Applied Econometrics SS 2008
Questions for review

First lecture

1. Ragnar Frisch's statement in the first Econometrica issue basically says that Econometrics is not (only) applied statistics and not (only) mathematics applied to economics. What is it, then?
2. How does the fundamental asset value evolve over time in the Glosten/Harris(1988) model. Which parts of the equation are associated with public and which parts with private information that influences the fundamental asset value.
3. When do we call an asset price "efficient"?
4. Explain the components of the equations that give the buy (bid) and the (ask) price in the Glosten/Harris model.
5. Explain how in the Glosten/Harris the market maker anticipates the impact that a trade event exerts on the fundamental asset price when setting buy and sell prices.
6. Why should it be interesting for a) an investor and b) for a stock exchange (like the New York Stock Exchange) to attach real numbers to the parameters of the Glosten/Harris (1988) model (z_0, z_1, c, μ).
7. What does the spread mean? Why is the spread associated with (implicit) transaction costs.
8. Which objects (variables and parameters) in the Glosten Harris model are a) observable to the market maker, but not to the econometrician b) observable to the econometrician.
9. The final equation of the Glosten/Harris model contains observable objects, unknown parameters and an unobservable component. What are these? What is the meaning of the unobservable variable in the final equation?
10. Why did we transform the equations for the market maker's buy and sell price that the market maker posts at point in time t?
11. In the "Mincer equation" from human capital theory (discussed in the lecture) what are the observable variables and what is the meaning of the unobservable component?
12. Why should the government be interested in estimating the parameters of the Mincer equation from human capital theory? Why do we need statistical hypothesis testing in that context?
13. Explain why we can conceive $ln(WAGE_i)$, S_i, $TENURE_i$, and $EXPR_i$ in the Mincer equation as random variables.

Second lecture

1. Discuss the key difference of experimental data and the economic data that is typically available.

2. Why is that difference important in the context of ceteris paribus/causal statement like "if x increases, ceteris paribus, then y increases".

3. Take the Mincer equation as an example. Why are the years of schooling "endogenous", i.e. the result of an economic decision process that is not made explicit in the equation. What is the role of the disturbance term ε in this context?

4. Provide another economic example (like the crime example or the liberalization example, but one that is not given in the lecture) where endogeneity is present. Cast your example in a CLRM. What is the economic interpretation of ε in your model?

5. There is a special name for model parameters which have an economic meaning. What do we call them?

6. Prove your algebraic knowledge and write the CLRM in matrix and vector notation. Indicate the dimensions of scalars, vectors and matrices explicitly!

7. Which objects in the CLRM are observable and which are not?

8. Why do we want that in the scatterplot (= points mapped in a ($x_i'b$,y_i) diagram) the points cluster closely around the 45 degree line?

9. We are in a CLRM setting, i.e. $y_i = x_i'\beta + \varepsilon_i$. Consider choosing $b = \beta$ in a diagram that takes y_i on the ordinate and $x_i'b$ on the abscissae. Sketch how the diagram would look like when a) $\varepsilon_i \sim N(0,1)$ b) $\varepsilon_i \sim N(0,20)$ and c) $\varepsilon_i = 0 \quad \forall \quad i$.

10. Imagine you choose as an estimate for β a b that yields a scatterplot where all points lie below the 45 degree line in the ($x_i'b$,y_i) scatterplot. Why is that troubling in the first place? How would you have to alter your choice of b to change that plot?

11. In the lecture we had a possible choice of $b = (X'X)^{-1}X'y$. Write out that expression in detail using the definition of the matrix X and the vector y. Can you write the right hand side expression using $x_i = x_{i1}, x_{i2}, \ldots, x_{iK}$ and y_i instead of the matrix X and the vector y? You have to use summations $\sum_{i=1}^{n}$ then!

Third lecture

1. Refresh your basics and derive the OLS estimator for $K = 2$. (one explanatory variable and one constant). This is statistics I!

2. Using observed data on explanatory variables and the dependent variable you can compute the OLS estimator as $b = (X'X)^{-1}X'y$. This yields a K dimensional column vector of real numbers. That's an easy task using a modern computer. Explain why we conceive the OLS estimator also as a vector of random variables (a random vector), and that the concrete computation of b yields a particular realization of this vector of random variables. Explain the source of randomness of b.

