Econometrics II

Fabian Waldinger (LMU Munich)

Lecture Structure

1. Recap from last lecture
2. Projection Matrix
3. Estimation of σ^2
4. Gauss-Markov Theorem: OLS is BLUE
5. GM3 contemporaneously uncorrelated
6. GM5 Normality assumption
7. t-tests

Recap from Last Lecture I

- GM Assumptions:
  1. The true model is linear in parameters: $y = \boldsymbol{X}\beta + \varepsilon$
  2. No Perfect Collinearity: the matrix $\boldsymbol{X}$ has rank k
  3. Zero Conditional Mean: $E(\varepsilon|\boldsymbol{X}) = 0$
  4. $Var(\varepsilon|\boldsymbol{X}) = \sigma^2 I$

Recap from Last Lecture II

- Under GM1-GM3 the OLS estimator is unbiased

$$\begin{aligned} E(\hat{\beta} \mid \boldsymbol{X}) &= \beta + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon|\boldsymbol{X}] \\ &= \beta + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E[\varepsilon|\boldsymbol{X}] \\ &= \beta \end{aligned}$$

- Additionally imposing GM4 we can show that

$$Var(\hat{\beta}|\boldsymbol{X}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

The Projection Matrix

- In the following, we will introduce the projection matrix
- This matrix is useful for many derivations

$$\hat{\varepsilon} = y - X\hat{\beta} = y - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'y$$
$$= \underbrace{[\boldsymbol{I_N} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']}_{\boldsymbol{M_X}} y$$

- This matrix has dimensions $N \times N$ and is a “residual maker:” if you premultiply y with $\boldsymbol{M_X}$ you get the OLS residuals

Properties of the Projection Matrix

1. Symmetric: $[\boldsymbol{I_N} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']' = [\boldsymbol{I_N'} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'] = [\boldsymbol{I_N} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']$
2. Idempotent: $\boldsymbol{M_X^q} = \boldsymbol{M_X}$
   e.g. $\boldsymbol{M_X}\boldsymbol{M_X} = \boldsymbol{M_X}[\boldsymbol{I_N} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']$
   $= \boldsymbol{M_X}\boldsymbol{I_N} - \boldsymbol{M_X}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'] = \boldsymbol{M_X}$
   The last equality follows because $\boldsymbol{M_X X = 0}$ see property 3
3. $\boldsymbol{M_X X} = \boldsymbol{0}$
   $\boldsymbol{M_X X} = \boldsymbol{I_N X} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X} \boldsymbol{= X - X = 0}_{N\times k}$
4. $\boldsymbol{M_X}\hat{\varepsilon} = \hat{\varepsilon}$

Estimation of σ^2

- The variance-covariance matrix $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ involves the disturbance variance σ^2 which is unknown
- It is reasonable to base an estimate on the RSS from the fitted regression
- We can use the projection matrix to derive this estimator:

$$\hat{\varepsilon} = \boldsymbol{M_x} y = \boldsymbol{M_x}(\boldsymbol{X}\beta + \varepsilon) = \boldsymbol{M_x}\varepsilon$$

- The last equality follows since $\boldsymbol{M_X X} = \mathbf{0}$ (see property 3 on the previous slide)
- Thus:

$$E(\hat{\varepsilon}'\hat{\varepsilon}) = E(\varepsilon' \boldsymbol{M_x' M_x} \varepsilon) = E(\varepsilon' \boldsymbol{M_x} \varepsilon)$$

- The last equality follows from property 2 on the previous slide

Estimation of σ^2

- (Remember in general $tr(A) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11} + a_{22} + a_{33}$)
- Now we use the fact that the trace of a scalar is the scalar $tr(a) = a$ if a is a scalar
- From this it follows that

$$
\begin{aligned}
E(\varepsilon' \boldsymbol{M_x} \varepsilon) &= E[tr(\varepsilon' \boldsymbol{M_x} \varepsilon)] \\
&= E[tr(\varepsilon' \boldsymbol{\varepsilon} \boldsymbol{M_x})] \\
&= \sigma^2 tr \boldsymbol{M_x} \\
&= \sigma^2 tr(\boldsymbol{I_N} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}') \\
&= \sigma^2 tr \boldsymbol{I_N} - \sigma^2 tr[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'] \\
&= \sigma^2 tr \boldsymbol{I_N} - \sigma^2 tr[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}] \\
&= \sigma^2 tr \boldsymbol{I_N} - \sigma^2 tr[\boldsymbol{I_K}] \\
&= \sigma^2 (N - k)
\end{aligned}
$$

Estimation of σ^2

- From this it follows that:

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(N-k)}$$

  is an unbiased estimator of σ^2
- This is the matrix equivalent of the formula of the first weeks of the semester:

$$\hat{\sigma}^2 = \frac{RSS}{(N-k)}$$

- Hence the estimated standard errorof $\hat{\beta}$ is: $\hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$

Gauss-Markov Theorem

- The Gauss Markov Theorem states that under GM1-4 the OLS estimator is the best linear unbiased estimator (BLUE)
- We have shown before that under GM1-3 OLS is a linear unbiased estimator (LUE)
- We now show that it is the best, i.e. most efficient estimator
- It is only the best if $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is a smaller variance than the variance of alternative linear estimators

Is There a Linear Estimator With Smaller Variance?

