Econometrics

Willi Mutschler

Winter 2017/2018 and Summer 2018

1 Introduction

See slides Slides 1-16

2 Review of basic statistics

- Random experiment (Zufallsexperiment)
- Sample space (Ergebnismenge)
- Event (Ereignis)
- Set operations (Verknüpfungen von Ereignissen)
- Partition (Partition oder vollständige Zerlegung)
- Probability (Wahrscheinlichkeit)
- Kolmogorov's axioms (Kolmogorovs Axiome)
- Conditional probability (bedingte Wahrscheinlichkeit)
- Total probability (Satz von der totalen Wahrscheinlichkeit)
- Bayes' theorem (Satz von Bayes)
- Independence (Unabhängigkeit)
- Random variables (Zufallsvariable)
  - Definition and intuition
  - Distribution function and quantile function (Verteilungsfunktion und Quantilfunktion)
  - Discrete and continuous random variables (diskrete und stetige Zufallsvariable)
  - Density function (Dichtefunktion)
  - Expectation (Erwartungswert)
  - Variance (Varianz)

- Special discrete distributions, e.g. Bernoulli, binomial, Poisson, geometric, hypergeometric, ...
- Special continuous distributions, e.g. normal, standard normal distribution, exponential, Pareto, χ^2,F, t, ...
- There are many more special distributions
- Which distribution can be used when?

Let X and Y be both (jointly) normally distributed: $X_i \sim N(\mu_x, \sigma_x^2)$ and $Y_i \sim N(\mu_y, \sigma_y^2)$, $corr(X,Y) = \rho$, then the joint density is

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)} exp\left\{-\frac{-\tilde{x}^2 - 2\rho\tilde{x}\tilde{y} + \tilde{y}^2}{2(1-\rho)^2}\right\}$$

where $\tilde{x} = (x-\mu_x)/\sigma_x$ and $\tilde{y} = (y-\mu_y)/\sigma_y$.

The conditional expectation is

$$E(Y|X) = \int_{-\infty}^{\infty} \underbrace{\frac{f(X,y)}{f(X)}}_{f(y|X)} y\ dy = \left(\mu_y - \mu_x\rho\frac{\sigma_y}{\sigma_x}\right) + \left(\rho\frac{\sigma_y}{\sigma_x}\right) X$$

.

The conditional variance is:

$$Var(Y|X) = \int_{-\infty}^{\infty} f(y|X)(y - E(Y|X))^2\ dy = \sigma_y^2(1-\rho^2)$$

.

What is the distribution of Y conditional on X?

$$Y|X \sim N(\alpha + \beta X, \sigma^2)$$

where $\alpha = \mu_y - \mu_x\rho\frac{\sigma_y}{\sigma_x}$ and $\beta = \rho\frac{\sigma_y}{\sigma_x}$ and $\sigma^2 = \sigma_y^2(1-\rho^2)$. Alternatively: $Y = \underbrace{\alpha + \beta X}_{E(Y|X)} + u$ with $u \sim N(0, \sigma^2)$.

3 Simple linear regression model

3.1 Specification

Tip example Example: Guests of a restaurant leave the waiter with more or less large amounts of tips.Since the waiter is always friendly to all guests he can't explain the difference in tips. The waiter collects data for 20 guests in a table. Slides 20-22

Economic Reasoning: Since the waiter treats all his guests in the same way, it can be assumed that the difference in the amount of gratuities y is essentially declared by the amount of the bill x (measured in euros):

Economic model: functional dependence (in general)

$$y = f(x)$$

and particularly

$$y = \alpha + \beta x$$

Focus on linear relationships. Slide with true relationship. α is intercept, β is slope of the line. R is the true linear relationship, we want to be as close as possible to it.

Now our work as econometrician begins! Econometric model:

$$y_t = \alpha + \beta x_t + u_t$$

y: endogenous variable, x is exogenous variable, u is error, α and β are true coefficients. Even though later on we will study the effects of several variables on another variable, right now we start with the simple linear regression model, i.e. only one exogenous and one endogenous variable. We focus on observed values, hence the subindex $t = 1, ..., T$, α and β are the same for all persons (assumption which we could relax). In Reality we won't see an exact relationship due to randomness, different moods of people, exogenous factors, therefore we add an error term u. Slide with Econometric model.

