DIW Berlin

Deutsches Institut
für Wirtschaftsforschung

www.diw.de

Consultancy to Develop and Implement a
Macroeconomic Model for Lesotho (DIMMoL)

Macro-Econom(etr)ic Modelling

Part 6

Dr. Stefan Kooths
DIW Berlin – Macro Analysis and Forecasting

Course program

- Introduction
- Outline of macroeconom(etr)ic models
- Macroeconomic framework
- **Econometric methodology (cont.)**
- Applied econometrics with EViews
- Lesotho case studies

Econometric methodology: Overview

- Fundamentals of probability
- Fundamentals of mathematical statistics
- Principles of regression analysis (cross sections)
- Time series regression models

Principles of regression analysis

- Population regression model
- Properties of OLS estimates
- Functional forms and data scaling
- Confidence intervals and hypothesis testing
- OLS asymptotics
- Goodness-of-fit and selection of regressors
- Specification and data problems

Population model and regression functions

population model („true" model)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

population regression function

using OLS estimation

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k$$

$\hat{\beta}_j \neq \beta_j$

sample regression function

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k + \hat{u}$$

Terminology

- Dependent variable (y)
  - explained variable
  - response variable
  - predicted variable
  - regressand
- Independent variables (x)
  - explanatory variables
  - control variables
  - predictor variables
  - regressors
- Fitted value ($\hat{y}$, speak: „y hat“)
- Error (u)
  - disturbance
  - „unobserved“ variables
- Residual ($\hat{u}$)

Gauss-Markov assumptions

- **Linearity in parameters**
  population model is characterized by a linear regression function and additive errors
- **Random sampling**
  random sample of n observations following the population model
- **No perfect collinearity**
  none of the independent variables is constant and no exact linear relationships among them
- **Zero conditional mean**
  error has an expected value of zero given any values of the independent variables
- **Homoskedasticity**
  error has the same variance given any value of the explanatory variables

OLS estimators are unbiased

OLS estimators are BLUE (Gauss-Markov Theorem)

Fitted values and residuals

OLS strategy

- Finding the β-vector that minimizes the sum of squared residuals (SSR)

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\hat{u}_i^2 \quad \rightarrow \quad \text{min!}$$

Goodness-of-fit: Mechanics

- Total sum of squares: SST
  (Σ squared deviations of y from the sample mean)
- Explained sum of squares: SSE
  (Σ squared deviations of yhat from the sample mean)
- Residual sum of squares: SSR
  (Σ squared residuals), minimized by OLS
- SST = SSE + SSR
- R^2 = SSE/SST = 1 – SSR/SST
  (coefficient of determination)

R^2 = square of the correlation coefficient between y and yhat

Goodness-of-fit: Interpretation

- R^2 is the proportion of the sample variation in the dependent variable explained by the independent variables
- R^2 never decreases when *any* variable is added to a regression
  - ⇨ makes it a poor tool for deciding whether a particular variable should be added to a model
  - ⇨ R^2 is no goddess of fit (especially in time series analysis)!

Adjusted R-squared (corrected R-squared)

$$\bar{R}^2 = 1 - \frac{\frac{SSR}{n-k-1}}{\frac{SST}{n-1}} = \bar{R}^2 = 1 - \frac{SSR}{SST}\frac{n-1}{n-k-1}$$

- Penalizes the number of regressors (= loss of degrees of freedom)
- Increases when t-statistic (F-statistic) of a single (group of) variable(s) is greater than 1

Interpreting the slope coefficients

- Simple (bivariate) regression

$$\hat{\beta}_1 = \frac{Cov(x,y)}{Var(x)} = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \mu_x)u_i}{\sum_{i=1}^{n}(x_i - \mu_x)^2} = \beta_1 + \frac{Cov(x,u)}{Var(x)}$$

- Multiple (multivariate) regression

$$\hat{\beta}_j = \frac{\sum_{i=1}^{n} r_{ij} \cdot y_i}{\sum_{i=1}^{n} r_{ij}^2} = \frac{Cov(r_j,y)}{Var(r_j)}$$

multicollinearity

⇨ partialling-out effect

⇨ omitted-variable bias

Variance of the slope coefficients

- Simple regression

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \mu_x)^2}$$

- Multiple regression

$$\mathrm{Var}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_{ij} - \mu_{x_j})^2 \cdot (1 - R_j^2)}$$

