
Iteration 8:  rho = 0.7820
Iteration 9:  rho = 0.7820
Iteration 10: rho = 0.7820
Cochrane-Orcutt AR(1) regression -- iterated estimates
      Source |       SS       df       MS              Number of obs =      55
-------------+------------------------------           F(  1,     53) =    5.09
       Model |  23.3857044     1  23.3857044           Prob > F      =  0.0282
    Residual |  243.353574    53  4.59157686           R-squared     =  0.0877
-------------+------------------------------           Adj R-squared =  0.0705
       Total |  266.739278    54  4.93961626           Root MSE      =  2.1428
------------------------------------------------------------------------------
         inf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |  -.6639584   .2942027    -2.26   0.028    -1.254054   -.0738626
       _cons |   7.287078   2.163178     3.37   0.001     2.948291    11.62586
-------------+----------------------------------------------------------------
         rho |   .7820093
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    0.801482
Durbin-Watson statistic (transformed) 1.600203

. predict inf_hat_CO, xb
```

Do you remember this coefficient??

Reasonable slope coeff.!

Here we see the DW d-stat. that we saw before


```
***
* plot the data points and the predicted regression line having used
* Cochrane-Orcutt
***
. twoway (line inf_hat_CO unem) (scatter inf unem), xmtick(0(2)10) ytitle("Inflation")
```



So what have we found, using the Cochrane-Orcutt method?

We have estimated **a static model**, but where we have allowed that the **shocks** u_t are serially correlated.

$y_t = \beta_0 + \beta_1 x_t + u_t$ where the shocks u_t are AR(1), $u_t = \rho u_{t-1} + e_t$, and e_t is white noise - $e_t \sim N(0, \sigma_e^2)$

where y in our case is inflation, and x is unemployment.

And we have found estimates of β_1 and ρ.

But we may also estimate our static model with serially correlated shocks using a **common factor** approach.

Again

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad t=1, 2...T$$
$$u_t = \rho u_{t-1} + e_t$$

Which may be transformed as follows

$$y_t = \beta_0 + \beta_1 x_t + u_t$$
$$\rho y_{t-1} = \rho\beta_0 + \rho\beta_1 x_{t-1} + \rho u_{t-1}$$

gives

$$y_t = \beta_0(1-\rho) + \beta_1 x_t - \rho\beta_1 x_{t-1} + \rho y_{t-1} + (u_t - \rho u_{t-1})$$

or unrestricted

$$y_t = \pi_0 + \pi_1 x_t + \pi_2 x_{t-1} + \pi_3 y_{t-1} + (u_t - \rho u_{t-1})$$

```
. reg inf unem l1.unem l1.inf, noheader
------------------------------------------------------------------------------
         inf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |
         --. |  -.6746869   .3591467    -1.88   0.066    -1.395704    .0463301
         L1. |   .5117387    .308468     1.66   0.103    -.1075366    1.131014
             |
         inf |
         L1. |   .7878368   .1199723     6.57   0.000     .5469823    1.028691
             |
       _cons |   1.668624   1.233995     1.35   0.182    -.8087242    4.145972
------------------------------------------------------------------------------
```

where

$\Pi_1 = -0.6747$ ($\approx \beta_1$)

$\Pi_3 = 0.7878$ ($\approx \rho$)

$\Pi_2 = 0.5117$ ($\approx -\rho\beta_1$) ($-\Pi_1 * \Pi_3 = -0.5315$)

The transformation shown on previous page;

$$y_t = \beta_0(1-\rho) + \beta_1 x_t - \rho\beta_1 x_{t-1} + \rho y_{t-1} + (u_t - \rho u_{t-1})$$
$$y_t = \pi_0 + \pi_1 x_t + \pi_2 x_{t-1} + \pi_3 y_{t-1} + (u_t - \rho u_{t-1})$$

```
. dis "-(PI1*PI3)= " -_b[unem]*_b[l1.inf] " should be close to PI2 = "_b[l1.unem] " if model is correct."
-(PI1*PI3)= .53154318 should be close to PI2 = .51173865 if model is correct.

. testnl -_b[unem]*_b[l1.inf] = _b[l1.unem]

  (1)  -_b[unem]*_b[l1.inf] = _b[l1.unem]

               chi2(1) =        0.01
           Prob > chi2 =        0.9361
```


What have we tested? We started from the assumption that the following model is the true one

$$y_t = \beta_0 + \beta_1 x_t + u_t \qquad t\text{=1, 2...T}$$
$$u_t = \rho u_{t-1} + e_t$$

Then we estimated with OLS the following

$$y_t = \pi_0 + \pi_1 x_t + \pi_2 x_{t-1} + \pi_3 y_{t-1} + (u_t - \rho u_{t-1})$$

And since we could not reject the H0, that $\Pi_2 = -\Pi_1 \cdot \Pi_3$, this indicates that the

$$y_t = \beta_0 + \beta_1 x_t + u_t \qquad t\text{=1, 2...T}$$
$$u_t = \rho u_{t-1} + e_t$$

is the true model.

But what if we had to reject H0? …. Think of another type of model, for instance an **autoregressive distributed lag model (ARDL)**, $y_t = \pi_0 + \pi_1 x_t + \pi_2 x_{t-1} + \pi_3 y_{t-1} + e_t$

ARDL; *autoregressive*; y_t is explained (in part) by lagged values of itself, *distributed lag*; successive lags of the "x-s".

A third way to analyse a static model with autocorrelation in the shocks: Here we use a regression method to estimate our static model with serially correlated shocks by putting **restrictions on the coefficients** (a non-linear regression method, i.e. non-linear in the parameters)

The restricted form

$$y_t = \beta_0(1-\rho) + \beta_1 x_t - \rho\beta_1 x_{t-1} + \rho y_{t-1} + (u_t - \rho u_{t-1})$$

```
. nl (inf = {b0}*(1-{rho}) + {b1}*unem - {rho}*{b1}*unem_1 + {rho}*inf_1) if _n > 1, initial (b0 1 b1 0
rho 0)
(obs = 55)
Iteration 0:  residual SS =  265.1197
Iteration 1:  residual SS =  243.7527
Iteration 2:  residual SS =   243.354
Iteration 3:  residual SS =  243.3536
Iteration 4:  residual SS =  243.3536
      Source |       SS       df       MS
-------------+------------------------------         Number of obs =        55
       Model |  246.963527     2  123.481764         R-squared     =    0.5037
    Residual |  243.353574    52  4.67987642         Adj R-squared =    0.4846
-------------+------------------------------         Root MSE      =  2.163302
       Total |  490.317101    54  9.07994631         Res. dev.     =  237.8783
------------------------------------------------------------------------------
         inf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /b0 |   7.287079   2.287763     3.19   0.002     2.696345    11.87781
        /rho |   .7820094   .0946834     8.26   0.000     .5920133    .9720056
         /b1 |  -.6639586    .330235    -2.01   0.050    -1.326624   -.0012936
------------------------------------------------------------------------------
  Parameter b0 taken as constant term in model & ANOVA table
```

55

Bottom line: Even if econometric theory says the least squares estimates are unbiased but inefficient in the presence of autocorrelation, this example has shown that in small samples we could experience that even the sign of the slope coefficient is wrong. Be careful…..

Autocorrelation is a sign of dynamic mis-specification!!!

That means, if you find signs of autocorrelation in the error term, perhaps you should consider to include lags of the explanatory variables, x_{t-1}, x_{t-2} … or lagged dependent variable y_{t-1}, y_{t-2} in your econometric model, and test again for autocorrelation.

5.16 Stationarity

Key 1: A key question is, "Does the *correlation between the variables at different time periods* tend to zero quickly enough?"

Key 2: *a stationary process* means that the mean and variance of the time series doesn't change over time ($E[y_t] = \mu$, $Var[y_t] = \sigma^2$), and that the covariance between two data points depend only on the time distance between them and not the point in time that they occur ($cov(y_t, y_{t-1}) = \gamma_s$). So if a shock occurs, stationary processes will return to its mean. Non-stationary time series, however, are slow-turning.

Key 3: *high persistence* means that a shock to the series tends to persist for long and the series drifts away from its historical mean path.

Key 4: *Integrated of order* n means that you need to difference the time series n times before it becomes stationary. It's often a question of $I(0)$ or $I(1)$, which in an AR(1) process means it's a question of $|\rho| < 1$ or $\rho = 1$

Key 5: *weakly dependent* means that you don't have to difference the time series for the process to be stationary—it already is. So $|\rho| < 1$ for an AR(1) process. This ensures that $\hat{\beta}_j$ is consistent (but not always unbiased)

Key 6: *unit root* means that you have to difference the time series *one* time for the process to become stationary. So $\rho = 1$ for an AR(1) process. The implication of this is that the series won't return to its origin, and such processes are hard to control.

Key 7: *white noise* is the well-behaved residual we're talking about: $e_t \equiv N(0, \sigma_e^2$. Our residual would be white noise if $\rho = 0$.

Key 8: *Cointegration* means that two non-stationary series cancel each other out such that the residual becomes stationary. This means that we can use the time series after all.

- Ideally, time series are stationary and weakly dependent.
- Non-stationary processes may give you **spurious regressions**: the regression may be really good and significant, but you cannot trust them. Many macroeconomic and financial time series are nonstationary.
- Check if we get a stationary process by removing the trend.

- The weak dependence is often needed for the central limit theorem. It means that two points in time get less related as the larger the time difference is between them (that $|\rho| < 1$ in a AR(1) model, for example).

- If the time series is highly persistent, we must be extremely cautious about using that time series in the model.

- Using first-differences in the model is often as informative as levels and it's often weakly dependent.

- An intuitive explanation of weak dependence: you get less dependent/heart-broken from your ex as time passes on—time heals all wounds. That is, $Corr(x_t, x_{t+h}) \to 0$ as $h \to \infty$ (asymptotically uncorrelated).

- Example of weak dependence: Let's study a *moving average process* of order m: $x_t = e_t + \sum_{lag=1}^{m} e_{t-lag}$. This is a weakly dependent process but terms in sequence of orer m are correlated but the rest is not.

- A stable AR process is crucial to achieve weak dependence (that is, that $|\rho| < 1$ in $y_t = \rho_1 y_{t-1} + e_t$.)

- But many economic time series are more $\rho = 1$, meaning that the value today is important for determining the value in the very distant future (that is, intuitively, that our best prediction of future values is the current value). This means that the process has an *unit root.*

- If $\rho = 1$ in the AR(1) process, then we have a *random walk*—a wander up and down with no real pattern (except potentially for drift)—and $\mathbb{E}[Y_t] = \mathbb{E}[Y_0]$ and $Var[Y_t] = \sigma_e^2 t$. So this process cannot be stationary since the variance changes over time. Therefore **Don't use data with random walks—they are hard to handle because discrete changes can have long-lasting effects.**

  - Random walk: $y_t = 1 \cdot y_{t-1} + e_t$

  - Random walk with drift: $y_t = \alpha + 1 \cdot y_{t-1} + e_t$

  - Examples: the stock price is constant before new infor arrives, and today's price is a function of all previous relevant information.

- Highly persistent behaviors can contain a trend, like a random walk with a drift.

- **Weakly dependent processes** are $I(0)$: you don't have to difference the process for it to become stationary. With AR(1): $y_t = \rho y_{t-1} + e_t$ is stationary because

$|\rho| < 1$.

- **Unit root processes**:
  - Random walk processes that are $I(1)$: you have to difference the process one time for it to become stationary: $y_t = \rho y_{t-1} + e_t$ is non-stationary because $\rho = 1$, but we can go $y_t - y_{t-1} = e_t \implies \Delta y_t = e_t$, which is a stationary process.
  - So unit roots are bad since they are non-stationary.
- If data is stationary and weakly dependent, the goodness-of-fit measures R^2 and $\bar{R}^2$ are valid even when autocorrelation exists.
- Cointegration is often used to see whether two products are in the same market.
- Be a bit selective on whether you should difference the time series or not because it limits the scope of questions we can answer.

5.16.1 Intuitive example of stationarity

- First of all, it is important to note that stationarity is a property of a process, not of a time series. You consider the ensemble of all time series generated by a process. If the statistical properties of this ensemble (mean, variance) are constant over time, the process is called stationary. Strictly speaking, it is impossible to say whether a given time series was generated by a stationary process (however, with some assumptions, we can take a good guess).
- More intuitively, stationarity means that there are no distinguished points in time for your process (influencing the statistical properties of your observation). Whether this applies to a given process depends crucially on what you consider as fixed or variable for your process, i.e., what is contained in your ensemble.
- A typical cause of non-stationarity are time-dependent parameters – which allow to distinguish time points by the values of the parameters. Another cause are fixed initial conditions.
- Examples:
  - The noise reaching my house from a single car passing at a given time is not a stationary process. E.g., the average amplitude is highest when the car is directly next to my house.

- The noise reaching my house from street traffic in general is a stationary process, if we ignore the time dependency of the traffic intensity (e.g., less traffic at night or on weekends). There are no distinguished points in time anymore. While there may be strong fluctuations of individual time series, these vanish when I consider the ensemble of all realisations of the process.
- If I we include known impacts on traffic intensity, e.g., that there is less traffic at night, the process is non-stationary again: The average amplitude varies with a daily rhythm. Every point in time is distinguished by the time of the day.

5.16.2 Identifying stationarity

Method 1: Graphical analysis

Method 2: Correlogram

- It let's us see how quickly the trend effect of a shock is disappearing.
- The correlogram plots the autocorrelations. The key is that the autocorrelations die away quickly for a stationary process but for a nonstationary process they do not.
- In the correlogram, we see the correlation over time.
- Rule of thumb: if a shock lasts for around 5 years or more, then we say the shock has "everlasting effects" (non-stationary)
- How to read the correlogram:
  - AC is the amount of autocorrelation we have from the basic autocorrelation function
    * the value for the *first* lag is the coefficient we would get on the *first* lag if we ran `reg lprice l1.price, noconst`
    * the value for the *second* lag is the coefficient we would get on the *second* lag if we ran `reg lprice l2.price, noconst`
    * and so on
  - PAC is the amount of autocorrelation we have when we also include the intercept in the function. This is a partial correlation since it measures the

correlation of y values that are k periods apart after removing the correlation from the intervening lags.

* the value for the *first* lag is the coefficient we would get on the *first* lag if we ran `reg lprice l1.price`. **If we see that 1 is in the confidence interval for `l1.price`, then we have a non-stationary process at this lag level.**
* the value for the *second* lag is the coefficient we would get on the *second* lag if we ran `reg lprice l2.price`. **If we see that 1 is in the confidence interval for `l2.price`, then we have a non-stationary process at this lag level.**
* and so on

– Q is the statistics value

– And then we have the p-value. If we're using a 5 % significance level, then everything from 5 % and down to 0 % indiciates autocorrelation. *I'm not completely sure about this one.*

Method 3: Unit root tests/Dickey–Fuller Tests

The standard tests are called Dickey–Fuller tests and first became available in 1979. They are also called unit root tests. Remember that unit roots are those processes that become stationary after differencing the y variable once. Null hypothesis is that it's still not stationary while the alternative hypothesis is that it is now stationary.

But here's the thing: there's three alternatives of the Dickey–Fuller test:

- 1) with sample average around zero: $\Delta y_t = \theta y_{t-1} + e_t$
- 2) with sample average not around zero: $\Delta y_t = \alpha + \theta y_{t-1} + e_t$
- 3) with sample average not around zero and with trend (useful if we think we got a trend in the data). : $\Delta y_t = \alpha + \delta t + \theta y_{t-1} + e_t$
- And you may get different conclusions with the different tests. It's best to start with the third alternative and work your way down to be on the safe side.
- It's important not to have autocorrelation when running Dickey–Fuller because you get biased standard errors, which fucks up the inference since we're using the t value. Then you should try adding more lags instead, and potentially also the trend. Anti–Nilsen likes starting with many lags (8 if monthly, 12 if quarterly, 5–6

if annual) and then work your way down. Then you check which of the longer ones are statistically significant.

- You could also use an **augmented version of the Dickey–Fuller test**, which involves **adding lags of the first difference**: For instance, $\Delta y_t = \alpha + \delta t + \gamma \boldsymbol{\Delta y_{t-1}} + \theta y_{t-1} + e_t$.
- The augmented test is useful because we can control for the possibility that the error term is autocorrelated.
- The t-values are not T-distributed anymore, so we need to use Dickey–Fuller's critical values. **If smaller than the critical value, we reject.**
- The original and the augmented test has the same critical values.
- H_0: non-stationary process ($\rho = 1 \iff \theta = \rho - 1 = 0$. So the time series is $I(2)$ or higher.
- H_1: stationary process ($\rho < 1 \iff \theta = \rho - 1 =< 0$.

Table 16.4 Critical Values for the Dickey–Fuller Test

Model	1%	5%	10%
$\Delta y_t = \gamma y_{t-1} + v_t$	−2.56	−1.94	−1.62
$\Delta y_t = \alpha_0 + \gamma y_{t-1} + v_t$	−3.43	−2.86	−2.57
$\Delta y_t = \alpha_0 + \alpha_1 t + \gamma y_{t-1} + v_t$	−3.96	−3.41	−3.13
Standard critical values	−2.33	−1.65	−1.28

Note: These critical values are taken from R. Davidson and J. G. MacKinnon (1993) *Estimation and Inference in Econometrics*, New York: Oxford University Press, p. 708.

Figure 48