Side note: Other functional forms are of course possible:

- Cobb-Douglas Production Function: $Y_t = \alpha L_t^{\beta} K_t^{\gamma} e^{u_t}$, take logs, then you get linear form. Of interest: is $\beta + \gamma = 1$
- logarithmic: $Y_t = a(X_t)^{\beta} e^{u_t}$ which may be rewritten $log(Y_t) = log(a) + \beta log(X_t) + u_t$ Then β is elasticity:$\beta = \frac{\partial Y_t / Y_t}{\partial X_t / X_t}$ very usefull as this is without dimension:

  $$\frac{\partial log(Y_t)}{\partial log(X_t)} = \frac{\partial log(Y_t)}{\partial y_t} \frac{\partial Y_t}{\partial X_t} \frac{\partial X_t}{\partial log(X_t)} = \frac{1}{Y_t} \partial Y_t \frac{X_t}{\partial X_t} = \frac{\partial Y}{\partial X_t} \frac{X_t}{Y_t} = \beta$$

  If X is raised by 1% then y changes by β %.
- quadratic (or polynomials) $Y_t = \alpha + \beta_1 X_t + \beta_2 X_t^2 + u_t$

The econometric model is specified using the A-, B- and C-assumptions

Functional specification (A-Assumptions)

Assumption a1: No relevant exogenous variable is omitted from the econometric model, and the exogenous variable included in the model is relevant [here: what about quality of the meal, we will consider these cases later on]

Assumption a2: The true functional dependence between x_t and y_t is linear

Assumption a3: The parameters α and β are constant for all T observations (x_t, y_t)

Error term specification (B-assumptions) Why do we need an error term:

- unsystematic measurement errors due to proxy variables
- some exogenous variables are not considered in the economic model since they are not observable
- human behavior is rather unpredictable and humans make errors

So the error term is supposed to fluctuate around 0, its average value. u is a random variable. Thougth experiment: Consider an infinite amount of samples of x and therefore y and u. Look at the distribution of u.

Assumption b1: $E(u_t) = 0$ for $t = 1, \ldots, T$ [no systematic measurement errors, very important, we loose efficiency]

Assumption b2: Homoskedasticity: $Var(u_t) = \sigma^2$ for $t = 1, \ldots, T$ [not so important, we can deal with heteroscedasticity]

Assumption b3: For all $t \neq s$ with $t = 1, 2, ..., T$ and $s = 1, 2, .., T$ we have

$$Cov(u_t, u_s) = 0$$

[no autocorrelation, if one persons gives a rather large tip and the person sitting at the next table follows that example, we have autocorrelation, we loose efficiency]

Assumption b4: The error terms u_t are (jointly) normally distributed.

b1 to b3: white noise assumptions

Compact notation of all B-assumptions:

b1-b4: $u_t \sim NID(0, \sigma^2)$ for $t = 1, \ldots, T$

Slide 23

Actually we have 20 of such distributions. Point A actually observed, u3 residual; expectation of all us must lie on R, have the same variance, bell shaped

Variable specification (C-assumptions) :

Assumption c1 The exogenous variable x_t is not stochastic, but can be controlled as in an experimental situation [highly unlikely, economic data is most often not from experimental situations, we will stick to this due to didactic reasons, later on we discuss what changes if we suppose that x is a random variable but not correlated with u]

Assumption c2 The exogenous variable x_t is not constant for all observations t: $S_{xx} > 0$ [we need some variation in x to explain the variation in y, this assumption also implies that we need at least two observations (which we also need to estimate two parameters)]

Of course, many (or even all?) of the A-, B-, and C-assumptions are restrictive and unrealistic :

- Number of exogenous variables, stochastic nature of variables
- Dynamics in variables, considering systems of equations
- Linear form
- Deterministic or stochastic trends

We will nevertheless suppose they are satisfied for the time being, and consider their violations later on

3.2 Point estimation

The simple (two-variable) linear regression model is

$$y_t = \alpha + \beta x_t + u_t$$

Estimation: Compute estimated values $\hat{\alpha}$ and $\hat{\beta}$ Intuitively: try to minimize the absolute values of the residuals or alternatively minimize quadratic deviations. OLS.R

We need to distinguish between true (α,β) and estimated values:

$$\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$$

How can we estimate the coefficients? (values of coefficients)

Least squares method (LS) Residual (Residuum): Difference between the observed value y_t and the estimated (predicted) value $\hat{y}_t$:

$$\begin{aligned} \hat{u}_t &= y_t - \hat{y}_t \\ &= y_t - \hat{\alpha} - \hat{\beta} x_t \end{aligned}$$

Idea: Choose $\hat{\alpha}$ and $\hat{\beta}$ such, that the sum of squared residuals

$$S_{\hat{u}\hat{u}} = \sum_{t=1}^{T} \hat{u}_t^2 = \sum_{t=1}^{T} (y_t - \hat{\alpha} - \hat{\beta} x_t)^2$$

is minimized!

