Econometrics II

Fabian Waldinger (LMU Munich)

Lecture Structure

1. Review of OLS with Matrix algebra
2. Example of OLS with one regressor
3. Second Order Condition and what it implies
4. Gauss-Markov Assumptions
5. Unbiasedness of OLS
6. Variance-Covariance Matrix of the OLS estimator

Recap from Last Lecture

- Linear Model:

$$y = \boldsymbol{X}\beta + \varepsilon$$

- We find the $\hat{\beta}$ that minimizes the RSS
- FOC:

$$-2\boldsymbol{X}'\hat{\varepsilon} = 0$$

$$\boldsymbol{X}'\hat{\varepsilon} = 0 \;\rightarrow\; \boldsymbol{X}'[y - \boldsymbol{X}\hat{\beta}] = 0 \rightarrow \hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'y$$

- For this last step we have to be able to invert $\boldsymbol{X}'\boldsymbol{X}$
- From the SOC we know that $\boldsymbol{X}'\boldsymbol{X}$ is positive definite $\rightarrow$ it is invertible
- $\boldsymbol{X}'\boldsymbol{X}$ has to have full rank to be invertible (i.e. it has rank k)

Rank of $\boldsymbol{X}'\boldsymbol{X}$

- $\boldsymbol{X}'\boldsymbol{X}$ has rank k if $\boldsymbol{X}$ has rank k
- For $\boldsymbol{X}$ to have rank k we require that there is no perfect linear combination between the columns of $\boldsymbol{X}$; e.g. we do not want the following X matrix (or other similar matrices)

$$\boldsymbol{X} = \begin{pmatrix} 1 & 5 \\ 1 & 5 \\ \vdots & \vdots \\ 1 & 5 \end{pmatrix}$$

- This condition ensures that there is no perfect multicollinearity
- This requires that N is at least as large as k (because the rank is at most the minimum of the two dimensions)

Example: Only One Regressor

- We now study the example of only one regressor ($y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i$). In that case:

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{12} \\ \vdots & \vdots \\ 1 & x_{N2} \end{pmatrix}_{N\times 2}$$

- Hence $\boldsymbol{X}'\boldsymbol{X} =$

$$\begin{pmatrix} 1 & \cdots & 1 \\ x_{12} & \cdots & x_{N2} \end{pmatrix}_{2\times N} \begin{pmatrix} 1 & x_{12} \\ \vdots & \vdots \\ 1 & x_{N2} \end{pmatrix}_{N\times 2} = \begin{pmatrix} N & \sum_{i=1}^{N} x_{i2} \\ \sum_{i=1}^{N} x_{i2} & \sum_{i=1}^{N} x_{i2}^2 \end{pmatrix}_{2\times 2}$$

Example: Only one Regressor

- Hence the OLS estimator $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'y$ is:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} N & \sum_{i=1}^{N} x_{i2} \\ \sum_{i=1}^{N} x_{i2} & \sum_{i=1}^{N} x_{i2}^2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & \cdots & 1 \\ x_{12} & \cdots & x_{N2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

$$= \begin{pmatrix} N & \sum_{i=1}^{N} x_{i2} \\ \sum_{i=1}^{N} x_{i2} & \sum_{i=1}^{N} x_{i2}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_{i2} y_i \end{pmatrix}$$

Example: Only one Regressor

- What is $(\boldsymbol{X}'\boldsymbol{X})^{-1}$?
- For a 2×2 matrix:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

- The inverse is defined as:

$$A^{-1} = \frac{1}{det} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Example: Only one Regressor

- Hence $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is:

$$\frac{1}{N\sum_{i=1}^{N}x_{i2}^2 - \sum_{i=1}^{N}x_{i2}\sum_{i=1}^{N}x_{i2}}\begin{pmatrix}\sum_{i=1}^{N}x_{i2}^2 & -\sum_{i=1}^{N}x_{i2} \\ -\sum_{i=1}^{N}x_{i2} & N\end{pmatrix}$$

- This can be simplyfied to:

$$\frac{1}{N\sum_{i=1}^{N}x_{i2}^2 - N^2\bar{x}_2^2}\begin{pmatrix}\sum_{i=1}^{N}x_{i2}^2 & -N\bar{x}_2 \\ -N\bar{x}_2 & N\end{pmatrix}$$

Example: Only one Regressor

- Hence $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'y$ is:

$$\frac{1}{N\sum_{i=1}^{N}x_{i2}^2 - N^2\bar{x}_2^2}\begin{pmatrix} \sum_{i=1}^{N}x_{i2}^2 & -N\overline{x}_2 \\ -N\overline{x}_2 & N \end{pmatrix}\begin{pmatrix} N\overline{y} \\ \sum_{i=1}^{N}x_{i2}y_i \end{pmatrix}$$

$$= \frac{1}{N\sum_{i=1}^{N}x_{i2}^2 - N^2\bar{x}_2^2}\begin{pmatrix} \sum_{i=1}^{N}x_{i2}^2 N\overline{y} - N\overline{x}_2\sum_{i=1}^{N}x_{i2}y_i \\ -N^2\overline{x}_2\overline{y} + N\sum_{i=1}^{N}x_{i2}y_i \end{pmatrix}$$