```
***
* the housing price index used before
***

***
* the first form of the DF (no constant)
***
. dfuller lprice, regress noconst

Dickey-Fuller test for unit root                   Number of obs   =        41

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.982            -2.634            -1.950            -1.606

------------------------------------------------------------------------------
    D.lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |
         L1. |  -.0530969   .0267925    -1.98   0.054    -.1072466    .0010529
------------------------------------------------------------------------------

***
* reject the H0 (that the lrpice coeff. is = 0, i.e rho = 1).
* That means here rho < 1 and we have a stationary process
***
```

Tells STATA to show extended output

The t-statistic is smaller than the critical value. We therefore reject H0.

H0: $\boldsymbol{\theta = 0}$ <=> $\boldsymbol{\rho = 1}$

Thus here; $\rho < 1$ (we have a stationary series)

```
***
* the second form of the DF (a constant)
***
. dfuller lprice, regress

Dickey-Fuller test for unit root                   Number of obs   =        41

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.332            -3.641            -2.955            -2.611
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.6142

------------------------------------------------------------------------------
    D.lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |
         L1. |   -.066086   .0495985    -1.33   0.190    -.1664084    .0342364
       _cons |  -.0017658   .0056471    -0.31   0.756     -.013188    .0096565
------------------------------------------------------------------------------

***
* do not reject the H0 (that the lrpice coeff. is = 0, i.e rho = 1)
* That means here non-stationary process since rho = 1
***
```

The *t*-statistic is larger than the critical value. We can therefore not reject H0, i.e. we accept H0. And H0 is that we have $\theta = 0$ or $\rho = 1$. Thus, we have a non-stationary series.

```
***
* the third form of the DF (a constant and a trend)
***

. dfuller lprice, regress trend

Dickey-Fuller test for unit root                   Number of obs   =        41

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.944            -4.233            -3.536            -3.202
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.6318

------------------------------------------------------------------------------
D.lprice     |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |
         L1. |  -.1792339   .0922179    -1.94   0.059    -.3659192    .0074514
      _trend |   .0007018   .0004848     1.45   0.156    -.0002798    .0016833
       _cons |  -.0272921   .0184948    -1.48   0.148    -.0647329    .0101487
------------------------------------------------------------------------------

***
* do not reject the H0 (that the lrpice coeff. is = 0, i.e rho = 1)
* That means here non-stationary process since rho = 1
***
```

The t-statistic is larger than the critical value. We can therefore not reject H0, i.e. we accept H0. And H0 is that we have $\theta = 0$ or $\rho = 1$. Thus, we have a non-stationary series.

We see that the outcomes of the Dickey-Fuller tests depend on whether we include a constant and not, and also whether we include a trend or not. What should we do??

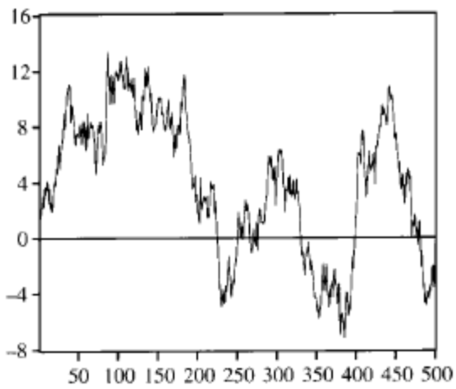
When to use what test..

When to use what test

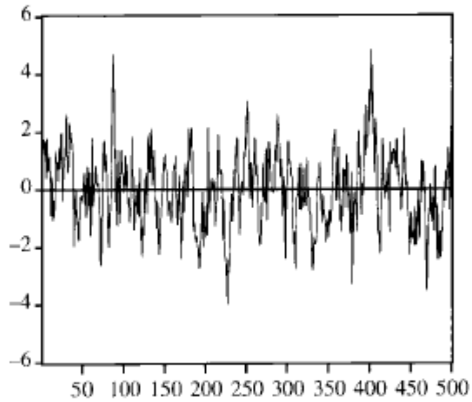
(1)

$$\Delta y_t = \theta y_{t-1} + e_t$$

If the series appears to be fluctuating around a sample average or zero.



(d) $y_t = y_{t-1} + v_t$

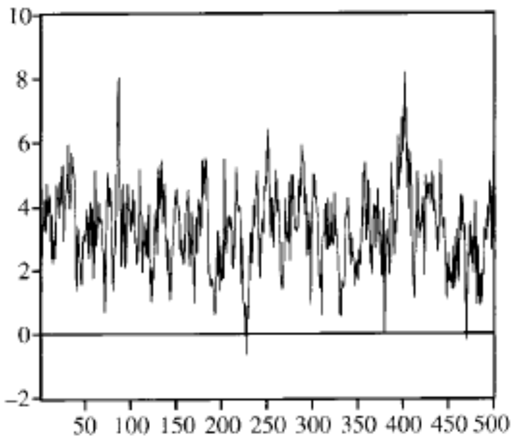


(a) $y_t = 0.7y_{t-1} + v_t$

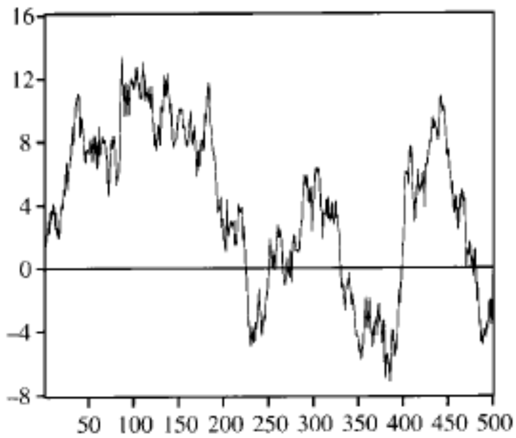
(2)

$$\Delta y_t = \alpha + \theta y_{t-1} + e_t$$

If the series appears to be fluctuating around a sample average which is nonzero.



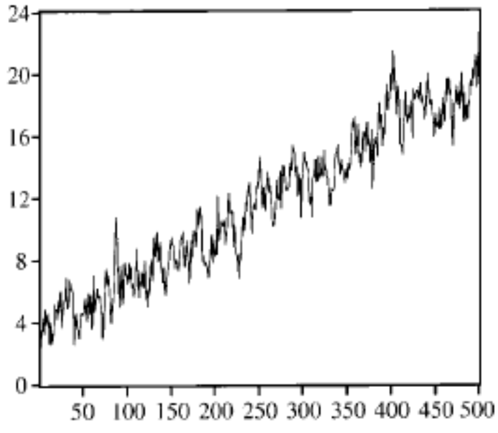
(*b*) $y_t = 1 + 0.7y_{t-1} + v_t$



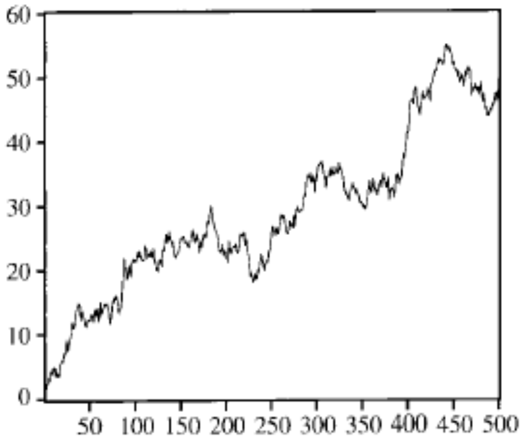
(*d*) $y_t = y_{t-1} + v_t$

(3)

$$\Delta y_t = \alpha + \delta \cdot t + \theta y_{t-1} + e_t$$



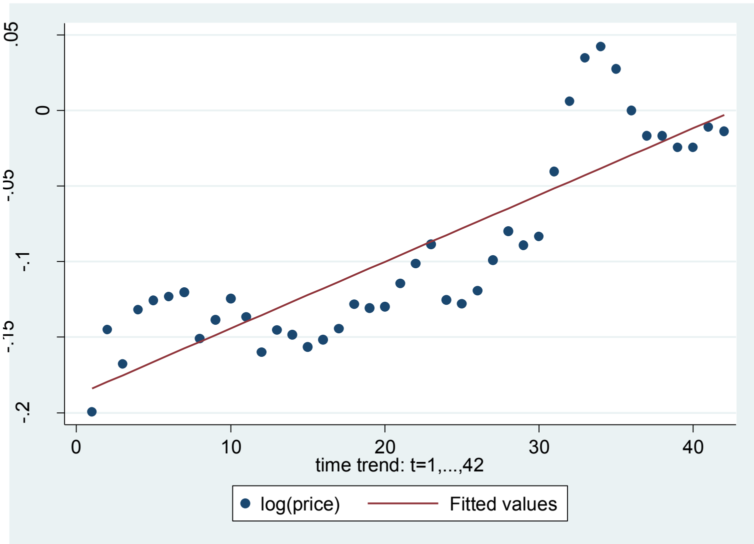
(c) $y_t = 1 + 0.01t + 0.7y_{t-1} + v_t$



(e) $y_t = 0.1 + y_{t-1} + v_t$

So what version of the DF test should we use for our example?

```
. twoway (scatter lprice t) (line prd_lpr t)
```



.

Augmented Dickey-Fuller test (if we believe we have autocorrelation in the error term)

(The standard Dickey Fuller)

$$y_t = \alpha + \rho y_{t-1} + e_t \qquad \text{AR(1)}$$
$$y_t - y_{t-1} = \alpha + (\rho - 1) y_{t-1} + e_t$$
$$\Delta y_t = \alpha + \theta y_{t-1} + e_t \qquad \text{where } \theta = (\rho - 1)$$

We can control for the possibility that e_t is autocorrelated by additional difference terms, Δy_{t-1}

$$\Delta y_t = \alpha + \theta y_{t-1} + \gamma \Delta y_{t-1} + e_t$$

This is called an augmented Dickey Fuller test.
More general[1]

$$\Delta y_t = \alpha + \theta y_{t-1} + \sum_{i=1}^{P} \gamma_i \Delta y_{t-i} + e_t$$

[1] The distribution of the parameters of γ_i follow approx. *t*-distribution.

(The critical values for ADF are the same as for the "standard" Dickey Fuller tests)

How to derive the augmented Dickey Fuller:

$$
\begin{aligned}
y_t &= \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + e_t \\
y_t - y_{t-1} &= \alpha + (\rho_1 - 1) y_{t-1} + \rho_2 y_{t-2} + e_t \\
\Delta y_t &= \alpha + (\rho_1 - 1) y_{t-1} + [\rho_2 y_{t-1} - \rho_2 y_{t-1}] + \rho_2 y_{t-2} + e_t \\
\Delta y_t &= \alpha + (\rho_1 + \rho_2 - 1) y_{t-1} - \rho_2 (y_{t-1} - y_{t-2}) + e_t \\
\Delta y_t &= \alpha + \theta y_{t-1} + \gamma \Delta y_{t-1} + e_t
\end{aligned}
$$

where

$$
\begin{aligned}
\theta &= (\rho_1 + \rho_2 - 1) \\
\gamma_1 &= -\rho_2 \\
\Delta y_{t-1} &= y_{t-1} - y_{t-2}
\end{aligned}
$$

```
. dfuller lprice, regress trend lags(2)

Augmented Dickey-Fuller test for unit root         Number of obs   =        39

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.409            -4.251            -3.544            -3.206
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3749