3. Explain in your own words what the terms unbiasedness, efficiency, and consistency mean. Could you give an intuitive explanation of how we can use these concepts to assess the quality of the OLS estimator $b = (X'X)^{-1}X'y$. (We will be more specific about this issue in later lectures, this question is just to make you aware of the issue.)

4. Multiply out $X'X$ and $X'y$ of the OLS estimator. How do the generic elements of this two objects look like? You can also write the OLS estimator as a function of sample means. How would you do that?

5. Linearity ($y_i = x_i'\beta + \varepsilon_i$) seems a restrictive assumption. Provide an example (not from lecture) where an initially nonlinear relation of dependent variable and regressors can be reformulated such that an equation linear in parameters result. Give another example (not in lecture) where this is not feasible.

6. What does it mean when we say that two random variables are orthogonal?

7. Strict exogeneity is a restrictive assumption. In another assignment you will be asked to prove the result (given in the lecture) that strict exogeneity implies that the regressors and the disturbance term are uncorrelated (or, equivalently, have zero covariance or, also equivalently, are orthogonal). Explain, using one of the economic examples given in the lecture (crime example or liberalization example), that assuming that error term and regressor(s) are uncorrelated is doubtful from an economic perspective (You have to give the ε an economic meaning like in the wage example where ε measured unobserved ability of an individual.)

Fourth lecture

1. How many random variables are contained in
   a) $Y = X\beta + \varepsilon$
   b) $Y = Xb + e$?

2. Show that in a semi-log model, such as
   $ln\ y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i$
   β_2 can be interpreted as a percentage change of y_i if x_{i2} increases by one unit.

3. Show that $E(\varepsilon_i|1, x_{i1}, x_{i2}) = 0$ implies

$$\begin{aligned} E(\varepsilon_i) &= 0 \\ E(\varepsilon_i x_{i1}) &= 0 \\ E(\varepsilon_i x_{i2}) &= 0 \end{aligned}$$

   Show that $E(\varepsilon_i) = 0$ implies that $Cov(\varepsilon_i, x_{i2}) = 0$.

4. Show the following results explicitly by writing out in detail the expectations for the case of continuous random variables.
   [a)] $E_X[E_{Y|X}(Y|X)] = E_Y(Y)$ (Law of Total Expectations `LTE`)
   [b)] $E_X[E_{Y|X}(g(Y)|X)] = E_Y(g(Y))$ (Double Expectation Theorem `DET`)
   [c)] $E_{XZ}[E_{Y|X,Z}(Y|X,Z)] = E_Y(Y)$ (`LTE`)
   [d)] $E_{Z|X}[E_{Y|X,Z}(Y|X,Z)|X] = E_{Y|X}(Y|X)$ (Law of Iterated Expectations `LIE`)
   [e)] $E_X[E_{Y|X}(g(X,Y)|X)] = E_{XY}(g(X,Y))$ (Generalized `DET`)
   [f)] $E_{Y|X}[g(X)Y|X] = g(X)E_{Y|X}(Y|X)$ (Linearity of Conditional Expectations)

   If you can't solve d)-f), don't give up. There will be an exercise on this.

5. Explain in your own words why the $rank(X)$ is a random variable.

6. Show by writing in detail for $K = 2$, $\beta = (\beta_1, \beta_2)'$ that
   $a'[var(\hat{\beta}|X) - var(b|X)]\ a \geq 0\ \forall a \neq 0$ implies that
   $var(\hat{\beta}_1|X) \geq var(b_1|X)$ for $a = (1, 0)'$.