Sources of variance

(1) error variance

(2) sample variance in x_j

(3) multicollinearity

(4) small sample size

Estimating the error variance

- Estimated error variance

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2 = \frac{SSR}{n-k-1}$$

  - k = number of regressors
  - n – k – 1 = degrees of freedom
- Standard error of the regression (SER)

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

  - root squared error
  - standard error of the estimate

Misspecification

- Overspecifying the model (including an irrelevant variable)
  - no effect on unbiasedness of OLS
  - multicollinearity increases the variances of the remaining OLS estimators
  - consumes degrees of freedom
- Underspecifying the model (excluding a relevant variable)
  - causes OLS to be biased if linearily correlated with the remaining independent variables
  - multicollinearity might decrease the variances of the remaining OLS estimators (bias vs. variability tradeoff)

Inference

- Hypothesis testing and confidence intervals depend on the variances of OLS estimators
- Error variance affects the variances of the OLS estimators
- Case 1: Classical Linear Model
  - Gauss-Markov + Normality assumption
  - Normality assumption: population error is normally distributed with zero mean and (constant!) variance σ^2

  ⇨ exact sampling distributions of the OLS estimators
- Case 2: OLS asymptotics
  - Gauss-Markov + large sample size
  - properties emerge as the sample size grows without bound

  ⇨ asymptotic properties of the OLS estimators (as in case 1)

CLM: Pro and cons

- Pro
  - Central Limit Theorem: many unobserved variables, each having a minor effect on the dependent variable have an aggregated average effect that is normally distributed
- Cons
  - CLM captures additive errors only
  - discrete values cannot be normally distributed
  - many economic variables are non-negative (but: often [logarithmic] transformations might restore normality)

Tests (overview)

- t-Test (and confidence intervals)
  - single population parameter
- F-Test
  - group of population parameters
- LM-Test
  - group of population parameters (asymptotic analysis)
- RESET Test
  - functional form
- Davidson-MacKinnon test
  - functional form for nonnested alternatives

The t-Test

- Testing hypotheses about a single population parameter (usually testing for $\beta = 0$)
- General setting (t statistic or t ratio)

$$t = \frac{\text{estimate} - \text{hypothesized value}}{\text{standard error}}$$

How many standard deviations is the estimated value away from the assumed (= tested) value?

- Regression parameters are („asymptotically") t-distributed with df = n–k–1

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

The t-Test: Rejection rules

- Two-sided test (H_1: $\beta \neq$ hypothesized value)

  Reject H_0 if: $|t| > t_c$

- One sided test (H_1: $\beta <$ hypothesized value)

  Reject H_0 if: $t < -t_c$

- One sided test (H_1: $\beta >$ hypothesized value)

  Recect H_0 if: $t > t_c$

- Alternative: Looking at respective p-values

Practical guidelines

- Check for statistical significance
- Check statisticially significant values for practical significance (magnitudes of the estimates); be careful about functional form and units of measurement
- Non-statistically significant values (at usual levels up to 10 %) might remain in the model if their economic influence in well-founded and if their magnitudes are important; p-values as large as 20 % might be acceptable in such cases
- Statistically insignificant variables whose parameters have the „wrong“ sign can be ignored
- Statistically significant variables with „wrong“ signs and a practically large effect indicate misspecification

Confidence intervals

- Regression parameters are („asymptotically") t-distributed with n–k–1 degrees of freedom

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

- Example: 95% confidence interval

$$\hat{\beta}_j \pm c \cdot se(\hat{\beta}_j)$$

c = 97,5th percentile in a t_{n-k-1} distribution

- Rule of thumb (df = n –k–1 > 50): c = 2

The F-Test

- Testing q multiple linear restrictions simultaneously (joint statistical significance)
  - unrestricted model: contains all independent variables
  - restricted model: contains q independent variables less than the unrestricted model
- Example for $k \geq 2$
  - $H_0 : \beta_1 = \beta_2 = 0$
  - $H_1 : H_0$ is not true
- Ratio of SSR_r and SSR_{ur} is F-distributed with $df_1 = q$ and $df_2 = n-k-1$

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/n-k-1}$$

The F-Test: Rejection rule

- Reject H_0 if: $F > c_\alpha$
- c depends on
  - nominator degrees of freedom (df_1)
  - denominator degrees of freedom (df_2)
  - signficance level α
- Alternative: Looking at p-value
- Remarks
  - Note: F-Test tests for joint statistical significance, i.e. at least one (but not necessarily all) of the restricted variables is (are) statistically significant
  - F-test for a single variable is equivalent to a two-sided t-test