------------------------------------------------------------------------------
D.lprice     |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |
         L1. |  -.2216337    .092006    -2.41   0.022    -.4086124   -.0346549
         LD. |    .327572   .1551807     2.11   0.042      .012207     .642937
        L2D. |   .1300876   .1491206     0.87   0.389     -.172962    .4331372
      _trend |    .000971   .0004867     1.99   0.054    -.0000182    .0019602
       _cons |   -.039384   .0190149    -2.07   0.046    -.0780269   -.0007412
------------------------------------------------------------------------------
```

The number of lags used when performing the ADF

Formally one should use an information criterion, AIC or BIC, to decide about the lag length (see the stata command *dfgls*). But I normally find the lag length by starting with long lags (for instance 8 when using quarterly data, perhaps as much as 12 with monthly data, and 5-6 with annual data). Then I see whether the longest ones are statistically significant. If yes, keep them, if not reduce the lag-length. One could also look at the correlogram (in STATA; *corrgram, ac* or *pac*) on the differenced series (since the dependent variable in an ADF is differenced). The point of several lags is to be sure that autocorrelation is not creating biased st.errors. And one should **see whether the conclusion about stationarity vs non-stationarity changes when one plays with the lag length (and/or trend and constant/noconstant)**.

5.16.3 Handling stationarity

Method 1: do nothing—i.e. don't use the data

Method 2: first differences

Method 3: co-integration

1. Regress $y = \beta_0 + \beta_1 x_t$ and obtain residuals, $\hat{e}$.
2. Check whether the residuals are unit root with the Dickey–Fuller test: $\Delta\hat{e}_t = \alpha_0 + \gamma\hat{e}_{t-1} + v_t$
3. If they are both non-stationary processes and we conclude that they are stationary together (**not** unit root), meaning that these time series are cointegrated $\iff$ the model is the long-term equilibrium.

Submethod 3.2: The Engle–Granger procedure

1. Regress $y = \beta_0 + \beta_1 x_t$ and obtain residuals, $\hat{e}$.
2. Plug into $\Delta y_t = \gamma_0 \Delta x_t - (1-\alpha_1)\hat{\varepsilon}_{t-1} + u_t$
3. Which can be rewritten to $\Delta y_t = \gamma_0 \Delta x_t - (1-\alpha_1)[\mathbf{y_{t-1}} - \hat{\beta}_{\mathbf{0}} - \hat{\beta}_{\mathbf{1}}\mathbf{x_t}] + u_t$

Note that when we use $\hat{\beta}_0$ and $\hat{\beta}_1$ instead of β_0 and β_1, then we call this the Engle–Granger two-step procedure.

Example:

ECN402 – Econometric Techniques

Highly persistent time-series (Hill et al. 12)
- stationary process; definition
- random walk/highly persistent time-series; definition
- correlogram; a way of detecting highly persistent time-series
- Dickey Fuller; a way of detecting highly persistent time-series
- Augmented Dickey Fuller (controlling for autocorrelation)

X-tra (if we have time) (Wooldridge 18.4)
- Co-integrated variables
- Error-correction

We will see what we mean by **a stationary series**, and then of course also what we mean by a **non-stationary time series**.

We will also see that using non-stationary time series might lead to **spurious results**. Thus, we will also look at ways of **detecting non-stationarity** ((*i*) graphical analysis (plot the data), (*ii*) correlogram, and (*iii*) unit root test/ Dickey Fuller tests).

Then we will also see what we can do if we have non-stationary series ((*i*) nothing, (*ii*) make the series stationary by **taking the difference**, (*iii*) **co-integration**).

Stationary process; definition

A time series y_t is stationary if

$E(y_t) = \mu$ (constant mean)

$Var(y_t) = \sigma^2$ (constant variance)

$\text{cov}(y_t, y_{t-s}) = \gamma_s$ (covariance depends on s, not t)

Also called a covariance stationary process.

Ex. of a stationary process

$$y_t = 0.5 + 0.5 y_{t-1} + e_t$$

(draw a graph)

Random walk/highly persistent time-series; definition

Ex. of a random walk

$$y_t = y_{t-1} + e_t \quad \text{(Note: slope coeff. 1 and intercept 0)}$$

Ex of a random walk with drift

$$y_t = 0.1 + y_{t-1} + e_t \quad \text{(Note: slope coeff. 1 and intercept } \neq 0\text{)}$$

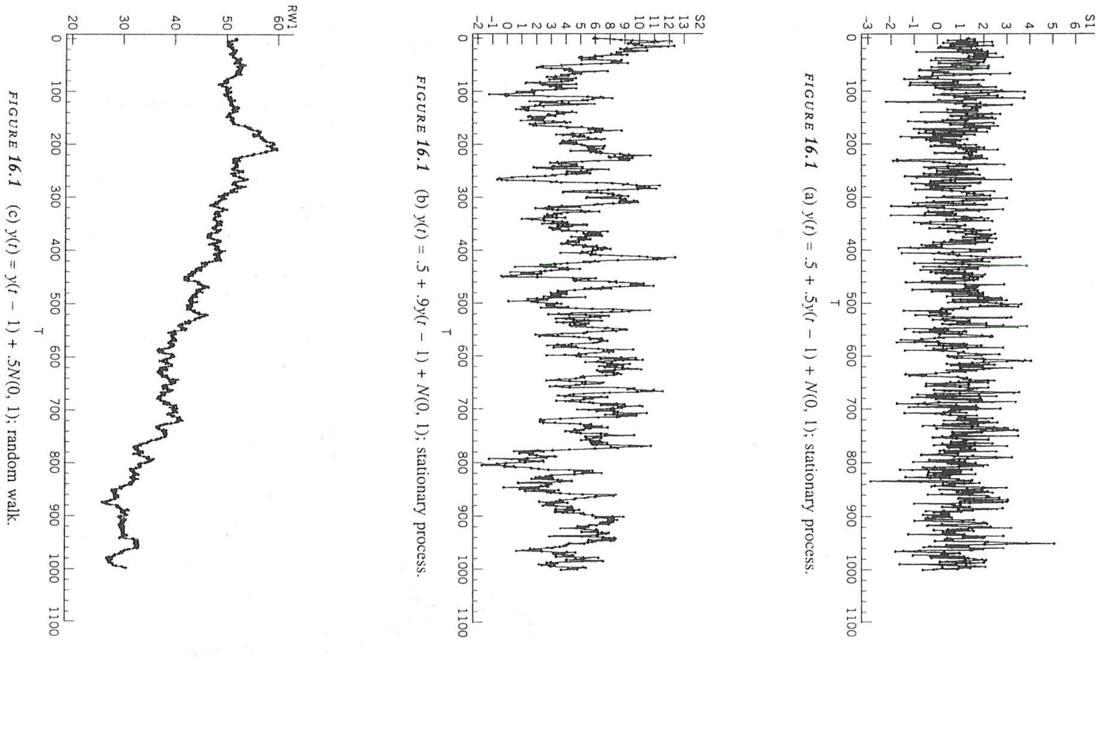
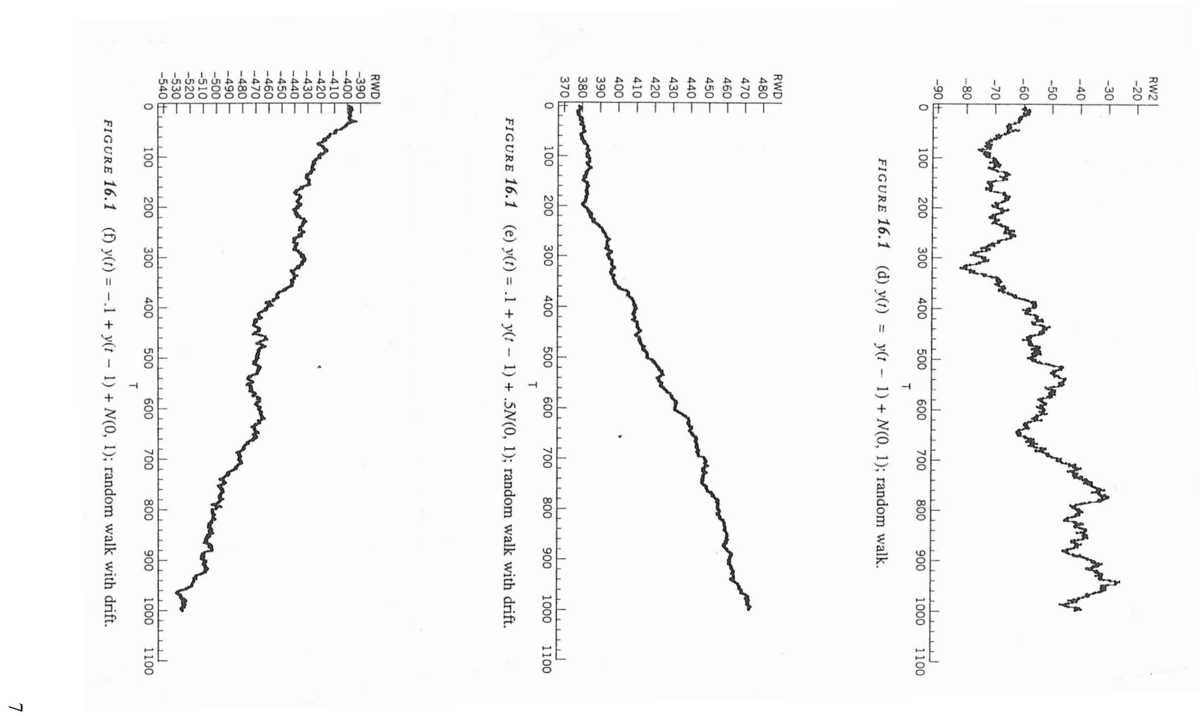


FIGURE 16.1 (a) $y(t) = .5 + .5y(t-1) + N(0, 1)$; stationary process.

FIGURE 16.1 (b) $y(t) = .5 + .9y(t-1) + N(0, 1)$; stationary process.

FIGURE 16.1 (c) $y(t) = y(t-1) + .5N(0, 1)$; random walk.



In general

(1)

$$y_t = \alpha + \rho y_{t-1} + e_t \qquad \text{AR(1) process}$$
$$y_t = \qquad y_{t-1} + e_t \qquad \text{Random walk}$$
$$y_t = \alpha + \; y_{t-1} + e_t \qquad \text{Random walk w/ drift}$$

What is special with a random walk process?

y_{t+1} depends strongly on y_t. But since y_t depends strongly on y_{t-1}, y_{t+1} will be highly correlated with y_{t-1}, y_{t-2} and so on. Thus, things that cause a discrete change in y_t can have long-lasting effects. (Think of an efficient capital market. Then the stock-price of a firm should stay constant until new information arrives. But, most importantly, today's price is a function all previous relevant information).

Random walk with drift

$$y_t = \alpha_0 + \rho y_{t-1} + v_t$$
$\rho = 1$ and then
$$y_t = \alpha_0 + y_{t-1} + v_t$$
$$y_1 = \alpha_0 + y_0 + v_1$$
$$y_2 = \alpha_0 + y_1 + v_2 = \alpha_0 + (\alpha_0 + y_0 + v_1) + v_2$$
$$y_3 = \alpha_0 + y_2 + v_3 = \alpha_0 + (2\alpha_0 + y_0 + v_1 + v_2) + v_3$$
$$\vdots$$
$$y_t = \alpha_0 \cdot t + y_0 + \sum_{s=1}^{t} v_s$$

We see that this series depends on all previous shocks (i.e. has a strong memory).

The term $\alpha \cdot t$ is called a *deterministic trend* (α is added each period), while $\sum_{s=1}^{t} v_s$ is called a *stochastic trend* (a stochastic component, v_t, is added each period).

If we think the random walk also have a standard time trend, the regression the looks like

$$y_t = \alpha_0 + \alpha_1 \cdot t + y_{t-1} + v_t$$
$$y_1 = \alpha_0 + \alpha_1 \cdot 1 + y_0 + v_1$$
$$y_2 = \alpha_0 + \alpha_1 \cdot 2 + y_1 + v_2 = \alpha_0 + \alpha_1 \cdot 2 + (\alpha_0 + \alpha_1 \cdot 1 + y_0 + v_1) + v_2 = \alpha_0 \cdot 2 + \alpha_1 \cdot 3 + y_0 + \sum_{s=1}^{2} v_s$$
$$\vdots$$
$$y_t = \alpha_0 t + \frac{t \cdot (t+1)}{2} \alpha_1 + y_0 + \sum_{s=1}^{t} v_s$$

And very, very important (see Wooldridge, chap 18.3):

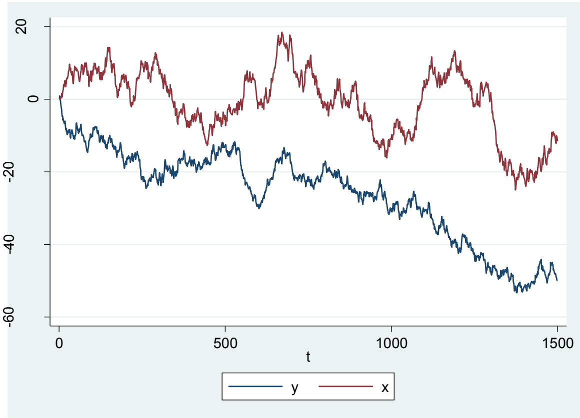
When non-stationary time series are used in a regression model the results indicate a significant relationship <u>even</u> when there is none => **SPURIOUS REGRESSION**

Example to show spurious results:
Make **two independent** series (y is not a function x, and x is not a function y)

Variance of the x-shocks = 1.0
Variance of the y-shocks = 0.5 (st.dev = 0.7)

```
set more off
clear all

set seed 54123

matrix C = ( 1.0, 0.00 \ 0.00, 0.5)
drawnorm u_x u_y , n(1500) cov(C) mean( 0 0) clear

gen int t = _n
tsset   t

gen     y = .
replace y = u_y if t == 1
gen     x = .
replace x = u_x if t == 1

forvalues i = 2(1)1500 {
  quietly replace  y = l1.y + u_y if `i' == t
  quietly replace  x = l1.x + u_x if `i' == t
}

summ
corr u_x u_y

twoway (line y t) (line x t)

reg y x, noheader
dis "R2 = " e(r2)
twoway (scatter y x)
```

$y(t) = y(t-1) + u\_y(t)$ where $u\_y(t) \sim N(0, 0.7^2)$
$x(t) = x(t-1) + u\_x(t)$ where $u\_x(t) \sim N(0, 1.0^2)$



We said we should have variance = 0.5 (st.dev = 0.70)

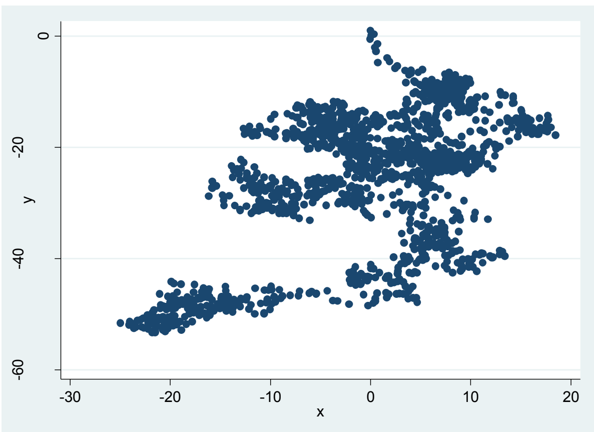
```
. sum u_y u_x
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         u_y |      1,500   -.0333204    .6999033  -2.009202   2.540901
         u_x |      1,500    -.007712    .9880102  -4.223562   3.020998
```

We said we should have variance = 1.0

```
. corr u_x u_y
             |      u_x      u_y
-------------+------------------
         u_x |   1.0000
         u_y |   0.0119   1.0000
```

No correlation between the two error terms


```
. reg y x, noheader
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .7390206   .0292103    25.30   0.000     .6817232    .7963179
       _cons |  -25.55825   .2712352   -94.23   0.000    -26.09029   -25.02621
------------------------------------------------------------------------------

. dis "R2 = " e(r2)
R2 = .29937532

. twoway (scatter y x)
```



Thus…. **BE CAREFUL IF YOU SUSPECT ONE OR ALL OF THE SERIES USED IN A REGRSSION TO BE NON-STATIONARY**.

Three ways of detecting highly persistent time series.

(1) Graphical analysis (plot the data)
(2) Correlogram
(3) Unit root test/ Dickey Fuller tests

Correlogram

For a stationary time-series we have

$$\text{cov}(y_t, y_{t+s}) = \gamma_s$$

The covariance between two observations in a series is only a function of distance *s* between the two, not time *t*. We can use this to construct the autocorrelation function

$$\rho_s = \frac{\text{cov}(y_t, y_{t+s})}{\text{var}(y_t)} = \frac{\gamma_s}{\gamma_0}$$

The estimated sample correlations are

$$\hat{\rho}_s = \frac{\text{cov}(y_t, y_{t+s})}{\text{var}(y_t)} = \frac{\hat{\gamma}_s}{\hat{\gamma}_0}$$

where

$$\hat{\gamma}_s = \frac{\sum (y_t - \bar{y})(y_{t+s} - \bar{y})}{T}$$

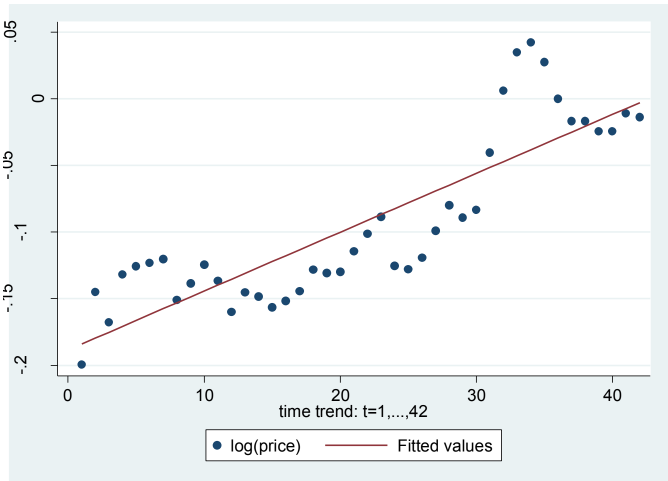
A plot of $\hat{\rho}_s$ against *s*, we get a **correlogram**

Ex: the housing prices again

```
. use "C:\Projects\Undervis\New_Wooldrige_Data\HSEINV.DTA"

. tsset year
        time variable:  year, 1947 to 1988
                delta:  1 unit

. reg lprice t

      Source |       SS       df       MS              Number of obs =      42
-------------+------------------------------           F(  1,    40) =  107.57
       Model |  .120404022     1  .120404022           Prob > F      =  0.0000
    Residual |  .044774132    40  .001119353           R-squared     =  0.7289
-------------+------------------------------           Adj R-squared =  0.7222
       Total |  .165178154    41  .004028735           Root MSE      =  .03346

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   .0044173   .0004259    10.37   0.000     .0035565    .0052781
       _cons |   -.188386   .0105121   -17.92   0.000    -.2096318   -.1671401
------------------------------------------------------------------------------

