Applied Econometrics

Bernd Süßmuth
IERE · Institute for Empirical Research in Economics

University of Leipzig

October 11, 2019
Teaching portfolio – IERE: Team Econometrics

**B.Sc. Wirtschaftswissenschaften**

- Ökonometrie
  - 10 LP
- Applied Econometrics
  - 5 LP

**M.Sc. Economics**

- Advanced Econometrics and Statistics
  - 10 LP
- Microeconometrics
  - 10 LP
- Times Series Analysis for Macroeconomics and Finance
  - 10 LP
- Quantitative Economic History
  - 5 LP
- Quantitative Economics of Education
  - 5 LP
- Seminar Applied Econometrics
  - 5 LP
Overview

1. Refresher: Classical linear regression model (CLRM)
2. Endogeneity and IV estimation
3. Panel data models: Random effects (RE) / Fixed effects (FE)
4. Box-Jenkins ARIMA: reloaded
5. GARCH models
6. VAR models and impulse response functions (IRFs)
7. Error corrections model (ECM / ARDL)
Classical linear regression model

1 Assumptions
2 Derivation and properties of the LS estimator
3 Tests in the CLRM context
4 Serial correlation and heteroskedasticity: WLS and Newey-West
a) Three central assumptions concerning the residuals

(i) \( E(u_t) = 0 \) for all \( t = 1, \ldots, T \)

The error term has **no** systematic impact on the dependent

(ii) \( Var(u_t) = \sigma^2 \) for all \( t = 1, \ldots, T \)

**constant** variance in error terms (homoskedasticity)

(iii) \( Cov(u_t, u_i) = 0 \) for all \( t \neq t' \)

Residuals are **not** serially correlated

\[ \rightarrow \text{no autocorrelation of residuals} \]
\[ \rightarrow \text{no serial correlation} \]
\[ (\Rightarrow \text{var/cov-matrix} = \sigma^2 I) \]
b) Two central assumptions concerning the explanatories

(iv) $\text{Cov} (X_t, u_t) = 0$ for all $t = 1, \ldots, T$

Residuals and (non-stochastic) exogenous variables are **not** correlated with each other.

(v) $X$ is of full rank

Each single exogenous must be able to vary independently of the others = **no** (quasi-)fixed relationship

(e.g. $x_{t,3} = x_{t,1} + x_{t,2}$: the 3 interests short/middle/long-termed do not vary independently as explanatories for dependent “GFCF”)

$\Rightarrow$ no multicollinearity

($X'X$ is invertible, i.e. finite or non-singular)
1 Classical linear regression model
- Assumptions
- Derivation and properties of the LS estimator
- Tests in the CLRM context
- Serial correlation and heteroskedasticity: WLS and Newey-West
a) Derivation of the normal equation system (NES)

Objective is to minimize sum of squared residuals:

$$\min_{\hat{\beta}} Q(\hat{\beta}) = \hat{u}u' = (y - X\hat{\beta})' (y - X\hat{\beta})$$

FOC: \[ \frac{\partial Q}{\partial \beta} \cdot \frac{\partial \hat{u}}{\partial \beta} = -2\hat{u}'X = 0 \iff \hat{u}'X = 0 \iff X'\hat{u} = 0 \]

\[ \Rightarrow X'\hat{u} = X' (y - X\hat{\beta}) = X'y - X'X\hat{\beta} = 0 \iff X'X\hat{\beta} = X'y \]

\[ \Rightarrow \hat{\beta} = (X'X)^{-1} X'y \]

NES
b) Properties of the regression (beta) coefficients

1) $\hat{\beta} =$ estimates, i.e. random variables

\[
\hat{\beta} = (X'X)^{-1} X'y = (X'X)^{-1} X'(X\beta + u) = (X'X)^{-1} X'X\beta + (X'X)^{-1} X'u = \beta + (X'X)^{-1} X'u
\]

for $u = 0 \Rightarrow \hat{\beta} = \beta$ (perfect fit)