7. Write the variance-covariance-matrix $Var(\varepsilon|X)$ explicitly. We require a variance covariance matrix to be symmetric and positive definite. Could you explain why the latter one is a sensible restriction? Hint: Think of computing the variance of a linear combination of random variables when the variance covariance matrix would not be positive definite. Definition of positive (semi) definite matrix:
   A is a symmetric matrix. x is a nonzero vector.
   1. if $x'Ax > 0$ for all nonzero x, then A is positive definite

2. if $x'Ax \geq 0$ for all nonzero x, then A is positive semi definite

8. Use the concept of positive (semi) definitness to proove the Gauss -Markov Theorem.

Fifth lecture

1. Refresh your basic Mathematics and Statistics:
   a) $v = a \cdot Z$. Z is a $(n \times 1)$ vector of random variables. a is $(1 \times n)$ vector of real numbers.
   b) $V = A \cdot Z$. Z is a $(n \times 1)$ vector of random variables. A is $(m \times n)$ matrix of real numbers.
   Compute for a) and b) the expectation and the variance covariance matrix.

2. Under which assumptions can the OLS estimator $b = (X'X)^{-1}X'y$ be called BLUE?

3. Why does the semi positive definiteness of $var(\hat{\beta}|X) - var(b|X)$ tell you something about the efficency of the OLS estimator $b = (X'X)^{-1}X'y$?

4. There are two main schools of statistics. Name them and explain them to an uninitiated person!

5. What is the null and the alternative hypothesis of the t-test presented in the lecture? How is the test statistic constructed?

6. Give examples for reasonable significance levels. Why are they called conventional significance levels? On which grounds could you criticize a significance level of 0.00000001?

7. Assume the value of a t-test statistic equals 3. What is your ad hoc interpretation of the result?

8. Assume the value of a t-test statistic equals 0.4. What is your ad hoc interpretation of the result?

9. Explain the following terms that are often used in a hypothesis testing framework:
   · What is a type 1 error and what is a type 2 error?
   · Explain the term "significance level".
   · What do we mean by the power of a test?

10. A bivariate normal density of two random variables is given by
    $f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{(1-\rho^2)}} \cdot \exp\left(-\frac{Q}{2(1-\rho^2)}\right)$
    with $Q = \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2 \cdot \rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2$
    show that if $\rho = 0$ follows that $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$, i.e. X and Y are independent.

11. Which assumption is used to establish that the OLS estimator is conditionally normally distributed. Which are the results from mathematical statistics that are employed to establish normality of the OLS estimator?

12. For which purpose is it important to know the distribution of the parameter estimate in the first place?

13. Why is b a random variable in the first place?

14. Given assumptions 1.1-1.5. Do we say something about the unconditional distribution of the OLS estimate b?

15. List all assumptions that are necessary for the result that the z-statistic

    $z_k = \frac{b_k - \bar{\beta}_k}{\sqrt{\sigma^2[(X'X)^{-1}]_{kk}}}$

    is standard normally distributed under the null hypothesis, i.e. $z_k \sim N(0,1)$.

16. What do we mean when we say a test statistic is "nuisance parameter free". Assume that we know σ^2. Is the z_k- statistic nuisance parameter free?

17. Is the t-statistic nuisance parameter free?

    $t_k = \frac{b_k - \bar{\beta}_k}{\sqrt{s^2[(X'X)^{-1}]_{kk}}}$

18. Conducting a t-test, one might argue that using the quantile table of the standard normal distribution instead of a quantile table of a t-distribution to determine the rejection area for a t-test is sufficient. When would you subscribe to that argument? Discuss.

19. One might also argue that using the biased estimator $\hat{\sigma}^2 = \frac{1}{n}\mathbf{e}'\mathbf{e}$ for the parameter σ^2 instead of the unbiased estimator $s^2 = \frac{1}{n-K}\mathbf{e}'\mathbf{e}$ is not too harmful. When would you subscribe to that argument? Discuss.

20. Give two examples for constructing the restrictions for a F-test. Define R and r. (for example Glosten/Harris model or Mincer equation).