. predict prd_lpr
(option xb assumed; fitted values)
```



```
. twoway (scatter lprice t) (line prd_lpr t)
```



```
. corrgram lprice

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.8959   0.9339   36.174  0.0000          |-------           |-------
2        0.8099  -0.1382   66.475  0.0000          |------           -|
3        0.7084  -0.0270   90.257  0.0000          |-----             |
4        0.6351   0.0642   109.87  0.0000          |-----             |
5        0.5682   0.0345      126  0.0000          |----              |
6        0.4940  -0.1140   138.53  0.0000          |---               |
7        0.4280   0.2476    148.2  0.0000          |---               |-
8        0.3389   0.0229   154.44  0.0000          |--                |
9        0.2598   0.1748   158.22  0.0000          |--                |-
10       0.1911   0.0569   160.33  0.0000          |-                 |
11       0.1317   0.2909   161.36  0.0000          |-                 |--
12       0.0706  -0.0430   161.67  0.0000          |                  |
13       0.0261  -0.0425   161.71  0.0000          |                  |
14      -0.0243  -0.1687   161.75  0.0000          |                 -|
15      -0.0877   0.0305   162.28  0.0000          |                  |
16      -0.1357   0.2226   163.59  0.0000         -|                  |-
17      -0.1769   0.0397    165.9  0.0000         -|                  |
18      -0.2030   0.5180   169.08  0.0000         -|                  |----
19      -0.2341   0.0352   173.48  0.0000         -|                  |
```

Alternatively, use the STATA command “ac lprice”


```
. reg lprice l1.lprice

      Source |       SS       df       MS              Number of obs =      41
-------------+------------------------------           F(  1,    39) =  354.55
       Model |  .138389375     1  .138389375           Prob > F      =  0.0000
    Residual |  .015222652    39  .000390324           R-squared     =  0.9009
-------------+------------------------------           Adj R-squared =  0.8984
       Total |  .153612026    40  .003840301           Root MSE      =  .01976

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |
         L1. |    .933914   .0495985    18.83   0.000     .8335916    1.034236
       _cons |  -.0017658   .0056471    -0.31   0.756     -.013188    .0096
------------------------------------------------------------------------------

. reg lprice l1.lprice l2.lprice

      Source |       SS       df       MS              Number of obs =      40
-------------+------------------------------           F(  2,    37) =  197.86
       Model |  .137711274     2  .068855637           Prob > F      =  0.0000
    Residual |   .01287635    37  .000348009           R-squared     =  0.9145
-------------+------------------------------           Adj R-squared =  0.9099
       Total |  .150587623    39  .003861221           Root MSE      =  .01866

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |
         L1. |   1.097732   .1512165     7.26   0.000      .791338    1.404126
         L2. |  -.1382114   .1492885    -0.93   0.361    -.4406987    .1642759
       _cons |  -.0011207   .0055322    -0.20   0.841      -.01233    .0100887
------------------------------------------------------------------------------
```

This we recognise from the *corrgram* command on the previous page: lag 1 – pac (partial autocorrelation)

1 is in the conf. interval; we cannot reject that it is =1, we have a non-stationary process

This we recognise from the *corrgram* command on the previous page: lag 2 – pac (partial autocorrelation)

```
. reg lprice l1.lprice t

      Source |       SS       df       MS              Number of obs =      41
-------------+------------------------------           F(  2,    38) =  183.30
       Model |  .139184729     2  .069592365           Prob > F      =  0.0000
    Residual |  .014427297    38  .000379666           R-squared     =  0.9061
-------------+------------------------------           Adj R-squared =  0.9011
       Total |  .153612026    40  .003840301           Root MSE      =  .01949

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |
         L1. |   .8207661   .0922179     8.90   0.000     .6340808    1.007451
           t |   .0007018   .0004848     1.45   0.156    -.0002798    .0016833
       _cons |  -.0279938   .0189577    -1.48   0.148    -.0663717    .0103841
------------------------------------------------------------------------------
```



```
***
* a non-stationary time-series
***
. corrgram gs6m

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.9634   0.9706   219.99  0.0000          |-------           |-------
2        0.9140  -0.3286   418.83  0.0000          |-------         --|
3        0.8666  -0.0631   598.36  0.0000          |------            |
4        0.8179  -0.0070   758.97  0.0000          |------            |
5        0.7725  -0.0523    902.9  0.0000          |------            |
6        0.7209  -0.1126   1028.8  0.0000          |-----             |
7        0.6773   0.1583   1140.4  0.0000          |-----             |-
8        0.6476  -0.0597   1242.8  0.0000          |-----             |
9        0.6229   0.0950   1338.1  0.0000          |----              |
10       0.6043  -0.0958   1428.1  0.0000          |----              |
11       0.5865   0.1582   1513.3  0.0000          |----              |-
12       0.5713  -0.0315   1594.5  0.0000          |----              |
13       0.5578  -0.0074   1672.2  0.0000          |----              |
14       0.5438   0.0435   1746.5  0.0000          |----              |
15       0.5276  -0.0499   1816.7  0.0000          |----              |
16       0.5113  -0.0490   1882.9  0.0000          |----              |
17       0.4954   0.0701   1945.3  0.0000          |---               |
18       0.4753  -0.0184   2003.1  0.0000          |---               |
19       0.4524  -0.0144   2055.7  0.0000          |---               |
20       0.4287   0.1059   2103.1  0.0000          |---               |
```

Unit root test / Dickey Fuller; a way of detecting highly persistent time-series

$$y_t = \alpha + \rho y_{t-1} + e_t \qquad \text{AR(1)}$$
$$y_t - y_{t-1} = \alpha + (\rho - 1) y_{t-1} + e_t$$
$$\Delta y_t = \alpha + \theta y_{t-1} + e_t \qquad \text{where } \theta = (\rho - 1)$$

If $\theta = 0$ <=> $\rho = 1$ => Unit root w/ drift
If $\theta < 0$ <=> $\rho < 1$ => Stationary process

H0: $\theta = 0$ <=> $\rho = 1$
H1: $\theta < 0$ <=> $\rho < 1$

If $\theta = 0$ ($\rho = 1$), then y_t follows a random walk and

$$\Delta y_t = \alpha + e_t$$

Could we just use a *t*-test to test whether θ=0? NO…

If θ=0, then the *t*-statistics has <u>not</u> a *t*-distribution. But we can use a **Dickey-Fuller** test.

Critical values depend on the form of the model

(1)
$$\Delta y_t = \theta y_{t-1} + e_t$$

(2)
$$\Delta y_t = \alpha + \theta y_{t-1} + e_t$$

(3)
$$\Delta y_t = \alpha + \delta \cdot t + \theta y_{t-1} + e_t$$

Table 16.4 **Critical Values for the Dickey–Fuller Test**

Model	1%	5%	10%
$\Delta y_t = \gamma y_{t-1} + v_t$	−2.56	−1.94	−1.62
$\Delta y_t = \alpha_0 + \gamma y_{t-1} + v_t$	−3.43	−2.86	−2.57
$\Delta y_t = \alpha_0 + \alpha_1 t + \gamma y_{t-1} + v_t$	−3.96	−3.41	−3.13
Standard critical values	−2.33	−1.65	−1.28

Note: These critical values are taken from R. Davidson and J. G. MacKinnon (1993) *Estimation and Inference in Econometrics*, New York: Oxford University Press, p. 708.


```
***
* the housing price index used before
***

***
* the first form of the DF (no constant)
***
. dfuller lprice, regress noconst

Dickey-Fuller test for unit root                   Number of obs   =        41

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.982            -2.634            -1.950            -1.606

------------------------------------------------------------------------------
    D.lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |
         L1. |  -.0530969   .0267925    -1.98   0.054    -.1072466    .0010529
------------------------------------------------------------------------------

***
* reject the H0 (that the lrpice coeff. is = 0, i.e rho = 1).
* That means here rho < 1 and we have a stationary process
***
```

Tells STATA to show extended output

The *t*-statistic is smaller than the critical value. We therefore reject H0.
H0: $\theta = 0$ <=> $\rho = 1$
Thus here; $\rho < 1$ (we have a stationary series)

```
***
* the second form of the DF (a constant)
***
. dfuller lprice, regress

Dickey-Fuller test for unit root                   Number of obs   =        41

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.332            -3.641            -2.955            -2.611
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.6142

------------------------------------------------------------------------------
    D.lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |
         L1. |   -.066086   .0495985    -1.33   0.190    -.1664084    .0342364
       _cons |  -.0017658   .0056471    -0.31   0.756     -.013188    .0096565
------------------------------------------------------------------------------

***
* do not reject the H0 (that the lrpice coeff. is = 0, i.e rho = 1)
* That means here non-stationary process since rho = 1
***
```

The *t*-statistic is larger than the critical value. We can therefore not reject H0, i.e. we accept H0. And H0 is that we have $\theta = 0$ or $\rho = 1$. Thus, we have a non-stationary series.


```
***
* the third form of the DF (a constant and a trend)
***

. dfuller lprice, regress trend

Dickey-Fuller test for unit root                   Number of obs   =        41

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.944            -4.233            -3.536            -3.202
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.6318

------------------------------------------------------------------------------
D.lprice     |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |
         L1. |  -.1792339   .0922179    -1.94   0.059    -.3659192    .0074514
      _trend |   .0007018   .0004848     1.45   0.156    -.0002798    .0016833
       _cons |  -.0272921   .0184948    -1.48   0.148    -.0647329    .0101487
------------------------------------------------------------------------------

***
* do not reject the H0 (that the lrpice coeff. is = 0, i.e rho = 1)
* That means here non-stationary process since rho = 1
***
```

The *t*-statistic is larger than the critical value. We can therefore not reject H0, i.e. we accept H0. And H0 is that we have $\theta = 0$ or $\rho = 1$. Thus, we have a non-stationary series.

We see that the outcomes of the Dickey-Fuller tests depend on whether we include a constant and not, and also whether we include a trend or not. What should we do??

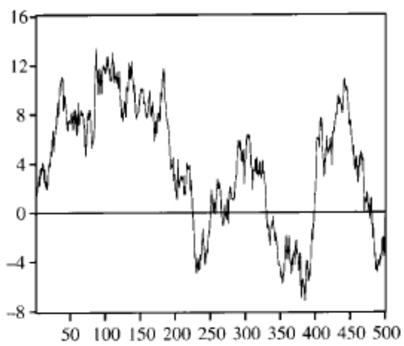
When to use what test..

When to use what test

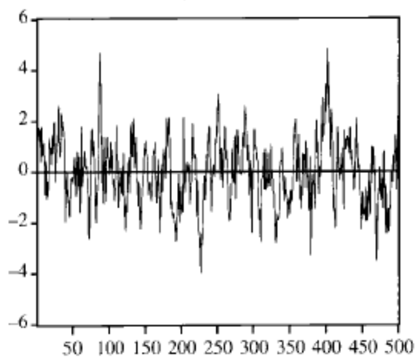
(1)

$$\Delta y_t = \theta y_{t-1} + e_t$$

If the series appears to be fluctuating around a sample average or zero.



(d) $y_t = y_{t-1} + v_t$

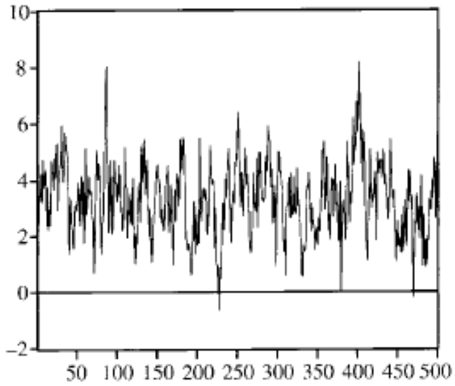


(a) $y_t = 0.7y_{t-1} + v_t$

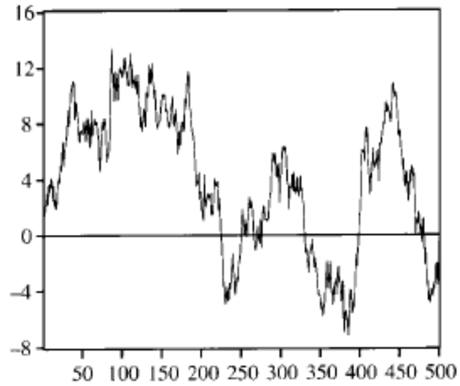
(2)

$$\Delta y_t = \alpha + \theta y_{t-1} + e_t$$

If the series appears to be fluctuating around a sample average which is nonzero.



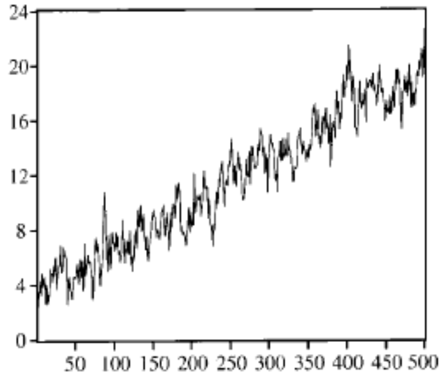
$(b)\ y_t = 1 + 0.7y_{t-1} + v_t$



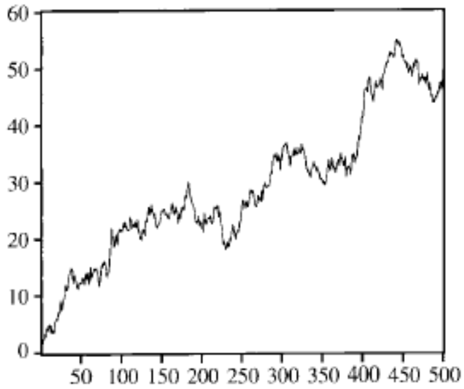
$(d)\ y_t = y_{t-1} + v_t$

(3)

$$\Delta y_t = \alpha + \delta \cdot t + \theta y_{t-1} + e_t$$



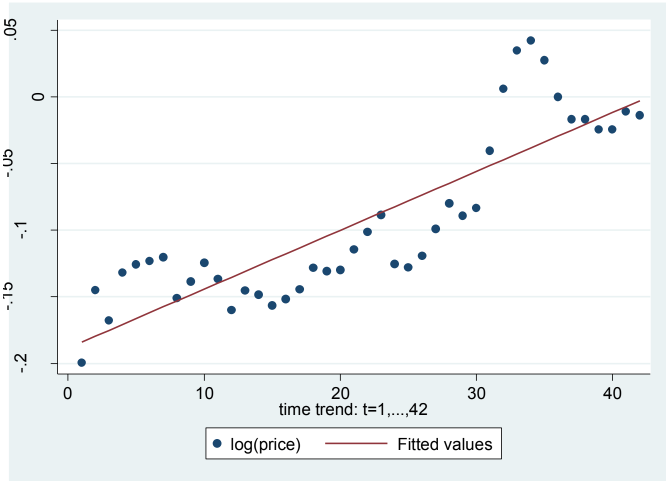
$(c)\ y_t = 1 + 0.01t + 0.7y_{t-1} + v_t$



$(e)\ y_t = 0.1 + y_{t-1} + v_t$

So what version of the DF test should we use for our example?