2) $\hat{\beta} =$ unbiased estimates: $E\left(\hat{\beta}\right) = \beta$

\[
E\left(\hat{\beta}\right) = E\left(\beta + (X'X)^{-1} X'u\right) = \beta + E\left((X'X)^{-1} X'u\right) = \beta + E\left(u\right)(X'X)^{-1} X' = \beta
\]

as long as $E\left(u\right) = 0$ holds.
b) Properties of the regression (beta) coefficients (cont'ed)

3) var/cov-matrix of $\hat{\beta} = \text{function of residual variance}$

$$Var \left( \hat{\beta} \right) = E \left( \hat{\beta} - E \left( \hat{\beta} \right) \right)^2$$

$$\sum_{\hat{\beta} \hat{\beta}} = E \left[ \left( \hat{\beta} - E \left( \hat{\beta} \right) \right) \left( \hat{\beta} - E \left( \hat{\beta} \right) \right)' \right]$$

$$= E \left[ \left( \hat{\beta} - \beta \right) \left( \hat{\beta} - \beta \right)' \right] = E \left( (X'X)^{-1} X' uu' X (X'X)^{-1} \right)$$

$$= (X'X)^{-1} X' E \left( uu' \right) X (X'X)^{-1} = (X'X)^{-1} \sigma^2 I X' (X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1} X' X (X'X)^{-1} = \sigma^2 (X'X)^{-1}$$

with $E \left( uu' \right) = \sum_{uu'} = \sigma^2 I$

$\Rightarrow$ the higher the residual variance, the higher $\sum_{\hat{\beta} \hat{\beta}}$ (standard errors!)
c) Further properties of the CLRM

4) Theoretical and empirical residual variance ≠ identical (usually)

\[ \text{Var} (u) = E (u - E (u))^2 = E ((u - 0) (u - 0)') \]

\[ = E (uu') \]

\[ = \sum_{uu'} = \sigma^2 I \]

⇒ theoretical residual variance (u unknown): \( \sigma^2 = \frac{1}{T} \sum_{t=1}^{T} u_t^2 \)

⇒ empirical residual variance (u unknown/estimated):

\[ \sigma^2 = \frac{1}{T-K} \sum_{t=1}^{T} \hat{u}_t^2 = \frac{1}{T-K} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 \]

with \( T = \) number of obs, \( K = \) number of regressors

⇒ \( T - K = \) degrees of freedom
c) Further properties of the CLRM (cont’ed)

5) Definition and critical aspects of goodness-of-fit measure $R^2$

5.1 Definition $\text{Var}(y) = \text{Var}(\hat{y}) + \text{Var}(\hat{u})$

i.e. total variance = explained variance + unexplained variance

$\Rightarrow R^2$ has a “correlation coefficient” structure:

$$R^2 = \frac{\text{Var}(\hat{y})}{\text{Var}(y)} = 1 - \frac{\text{Var}(\hat{u})}{\text{Var}(y)},$$

measuring how strongly $y$ correlates with $\hat{y}$: $0 \leq R^2 \leq 1$

$\Rightarrow$ Adjusted $R^2 \approx \frac{T}{T-K} R^2 - \frac{K}{T-K} \left(1 - \frac{T-1}{T-K} (1 - R^2)\right)$

5.2 Three crucial points of weakness of $R^2$

★ purely technical measure (no theoretical basis / motivation)
★ depending on data / regression properties (trend, OVB, etc.)
★ not invariant to variable transformations (changes in $u$-structure)
# Contents I

1. Classical linear regression model
   - Assumptions
   - Derivation and properties of the LS estimator
   - Tests in the CLRM context
   - Serial correlation and heteroskedasticity: WLS and Newey-West
a) Three examples for classical two-sided tests

- **Macro-consumption function: does real interest matter?**
  \[ H_0 : \beta_r = 0 \]

- **Wage equation: do subjects suffer from money illusion?**
  \[ H_0 : \beta_{\pi e} = 1 \]

