GMM, Indirect Inference and Bootstrap

Prerequisites: Probability theory, statistical inference, multiple linear regression

Willi Mutschler

TU Dortmund

Winter 2015/2016

Random experiments and probability

- Random experiments
  (sample space, outcomes, events and their combinations)
- Probabilities
  (definition of probability)
- Conditional probability and independence
  (conditional probability, total probability, Bayes' formula, independence)

Random variables and distributions

- Basics
  (cumulative distribution function, quantile function, discrete random variables, continuous random variables, linear transformations)
- Parameters of distributions
  (expectation, variance, moments)
- Important continuous distributions
  (normal or Gaussian, exponential, log-normal, χ^2, t, F)

Prerequisites

Probability theory and statistical inference

Joint distributions and limit theorems

- Joint distribution of two (or n) random variables
- Limit theorems
  (law of large numbers, central limit theorem)

Random samples and statistics

- Random samples
- Statistics
- Statistics of normally distributed samples

Estimation

- Point estimation
  (unbiasedness, consistency, estimation of expectations, estimation of probabilities and proportions, estimation of variances, of quantiles, and of correlation coefficients)
- Interval estimation

Hypothesis testing

- Basics (null and alternative hypothesis, test statistic, critical area, error of the first and second kind, significance level, power, p-value)
- Tests for expectations
  (test for a single expectation, comparison of two expectations)
- Tests for probabilities and proportions

- Model specification

$$y_t = \alpha + \beta_1 x_{1t} + \ldots + \beta_K x_{Kt} + u_t \quad \text{for } t = 1, \ldots, T$$

- In matrix notation

$$\begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{K1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1T} & \ldots & x_{KT} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix}$$

or

$$y = X\beta + u$$

Model assumptions

- A1 There are no omitted and no redundant exogenous variables.
- A2 The relation between y and X is linear.
- A3 The parameters β are constant over all T observations.
- B1-4 Error terms
$$u \sim N\left(0, \sigma^2 I_T\right).$$
- C1 The matrix X is not stochastic.
- C2 No multicollinearity, $rank(X) = K + 1$.

Prerequisites

Multiple linear regression

- OLS estimators
$$\hat{\beta} = \left(X'X\right)^{-1} X'y$$
- Estimated model
$$\hat{y} = X\hat{\beta}$$
- Residuals
$$\hat{u} = y - \hat{y}$$
- Measure of fit
$$R^2 = \frac{\sum_t \left(\hat{y}_t - \bar{y}\right)^2}{\sum_t \left(y_t - \bar{y}\right)^2} = 1 - \frac{\sum_t \hat{u}^2}{\sum_t \left(y_t - \bar{y}\right)^2}$$

- Distribution of the OLS estimator

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

- Distribution of linear combination

$$r'\hat{\beta} \sim N\left(r'\beta, \sigma^2 r' \left(X'X\right)^{-1} r\right)$$

where r is a column vector

- The t test is used to test the hypotheses

$$\begin{aligned} H_0 &: \quad r'\beta = q \\ H_1 &: \quad r'\beta \neq q \end{aligned}$$

- The test statistic is

$$t = \frac{r'\hat{\beta} - q}{\sqrt{\sigma^2 r' \left(X'X\right)^{-1} r}}$$

- Reject H_0 if $|t| > c$ where c is the $(1 - \alpha/2)$-quantile of the t-distribution with $n - K - 1$ degrees of freedom

- The F test is used to test the hypotheses

$$\begin{aligned} H_0 &: \quad R\beta = q \\ H_1 &: \quad R\beta \neq q \end{aligned}$$

- The test statistic is

$$F = \frac{\left(R\hat{\beta} - q\right)' \left[R\left(X'X\right)^{-1} R'\right]^{-1} \left(R\hat{\beta} - q\right)/L}{\hat{u}'\hat{u}/(T-K-1)}$$

- Reject H_0 if $F > c$ where c is the $(1-\alpha)$-quantile of the F-distribution with L and $T-K-1$ degrees of freedom

- Generalized Least Squares assumption about the error term vector

$$u \sim N\left(0, \Omega\right)$$

- The GLS estimator

$$\hat{\beta}^{GLS} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}y$$

  is efficient
- GLS is a unifying framework for autocorrelation and heteroscedasticity

- Stochastic exogenous variables
- Case 1: The error term u is independent of X
- Case 2: The error term and the exogenous variables are contemporaneously uncorrelated, i.e. for $t = 1, \dots, T$

$$Cov(X_{kt}, u_t) = 0$$

- Case 3: The error term and the exogenous variables are contemporaneously correlated