Sixth lecture

1. Suppose you have estimated a parameter vector $\mathbf{b} = (\ 0.55 \quad 0.37 \quad 1.46 \quad 0.01\ )'$ with an estimated variance-covariance matrix

$$\widehat{Var(\mathbf{b}|\mathbf{X})} = \begin{bmatrix} 0.01 & 0.023 & 0.0017 & 0.0005 \\ 0.023 & 0.0025 & 0.015 & 0.0097 \\ 0.0017 & 0.015 & 0.64 & 0.0006 \\ 0.0005 & 0.0097 & 0.0006 & 0.001 \end{bmatrix}$$

   a) Compute the 95% confidence interval each parameter b_k.
   b) What does the specific confidence interval computed in a) tell you?
   c) Why are the bounds of a confidence interval for β_k random variables?
   d) Another estimation yields an estimated b_k with the corresponding standard error $se(b_k)$. You conclude from computing the t-statistic $t_k = \frac{\beta_k - \bar{\beta}_k}{se(b_k)}$ that you can reject the null hypothesis $H_0 : b_k = \bar{\beta}_k$ on the α% significance level. Now, you compute the $(1-\alpha)$% confidence interval. Will $\bar{\beta}_k$ lie inside or outside the confidence interval?

2. Suppose, computing the lower bound of the 95% confidence interval yields $b_k - t_{\alpha/2}(n-K)se(b_k) = -0.01$. The upper bound is $b_k + t_{\alpha/2}(n-K)se(b_k) = 0.01$ Which of the following statements are correct?

   (a) With probability of 5% the true parameter β_k lies in the interval -0.01 and 0.01.
   (b) The null hypothesis $H_0 : \beta_k = \bar{\beta}_k$ cannot be rejected for values $(-0.01 \leq \bar{\beta}_k \leq 0.01)$ on the 5% significance level.
   (c) The null hypothesis $H_0 : \beta_k = 1$ can be rejected on the 5% significance level.
   (d) The true parameter β_k is with probability $1-\alpha = 0.95$ greater than -0.01 and smaller than 0.01.
   (e) The stochastic bounds of the $1-\alpha$ confidence interval overlap the true parameter with probability $1-\alpha$.
   (f) If the hypothesized parameter value $\bar{\beta}_k$ falls within the range of the $1-\alpha$ confidence interval computed from the estimates b_k and $se(b_k)$ then we do not reject $H_0 : \beta_k = \bar{\beta}_k$ at the significance level of 5%.

3. a) Show that if the regression includes a constant:

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i$$

   then the variance of the dependent variable can be written as:

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

Hint: $\bar{y}=\bar{\hat{y}}$

b) Take your result from a) and formulate an expression for the coefficient of determination R^2.

c) Suppose, you estimated a regression with an $R^2 = 0.63$. Interpret this value.

d) Suppose, you estimate the same model as in c) without a constant. You know that you cannot compute a meaningful centered R^2. Therefore, you compute the uncentered R^2_{uc}:

$$R^2_{uc}=\frac{\hat{\mathbf{y}}'\hat{\mathbf{y}}}{\mathbf{y}'\mathbf{y}}=0.84$$

Compare the two goodness of fit measures in c) and d). Would you conclude that the constant can be excluded because $R^2_{uc} > R^2$?

4. Regression with EViews:

In a hedonic price model the price of an asset is explained with its characteristics. In the following we assume that housing pricing can be explained by its size $sqrft$ (measured as *square feet*), the number of bedrooms $bdrms$ and the size of the lot $lotsize$ (also measured as *square feet.* Therefore, we estimate the following equation with OLS:

$$log(price)=\beta_0+\beta_1 log(sqrft)+\beta_2 bdrms+\beta_3\log(lotsize)+\varepsilon$$

Results of the estimation can be found in the following table:

Dep. Variable: LPRICE
Incl. observations: 88

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.29704	0.65128	-1.99152	0.0497
LSQRFT	0.70023	—	7.54031	0.0000
BDRMS	0.03696	0.02753	—	—
LLOTSIZE	0.16797	0.03828	4.38771	0.0000
R-squared	0.64297	Mean dependent var		5.6332
Adjusted R-squared	0.63021	S.D. dependent var		0.3036
S.E. of regression	0.18460	Akaike info criterion		-0.4968
Sum squared resid	2.86256	Schwarz criterion		-0.3842
Log likelihood	25.86066	F-statistic		—
Durbin-Watson stat	2.08900	Prob(F-statistic)		0.0000

(a) Interpret the estimated coefficients $\hat{\beta}_1$ und $\hat{\beta}_2$.

(b) Compute the missing values for *Std. Error* and *t-Statistic* in the table and comment on the statistical significance of the estimated coefficients (H_0: $\beta_j = 0$ vs. H_1: $\beta_j \neq 0$, $j = 0, 1, 2, 3$).