- **Cobb Douglas production function: linear homogeneity justified?**
  - **log-linearize**
  - **replace** \( \alpha_1 \) **by** \( (k - \alpha_2) \)
  - **\( \Leftrightarrow \) for** \( k = 1 \): \( \alpha_1 + \alpha_2 = 1 \) **holds**

  \[ H_0 : k = 1 \]
b) Definitions of error type and level of significance

- **Type I Error**
  
  \( H_0 \) is rejected, although \( H_0 \) is true

- **Type II Error**
  
  \( H_0 \) is *not* rejected, although \( H_1 \) \( (H_0) \) is true (false)

- **Level of significance**

  Probability of Type I Error

  ⇒ Both error types cannot be minimized at a time: “trade-off”

  ⇒ Minimization of Type I Error is convenient
c) Test on more than one regression coefficient

- Refresh: Case 1 – test on single parameter: t-test

- Problems if testing hypotheses with Student’s t-test
  
  1. Eventually neither $H_0$ nor its polar opposite is confirmed
  
  2. Dependence of hypothesis testing on standard deviation of estimated coefficients and, thus, from sample size

- Refresh: Case 2 – simultaneous test on $>1$ parameter: F-test
Classical linear regression model

- Assumptions
- Derivation and properties of the LS estimator
- Tests in the CLRM context
- Serial correlation and heteroskedasticity: WLS and Newey-West
Heteroskedasticity

- Basic definition

\[ y = X\beta + \epsilon \quad \rightarrow \quad \text{single vector notation:} \quad y_i = x_i' \beta + \epsilon \]
\[ \rightarrow \quad \text{but now:} \quad \sigma_i \neq \sigma_j \]

- Refresh: heterosked. & WLS (e.g. Komlos/Süssmuth, p. 91-94)
Serial correlation I

Similar to heteroskedasticity problem

h1 in case of known weights (known rel. of variance) and

h2 in case of unknown weights, we may also discriminate 2 cases in the serial correlation context

a1 for adequate linear model and

a2 for inadequate linear model (misspecification)

Example: serial correlation due to misspecification
Refresh: Serial correlation (e.g. Komlos/Süssmuth, p. 83-91)

In rare cases h1 and a1, we may work with WLS and Cochrane-Orcutt-Transformation (COT)

To identify symptoms of heteroskedasticity and serial correlation and to check, if –for cases h1 and a1– WLS bzw. COT “did well,” we use the standard tests:

- Goldfeldt-Quandt-, Breusch-Pagan-, Harvey-Godfrey-, White-test
- Durbin-Watson-test

In all other cases, we resort to so-called heteroskedasticity and autocorrelation consistent Newey-West (“robust”) standard errors

Coefficient estimates = untouched: unbiasedness (!)
Newey-West standard errors

“HAC \equiv \text{heteroskedasticity and autocorrelation consistent}”

- Lagged endogenous values as in AR(1) model (here as OLS sys):

\[ Y_t = \phi_0 + \phi_1 Y_{t-1} + \epsilon_t \equiv \phi_0 + \phi_1 X_t + \epsilon_t \]  

(1)

are stochastic and depend on past shocks

- Shocks \( \epsilon_t \) are by definition serially correlated

⇒ Sample variance of LS estimator is concerned / biased

⇒ Requires alternative computation of unbiased standard errors

- Generalization for AR(\( p \)) via \( X = \{ Y_{t-1}, ..., Y_{t-p} \} \)
The standard LS-estimator for an AR($p$) is $\hat{\phi} = (X'X)^{-1} X'Y_t$

Analogously, it holds for an AR(1) as in (1)

$$\hat{\phi}_1 = \frac{1}{T} \sum_{t=1}^{T} (X_t - \bar{X}) (Y_t - \bar{Y}) = \frac{\sigma_{XY}}{\sigma_X^2}. $$

What is the difference between estimated and actual $\phi$ for an AR($p$)?