Derivation of the OLS estimators Necessary conditions: Differentiate - Set FOC to zero - Reorder Sufficient conditions: Check second-order derivatives

Differentiate

$$S_{\hat{u}\hat{u}} = \sum_{t=1}^{T} \left(y_t - \hat{\alpha} - \hat{\beta} x_t \right)^2$$

with respect to $\hat{\alpha}$ and $\hat{\beta}$:

$$\begin{aligned}
\frac{\partial S_{\hat{u}\hat{u}}}{\partial \hat{\alpha}} &= \sum_{t=1}^{T} 2\left(y_t - \hat{\alpha} - \hat{\beta} x_t\right)(-1) \\
\frac{\partial S_{\hat{u}\hat{u}}}{\partial \hat{\beta}} &= \sum_{t=1}^{T} 2\left(y_t - \hat{\alpha} - \hat{\beta} x_t\right)(-x_t)\,.
\end{aligned}$$

Setting the derivatives to zeros and reordering yields

$$\begin{aligned}
\sum_{t=1}^{T}\left(y_t - \hat{\alpha} - \hat{\beta} x_t\right) &= 0 \\
\sum_{t=1}^{T}\left(x_t y_t - x_t\hat{\alpha} - \hat{\beta} x_t^2\right) &= 0
\end{aligned}$$

or the *normal equations* (Normalgleichungen)

$$\hat{\alpha} T + \hat{\beta}\sum_{t=1}^{T} x_t = \sum_{t=1}^{T} y_t \tag{1}$$

$$\hat{\alpha}\sum_{t=1}^{T} x_t + \hat{\beta}\sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T} x_t y_t. \tag{2}$$

Different ways to solve this system of equations, e.g. in matrix form:

$$\begin{aligned}
\begin{pmatrix} T & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} x_t^2 \end{pmatrix}\begin{pmatrix}\hat{\alpha} \\ \hat{\beta}\end{pmatrix} &= \begin{pmatrix}\sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T} x_t y_t\end{pmatrix} \\
\begin{pmatrix}\hat{\alpha} \\ \hat{\beta}\end{pmatrix} &= \begin{pmatrix} T & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} x_t^2 \end{pmatrix}^{-1}\begin{pmatrix}\sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T} x_t y_t\end{pmatrix}
\end{aligned}$$

Compute the inverse of a 2×2 matrix is easy:

$$\begin{aligned}
\begin{pmatrix} T & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} x_t^2 \end{pmatrix}^{-1} &= \frac{1}{\det\begin{pmatrix} T & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} x_t^2 \end{pmatrix}}\begin{pmatrix}\sum_{t=1}^{T} x_t^2 & -\sum_{t=1}^{T} x_t \\ -\sum_{t=1}^{T} x_t & T\end{pmatrix} \\
&= \frac{1}{T\sum_{t=1}^{T} x_t^2 - \left(\sum_{t=1}^{T} x_t\right)^2}\begin{pmatrix}\sum_{t=1}^{T} x_t^2 & -\sum_{t=1}^{T} x_t \\ -\sum_{t=1}^{T} x_t & T\end{pmatrix}
\end{aligned}$$

Therefore:

$$\begin{pmatrix}\hat{\alpha} \\ \hat{\beta}\end{pmatrix} = = \frac{1}{T\sum_{t=1}^{T} x_t^2 - \left(\sum_{t=1}^{T} x_t\right)^2}\begin{pmatrix}\sum_{t=1}^{T} x_t^2 & -\sum_{t=1}^{T} x_t \\ -\sum_{t=1}^{T} x_t & T\end{pmatrix}\begin{pmatrix}\sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T} x_t y_t\end{pmatrix}$$

Let's focus on $\hat{\beta}$:

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{t=1}^{T} x_t y_t - \frac{1}{T}\sum_{t=1}^{T} x_t \sum_{t=1}^{T} y_t}{\sum_{t=1}^{T} x_t^2 - \frac{1}{T}\left(\sum_{t=1}^{T} x_t\right)^2} \\ &= \frac{\sum_{t=1}^{T}\left(x_t-\bar{x}\right)\left(y_t-\bar{y}\right)}{\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2} \\ &= \frac{S_{xy}}{S_{xx}}. \qquad (3)\end{aligned}$$