(c) Test the null hypothesis H_0: $\beta_1 = 1$ vs. H_1: $\beta_1 \neq 1$.

(d) Estimate the p-value for $\hat{\beta}_2$ as close as possible and interpret.

(e) What is the null hypothesis of this specific $F - Statistic$? Compute the missing value and interpret the result.

(f) Interpret the value of *R-squared*.

(g) An alternative specification of the model that excludes the lot size as an explanatory variable provides you with values for the Akaike information criterion of -0.313 and a Schwartz criterion of -0.229. Which specification would you prefer?

Seventh lecture

1. How is the theorem from multivariate statistics $(z-\mu)'\Omega^{-1}(z-\mu) \sim \chi^2(\#r)$ with $z \sim MVN(\sigma, \Omega)$ used to construct the F (or Wald) statistic to test linear hypotheses about the parameters. Explain.

2. What is the difference between the standard deviation of a parameter estimate b_k and the standard error of the estimate?

3. Discuss the pros and contras of alternative ways to present the results for a t-test:

    a) parameter estimate and

    *** for significant parameter estimate (at $\alpha = 1$ %)

    ** for significant parameter estimate (at $\alpha = 5$ %)

    * for significant parameter estimate (at $\alpha = 10$ %)

    b) parameter estimate and p-value

    c) parameter estimate and t-statistic

    d) parameter estimate and parameter standard error

    e) your preferred choice

4. Look up in the library a leading economic/finance/marketing journal (e.g. American Economic Review, Journal of Finance, Review of Financial Studies, Econometrica, Journal of Marketing Research) and check, how estimation results along with t-test type statistics are presented. The results may not necessarily come from an OLS regression, but the presentation of the estimation results is often very similar (estimate and standard error of estimate, or estimate and p-value,...). You should get an idea how to interpret such estimation results tables.

5. What can you do if you want to narrow the confidence bounds for the parameter β_k? Discuss the possibilities:

- Increase α

- Increase n (sample size)

  Can you explain how the effect of an increase of the sample size on the confidence bounds works?

Eighth lecture

1. When would you use the uncentered R^2_{uc} and when would you use the centered R^2. Why is the uncentered R^2_{uc} higher than a centered R^2? What is the range of the R^2_{uc} and R^2 ?

2. How would you interpret a R^2 of 0.38?

3. Why would you use an adjusted R^2_{adj}? What is the idea behind the adjustment of the R^2_{adj}? Which values can the R^2_{adj} take?

4. What is the intuition behind the computation of AIC and SBC?

5. Use the data set `dcx_gh.wf1` of the Glosten/Harris(1988) model. Estimate the OLS regression with EViews (Assignment Sheet 2 Task 1). Estimate a restricted version of the Glosten/Harris(1988) model that sets the parameters $\mu = 0$ $z_0 = 0$ and $z_1 = 0$. Compare the Akaike criterion of the restricted model version with the unrestricted model specification. Which model is chosen by the Akaike criterion? Do the same model selection procedure again, this time using the Schwarz criterion. Which model is chosen according to the Schwarz Criterion? Also comment on the R^2_{adj} and R^2. (Hint: Akaike and Schwarz criterion, R^2_{adj} and R^2 are automatically produced in Eviews when you run an OLS regression. The values can found in the lower panel in the OLS Eviews output with the names Akaike info criterion and Schwarz criterion R^2_{adj} and R^2.)

6. Why is the Schwarz criterion higher than the Akaike criterion?

7. Explain your own in words: What does convergence in probability mean? What does convergence almost surely mean? Which concept is stronger?

8. A researcher conducts an OLS regression with a computer software (NOT EVIEWS) that is unfortunately not able to report p-values. Besides the four coefficients and their standard errors the program only reports the t-statistics that test the null hypothesis $H_0 : \beta_k = 0$ for k = 1,2,3,4. Interpret the t-statistics below and compute the associated p-values. (Interpret the p-values for a reader who works with a significance level $\alpha = 5\%$.)