$$\hat{\phi} = (X'X)^{-1} X'Y_t = (X'X)^{-1} X' (X\phi + \epsilon) = \phi + (X'X)^{-1} X'\epsilon$$

$$\Leftrightarrow \hat{\phi} - \phi = (X'X)^{-1} X'\epsilon$$

Analogously, in the simple AR(1) case

$$\hat{\phi}_1 - \phi_1 = \frac{1}{T} \sum_{t=1}^{T} (X_t - \bar{X}) \epsilon_t = \frac{1}{T} \sum_{t=1}^{T} \nu_t \sigma_X^2 $$

(2)
From (2), it follows that

$$\text{var} \left( \hat{\phi}_1 \right) = \text{var} \left( \frac{1}{T} \sum_{t=1}^{T} v_t \right) \left( \sigma_X^2 \right)^{-2}. \quad (3)$$

⇒ Focusing only on the variance in the numerator of (3) and setting $T = 2$

$$\text{var} \left( \frac{1}{T} \sum_{t=1}^{T} v_t \right) = \text{var} \left[ \frac{1}{2} (v_1 + v_2) \right].$$

⇒ Since $\text{var}(aX) = a^2 \text{var}(X)$, it follows from the variance decomposition theorem

$$\text{var} \left( \frac{1}{T} \sum_{t=1}^{T} v_t \right) = \frac{1}{4} \left[ \text{var} \left( v_1 \right) + \text{var} \left( v_2 \right) + 2\text{cov} \left( v_1, v_2 \right) \right].$$
As $\rho_{X,Y} = \text{corr}(X, Y) = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$, we may write:

$$\text{var} \left( \frac{1}{T} \sum_{t=1}^{T} v_t \right) = \frac{1}{2} \sigma_v^2 + \frac{1}{2} \rho_1 \sigma_v^2 = \frac{1}{2} \sigma_v^2 \cdot f_2,$$

where $\rho_1 = \text{corr}(v_1, v_2)$ and $f_2 = (1 + \rho_1)$.

- In the i.i.d. idealistic case: $\rho_1 = 0 \iff f_2 = 1$ and $\text{var}(\hat{\phi}_1)$ is unbiased.
- However, if $\rho_1 \neq 0$, this is not the case!

In general, i.e. for $T = T$,

$$\text{var} \left( \frac{1}{T} \sum_{t=1}^{T} v_t \right) = \frac{\sigma_v^2}{T} \cdot f_T \Rightarrow \text{var}(\hat{\phi}_1) = \frac{\sigma_v^2}{T (\sigma_X^2)^2} \cdot f_T.$$
\[ \text{SE} \left( \hat{\phi}_1 \right) = \sqrt{\frac{\sigma_v^2}{T \left( \sigma_X^2 \right)^2} \cdot f_T}, \text{ where } f_T = 1 + 2 \sum_{j=1}^{T-1} \left( \frac{T - j}{T} \right) \rho_j \] (4)

- Conventional standard errors \( \neq \) correct for autocorrelated \( \epsilon_t \)
- They are off by a factor \( f_T \) as can be seen from (4)
- We require a SE-adjustment!

We conclude:

- Necessity of “robust” computation of SE (warranting HAC errors)
- If we would know \( f_T \), we could straightforwardly adjust, but usually
\[ \Rightarrow f_T \] must be estimated explicitly
The most common estimator for $f_T$ is

$$\hat{f}_T = 1 + 2 \sum_{j=1}^{m-1} \left( \frac{m-j}{m} \right) \tilde{\rho}_j$$

- where $\tilde{\rho}_j$ is an estimate for $\rho_j$

- $m$ is referred to as “truncation parameter”:
  - Newey and West (1987) set $m = 4 \left( \frac{T}{100} \right)^{2/9}$
  - A convenient rule of thumb is $m = 0.75 T^{1/3}$

- In Stata: `newey` or `robust` option (check out: `help newey`)
- In EViews: `Estimate` → `Options` → `HAC (Newey-West)`