Econometrics II

Fabian Waldinger (LMU Munich)

OLS Using Matrix Notation

- In matrix notation the OLS regression equation can be written as:

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

- which is short for:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}_{N\times 1} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{pmatrix}_{N\times k} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}_{k\times 1} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}_{N\times 1}$$

Residuals and RSS

- Residual of observation i:

$$\begin{aligned}\hat{\varepsilon}_i &= y_i - [\hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_k x_{ik}] \\ &= y_i - \hat{y}_i \\ &= \text{"observed } y\text{"} - \text{"fitted value"}\end{aligned}$$

- We obtain the Ordinary Least Squares (OLS) estimator by minimizing the residual sum of squares (RSS)
- The sum of squared residuals (RSS) is:

$$\sum_{i=1}^{N} \hat{\varepsilon}_i^2 \text{ or in matrix notation } \hat{\varepsilon}'\hat{\varepsilon} = (y - \boldsymbol{X}\hat{\beta})'(y - \boldsymbol{X}\hat{\beta})$$

- where $\hat{\varepsilon}$ is a $N \times 1$ vector of residuals, by multiplying
- $\hat{\varepsilon}$ by its transpose we get the RSS

Minimizing RSS - Matrix Nation

- In matrix notation the FOCs are:

$$-2 \underset{k\times N}{\boldsymbol{X}'} \underset{N\times 1}{\hat{\varepsilon}} = \underset{k\times 1}{0}$$

- From the FOC we can derive the OLS estimator:

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

GM-Assumptions 1-3

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

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

Unbiasedness of OLS

- These 3 assumptions need to hold to ensure that the OLS estimator is unbiased
- A particular focus is on GM assumption 3.: $E(\varepsilon|X) = 0$
- There are also stronger or weaker versions of GM assumption 3
- Under stronger assumptions OLS is still unbiased and 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 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:

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

- Taking expectations:

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

GM4

4.

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

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

Variance-Covariance Matrix of OLS Estimator

- Under GM1-4 we can derive the variance-covariance matrix of the OLS estimator (note this is skipping some steps see lecture 2):

$$
\begin{aligned}
Var(\hat{\beta}|\boldsymbol{X}) &= E\{(\hat{\beta} - E[\hat{\beta}|\boldsymbol{X}])(\hat{\beta} - E([\hat{\beta}|\boldsymbol{X}])'|\boldsymbol{X}\} \\
&= E\{(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|\boldsymbol{X}\} \\
&= E\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon)'|\boldsymbol{X}\} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E\{\varepsilon\varepsilon'|\boldsymbol{X}\}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}
$$

GM-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

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)

Violations of GM-Assumptions

- GM2: OLS cannot be estimated (e.g. dummy variable trap)
- GM4: OLS is no longer most efficient estimator, s.e. are wrong but OLS remains unbiased/consistent

Violations of GM3

- Leading causes for violations of GM3:
  1. Omitted variable bias
  2. Measurement Error in X
  3. Simultaneity (next lecture)

Omitted Variable Bias

- We can show that the omitted variable bias is:

$$E(\tilde{\beta}_2|\boldsymbol{X}) = \beta_2 + \beta_3 \frac{\sum(x_{2i} - \bar{x}_2)(x_{3i} - \bar{x}_3)}{\sum(x_{2i} - \bar{x}_2)^2} = \beta_2 + \beta_3 \frac{Cov(x_2, x_3)}{Var(x_2)}$$

Measurement Error

- Classical measurement error leads to attenuation bias
- With measurement error the plim of $\hat{\beta}_2$is:

$$plim\hat{\beta}_2 = \beta_2 \left(\frac{\sigma_Z^2}{\sigma_Z^2 + \sigma_w^2} \right)$$

- The last term in parentheses is < 1
- If the measurement error is not classical (i.e. $E(w) \neq 0$) OLS will also be biased

How Do We Overcome Violations of GM3?

- Solutions for improving basic OLS regressions
  - add control variables (overcomes omitted variable bias)
    particularly powerful control variables are fixed effects (see panel data models)
  - get variables without measurement error
- Design identification strategies that directly address the violation of GM3
  1. Randomize the variable of interest
  2. Find quasi-random variation:
     1. IV
     2. DiD
     3. RD

Randomized Experiments

- Randomized experiments address violations of GM3
- They can be evaluated with the following regression:

$$Y_i = \beta_1 + \beta_2 D_i + \varepsilon_i$$

- This will estimate the causal effect of the treatment because the treatment is randomly assigned

Instrumental Variables

- IV also allows to address violations of GM3
- 2 important conditions for a valid IV:
  1. $Cov(S, Z) \neq 0$ (first stage exists)
  2. $Cov(Z, \varepsilon) = 0$ (exclusion restriction: Z is uncorrelated with any other determinants of the dependent variable)
- For a simple model with one regressor the IV estimator is:

$$\beta_2 = \frac{Cov(z, y)}{Cov(z, x)}$$

- Or in matrix notation:

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

IV Jargon

- Causal relationship of interest:

$$y = \beta_1 + \beta_2 x + \varepsilon$$

- First-Stage regression:

$$x = \gamma_1 + \gamma_2 z + \mu$$

- Second-Stage regression:

$$y = \beta_1 + \beta_2 \hat{x} + \upsilon$$

- Reduced form:

$$y = \lambda_1 + \lambda_2 z + \upsilon$$

IV with Heterogeneous Treatment Effects

- When treatment effects are heterogeneous, one needs the following assumptions for IV:
  1. Independence assumption:
     The IV is independent of the vector of potential outcomes and potential treatment assignments (i.e. as good as randomly assigned):
  2. Exclusion restriction
  3. First stage
  4. Monotonicity
     Either $\gamma_{1i} \geq 0$ for all i or $\gamma_{1i} \leq 0$ for all i

IV Estimates LATE

- If all 4 assumptions are satisfied, IV estimates LATE (Local Average Treatment Effect)
- LATE is the average effect of X on Y for those whose treatment status has been changed by the instrument Z (the compliers)

Panel Data

- Models with panel data can be written as:

$$y_{it} = \boldsymbol{x_{it}}\beta + c_i + u_{it}$$

- where $\boldsymbol{x}_{it}$ has dimensions $1 \times K$
- interest lies in estimating the $K \times 1$ vector β
- c_i is unobserved and hence enters the estimation error if we just regress y on x
- u_{it} is an error term that satisfies GM3

Panel Data Methods

- If c_i was independent of all explanatory variables in all time periods, there would be no violation of GM3 if we ignore the c_i component of the error term and simply regressed y on x:

$$y_{it} = \boldsymbol{x_{it}}\beta + \underbrace{c_i + u_{it}}_{\nu_{it}}$$

- Panel data methods allow us to obtain consistent estimates of β even if c_i is not independent of all $\boldsymbol{x}_{it}$
- We will cover two main methods for addressing the correlation of $\mathbf{x}$ and c_i
  1. Random effects
  2. Fixed effects
     1. within estimator
     2. dummy variables
     3. first differences
- The methods differ on the assumptions that we make on the relationship between $\mathbf{x}$ and c_i

Differences-in-Differences

- Differences-in-differences:

$$E[Y_{it} \mid i = Treat, t = Post] - E[Y_{it} \mid i = Treat, t = Pre]$$

$$-(E[Y_{it} \mid i = Control, t = Post]) - E[Y_{it} \mid i = Control, t = Pre]) = \delta$$

- Regression differences-in-differences:

$$Outcome_{it} = \beta_1 + \beta_2 Treat_i + \beta_3 Post_t + \beta_4 (Treat * Post)_{it} + \varepsilon_{it}$$

- Key assumption: the outcome in treatment and control group would follow the same time trend in the absence of the treatment

Regression Discontinuity Design

- Regression discontinuity research designs exploit the fact that some rules are quite arbitrary and therefore provide good quasi-experiments if you compare people (or cities, firms, countries,...) who are just affected by the rule with people who are just not affected by the rule
- There are 2 types of RD designs:
  1. Sharp RD: treatment is a deterministic function of a covariate x
  2. Fuzzy RD: exploits discontinuities in the probability of treatment conditional on a covariate x (the discontinuity is then used as an IV)
- RD captures the causal effect by distinguishing the nonlinear and discontinuous function, $1(x_i \geq x_0)$ from the smooth function $f(x_i)$

Sharp RD

$$
\begin{aligned}
E[Y_{0i} \mid x_i] =& \alpha + \beta x_i \\
Y_{1i} =& Y_{0i} + \rho
\end{aligned}
$$

- This leads to the regression:

$$Y_i = \alpha + \beta x_i + \rho D_i + \eta_i$$

- Key identifying assumption:

$$E[Y_{0i} \mid x_i] \text{ and } E[Y_{1i} \mid x_i] \text{ are continuous in } x_i \text{ at } x_0$$

Exam

- Exam time will be 90 minutes
- 90 points in total (i.e. 1 point a minute)
- 3 Questions with sub questions
- All material from lecture slides is relevant
- You will also need to read/understand R code

Practical Exam Tips

- Answer only the question!
- Don't write down material that is somewhat related but not needed to answer the question
- Write down the math and or graphs if it helps to answer the question
- Write as clearly as possible
- Don't talk about problems in generic terms "There is omitted variable bias" but be specific "Because w is omitted from the regression there will be an upward bias."
- Try to answer all sub-questions