- Any linear estimator will be

$$\widetilde{\beta} = \underset{k \times N}{\boldsymbol{A}(\boldsymbol{X})} y = \boldsymbol{A}(\boldsymbol{X})[\boldsymbol{X}\beta + \varepsilon]$$

- If $\boldsymbol{A}(\boldsymbol{X})$ is unbiased it must be that $\boldsymbol{A}(\boldsymbol{X})\boldsymbol{X} = \boldsymbol{I}$
- The variance of the alternative estimator would be

$$Var(\widetilde{\beta}|\boldsymbol{X}) = E(\boldsymbol{A}\varepsilon\varepsilon'\boldsymbol{A}'|\boldsymbol{X})$$

- Using GM4 this simplifies to:

$$Var(\widetilde{\beta}|\boldsymbol{X}) = \boldsymbol{A}E(\varepsilon\varepsilon'|\boldsymbol{X})\boldsymbol{A}' = \sigma^2\boldsymbol{A}\boldsymbol{I}\boldsymbol{A}' = \sigma^2\boldsymbol{A}\boldsymbol{A}'$$

Is There a Linear Estimator With Smaller Variance?

- If the alternative estimator had smaller variance we would have:

$$Var(\widetilde{\beta}|\boldsymbol{X}) - Var(\hat{\beta}|\boldsymbol{X}) < 0$$

$$Var(\widetilde{\beta}|\boldsymbol{X}) - Var(\hat{\beta}|\boldsymbol{X}) = \sigma^2 \boldsymbol{AA'} - \sigma^2(\boldsymbol{X'X})^{-1}$$

- Now using the fact that $\boldsymbol{A}(\boldsymbol{X})\boldsymbol{X} = \boldsymbol{I}$ we can rewrite this as

$$\begin{aligned} &= \sigma^2 \boldsymbol{AA'} - \sigma^2 \boldsymbol{AX}(\boldsymbol{X'X})^{-1}\boldsymbol{X'A'} \\ &= \sigma^2 \boldsymbol{A} \underbrace{[\boldsymbol{I_N} - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'}]}_{=\boldsymbol{M_X}} \boldsymbol{A'} \end{aligned}$$

- Because M_X is symmetric and idempotent $\boldsymbol{AM_XA'}$ is positive semidefinite
- Hence, OLS is BLUE
- Note: Wooldridge also shows this proof without matrix algebra in section 3A.6. It is much more cumbersome without matrix algebra

GM Assumption 3 - Contemporaneously Uncorrelated

- Last week we showed that GM3: $E(\varepsilon|\boldsymbol{X}) = 0$ ensures that OLS is an unbiased estimator
- What happens if we assume a weaker version of GM3?
- In particular, consider GM3-contemporraneously uncorrelated (cu)

$$E(\varepsilon_i x_{ij}) = corr(\varepsilon_i, x_{ij}) = 0$$

- for all i and j

GM3cu - Unbiased?

- Would OLS remain unbiased under this weaker GM3cu?
- From before we know that:

$$\hat{\beta} = \beta + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon$$

- If we take expectations we get:

$$E[\hat{\beta}] = \beta + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon]$$

- $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is a function of all x_{ij} and not just a function of a single i hence $E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon] \neq 0$
- OLS would be biased in this case
- However, it can be shown that under this weaker GM3cu $\hat{\beta}$ is "aymptotically unbiased" i.e. consistent as $N \to \infty$
- The proof of this is not trivial (wait for an MSc level econometrics course)

Normality Assumption

- We add the final classical linear model assumption: $\hat{\beta}$ has a multivariate normal distribution:

$$\underset{N\times 1}{\varepsilon}|\boldsymbol{X} \sim \underset{\substack{\uparrow \\ GM5}}{N}(\underset{\substack{N\times 1 \\ \uparrow \\ GM3}}{0}, \sigma^2 \underset{\substack{N\times N \\ \uparrow \\ GM4}}{\boldsymbol{I}})$$

- This implies:

$$\hat{\beta}|\boldsymbol{X} \sim N(\beta, \sigma^2(\boldsymbol{X'X})^{-1})$$

- This assumptions allows us to carry out hypothesis tests and construct confidence intervals
- While this is a strong assumption, it can be shown that if $N \to \infty$ the distribution of the error term will converge to a Normal (proof not done in this course)

t-Tests

- Assumption GM5 also implies that:

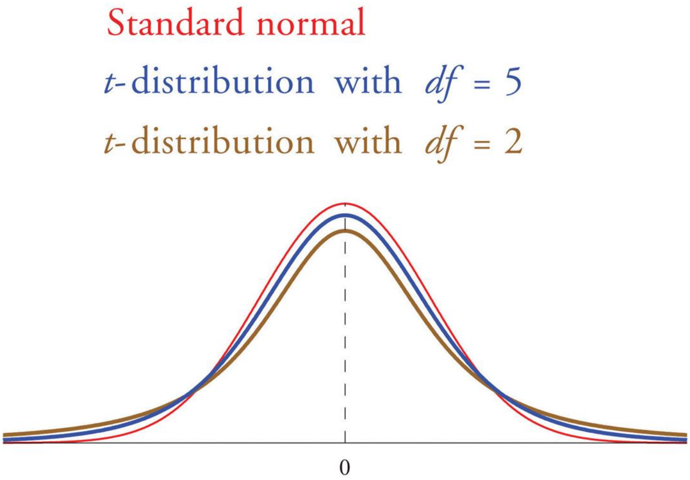
$$\hat{\beta} - \beta \sim N(0, \sigma^2(\boldsymbol{X'X})^{-1})$$

- In practice we do not know σ^2 but can estimate $\hat{\sigma}^2$
- This, however, messes up the normality assumption:

$$\hat{\beta} - \beta \nsim N(0, \hat{\sigma}^2(\boldsymbol{X'X})^{-1})$$

- What it the distribution in that case? We can show that it is distributed as a t-distribution and hence we can use t-tests to test hypotheses about β

t-Distribution vs. Normal Distribution



t-Distribution

- In general, suppose you have two *independent* random variables u and w with the following properties:

$$u \sim N(0, v_u)$$

$$w \sim \chi^2(df)$$

- Then:

$$\frac{\frac{u}{\sqrt{v_u}}}{\sqrt{\frac{w}{df}}} \sim t(df)$$

χ^2-Distribution

- In general, if we sum the squares of N independent standard normal random variables then this sum is distributed as a chi-square distribution:
- e.g. if $v \sim N(0, I_N)$ then:

$$v'v \sim \chi^2(N)$$

- If $v \sim N(0, I_N)$ and $\boldsymbol{A}$ is an idempotent matrix with $rank(\boldsymbol{A}) = q$ then:

$$v'\boldsymbol{A}v \sim \chi^2(q)$$

Distribution of the Test Statistic

- Now we show that the standard test statistic for t-test is distributed as a t-Distribution
- Under GM5:

$$\varepsilon | \boldsymbol{X} \sim N(0, \sigma^2 \boldsymbol{I}) \rightarrow \frac{\varepsilon}{\sigma} \sim N(0, \boldsymbol{I})$$

- Hence:

$$\frac{\varepsilon' \boldsymbol{M_x} \varepsilon}{\sigma^2} \sim \chi^2(\underset{\underset{rank\ \boldsymbol{M_X}}{\uparrow}}{N - k})$$

- From above we also know:

$$\hat{\beta} - \beta \sim N(0, \sigma^2(\boldsymbol{X'X})^{-1})$$

Distribution of the Test Statistic

- Now we define the test statistic as follows

$$\frac{\frac{\hat{\beta}-\beta}{\sqrt{\sigma^2(\mathbf{X'X})^{-1}}}}{\sqrt{\frac{\frac{\varepsilon' \mathbf{M_x} \varepsilon}{\sigma^2}}{N-k}}} = \frac{\frac{\hat{\beta}-\beta}{\sqrt{\sigma^2(\mathbf{X'X})^{-1}}}}{\sqrt{\frac{\hat{\varepsilon}'\hat{\varepsilon}}{\sigma^2(N-k)}}} = \frac{\frac{\hat{\beta}-\beta}{\sqrt{\sigma^2(\mathbf{X'X})^{-1}}}}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}} = \frac{\frac{\hat{\beta}-\beta}{\sqrt{(\mathbf{X'X})^{-1}}}}{\sqrt{\hat{\sigma}^2}} \sim t(N-k)$$

- The first equality follows because $\mathbf{M_x}\varepsilon = \mathbf{M_x}(y - \mathbf{X}\beta) = \mathbf{M_x}y - \mathbf{M_x X}\beta = \mathbf{M_x}y = \hat{\varepsilon}$, hence $\varepsilon' \mathbf{M_x} \varepsilon = \varepsilon' \mathbf{M'_x M_x} \varepsilon = \hat{\varepsilon}'\hat{\varepsilon}$
- We can rewrite this as:

$$\frac{\hat{\beta}-\beta}{\sqrt{\hat{\sigma}^2(\mathbf{X'X})^{-1}}} \sim t(N-k)$$

- This is the standard formula for the t-test: $\frac{\hat{\beta}-\beta^0}{se(\hat{\beta})}$