Example: $\hat{\beta}_2$

- Hence

$$\hat{\beta}_2 = \frac{-N^2\overline{x}_2\overline{y} + N\sum_{i=1}^{N} x_{i2}y_i}{N\sum_{i=1}^{N} x_{i2}^2 - N^2\bar{x}_2^2} = \frac{\frac{1}{N}\sum_{i=1}^{N} x_{i2}y_i - \overline{x}_2\overline{y}}{\frac{1}{N}\sum_{i=1}^{N} x_{i2}^2 - \bar{x}_2^2}$$

The Second Order Condition Simple Regression

- We will look at the SOC in the simple linear regression case (only one x variable)
- Recall the SOC from the last lecture: $2(\boldsymbol{X}'\boldsymbol{X})$
- We therefore have to understand whether $(\boldsymbol{X}'\boldsymbol{X})$ is positive definite (would imply a minimum of the RSS) or negative definite (would imply a maximum)
- In the simple linear regression example:

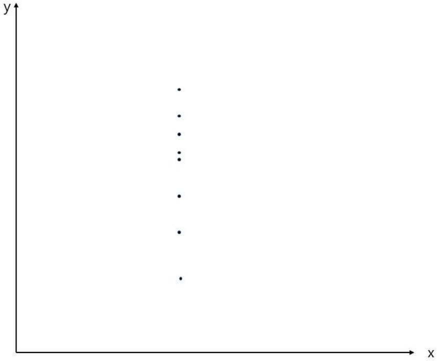
$$(\boldsymbol{X}'\boldsymbol{X}) = \begin{pmatrix} N & \sum_{i=1}^{N} x_{i2} \\ \sum_{i=1}^{N} x_{i2} & \sum_{i=1}^{N} x_{i2}^2 \end{pmatrix}_{2\times 2}$$

The Second Order Condition Simple Regression

- How to check whether $(\boldsymbol{X}'\boldsymbol{X})$ is positive or negative definite?
- Recall that a 2x2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is positive definite iff the two leading principle minors are positive, i.e.:
  1. $a > 0$
  2. $det \begin{pmatrix} a & b \\ c & d \end{pmatrix} > 0$
- In our example:
  1. $a = N > 0$
  2. $det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{N} \sum_{i=1}^{N} x_{i2}^2 - \bar{x}^2 > 0$
- The latter expression is the expression of the sample variance of x which has to be > 0

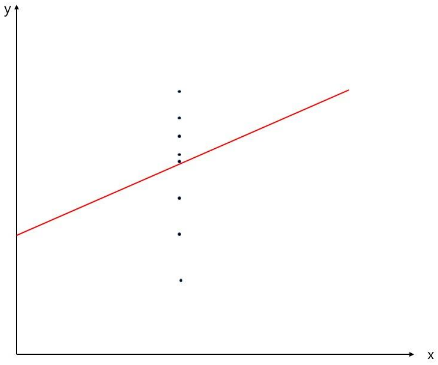
What Would Happen if $s_{x^2} = 0$?

- What would happen if $s_{x^2} = 0$?
- The data would look as follows:



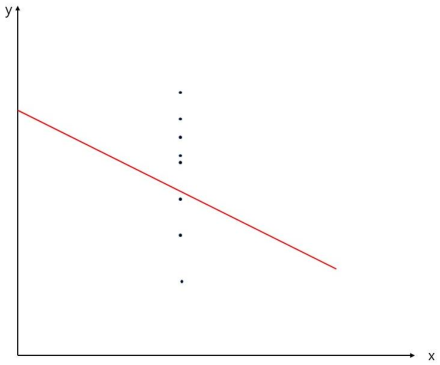
What Would Happen if $s_{x^2} = 0$?

- What would happen if $s_{x^2} = 0$?
- The data would look as follows:



What Would Happen if $s_{x^2} = 0$?

- What would happen if $s_{x^2} = 0$?
- The data would look as follows:



What Would Happen if $s_{x^2} = 0$?