```
. twoway (scatter lprice t) (line prd_lpr t)
```



Augmented Dickey-Fuller test (if we believe we have autocorrelation in the error term)

(The standard Dickey Fuller)

$$y_t = \alpha + \rho y_{t-1} + e_t \qquad \text{AR(1)}$$
$$y_t - y_{t-1} = \alpha + (\rho - 1) y_{t-1} + e_t$$
$$\Delta y_t = \alpha + \theta y_{t-1} + e_t \qquad \text{where } \theta = (\rho - 1)$$

We can control for the possibility that e_t is autocorrelated by additional difference terms, Δy_{t-1}

$$\Delta y_t = \alpha + \theta y_{t-1} + \gamma \Delta y_{t-1} + e_t$$

This is called an augmented Dickey Fuller test.
More general[1]

$$\Delta y_t = \alpha + \theta y_{t-1} + \sum_{i=1}^{P} \gamma_i \Delta y_{t-i} + e_t$$

[1] The distribution of the parameters of γ_i follow approx. t-distribution.

(The critical values for ADF are the same as for the "standard" Dickey Fuller tests)

How to derive the augmented Dickey Fuller:

$$
\begin{aligned}
y_t &= \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + e_t \\
y_t - y_{t-1} &= \alpha + (\rho_1 - 1) y_{t-1} + \rho_2 y_{t-2} + e_t \\
\Delta y_t &= \alpha + (\rho_1 - 1) y_{t-1} + [\rho_2 y_{t-1} - \rho_2 y_{t-1}] + \rho_2 y_{t-2} + e_t \\
\Delta y_t &= \alpha + (\rho_1 + \rho_2 - 1) y_{t-1} - \rho_2 (y_{t-1} - y_{t-2}) + e_t \\
\Delta y_t &= \alpha + \theta y_{t-1} + \gamma \Delta y_{t-1} + e_t
\end{aligned}
$$

where

$$
\begin{aligned}
\theta &= (\rho_1 + \rho_2 - 1) \\
\gamma_1 &= -\rho_2 \\
\Delta y_{t-1} &= y_{t-1} - y_{t-2}
\end{aligned}
$$


```
. dfuller lprice, regress trend lags(2)

Augmented Dickey-Fuller test for unit root         Number of obs   =        39

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.409            -4.251            -3.544            -3.206
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3749

------------------------------------------------------------------------------
D.lprice     |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |
         L1. |  -.2216337    .092006    -2.41   0.022    -.4086124   -.0346549
         LD. |    .327572   .1551807     2.11   0.042      .012207     .642937
        L2D. |   .1300876   .1491206     0.87   0.389     -.172962    .4331372
      _trend |    .000971   .0004867     1.99   0.054    -.0000182    .0019602
       _cons |   -.039384   .0190149    -2.07   0.046    -.0780269   -.0007412
------------------------------------------------------------------------------
```

The number of lags used when performing the ADF

Formally one should use an information criterion, AIC or BIC, to decide about the lag length (see the stata command *dfgls*). But I normally find the lag length by starting with long lags (for instance 8 when using quarterly data, perhaps as much as 12 with monthly data, and 5-6 with annual data). Then I see whether the longest ones are statistically significant. If yes, keep them, if not reduce the lag-length. One could also look at the correlogram (in STATA; *corrgram, ac* or *pac*) on the differenced series (since the dependent variable in an ADF is differenced). The point of several lags is to be sure that autocorrelation is not creating biased st.errors. And one should **see whether the conclusion about stationarity vs non-stationarity changes when one plays with the lag length (and/or trend and constant/noconstant)**.

Non-stationary time-series and regressions

General rule:
If time-series are not stationary, don't use them. You might end up with spurious results.

There is one exception to the general rule

If two series are both integrated of order 1 (i.e. we have to take the first-difference to get the series stationary – see next page), we are allowed to use them given that the residual form the regression model are stationary.

What if the residual form the cointegration model is non-stationary?

Take the first-difference of the series. But ..... be careful with the interpretation. Such a regression gives information on the relationship between the changes in the variables, not the levels.

What if a time series is **integrated of order 1**?

What does it mean that a series is integreated of order 1?

We start with an AR(1) process

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

but we now assume $\theta = 1$. Then we might get a well-behaved series by taking the first-difference

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

We call this series **integrated of order one since**, we **get a stationary series after taking the difference once**. If we have to take the first-difference of the first-differences series, we call it integrated of order 2 (and so on…).

It is pointed out that when non-stationary time series are used in a regression model the results may indicate a significant relationship <u>even</u> when there is none => SPURIOUS REGRESSION

Thus a general rule is: If you might think that one or more of the time series that you plan to use in a regression model is nonstationary => DON'T USE THEM.

X-tra

Cointegration

There is one exception to the general rule.

We have two non-stationary time-series

$$Y_t = \alpha + \beta X_t + \varepsilon_t$$
$$\Updownarrow$$
$$\varepsilon_t = Y_t - \alpha - \beta X_t$$

If both Y and X are unit roots, it might that they "cancel each other out", i.e. that we get a series that is stationary. The idea is that if these two variables are cointegrated (share similar stochastic trends), they never diverge too far from each other.

(Cointegration if used in competition economics to see whether two products are in the same market. If so, this might have consequences for the outcome of a M&A application)

Ex:

Prices of two related products; interest rates of various T-bills

1) Run Y on an intercept and X
2) Get the residual
3) If the residual is not unit root, i.e. the residuals is stationary => series are cointegrated (own cointegration tables)

If you run

$$Y_t = \alpha + \beta X_t + \varepsilon_t$$
$$\Updownarrow$$
$$\varepsilon_t = Y_t - \alpha - \beta X_t$$

TABLE 18.4

Asymptotic Critical Values for Cointegration Test: No Time Trend

Significance level	1%	2.5%	5%	10%
Critical value	−3.90	−3.59	−3.34	−3.04

If we run include a time-trend in the cointegration regression

$$Y_t = \alpha + \beta X_t + \eta \cdot t + \varepsilon_t$$

$$\Updownarrow$$

$$\varepsilon_t = Y_t - \alpha - \beta X_t - \eta \cdot t$$

TABLE 18.5

Asymptotic Critical Values for Cointegration Test: Linear Time Trend

Significance level	1%	2.5%	5%	10%
Critical value	−4.32	−4.03	−3.78	−3.50

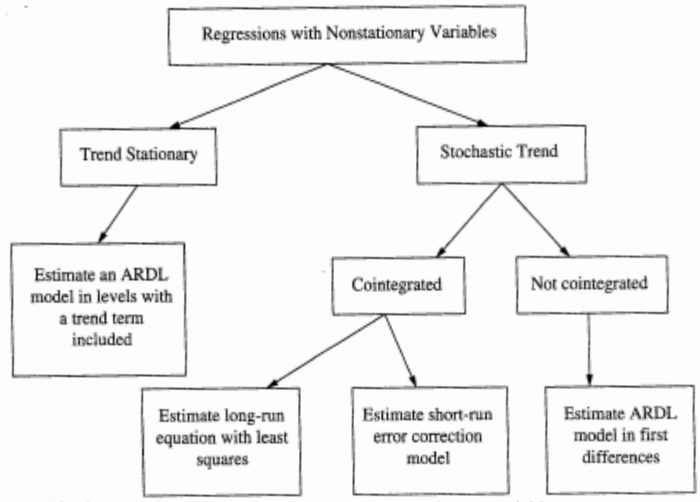


FIGURE 12.4 Regression with time-series data: nonstationary variables.

Error correction

In an "error correction" model specification is the change in one variable related to the change in another variable, as well as the gap between the variables in the previous period.

$$y_t = \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \alpha_1 y_{t-1} + u_t$$

$$y_t - y_{t-1} = \alpha_0 + \gamma_0 x_t - \gamma_0 x_{t-1} + \gamma_1 x_{t-1} + \gamma_0 x_{t-1} + (\alpha_1 - 1) y_{t-1} + u_t$$

$$y_t - y_{t-1} = \alpha_0 + \gamma_0 (x_t - x_{t-1}) + (\gamma_0 + \gamma_1) x_{t-1} + (\alpha_1 - 1) y_{t-1} + u_t$$

$$y_t - y_{t-1} = \frac{(1-\alpha_1)}{(1-\alpha_1)} \alpha_0 + \gamma_0 (x_t - x_{t-1}) + \frac{(1-\alpha_1)}{(1-\alpha_1)} (\gamma_0 + \gamma_1) x_{t-1} - (1-\alpha_1) y_{t-1} + u_t$$

$$y_t - y_{t-1} = \gamma_0 (x_t - x_{t-1}) - (1-\alpha_1) \left[y_{t-1} - \frac{\alpha_0}{(1-\alpha_1)} - \frac{(\gamma_0 + \gamma_1)}{(1-\alpha_1)} x_{t-1} \right] + u_t$$

$$\Delta y_t = \gamma_0 \Delta x_t - (1-\alpha_1) \underbrace{[y_{t-1} - \beta_0 - \beta_1 x_{t-1}]}_{\varepsilon_{t-1}} + u_t$$

We start from an ARDL model

Subtract y_{t-1} from the LHS and from the RHS

Collect terms to get the difference $x_t - x_{t-1}$

Multiply a couple of terms with $(1-\alpha_1) / (1-\alpha_1)$

Rearrange inside the square brackets (see the long-run multiplier of x on y)

Advantages:

- Incorporates short-run, and long-run effects
- Are using first-differenced variables (making terms stationary, including the error-correction term, ε_{t-1}).
- Adjustment speed; $(1-\alpha_1)$ if deviation from the long-run relationship – how quickly do we get back to the long-run steady-state?

1) Run the cointegration regression

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

2) Do a formal cointegration test
3) Generate the lagged residual

$$\hat{\varepsilon}_t$$

4) Include the lagged residual in the new regression

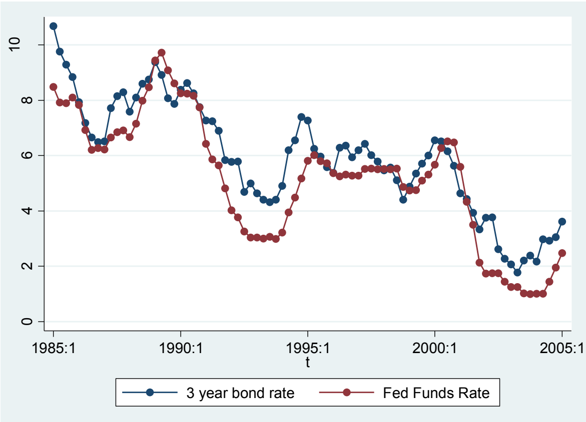
$$\Delta y_t = \gamma_0 \Delta x_t - (1-\alpha_1)\left[y_{t-1} - \hat{\beta}_0 - \hat{\beta}_1 x_t\right] + u_t$$
$$\Delta y_t = \gamma_0 \Delta x_t - (1-\alpha_1)\hat{\varepsilon}_{t-1} + u_t$$

Note: When we use $\hat{\beta}_0$ and $\hat{\beta}_1$ instead of use β_0 and β_1, we refer to this as Engle-Granger two-step procedure.

An example:

The dataset *usa.dta* include quarterly US data for the Federal Funds rate, *F,* and 3 year bond rate, *B,* for the period 1985:1 - 2005:1.

```
. use usa.dta
. tset t
        time variable:  t, 1985:1 to 2005:1
                delta:  1 quarter

. twoway (connected B t) (connected F t)
```



```
. corrgram F

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.9568   0.9741   76.933  0.0000          |-------           |-------
2        0.8865  -0.6725   143.82  0.0000          |-------      -----|
3        0.7978   0.0752   198.67  0.0000          |------            |
4        0.6961  -0.0810   240.98  0.0000          |-----             |
5        0.5942   0.0593    272.2  0.0000          |----              |
6        0.5000  -0.0797   294.62  0.0000          |----              |
7        0.4127  -0.0613   310.09  0.0000          |---               |
8        0.3289   0.0433   320.06  0.0000          |--                |
9        0.2531   0.1615   326.04  0.0000          |--                |-
10       0.1828  -0.0764    329.2  0.0000          |-                 |
11       0.1211   0.0619   330.61  0.0000          |                  |
12       0.0685   0.0257   331.07  0.0000          |                  |
13       0.0251  -0.0465   331.13  0.0000          |                  |
14      -0.0089   0.2971   331.14  0.0000          |                  |--
15      -0.0294   0.0897   331.22  0.0000          |                  |
16      -0.0370   0.1873   331.37  0.0000          |                  |-
     :
     :
38       0.1788   0.2902   351.61  0.0000          |-                 |--
```



```
. corrgram B

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.9235   0.9396   71.677  0.0000          |-------           |-------
2        0.8342  -0.3151    130.9  0.0000          |------          --|
3        0.7458   0.0172   178.85  0.0000          |-----             |
4        0.6542  -0.1675   216.21  0.0000          |-----            -|
5        0.5783   0.2208    245.8  0.0000          |----              |-
6        0.5206  -0.0504   270.09  0.0000          |----              |
7        0.4667  -0.0419   289.88  0.0000          |---               |
8        0.4217   0.1482   306.25  0.0000          |---               |-
9        0.3811  -0.0440   319.82  0.0000          |---               |
10       0.3298   0.1707   330.11  0.0000          |--                |-
11       0.2890   0.1220   338.14  0.0000          |--                |
12       0.2510  -0.1373   344.28  0.0000          |--               -|
13       0.2217   0.0710   349.13  0.0000          |-                 |
14       0.1937   0.2105    352.9  0.0000          |-                 |-
15       0.1675   0.1074   355.76  0.0000          |-                 |
     :
     :
38       0.0668   0.1990   378.05  0.0000          |                  |-
```

```
. dfuller F, regress noconst

Dickey-Fuller test for unit root                     Number of obs   =        80

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.754            -2.608            -1.950            -1.610

------------------------------------------------------------------------------
         D.F |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |
         L1. |  -.0162724   .0092766    -1.75   0.083    -.0347371    .0021923
------------------------------------------------------------------------------
```

Is this the correct DF test?

```
. dfuller F, regress
Dickey-Fuller test for unit root                     Number of obs   =        80

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.113            -3.538            -2.906            -2.588
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.7097

------------------------------------------------------------------------------
         D.F |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |
         L1. |  -.0259003    .023264    -1.11   0.269    -.0722152    .0204147
             |
       _cons |   .0596832   .1321246     0.45   0.653    -.2033566     .322723
------------------------------------------------------------------------------
```

What about this; is this the correct DF test?


```
. dfuller F, regress trend
Dickey-Fuller test for unit root                     Number of obs   =        80
                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.393            -4.084            -3.470            -3.162
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.8630
------------------------------------------------------------------------------
D.F          |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |
         L1. |  -.0489295   .0351312    -1.39   0.168    -.1188847    .0210256
      _trend |  -.0030329   .0034628    -0.88   0.384    -.0099283    .0038625
       _cons |   .3023439   .3070366     0.98   0.328    -.3090442    .9137319
------------------------------------------------------------------------------

. dfuller F, regress trend lags(1)
Augmented Dickey-Fuller test for unit root           Number of obs   =        79
                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -3.140            -4.086            -3.471            -3.163
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0971
------------------------------------------------------------------------------
D.F          |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |
         L1. |  -.0812005   .0258611    -3.14   0.002    -.1327184   -.0296825
         LD. |   .6946459   .0836366     8.31   0.000     .5280333    .8612584
      _trend |  -.0058423   .0025498    -2.29   0.025    -.0109217   -.0007629
       _cons |   .6470399   .2270279     2.85   0.006     .1947772    1.099303
------------------------------------------------------------------------------
```

Here we see several things

i) The df-test statistics depend on whether we include lagged differences or not
ii) The significance of the trend depends on whether we include lagged differences or not
iii) The conclusion about stationarity seems to be unaffected independent of whether we include/exclude lags/trend.

```
. *** The first command below shows how you could run the DF-regression yourself
. reg d.F l1.F l1.d1.F t, noheader

------------------------------------------------------------------------------
         D.F |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |
         L1. |  -.0812005   .0258611    -3.14   0.002    -.1327184   -.0296825
         LD. |   .6946459   .0836366     8.31   0.000     .5280333    .8612584
             |
           t |  -.0058423   .0025498    -2.29   0.025    -.0109217   -.0007629
       _cons |   1.231273   .4700809     2.62   0.011     .2948232    2.167722
------------------------------------------------------------------------------

. dfuller F, regress trend lags(1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        79
                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -3.140            -4.086            -3.471            -3.163
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0971

------------------------------------------------------------------------------
D.F          |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |
         L1. |  -.0812005   .0258611    -3.14   0.002    -.1327184   -.0296825
         LD. |   .6946459   .0836366     8.31   0.000     .5280333    .8612584
      _trend |  -.0058423   .0025498    -2.29   0.025    -.0109217   -.0007629
       _cons |   .6470399   .2270279     2.85   0.006     .1947772    1.099303
------------------------------------------------------------------------------
```

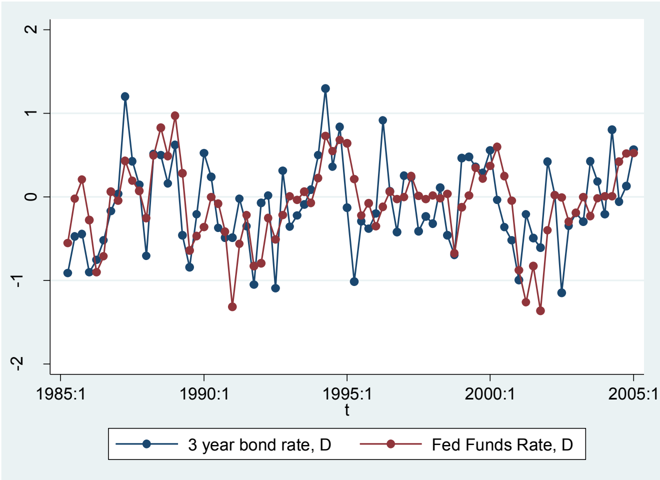
- The F series is non-stationary
- It seems relevant to include a trend
- It seems relevant to include lags


```
. gen dF = d1.F

. dfuller dF, regress noconst lag(1)
Augmented Dickey-Fuller test for unit root         Number of obs   =        78
                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -3.965            -2.609            -1.950            -1.610

------------------------------------------------------------------------------
        D.dF |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          dF |
         L1. |  -.3729292   .0940624    -3.96   0.000    -.5602708   -.1855876
         LD. |   .1254171   .1131793     1.11   0.271    -.0999989    .3508332
------------------------------------------------------------------------------

. gen dB = d1.B
. dfuller dB, regress noconst lag(1)
Augmented Dickey-Fuller test for unit root         Number of obs   =        78

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -5.315            -2.609            -1.950            -1.610
------------------------------------------------------------------------------
        D.dB |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          dB |
         L1. |  -.7065167    .132938    -5.31   0.000    -.9712856   -.4417477
         LD. |   .0293653   .1132489     0.26   0.796    -.1961894    .2549201
------------------------------------------------------------------------------
```

Here we take the first difference of the two series, to test whether they are both integrated of order 1 (i.e. becomes stationary after having taken the first-difference)

- Both of the series becomes stationary after having taken the first-difference once. That means, they are both integrated of order 1.
- Why are we using the noconst option?