   [a)] t_1 = -1.99

   [b)] t_2 = 0.99

   [c)] t_3 = -3.22

   [d)] t_4 = 2.3

Ninth lecture

1. Illustrate graphically the concept of convergence in probability. Illustrate graphically a random sequence that does not converge in probability.
2. Explain in your own words: What does convergence in mean square mean? Does convergence in mean square imply convergence in probability? Or does convergence in probability imply convergence in mean square?
3. Illustrate graphically the concept of convergence in distribution. What does convergence in distribution mean? Think of an example and provide a graphical illustration where the c.d.f. of the sequence of random variables does not converge in distribution.
4. Explain in your own words the weak law of large numbers. What assumptions have to be fullfilled that you can apply Khinchin's WLLN?
5. Explain in your own words the Lindeberg-Levy (LL) central limit theorem. What assumptions have to be fullfilled that you can apply the LL CLT?
6. Name the concept that is associated to the following short hand notations and explain their meaning:

$z_n \underset{d}{\rightarrow} N(0,1)$	$z_n \underset{d}{\rightarrow} z$	$\overline{z}_n \overset{a}{\sim} N(\mu, \frac{\sigma^2}{n})$
$\operatorname*{plim}_{n\rightarrow\infty} z_n = \alpha$	$\sqrt{n}(\underline{z}_n - \underline{\mu}) \underset{d}{\rightarrow} N(\underline{0}, \Sigma)$	$z_n \underset{a.s.}{\rightarrow} \alpha$
$z_n \underset{m.s.}{\rightarrow} \alpha$	$y_n \underset{p}{\rightarrow} \alpha$	$z_n^2 \underset{d}{\rightarrow} \chi^2(1)$

Tenth lecture

1. Apply the "useful lemmas" of the lecture:

- $\underline{z}_n \underset{p}{\to} \alpha$
  $\ln(\underline{z}_n) \underset{p}{\to} ?$

- $z_n \underset{d}{\to} z$, $z \sim N(0,1)$, $z^2 \sim \chi^2(1)$
  $z_n \underset{d}{\to} ?$, $z_n^2 \underset{d}{\to} ?$

- $\underline{z}_n \underset{d}{\to} \underline{z}$, $a(\underline{z}) = \underline{Az}$, $\underline{z} \sim MVN(\underline{\mu}, \Sigma)$, $\underline{Az} \sim MVN(\underline{A\mu}, \underline{A}\Sigma\underline{A}')$
  $\underline{Az}_n \underset{d}{\to} ?$

- $\underline{x}_n \underset{d}{\to} N(0,1)$, $\underline{y}_n \underset{p}{\to} \underline{\alpha}$
  $\underline{x}_n + \underline{y}_n \underset{d}{\to} ?$

- $\underline{x}_n \underset{d}{\to} \underline{x}$, $\underline{y}_n \underset{p}{\to} 0$
  $\underline{x}_n + \underline{y}_n \underset{d}{\to} ?$

- $\underline{x}_n \underset{d}{\to} \underline{x}$, $\underline{y}_n \underset{p}{\to} 0$
  $\underline{x}_n \cdot \underline{y}_n \underset{p}{\to} ?$

- $\underline{x}_n \underset{d}{\to} MVN(0, \Sigma)$
  $\underline{A}_n \cdot \underline{x}_n \underset{d}{\to} ?$

- $\underline{x}_n \underset{d}{\to} \underline{x}$, $\underline{A}_n \underset{p}{\to} \underline{A}$
  $\underline{x}_n' \underline{A_n}^{-1} \underline{x}_n \underset{d}{\to} ?$

- $\underline{x}_n \underset{d}{\to} \underline{x}$, $\underline{y}_n \underset{p}{\to} \underline{\alpha}$
  $\underline{x}_n + \underline{y}_n \underset{d}{\to} ?$

2. When large sample theory is used to derive the properties of the OLS estimator the set of assumptions for the finite sample properties of the OLS estimator are altered. Explain. What assumptions are retained?