- If there was no variation in X we could not estimate a regression line
- This would be a violation of SLR.3 from last term
- Here we see mathematically why SLR.3 needs to hold

Gauss-Markov Assumptions

- In the lectures with Joachim Winter you covered the Gauss-Markov assumptions
- We now look at GM assumptions that have been adapted for the use of matrix Algebra
  1. The true model is linear in parameters:

$$y = \boldsymbol{X}\beta + \varepsilon$$

  1. No Perfect Collinearity
     The matrix X has rank k
  2. Zero Conditional Mean

$$E(\varepsilon|\boldsymbol{X}) = 0$$

- Note: assumption 3 is implied by MLR.4. under the random sampling assumption MLR.2 from last term

Unbiasedness of OLS Estimator

- These 3 assumptions need to hold to ensure that the OLS estimator is unbiased
- A particular focus is on GM assumption 3: $E(\varepsilon|X) = 0$
- This assumption ensures that the value of X does not tell us anything about the mean of ε (but may tell us something about the variance...)
- There are also stronger or weaker versions of GM assumption 3
- Under stronger assumptions OLS is still unbiased and under the weaker of assumption $[corr(\varepsilon_i, x_{ik}) = 0]$ we can show that OLS is consistent (but not unbiased)

Proof that OLS is Unbiased

- We now show that the OLS estimator is unbiased if GM1-GM3 hold

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

- We now plug in the true model for y:

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

- This can be rewritten as:

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

Proof that OLS is Unbiased

- What is the expected value of the expression on the previous slide?
- We use assumption GM3

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

- The last step follows from GM3
- This shows that $\hat{\beta}_{OLS}$ is an unbiased estimator for β

Variance-Covariance Matrix of the Error Term

- The variance-covariance matrix of $\varepsilon|\boldsymbol{X}$is defined as follows:

$$
\begin{aligned}
Var(\varepsilon|\boldsymbol{X}) &\equiv \begin{pmatrix} var(\varepsilon_1|\boldsymbol{X}) & cov(\varepsilon_1,\varepsilon_2|\boldsymbol{X}) & \cdots & cov(\varepsilon_1,\varepsilon_N|\boldsymbol{X}) \\ & var(\varepsilon_2|\boldsymbol{X}) & & \vdots \\ \vdots & & & \\ & \cdots & & var(\varepsilon_N|\boldsymbol{X}) \end{pmatrix}_{N\times N} \\
&\equiv E[(\varepsilon - E\varepsilon|\boldsymbol{X})(\varepsilon - E\varepsilon|\boldsymbol{X})'|\boldsymbol{X}]
\end{aligned}
$$

GM Assumption 4

- We now introduce GM assumption 4:

$$Var(\varepsilon|\boldsymbol{X}) = \sigma^2 I$$

- This is short for

$$Var(\varepsilon|\boldsymbol{X}) = \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2 \end{pmatrix}$$

- This assumptions implies two important characteristics:
  1. No Heteroscedasticity
  2. No Autocorrelation

The Variance-Covariance Matrix of the OLS Estimator

- Under GM1-GM4 we can derive the variance-covariance matrix of the OLS estimator
- Under GM1-GM2 we have shown that:

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

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

The Variance-Covariance Matrix of the OLS Estimator

- What is the variance-covariance matrix of $\hat{\beta}$?

$$Var(\hat{\beta}|\boldsymbol{X}) = E\{(\hat{\beta} - E[\hat{\beta}|\boldsymbol{X}])(\hat{\beta} - E([\hat{\beta}|\boldsymbol{X}])'|\boldsymbol{X}\}$$

- The above is just the formula for the variance-covariance matrix in matrix notation
- From assumptions GM1-GM3 we know that $E[\hat{\beta}|\boldsymbol{X}] = \beta$, hence:

$$E\{(\hat{\beta} - E[\hat{\beta}|\boldsymbol{X}])(\hat{\beta} - E([\hat{\beta}|\boldsymbol{X}])'|\boldsymbol{X}\} = E\{(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|\boldsymbol{X}\}$$

- Now we substitute $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon$ for $\hat{\beta} - \beta$

$$E\{((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon)((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon)'|\boldsymbol{X}\}$$

The Variance-Covariance Matrix of the OLS Estimator

- Now using matrix algebra:

$$E\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon)'|\boldsymbol{X}\}$$

$$E\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon)(\varepsilon'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1'})|\boldsymbol{X}\}$$

- Because $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is symmetric: $(\boldsymbol{X}'\boldsymbol{X})^{-1'}=(\boldsymbol{X}'\boldsymbol{X})^{-1}$

$$= E\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon)(\varepsilon'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1})|\boldsymbol{X}\}$$

$$= E\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon\varepsilon'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}|\boldsymbol{X}\}$$

The Variance-Covariance Matrix of the OLS Estimator

- Because $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$ are fixed with respect to $\boldsymbol{X}$ we can rewrite the expression on the previous slide as:

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E\{\varepsilon\varepsilon'|\boldsymbol{X}\}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- Because of GM4 $E\{\varepsilon\varepsilon'|\boldsymbol{X}\} = \sigma^2 I$, hence the expression above simplifies to:

$$\begin{aligned} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2 I\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'I\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \end{aligned}$$

- The last expression is the formula for the variance-covariance matrix of the OLS estimator