```
. twoway (connected d1.B t) (connected d1.F t)
```




```
. reg B F, noheader

------------------------------------------------------------------------------
           B |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |   .8325052   .0344759    24.15   0.000     .7638826    .9011279
       _cons |   1.643733    .194819     8.44   0.000     1.255956    2.031511
------------------------------------------------------------------------------

. predict u_hat, resid

***
* Since u_hat is a residual, the mean is zero. Thus, use the noconstant option
***
. dfuller u, lag(3) noconst regress
Augmented Dickey-Fuller test for unit root         Number of obs   =        77
                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -4.386            -2.609            -1.950            -1.610
------------------------------------------------------------------------------
     D.u_hat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       u_hat |
         L1. |  -.3777153   .0861227    -4.39   0.000    -.5493575   -.2060731
         LD. |   .3739371   .1092245     3.42   0.001      .156253    .5916212
        L2D. |  -.0162649   .1107365    -0.15   0.884    -.2369625    .2044327
        L3D. |    .221658   .1087306     2.04   0.045     .0049582    .4383579
------------------------------------------------------------------------------

. ***
. * No constant, no time trend; t-value = -4.39 < -3.90 = critical value (1% sign. level)
. *          from the cointegration table, i.e. we reject H0 (random walk)
. *    Thus the residual is stationary and we can believe in the regression
. *    results. That means we have series that are cointegrated.
. ***
```

Compare these critical values to the ones (and correct ones) on the next page)

```
***
* Using the egranger command (download it), you will get the correct critical values
***
egranger B F, lag(3) regress

*** The t-value of the two tests is identical, but the critical values are different.
. egranger B F, lag(3) regress

Augmented Engle-Granger test for cointegration        N (1st step)  =        81
Number of lags   =  3                                 N (test)      =        77
------------------------------------------------------------------------------
                  Test          1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -4.386            -4.037            -3.414            -3.098

Critical values from MacKinnon (1990, 2010)
------------------------------------------------------------------------------
Engle-Granger 1st-step regression
------------------------------------------------------------------------------
           B |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |   .8325052   .0344759    24.15   0.000     .7638826    .9011279
       _cons |   1.643733    .194819     8.44   0.000     1.255956    2.031511
------------------------------------------------------------------------------
Engle-Granger test regression
------------------------------------------------------------------------------
  D._egresid |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _egresid |
         L1. |  -.3777153   .0861227    -4.39   0.000    -.5493575   -.2060731
         LD. |   .3739371   .1092245     3.42   0.001      .156253    .5916212
        L2D. |  -.0162649   .1107365    -0.15   0.884    -.2369625    .2044327
        L3D. |    .221658   .1087306     2.04   0.045     .0049582    .4383579
------------------------------------------------------------------------------
```

Here are the correct critical values for a cointegration test

Conclusion:

The two series B and F are cointegrated, and we can therefore trust the slope coefficient even if both B and F are non-stationary.

Cointegration means that there is a long-run equilibrium relationship between the two variables (they do not deviate too much from each other).

What does it mean economically?

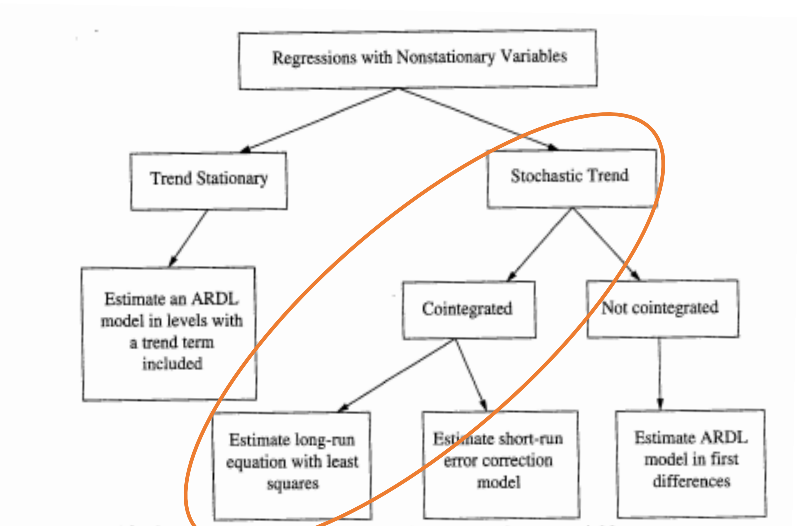


FIGURE 12.4 Regression with time-series data: nonstationary variables.

```
***
* the error correction model
***
. reg d1.B d1.F l1.u_hat

      Source |       SS       df       MS              Number of obs =      80
-------------+------------------------------           F(  2,    77) =   24.50
       Model |  8.68553471     2  4.34276736           Prob > F      =  0.0000
    Residual |  13.6464097    77    .1772261           R-squared     =  0.3889
-------------+------------------------------           Adj R-squared =  0.3731
       Total |  22.3319444    79  .282682841           Root MSE      =  .42098

------------------------------------------------------------------------------
         D.B |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |
         D1. |   .7529687   .1080218     6.97   0.000     .5378699    .9680675
             |
       u_hat |
         L1. |  -.2357887   .0723679    -3.26   0.002    -.3798916   -.0916858
             |
       _cons |  -.0315012   .0477661    -0.66   0.512    -.1266156    .0636133
------------------------------------------------------------------------------

***
* what is the interpretation of the coeff. in front of l1.u_hat?
***
```

$$y_t = \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \alpha_1 y_{t-1} + u_t$$

$$\Updownarrow$$

$$\Delta y_t = \gamma_0 \Delta x_t - (1-\alpha_1)\left[y_{t-1} - \beta_0 - \beta_1 x_{t-1}\right] + u_t$$

Is it correct to include a constant here?

slope = 0.753; a change in the fund rate F is relatively quickly reflected in the Bond rate B (i.e. monetary policy works, and consistent with the Engle-Granger test results slope = .8325052)

coeff. = -0.236; 0.24 of the discrepancy in the two rates in the previous month is eliminated this month.

Are we sure that there are no autocorrelation problems? If so, what should we do?

5.17 Short examples

What I want to test: Are house prices determined by the same model in university and non-university towns?

$$price = \beta_{+}\beta_1 sqft + \beta_2 yearbuilt + \delta_1 view + u$$

What I want to test: are wages different between men and women, and how does education affect the situation?

$$\ln(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 female \cdot educ + u$$

$H_0 : \delta_0 = \delta_1 = 0$ (average wages are identical for men and women who have the same levels of education. Run F-test.

What I want to test: how effective was the job training program?

$$wage = \beta_0 + \beta_1 age + \beta_2 educ + \delta_1 train + u$$

What I want to test: how effective was the job training program, and will the educational level for those who participate also affect wages?

$$wage = \beta_0 + \beta_1 age + \beta_2 educ + \delta_1 train + \gamma_1 educ \cdot train + u$$

What I want to test: is there a non-linear effect of age on wages?

$$\ln(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \beta_2 age + \beta_3 age^2 + u$$

What I want to test: Does the return to education depend on gender?

$$\ln(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \beta_2 age + \delta_1 educ \cdot female + u$$

What I want to test: Are house prices determined by the same model in university and non-university towns?

$$price = \beta_0 + \beta_1 sqft + \beta_2 yearbuilt + \delta_1 view + u$$

What I want to test: Does the return to education depend on age?

$$\ln(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \beta_2 age + \beta_3 educ \cdot age + u$$

What I want to test: Are all coefficients equal across groups?

Restricted model:

$$price = \beta_0 + \beta_1 sqft + \beta_2 pool + \beta_3 age + u$$

Unrestricted model:

$$price = \beta_0 + \beta_1 sqft + \beta_2 pool + \beta_3 age + \delta_0 Utown + \delta_1 sqft \cdot Utown + \delta_2 pool \cdot Utown + \delta_3 age \cdot Utown + u$$

What I want to test: What is the effect of another dollar of law enforcement on crime? Including lagged crime helps us measure that effect properly.

$$crim = \beta_0 + \beta_1 unem + \beta_2 expend + \beta_3 crime_{-1} + u$$

What I want to test: How consumption depends on income.

I'll likely see that I'm a bit careful with my consumption although my income is increasing.

$$C_t = \beta_0 + \beta_1 y_t + \gamma_{t-1} + u_t$$

What I want to test: relationship between unemployment and inflation.

$$\pi_t + \beta_0 + \beta_1 unempl_t + u_t$$

1. Run the regression as normal.
2. Get spurious results so de-trend the data by adding t.
3. Still unexpected results so run tests for identifying whether we have autocorrelation, like Durbin–Watson. Should preferably run several tests to be sure, so examine graphically, test with lags on the error term, and check a correlogram as well.

4. If the tests indicates autocorrelation, correct it, for instance with the quasi-difference data method or the Cochrane–Orcutt method.

5.18 Wider exam questions

5.18.1 V18 Q2

We have data on the production of oil and gas from 264 counties in 5 US-states (Idaho, Montana, Colorado, North Dakota and South Dakota) for the year 1980. *Fields* is the number of producing oil and gas fields in a county. *WagePP* is the county-level aggregate annual per capita wage, measured in US dollars. The average number of oil and gas fields in the data in 1980 is 10, and the average wage per capita is \$3800. We are interested in estimating the effect of oil and gas fields on wages, and estimate the following regression:

$$wagePP_i = \beta_0 + \beta_1 fields_i + u_i$$

```
. reg wagePP fields
------------------------------------------------------------------------------
      wagePP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      fields |   11.09928   5.909638     1.88   0.061    -.5371498    22.73571
       _cons |   3695.216   180.4633    20.48   0.000     3339.873    4050.559
------------------------------------------------------------------------------
```

Figure 49

a) Interpret the estimated coefficients. Is the sign of the estimated coefficient consistent with your economic expectation?

Proceed as follows (it's a bit overkill compared to the solutions manual, but it covers things well):

- **Interpret literally**: Use terms like "one additional unit of x *is associated with...*"/"the coefficient suggest that one additional unit of x gives...".
- **Talk about what a literal interpretation like this requires**:
  - **p-value**: is it significantly different from zero?
  - **constant and the origin**: do we observe the origin in the data set?
  - **assumptions need to hold**: if not, we may for instance get biased estimators, leading to erroneous estimates.
- Based on the bullet points above, potentially summarize with a quick conclusion.

So:

- The output states that one additional oil and gas field in a county is associated with a \$11.10 higher aggregate per capita wage per year in that county, everything else held equal.
- The output also states that the aggregate per capita wage per year in a county without any oil and gas fields is \$3695.
- However, this may be a too literal interpretation of the output, and there are some issues here. For instance, the fields estimate is significant only at a 10 % level. If we use a significance level of 5 %, we cannot really conclude that the estimate is significantly different from zero. On the other hand, the p-value is close to 5 % and the sign of the estimated coefficient makes sense—that additional oil and gas fields are associated with higher per-capita wages across the county—so it seems reasonable to not focusing too much on the p-value.
- In addition, the constant can be interpreted literally only if the origin is among the sample since we don't know whether there exists a linear relationship outside of the domain we studied in our sample. It's reasonable to assume that at least some of the 264 counties did not have an oil or gas field, meaning that we can likely interpret this statistically significant estimate in a literal way.
- And finally, to have a causal interpretation, we need the regression assumptions to hold. This is however unlikely, in part because we have too few variables in the model. This creates an *omitted variable bias*, violating the assumptions (and especially the zero conditional mean assumption).

From the solution:

> We see that one additional oil and gas field is associated with an increase of approximately \$11 in per capita wages. This is statistically significant at the 10% level. The sign suggests that an increase in oil and gas fields would increase wages. This makes sense as the opening of a new oil/gas field would lead to increased economic opportunity and potentially an increase in wages required to attract new additional workers into employment in the oil/gas field. We see from this simple model that a county with 0 oil and gas fields would have a per capita wage of \$3695.

b) The researcher has now added five additional control variables: idaho, montana, colorado, and north_ dakota, binary variables equal to 1 if a county is located in one of these states as well as totEMP, the total number of employees in the county. Interpret the estimated coefficient β_{idaho}*. Discuss how the coefficient on the variable fields changes when including* *the additional control variables. Discuss what may be a reason for such a change.*