3. What are the properties of $b = (X'X)^{-1}X'y$ under the "new" assumptions?

4. Where does the WLLN come into play when analyzing the distributional properties of $b = (X'X)^{-1}X'y$?

5. If $\sqrt{n}(z_n - \mu) \underset{d}{\to} N(0, \sigma^2)$ what is the "approximate" distribution of z_n, i.e. $z_n \overset{a}{\sim} ?$

Eleventh lecture

1. Consider the following assumptions:

   (a) linearity

   (b) rank condition: $K \times K$ matrix $E(\mathbf{x}_i \mathbf{x}_i') = \mathbf{\Sigma_{xx}}$ is nonsingular

   (c) predetermined regressors: $E(\mathbf{g}_i) = 0$ where $\mathbf{g}_i = \mathbf{x}_i \cdot \varepsilon_i$

   (d) $\mathbf{g}_i$ is a martingale difference sequence with finite second moments

   i) Show, that under those assumptions, the OLS estimator is distributed asymptotically normal:

   $$\sqrt{n}(\mathbf{b} - \boldsymbol{\beta}) \underset{d}{\rightarrow} N(0, \mathbf{\Sigma_{xx}}^{-1} E(\varepsilon_i^2 \mathbf{x}_i \mathbf{x}_i') \mathbf{\Sigma_{xx}}^{-1})$$

   ii) Further, show that assumption 4 implies that the ε_i are serially uncorrelated or $E(\varepsilon_i \varepsilon_{i-j}) = 0$.

2. Show, that the test statistic

   $$t_k \equiv \frac{\sqrt{n}(b_k - \beta_k)}{\sqrt{[\widehat{Avar(b_k)}]}} \underset{d}{\rightarrow} N(0,1)$$

   converges in distribution to a standard normal distribution. Note, that b_k is the k-th element of $\mathbf{b}$ and $Avar(b_k)$ is the (k,k) element of the $K \times K$ matrix $Avar(\mathbf{b})$. Use the facts, that $\sqrt{n}(b_k - \beta_k) \underset{d}{\rightarrow} N(0, Avar(b_k))$ and $\widehat{Avar(b)} \underset{p}{\rightarrow} Avar(b)$. Use Lemma 2.4(c) for argumentation.

3. Show, that the test statistic

   $$W \equiv n(\mathbf{Rb} - \mathbf{r})'[\mathbf{R}\widehat{Avar(\mathbf{b})}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{r}) \underset{d}{\rightarrow} \chi^2(\#\mathbf{r})$$

   converges in distribution to a Chi-square with $\#\mathbf{r}$ degrees of freedom. As a hint, rewrite the equation above as $W \equiv \mathbf{c}_n' \mathbf{Q}_n^{-1} \mathbf{c}_n$. Use Lemma 2.4(d) and the footnote on page 41 for argumentation.

Twelfth lecture

1. At which stage of the derivation of the consistency property of the OLS estimator do we have to invoke a WLLN?
2. What does it mean when an estimator has the CAN property?
3. At which stage of the derivation of the asymptotic normality of the OLS estimator do we have to invoke a WLLN and when a CLT?
4. Which of the useful lemmas 1-6 is used at which stage of a) consistency proof and b) asymptotic normality proof?
5. Explain the difference of the assumptions regarding the variances of the disturbances in the finite sample context and using asymptotic reasoning.
6. There is a special case when the finite sample variances of the OLS estimator based on finite sample assumptions and based on large sample theory assumptions (almost) coincide. When does this happen?
7. What would you counter an argument of someone who says that working with the variance covariance estimate $s^2(X'X)^{-1}$ is quite OK as it is mainly consistency of the parameter estimates that counts?
8. What do we need an estimate $\widehat{Var(b)}$ for anyway?
9. In which way does replacing the assumption $E(\varepsilon_i|X) = 0$ by $E(\varepsilon_i x_i) = 0$ help removing a severe restriction and which way do the problems associated with the strict exogeneity assumption remain? Give an example.

Thirteenth lecture

1. What is an ensemple mean? Explain.

2. Why do we need the concept of stationarity in the first place?

3. Explain the concepts of weak stationarity and strong stationarity. Explain:

a) In your own words.

b) By formulas.

4. Go with open eyes through the world and look at every time series that crosses your way. Think about the possible stationary or non-stationary character of the time series. ;-)

5. When is a stationary process ergodic? Explain:

a) In your own words.

b) By formulas.