The LM-Test (Lagrange-Multiplier Test)

- Step1:
  Estimate the restricted model (with q restrictions) and save the residuals u_r
- Step 2:
  Regress u_r on all of the independent variables and obtain the R^2 as UR^2
- Step 3:
  Compute $LM = n \cdot UR^2$
- Step 4:
  LM follows a Chi-Square distribution with df = q; reject H_0 if LM > c (alternatively, look at p-values)

The RESET Test

- RESET = regression specification error test
- Tests for functional form misspecification
  - not a general test for misspecification (i.e. linearly dependent omitted variables cannot be detected)
  - if functional form is properly specified, heteroscedasticity is not detected
- Strategy:
  - Add p polynomials in the OLS fitted values to the original (= tested) estimation equation (here: p = 2):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + e$$

  - F-test for signficance of the δ-parameters; test statistic is $F_{p,n-k-1-p}$ distributed

Tests against nonnested alternatives

- Strategy 1: Comprehensive model approach
  - construct a comprehensive model that contains each model as a special case
  - testing the restrictions that lead to each of the models via F-tests
- Strategy 2: Davidson-MacKinnon test
  - estimate each model seperately
  - check, whether the fitted values of alternative 1 are significant when added as a regressor in alternative 2 and v.v.
- Problems
  - a clear winner need not emerge (if none of the special models can be rejected, use adjusted R-squared as creterion)
  - only relative performance is tested, none of the alternatives needs to be the correct model

Model selection criteria

- Nested models
  - t-Tests for significance of a single variable
  - F-Tests for joint significance of a group of variables
- Nonnested models
  - Davidson-MacKinnon + adjusted R-squared
    (BUT: not to be used for functional form of the dependent variable!)
  - Akaike Information Criterion (AIC)
    $AIC = n \cdot \ln(SSR) + 2(k+1)$
  - Schwartz Baysian Criterion (SBC)
    $SBC = n \cdot \ln(SSR) + (k+1) \cdot \ln(n)$

  } smaller value is prefered
  (different implementations exist)
- General rule: Parsimony is buitiful

Functional forms involving logarithms

- **level-level model: regressing y on x**
  $\Delta y = \beta_j \Delta x_j$

- **level-log model: regressing y on log(x)**
  $\Delta y = (\beta_j / 100)\% \Delta x_j$

- **log-level model: regressing log(y) on x**
  $\%\Delta y = (100\beta_j)\Delta x_j \Rightarrow 100\beta_j$ = semi-elasticity

- **log-log model: regressing log(y) on log(x)**
  $\%\Delta y = \beta_j \%\Delta x_j \Rightarrow \beta_j$ = elasticity

Rules of thumb for using logarithms

- Strictly positive variables often tend to be heteroskedastic or skewed ⇨ taking logs often mitigates/eliminates these problems
- Taking logs narrows the range of the variable ⇨ makes them less sensitive to outlying observations
- Taking logs works for strictly positive variables only zero observations in y ⇨ log(1+y) may work
- Positive dollar amount or large integers ⇨ try logs
- Variables that are measures in years ⇨ try levels
- Variables that are proportions ⇨ try rather levels

Functional form involving quadratic terms

- Can capture increasing or diminishing marginal effects …
- … but might also indicate functional form misspecification (e.g. levels instead of logs or vice versa)
- Note: Marginal effects are no longer constant, i.e. they depend on the value of the respective variable

Functional form involving dummy variables

- Capture qualitative information
- g different groups ⇨ g–1 dummies
- Stand-alone dummies for group-specific intercepts
- Interaction terms for group-specific slope parameters
- BUT: Each observation is somewhat unique
  - risk of over-dummying the model
  - ⇨ each dummy must have an economically justified interpretation

Units of measurements

- No effect
  - on significance of parameters
  - on goodness-of-fit
- Reflected in the magnitudes of the regression parameters
- Special case: log(y)-models
  - ⇨ nothing happens to the regression parameters if the units of measurement of the dependent variable are changed

Heteroskedasticity

- Does not cause bias or inconsistency in OLS estimators
- BUT: The usual standard errors and test statistics are no longer valid (OLS estimators are no longer BLUE)
- Tests: Regressing the squared OLS residuals …
  - … on the independent variables (Breusch-Pagan)
  - … on the independent variables plus their squares and all cross products (White)
  - … on the fitted and squared fitted values (special White)
- Solution
  - Weighted least squares
  - constructing heteroskedasticity-robust statistics