```
------------------------------------------------------------------------------
      wagePP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      fields |   5.517378   6.015098     0.92   0.360    -6.327779    17.36254
      totEMP |   .0631531   .0156362     4.04   0.000     .0323618    .0939445
       idaho |   1736.371   530.3445     3.27   0.001     691.9966    2780.745
     montana |   857.3935   482.2863     1.78   0.077    -92.34278     1807.13
    colorado |   1376.785   515.3574     2.67   0.008     361.9242    2391.646
north_dakota |   789.5051   487.9313     1.62   0.107    -171.3476    1750.358
       _cons |   2464.249   332.6954     7.41   0.000     1809.093    3119.405
------------------------------------------------------------------------------
```

Figure 50

- *Fields* get less significant since the coefficient fell substantially, and we cannot conclude that this is significantly different from 0. Therefore, the number of fields may not be driving wages.
- Instead, total employees may be more important than the number of fields. Fields create jobs directly and indirectly, but then it's important to consider the size of those fields. And size is likely more important to wages than the number of fields. This is likely captured better in *TotEmp* because size and jobs created from field are generally positively correlated, and jobs created affect *TotEmp*.
- Furthermore, the *Idaho* variable's estimated coefficient states that Idaho has $1736 higher wages than the base group (South Dakota), everything else held equal.
- In conclusion, controlling for state-dependent factors as well as the number of employees in the county makes the number of fields less significant, and this indicates that the simple regression model in a) did indeed suffer from omitted variable bias.

From the solution:

> We see that the coefficient on Idaho is positive and statistically significant. This suggests higher per capita wages in Idaho than in the excluded base state, South Dakota. The coefficient suggests that a county in Idaho has $1736 higher per capita wages than a county in South Dakota. Note that it is not sufficient to simply say there are higher wages in Idaho without mentioning the fact that South Dakota is the excluded state, as the question is clear there is data on 5 states only. The estimated coefficient of fields changes considerably when adding these additional control variables. We see the magnitude of the coefficient declines from 11 to 5.5, and also becomes insignificant (though we can see 5.5 is well within the 95% CI of question a). A likely reason for this

large change is by including these state dummy variables (as well as a control variable for total employment) the regression in question b is reducing the omitted variable bias compared to the regression in question a. We think the estimated coefficient in question a) was biased. In particular, there are a lot of state-specific factors which are being captured by the state dummy variables.

c) Do you think the estimation in question b above represents the causal effect of oil and gas fields on county-level wages per capita? If so, why? If not, why not?

- A causal interpretation requires that the assumptions for the model hold.
- Such a large topic as aggregate per-capita wages in a county will likely be affected by much more than the number of oil and gas fields, total employees in the county, and state-specific dummies, such as:
  - The ratio of white- versus blue-collar workers. A predominantly blue-collar county could potentially see more increased wages than a white-collar county.
  - The oil price is not included in the model, and the revenues from the oil production will likely be important for the wage level as well.
  - The taxation level of the profits from oil and gas production may also be an important driver because higher taxes might create more jobs indirectly from the production than low taxes would (since more money would end up at the owners of the oil and gas companies).
- Therefore, it's unlikely that we have a causal effect from this model.
- Additionally, we could be facing **reverse causality** here: that higher wages lead to opening of new fields, which is the opposite of what we are studying in our model.

It seems unlikely that this is the causal effect of oil and gas fields on county level wages. While we have included these additional variables in the previous question, there are many other factors not included. We think that the zero conditional mean assumption does not hold. This should be stated with respect to the relevant information in the question, here we assume: $\mathbb{E}(u|fields, totEMP, idaho, \ldots, north\_dakota) = 0$.

This is likely to be violated by the importance of additional relevant county level factors which are omitted. One example of an omitted factor could be employment composition of counties. Employment composition of the county likely impacts the decision to start production on a new oil/gas field and also affects wages. For example, if there is a county with a high composition of white-collar jobs, then perhaps an additional oil field has a very different effect on wages here than in a county with a high composition of blue-collar jobs. This is only one possible example - the point is that there are many other additional economic factors besides total employment in the county which might impact the production decision and affect wages which are omitted from the model. Additionally, you might argue that an increase in the county wage level leads to the opening of new oil and gas fields, a story of reverse causality.

If there is a valid argument for why you think this is the causal effect, this can be acceptable but this must be clear why exactly this is the case and well argued.

d) Below you see a scatter plot of the residuals and fitted values from the regression in question b. Does the scatterplot suggest any problems with the estimation results? Explain.

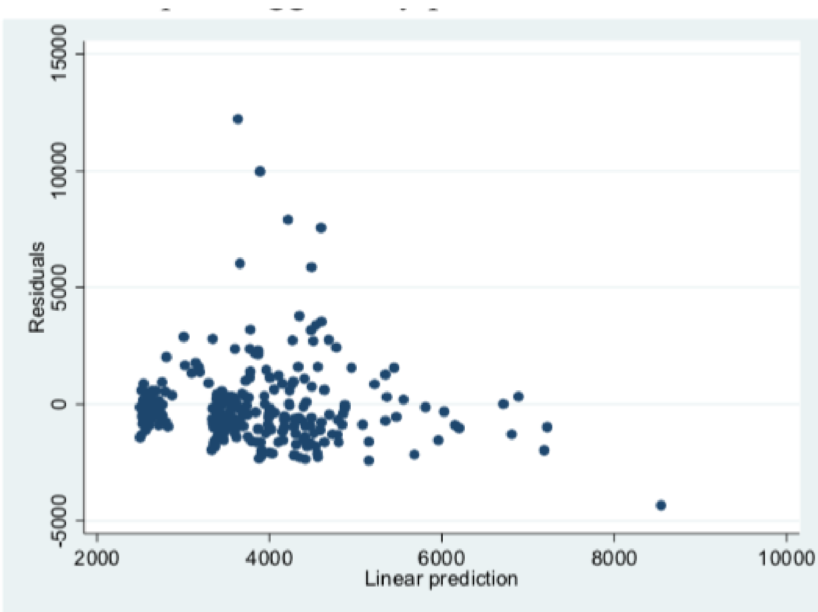


Figure 51

This is an informal diagnostic test for the presence of heteroscedasticity. We expect a definition of heteroscedasticity and a short statement of its consequences. We think that if there is no pattern in the residual plot, then this suggests homoskedasticity, however, if we can see a pattern in the residual plot, this suggests heteroskedasticity. We see that at higher values of y_hat, residuals tend are more concentrated around 0, while at lower values, they are more dispersed. This suggests heteroskedasticity may be a problem, and the researcher should account for this by using heteroskedastic robust standard errors. In addition, the researcher could also perform a statistical test (Breusch-Pagan or White test) to further inform this decision.

e) Suppose the researcher collected data on the average global price of oil in 1980 (avg_oil_price) and were to estimate the following regression: $wagePP_i = \beta_0 + \beta_1 fields_i + \beta_2 avg\_oil\_price_i + u_i$. *What would be the consequences of including such a control variable in the regression in question a on the estimated coefficient of the variable fields? Briefly explain your answer.*

- Average global oil price in 1980 will be the same in all counties. This means that we get a column in our dataset that is identical in every row. We cannot use such data because we can't estimate coefficients with it since every observation will always be equal to the average, which will give us 0 in the denominator of the beta calculations. So this would violate assumption 3 of having variance in the variable.
- **Since we cannot have 0 in the denominator, the variable will be excluded all-together in the statistics software. Therefore, including the variable will have no effect on any of the estimated coefficients in the model.**
- If we instead had included oil prices within each county—which would likely have some variation because of, for instance, transaction costs, different quality of the oil and gas, etc.—then we could control for the oil prices.
- If county-specific oil prices were included, and given that they had some variation, then we would likely change the omitted variable bias. Specifically, since correlation between oil price and the number of fields is likely positive, and since the the true parameter value of oil price will likely be positive—a higher oil price may create more benefits for an oil-producing county (which it seems like this sample is specifically looking at) than downsides so that per-capita wages could increase at the aggregate level—then excluding this variable creates a positive bias. Including this variable could therefore have reduced the bias caused by omitted variables.

From the solution:

> If the researcher includes the global price of oil in 1980, this will have no impact on the estimated coefficient of the variables fields, and the estimated coefficient would be the exact same as in question a. This is because in each county, the global price of oil is identical and we cannot include a variable with no variation in our regression. Put differently, if we were to include this variable, it would be omitted from the regression. This is a violation of our assumption 3 as the variable avg_oil_price is the same across all counties.

Note that if the researcher had collected county-level data on the price of oil in each county, then the inclusion of this variable would impact the results as all counties have different prices of oil (provided the county-level price of oil was not the same for all counties in the sample). Students discussing omitted variable bias and the expected direction of the bias were awarded some credit depending on the quality of such discussion.

5.18.2 V17 Q2

We are interested in the macro consumption function, in this case the relationship between household consumption and household income. We have annual data of these variables, called consumption and income, for the period 1978-2016 for Norway. The variables are log transformed and plotted in the figure below.

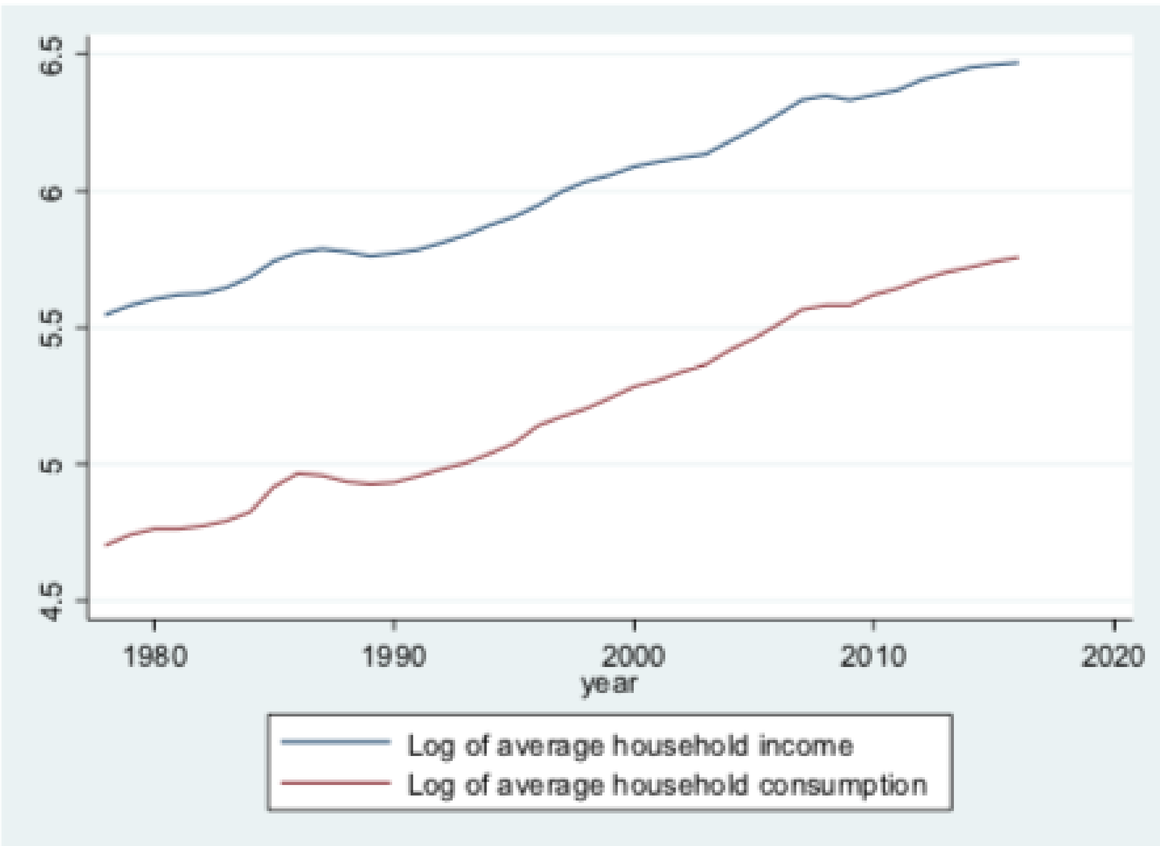


Figure 52

Use the log file from Stata for question 2 in order to answer the following questions: (Note that you should not comment and explain every line of commands in the log file, use only what you need in order to explain your answers.)

a) Interpret the estimated slope coefficient in column 1.

```
. estimates table col*, stats(N r2_a)  star b(%7.3f)
------------------------------------------------------------------
    Variable     |    col1         col2          col3         col4
-----------------+------------------------------------------------
       income    |  1.146***      0.815       1.078***     -3.898
           y2    |               -0.044                     0.821
           y3    |                0.009                    -0.048
after_credit_lib |                           -0.737***      3.343
afterD_income    |                            0.123**      -0.559
        _cons    | -1.675***      0.209      -1.285***     13.136
-----------------+------------------------------------------------
            N    |      39           39           39           39
         r2_a    |   0.998        0.999        0.999        0.999
------------------------------------------------------------------
                            legend: * p<0.05; ** p<0.01; *** p<0.001
```

Figure 53

- The text states that these variables has already been log-transformed (but the author just chose not to include an indication of that in the variable names). Since both variables are log-transformed, we have an elasticity interpretation.
- Specifically, increasing income by 1 % is associated with an 1.146 % increase in consumption, everything else held equal.
- However, this doesn't seem plausible since this would mean that, at the macro level in Norway, people would use all their additional income *and then some* to consume goods. Meaning that they would tap into their savings even more every time their wages increase.
- It's reasonable that consumption at the macro level increases with income, but that it increases less than the increase in income. We therefore expect the estimated coefficient to be strictly positive but less than 1.
- The likely reason for this strange result is the omitted variable bias that occurs when we include only one variable on such a large topic as consumption. This bias violates assumption 4 of zero conditional mean, and we therefore cannot interpret the relationship causally since all assumptions must be in place.

From the solution:

> The elasticity of consumption with respect to income is 1.14; a 1% increase in income increases consumption with 1.14 %.

6 STATA commands

- You can use the full command or short-hand commands in Stata (like `quietly` or `qui`).
- Everything following a comma in a command is an option. There you can specify more.
- For panel data, a_i is called `sigma_u`.
- For panel data, u_{it} is called `sigma_e`.

6.1 **generate**, **gen**

Generates a new variable.

Listing 3: Scaling

```
gen wage_bitcoin = wage * 0.00015
```

6.2 **regress**, **reg**

Do a regression.

Listing 4: Generate $\widehat{wage} = \hat{\beta}_0 + \hat{\beta}_1 educ$

```
reg wage educ
```

Listing 5: Generate $\widehat{wage} = \hat{\beta}_0 + \hat{\beta}_1 educ + \hat{\beta}_2 IQ$

```
reg wage educ IQ
```

```
reg lwage educ IQ

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  2,   932) =   69.42
       Model |  21.4779447     2  10.7389723           Prob > F      =  0.0000
    Residual |  144.178339   932  .154697788           R-squared     =  0.1297
-------------+------------------------------           Adj R-squared =  0.1278
       Total |  165.656283   934  .177362188           Root MSE      =  .39332

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0391199   .0068382     5.72   0.000     .0256998      .05254
          IQ |   .0058631   .0009979     5.88   0.000     .0039047    .0078215
       _cons |   5.658288   .0962408    58.79   0.000     5.469414    5.847162
------------------------------------------------------------------------------
```

Figure 54: Regression print

6.2.1 Interpreting the OLS regression output


```
reg lwage educ IQ

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  2,   932) =   69.42
       Model |  21.4779447     2  10.7389723           Prob > F      =  0.0000
    Residual |  144.178339   932  .154697788           R-squared     =  0.1297
-------------+------------------------------           Adj R-squared =  0.1278
       Total |  165.656283   934  .177362188           Root MSE      =  .39332

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0391199   .0068382     5.72   0.000     .0256998      .05254
          IQ |   .0058631   .0009979     5.88   0.000     .0039047    .0078215
       _cons |   5.658288   .0962408    58.79   0.000     5.469414    5.847162
------------------------------------------------------------------------------
```


Figure 55: Regression print

In the upper left corner, we got the ANOVA table.

- **Source**: looking at the breakdown of variance in the outcome variable, these are the categories we will examine: Model, Residual, and Total. The Total variance is partitioned into the variance which can be explained by the independent variables (Model) and the variance which is not explained by the independent variables (Residual, sometimes called Error).

- **SS**: these are the Sum of Squares associated with the three sources of variance, Total, Model and Residual.

- **df**: these are the degrees of freedom associated with the sources of variance. The total variance has N-1 degrees of freedom. The model degrees of freedom corresponds to the number of coefficients estimated minus 1. Including the intercept, there are 3 coefficients, so the model has 3-1=2 degrees of freedom. The Residual degrees of freedom is the DF total minus the DF model, $934 - 2 = 932$.

- **MS**: these are the Mean Squares, the Sum of Squares divided by their respective DF.

In the upper right corner, we got the overall model fit

- **Number of obs**: this is the number of observations used in the regression analysis.
- **F(2,932)**: this is the F-statistic, which is the Mean Square Model (10.7389723) divided by the Mean Square Residual (.154697788), yielding F=69.42. The numbers in parentheses are the Model and Residual degrees of freedom from the ANOVA table.
- **Prob > F**: this is the p-value associated with the above F-statistic. It is used in testing the null hypothesis that all of the model coefficients are 0.
- **R-squared**: R-Squared is the proportion of variance in the dependent variable (`lwage`) that can be explained by the independent variables (`educ`, `IQ`). This is an overall measure of the strength of association and does not reflect the extent to which any particular independent variable is associated with the dependent variable.
- **Adj R-squared**: this is an adjustment of the R-squared that penalizes the addition of extraneous predictors to the model. Adjusted R-squared is computed using the formula $1 - (1 - R^2)\frac{n-1}{n-k-i}$, where k is the number of x variables.
- **Root MSE**: root MSE is the standard deviation of the error term, and is the square root of the Mean Square Residual (or Error).

At the bottom, we got the parameter estimates

- **lwage**: this column shows the dependent variable at the top (`lwage`) with the predictor variables below it (`educ`, `IQ`, and `_cons`). The last variable (`_cons`) represents the constant or intercept.
- **Coef.**: these are the values for the regression equation for predicting the dependent variable from the independent variable. The column of estimates provides the values for $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ for this equation.
- **Std. Err.**: these are the standard errors associated with the coefficients.
- **t**: these are the t-statistics used in testing whether a given coefficient is significantly different from zero.

- **P>|t|**: this column shows the *2-tailed* p-values used in testing the null hypothesis that the coefficient (parameter) is 0.

- **[95% Conf. Interval]**: these are the 95% confidence intervals for the coefficients. The confidence intervals are related to the p-values such that the coefficient will not be statistically significant at alpha = .05 if the 95% confidence interval includes zero. These confidence intervals can help you to put the estimate from the coefficient into perspective by seeing how much the value could vary.

6.3 `ln()`

Transform the variable to logarithmic form.

Listing 6: Generate $\ln(\widehat{wage}) = \hat{\beta}_0 + \hat{\beta}_1 educ$

```
gen ln_wage = ln(wage)
reg ln_wage educ
```

```
reg ln_wage educ
------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0598392   .0059631    10.03   0.000     .0481366    .0715418
       _cons |   5.973063   .0813737    73.40   0.000     5.813366    6.132759
------------------------------------------------------------------------------
```

Figure 56

6.4 `*c`

Scale a variable.