6. Which assumptions have to be fulfilled to apply Kinchin's WLLN? Which of these assumptions are weakend by the ergodic theorem? Which assumption is used instead?

7. What assumption has to be fulfilled to apply the Lindberg-Levy CLT? Is the stationarity and ergodicity assumption enough to apply the CLT? What property of the sequence $\{g_i\} = \{\varepsilon_i x_i\}$ do we assume to apply a CLT?

8. How do you call a stochastic process for which $E(g_i|g_{i-1}, g_{i-2}, ...) = 0$?

9. Show that if a constant is included in the model it follows from $E(g_i|g_{i-1}, g_{i-2}, ...) = 0$, that $\text{cov}(\varepsilon_i, \varepsilon_{i-j}) = 0 \qquad \forall j \neq 0$

10. When using the White estimate of the covariance, $\widehat{Var(b)} = \Sigma^{-1}E(\varepsilon_i x_i x_i')\Sigma^{-1}$, which assumption regarding the covariances of the errors ε_i and ε_{i-j} do we make?

Fourteenth lecture

1. When and why would you use GLS? Describe the limiting nature of the GLS approach.

2. A special case of the GLS approach is weighted least square (WLS). What difficulties could arise in a WLS estimation? How are the weights constructed?

3. Discuss the problem of an omitted variable in a linear regression using an example. When does an omitted variable not cause biased estimates?

4. When does exact multicollinearity occur? What happens to the OLS estimator in this case?

5. How is the OLS estimator and its standard error affected by (not exact) multicollinearity?

6. Which steps can be taken to overcome the problem of multicollinearity?

Fiveteenth lecture

1. Explain the term endogeniety bias. Give a practical economic example when an endogeniety bias occurs.

2. Derive the following equations of a demand and supply market equilibrium model. Hint: Hayashi Chapter 3 p. 187-191

$$p_i = \frac{\beta_0-\alpha_0}{\alpha_1-\beta_1} + \frac{v_i-u_i}{\alpha_1-\beta_1}$$

$$q_i = \frac{\alpha_1\beta_0-\alpha_0\beta_1}{\alpha_1-\beta_1} + \frac{\alpha_1 v_i-\beta_1 u_i}{\alpha_1-\beta_1}$$

$$Cov(p_i,u_i) = -\frac{Var(u_i)}{\alpha_1-\beta_1} \qquad Cov(p_i,v_i) = \frac{Var(v_i)}{\alpha_1-\beta_1}$$

$$Cov(p_i,q_i) = \alpha_1 Var(p_i) + Cov(p_i,u_i)$$

$$p_i = \frac{\beta_0-\alpha_0}{\alpha_1-\beta_1} + \frac{\beta_2}{\alpha_1-\beta_1}x_i + \frac{\zeta_i-u_i}{\alpha_1-\beta_1}$$
$$q_i = \frac{\alpha_1\beta_0-\alpha_0\beta_1}{\alpha_1-\beta_1} + \frac{\alpha_1\beta_2}{\alpha_1-\beta_1}x_i + \frac{\alpha_1\zeta_i-\beta_1 u_i}{\alpha_1-\beta_1}$$

$$Cov(x_i,p_i) = \frac{\beta_2}{\alpha_1-\beta_1}Var(x_i)$$
$$\alpha_1 = \frac{Cov(x_i,q_i)}{Cov(x_i,p_i)}$$

3. Discuss the problem of measurement errors or errors-in-variables. Give an economic example.

4. What is solution to the endogeneity problem in a linear regression framework.?

5. Use the same tools used to derive the CAN property of the OLS property to derive the CAN property of the IV estimator. Start with the sampling error and make your assumptions on the way. (applicability of WLLN, CLT...)

6. When would you use an IV estimator instead of an OLS estimator? (Hint: Which assumption of the OLS estimation is violated and what is the consequence.)

7. Describe the basic idea of instrumental variables estimation. How are the unknown parameters related to the data generating process?

8. Which assumptions are necessary to derive the instrumental variables estimator and the CAN property of the IV estimator?

9. Where do the assumptions enter the derivation of the instrumental variables estimator and the CAN property of the IV estimator?

10. Show that the OLS estimator can be conceived as a special case of the IV estimator.