```
gen wage_1000 = wage/1000
reg wage_1000 educ
```

```
gen annual_wage = wage*12
reg annual_wage educ_month
```

6.5 quietly, qui

Run a regression without outputting the ANOVA table or overall model fit table—that is, only getting the parameter estimation table

```
quietly reg lwage educ IQ
```

```
qui reg lwage educ IQ
```

6.6 _b[]

Retrieve the estimated parameter of a specific independent variable.

```
quietly reg lwage educ
gen beta1=_b[educ]
```

6.7 display,disp

Print function. Can print both variables and calculations.

```
quietly reg lwage educ
gen beta1=_b[educ]
gen beta2=_b[IQ]
display beta1
display beta1 + beta2
```

6.8 if

Filters the dataset

```
reg price sqft pool age if utown==1
```

6.9 `keep`

`Keep` will keep only the observations that passes the `if` and drops the rest. Alternatively, one could specify only the variables vi want to delete or keep.

Listing 7: Keep only observations that doesn't have missing values in father's or mother's education

```
keep feduc!=. & meduc!=.
reg lwage educ IQ age feduc
```

6.10 `corr`

Print a correlation table of the selected variables

```
corr feduc meduc
```

6.11 `e()`

Obtain particular values from the regression. So it seems like it is similar to `attributes` in R.

```
reg annual_wage eduq IQ
display e(rss)
display e(N)
display e(r2)
```

6.12 `exp()`

Exponential of Euler's number, e.

```
reg log_wage eduq IQ
display 100*(exp(_b[educ]-1)
```

6.13 **describe**, **desc**

FYLL INN

```
FYLL INN
```

6.14 **summarize**, **sum**, **su**

```
. summ  cigsA lincA lcigpA educA ageA agesqA restaurnA constA

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       cigsA |       807    .9661919    1.574979          0   11.41837
       lincA |       807    1.286915    .7200436   .5829876   10.00486
      lcigpA |       807    .5529577     .333062   .2296518   4.085199
       educA |       807    1.667371    1.022053   .3393095   15.52803
        ageA |       807    6.028058    7.185146   2.157312   85.40414
-------------+--------------------------------------------------------
      agesqA |       807    329.5523    615.1486    41.5149   7515.564
   restaurnA |       807    .0401264    .0790122          0   .5765851
      constA |       807    .1349764    .0808987   .0565516   .9705016
```

Figure 57

```
summ cigsA lincA lcigpA educA ageA agesqA resurnA constA
```

6.15 **detail**

Prints out a more detailed description about the variable.

```
sum log_wage, detail
```

6.16 **invttail()**

Compute the critical t-value

```
display invttail(405, .05) * Output: 1.645
display invttail(405, .95) * Output: -1.645
display invttail(405, .05/2) * Output: 1.96
```

6.17 **test**

Run a hypothesis test. Displays the F-distribution. This is the easiest way to run an F-test. Run it to compare the nested and non-nested model.

```
test educ=0
```

Adding $= 0$ is not necessary as this is the default.

```
test educ
```

```
test educ=55
```

6.18 **ttest**

Run a two-sample t test with equal variances to test the significance of the difference.

```
ttest unem78, by(train)
```

```
.         * test the significance of the difference in unemployment in 1978 by job training status
.                 ttest unem78, by(train)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     260    .3538462    .0297116    .4790844    .2953392    .4123531
       1 |     185    .2432432    .0316293     .430205    .1808405     .305646
---------+--------------------------------------------------------------------
combined |     445    .3078652    .0219071      .46213    .2648108    .3509196
---------+--------------------------------------------------------------------
    diff |            .1106029    .0441888                .0237572    .1974486
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   2.5030
Ho: diff = 0                                     degrees of freedom =      443

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9937         Pr(|T| > |t|) = 0.0127          Pr(T > t) = 0.0063
```

Figure 58

6.19 **ttail()**

Retrieve the p-value

```
display ttail(30, 1.0) * Output: 0.0335
```

6.20 predict

Obtain fitted values

```
predict u_hat, resid
```

6.21 not_...

Gives us the opposite. Say that we have three dummies: D1, D2, D3. If we want D2 and D3, we could write

```
not_D1
```

6.22 nocons

Regress without the constant.

```
. *** Regression on transformed variables, FGLS

. reg cigsA lincA lcigpA educA ageA agesqA restaurnA constA, noconst

      Source |       SS       df       MS              Number of obs =     807
-------------+------------------------------           F(  7,   800) =   43.50
       Model |  758.856198     7  108.408028           Prob > F      =  0.0000
    Residual |   1993.8312   800    2.492289           R-squared     =  0.2757
-------------+------------------------------           Adj R-squared =  0.2693
       Total |  2752.68739   807  3.41101288           Root MSE      =  1.5787

------------------------------------------------------------------------------

                                                                                20

       cigsA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lincA |    1.29524   .4370118     2.96   0.003     .4374146    2.153065
      lcigpA |  -2.940292   4.460144    -0.66   0.510    -11.69526    5.814676
       educA |  -.4634465   .1201587    -3.86   0.000     -.69931    -.2275829
        ageA |   .4819481   .0968082     4.98   0.000     .2919199    .6719762
      agesqA |  -.0056272   .0009395    -5.99   0.000    -.0074713   -.0037831
   restaurnA |  -3.461065   .7955049    -4.35   0.000    -5.022588   -1.899541
      constA |   5.635378   17.80314     0.32   0.752      -29.311    40.58176
------------------------------------------------------------------------------
```

Figure 59

```
reg cigsA lincA lcigpA educA ageA agesqA restaurnA constA,
   noconst
```

6.23 use

```
Use "/Users/christianbraathen/Sync/NHH/9.
   semester/ECN402/Class 3/wage_crosssection.dta", clear
```

6.24 clear

```
Use "/Users/christianbraathen/Sync/NHH/9.
   semester/ECN402/Class 3/wage_crosssection.dta", clear
```

6.25 aweight=...

Weighted least squares

```
reg cigs lincome lcigpric educ age agesq restaurn
   [aweight=1/h_hat]
```

6.26 robust

Robust standard errors

```
reg cigs lincome lcigpric educ age agesq restaurn, robust
```

6.27 estat

Postestimation statistics

```
reg cigs lincome lcigpric educ age
estat ovtest
```

6.28 **ovtest**

Do the RESET test directly in stata with 2^{nd}, 3^{rd}, and 4^{th} degree polynomial.

```
reg cigs lincome lcigpric educ age
estat ovtest
```

```
qui reg q pb pl pr i

estat ovtest

Ramsey RESET test using powers of the fitted values of q
       Ho:  model has no omitted variables
                  F(3, 22) =       3.30
                  Prob > F =       0.0392
```

Figure 60

6.29 **L1.**

Lag a variable one period

```
reg i3 inf L1.inf def
```

6.30 **D1.**

First-difference a variable

```
gen df = d1.lprice
```

6.31 **e(sample)**

Use the same observations as in the last regression. Useful when using lags since they change the number of observations we can use.

```
reg i3 inf L1.inf def
```

6.32 tsline

Useful to get a line graph across time.

```
FYLL INN
```

6.33 tsset

Define which variable that keeps track of the time dimension.

```
tsset year
```

6.34 twoway (scatter ...)

Create a regression line in a graph (`twoway`) and add a scatter plot (`scatter`)

```
twoway (scatter linvpc lprice)
```

6.35 noheader

```
reg y x, noheader
```

6.36 corrgram

Get a correlogram of the time series

```
corrgram lprice
```

6.37 dfuller

Run Dickey–Fuller tests

```
dfuller lprice, regress noconst
```

```
dfuller lprice, regress
```

```
dfuller lprice, regress trend
```

Augmented Dickey–Fuller (if we think we have autocorrelation in the error term):

```
dfuller lprice, regress trend lags(2)
```

6.38 xtset

Specify which variables keep track of panel observations and time observations

```
xtset id year
```

6.39 xtreg

Perform panel data estimation

6.40 cap drop

Remove variable, equivalent to `rm` from R.

```
cap drop lwage
```

6.41 i.year

```
reg ln_y ln_k ln_n i.year, noheader
```

6.42 noheader

```
reg ln_y ln_k ln_n i.year, noheader
```

6.43 **vce(cluster ...**

Run OLS with cluster robust standard errors

```
reg ln_y ln_k ln_n i.year, vce(cluster id) noheader
```

6.44 **fe**

Fixed effects

```
xtreg ln_y ln_k ln_n i.year, fe vce(cluster id)
```

6.45 **i.id**

Least-squared dummy variables

```
set matsize 800
reg ln_y ln_k ln_n i.year i.id, vce(cluster id) noheader
```

6.46 **re**

Random effects

```
xtreg ln_y ln_k ln_n i.year, re vce(cluster id)
```

6.47 **hausman**

Run a Hausman test.

```
quietly xtreg ln_y ln_k ln_n i.year, fe
est store my_fe

quietly xtreg ln_y ln_k ln_n i.year, re
est store my_re
```

```
hausman my_fe my_re, sigmamore
```

6.48 if ... != .

For IVs: included to make sure that we only get the relevant observations when using the dataset.

```
reg educ fatheduc if lwage != .
```

6.49 ivregress 2sls

Run IV estimation. To get the correct standard errors, we let STATA do the job for us. **Note that the parenthesis shows the second regression**

```
ivregress 2sls lwage (edic = fatheduc)

# another example, showing that we must include all
   exogeneous explanatory variables in the IV estimation too:
ivregress 2sls lwage (educ = fatheduc motheduc exper expersq)
   exper expersq
```

6.50 overid

Tests if overidentifying restrictions

```
estat overid
```

7 Summary

7.1 Terminology

- Left-hand side variable is called the y variable, the dependent variable, the explained variable, the response variable, and the regressand.
- Right-hand side variable is called the x variable, the independent variable, the explanatory variable, the control variable, and the regressor.
- Estimator: the function that maps samples into the parameter space.
- Estimate: the value the estimator outputs.
  - So the *input* is the sample, the *processor* is the estimator function, and the *output* is the estimate.
  - *BLUE*: Best Linear Unbiased Estimator. This is true if assumption 1–5 holds for OLS because the OLS estimator is then both unbiased and efficient.
- R^2: the fraction of sample variance in y that is explained by $\vec{x}$.
- Coefficient interpretation:
  - One additional unit of x is associated with ..., everything else held equal
  - The coefficient suggests that one additional unit of x leads to..., everything else held equal
  - one (unit/%) change in x is leasing to an (...) (unit/%) change in y, all else equal
- "We regressed y on $x_1, ..., x_k$"
- "Once (*newly included variable*) is accounted for, a change in the (*previously included variable*) is (a more/not a more) strong predictor of (y)".
- "We have controlled for the variables $x_2, ..., x_k$ when estimating the effect of x_1 on y"
- Keeping H_0: "We don't have sufficient evidence to discard H_0 in favor of H_A".
- Keeping H_0: "We fail to reject H_0 at the α% significance level".
- Keeping H_0: "x_j is statistically insignificant at the 5% level."

- Keeping H_0: "Once <put in the names of the other variables but the one you tested> has been accounted for, x_j has no effect on the expected value of y."
- Going for H_A: "there is evidence against the null hypothesis at a $100 \cdot \alpha\%$ significance level."
- Going for H_A: "$\hat{\beta}_j$ is statistically different from 0 at the 5 % significance level."
- Going for H_A: "x_j is statistically significant at the 5% level."
- Going for H_A with p-value $< 1\%$: "x_j is statistically significant at any conventional significance level."
- Going for H_A in an F-test with multiple variables: "<variables> are *jointly* statistically significant."
- The dummy as a stand-alone term ($\delta_1 D_1$) if you want the dummy to affect the intercept. Interpretation: "The difference in y between <what $D_1 = 1$ is> and <what $D_1 = 0$ is> *given* the same $\vec{x}$ values (and u). The differences occurs either because of <what the dummy stands for> or because of factors that are associated with the dummy and that we haven't controlled for."
- On the basis of <test>, ... For example: "On the basis of RESET, this model is preferred."

7.2 Various Notes

- Time series and panel data involves time, so they need special treatment because of the correlation across time.
- Dealing with the error term is perhaps the most important component of any econometric analysis.
- You need a solid model and your assumptions must be well-covered
- You also need to be able to *sell* the story you're telling.
- $Var(X) \equiv \mathbb{E}[(X-\mu)^2]$ (and for a sample: $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2)$
- $Cov(X,Y) \equiv \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$ (and for a sample: $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}))$
- Probability distributions used in this course, all used for inference:

- The normal distribution
- The standard normal distribution
- The Chi-Square distribution
- The t-distribution
- The F-distribution

- Economic theory: qualitative results. Econometric theory: quantitative results.
- If the time series is non-stationary, don't use the data—we could otherwise get spurious regressions.
- If estimator is unbiased, only then can we give a causal interpretation to the estimated coefficient. For simple OLS, this means that assumption 1–4 must hold.
- Don't think about significance as "good": if we have bias, we may get the wrong significance (significant although it wasn't).
- The three big ones in time series analysis is *trend*, *autocorrelation*, and *stationarity*.

8 Exam

Question 2 (25%)

We have data on the production of oil and gas from 264 counties in 5 US-states (Idaho, Montana, Colorado, North Dakota and South Dakota) for the year 1980. *Fields* is the number of producing oil and gas fields in a county. *WagePP* is the county-level aggregate annual per capita wage, measured in US dollars. The average number of oil and gas fields in the data in 1980 is 10, and the average wage per capita is \$3800. We are interested in estimating the effect of oil and gas fields on wages, and estimate the following regression:

$$wagePP_i = \beta_0 + \beta_1 fields_i + u_i$$

. reg wagePP fields

wagePP	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
fields	11.09928	5.909638	1.88	0.061	-.5371498	22.73571
_cons	3695.216	180.4633	20.48	0.000	3339.873	4050.559

a. Interpret the estimated coefficients. Is the sign of the estimated coefficient β_1 consistent with your economic expectation?

4

b. The researcher has now added five additional control variables: *idaho, montana, colorado,* and *north_dakota,* binary variables equal to 1 if a county is located in one of these states as well as *totEMP,* the total number of employees in the county. Interpret the estimated coefficient β_{idaho}. Discuss how the coefficient on the variable *fields* changes when including the additional control variables. Discuss what may be a reason for such a change.

wagePP	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
fields	5.517378	6.015098	0.92	0.360	-6.327779	17.36254
totEMP	.0631531	.0156362	4.04	0.000	.0323618	.0939445
idaho	1736.371	530.3445	3.27	0.001	691.9966	2780.745
montana	857.3935	482.2863	1.78	0.077	-92.34278	1807.13
colorado	1376.785	515.3574	2.67	0.008	361.9242	2391.646
north_dakota	789.5051	487.9313	1.62	0.107	-171.3476	1750.358
_cons	2464.249	332.6954	7.41	0.000	1809.093	3119.405

c. Do you think the estimation in question b above represents the causal effect of oil and gas fields on county-level wages per capita? If so, why? If not, why not?

Figure 61