Divide the first normal equation by T, then

$$\hat{\alpha} + \hat{\beta}\bar{x} = \bar{y}$$

or

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}. \qquad (4)$$

$$\begin{aligned}\hat{\beta} &= S_{xy}/S_{xx} \\ \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x}\end{aligned}$$

with

$$\begin{aligned}S_{xx} &= \sum\left(x_t-\bar{x}\right)^2 = \sum x_t^2 - T\bar{x}^2 \\ S_{xy} &= \sum\left(x_t-\bar{x}\right)\left(y_t-\bar{y}\right) = \sum y_t x_t - T\bar{x}\bar{y}\end{aligned}$$

These are variations, not variances or covariances as we are not dividing by $T-1$.

Interpretation of the estimators or estimates? Do for our example! OLS.R

Numeric illustration for a three points example

t	$\mathbf{x_t}$	$\mathbf{y_t}$
1	10	2
2	30	3
3	50	7

Calculate

$$\begin{aligned}&\hat{\alpha}, \hat{\beta} \\ &\hat{y}_1, \hat{y}_2, \hat{y}_3 \\ &\hat{u}_1, \hat{u}_2, \hat{u}_3 \\ &S_{\hat{u}}, S_{\hat{u}\hat{u}}\end{aligned}$$

Compute $\bar{x} = 30$ and $\bar{y} = 4$ and

$$\begin{aligned}S_{xx} &= \sum\left(x_t-\bar{x}\right)^2 = 400 + 0 + 400 = 800 \\ S_{xy} &\equiv \sum\left(x_t-\bar{x}\right)\left(y_t-\bar{y}\right) \\ &= (-20)\cdot(-2) + 0\cdot 1 + 20\cdot 3 = 100\end{aligned}$$

Estimates (Schätzwerte)

$$\begin{aligned}\hat{\beta} &= S_{xy}/S_{xx} = 100/800 = 0.125 \\ \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} = 4 - 0.125 \cdot 30 = 0.25\end{aligned}$$

With each additional Euro on the bill, the waiter receives an additional tip of 0.125 Euro. $\hat{\alpha}$ is a level parameter, mostly no good interpretation (no bill, but still 0.25 Euros tip does not make sense). Always stick to the interval where you have observations!

Estimated model

$$\hat{y}_t = 0.25 + 0.125 \cdot x_t$$

If the billing amounts are $x_1 = 10$, $x_2 = 30$ and $x_3 = 50$ we predict the following gratuities

$$\begin{aligned}\hat{y}_1 &= 0.25 + 0.125 \cdot 10 = 1.5 \\ \hat{y}_2 &= 0.25 + 0.125 \cdot 30 = 4 \\ \hat{y}_3 &= 0.25 + 0.125 \cdot 50 = 6.5\end{aligned}$$

Estimated error terms (residuals)

$$\begin{aligned}\hat{u}_1 &= y_1 - \hat{y}_1 = 2 - 1.5 = 0.5 \\ \hat{u}_2 &= y_2 - \hat{y}_2 = 3 - 4 = -1 \\ \hat{u}_3 &= y_1 - \hat{y}_1 = 7 - 6.5 = 0.5\end{aligned}$$

Sum of residuals: $S_{\hat{u}} = 0.5 - 1 + 0.5 = 0$. (This is always the case!)

Sum of squared residuals

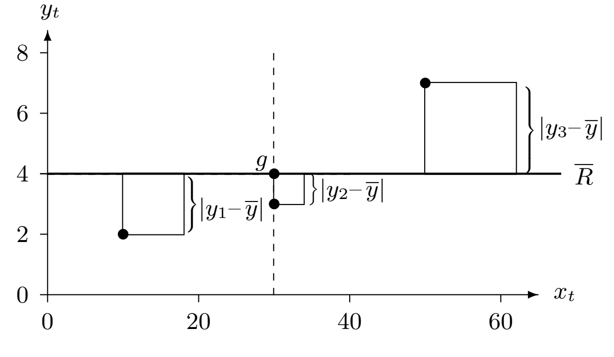
$$S_{\hat{u}\hat{u}} = \sum \hat{u}_t^2 = 0.5^2 + (-1)^2 + 0.5^2 = 1.5$$

3.3 The coefficient of determination R^2

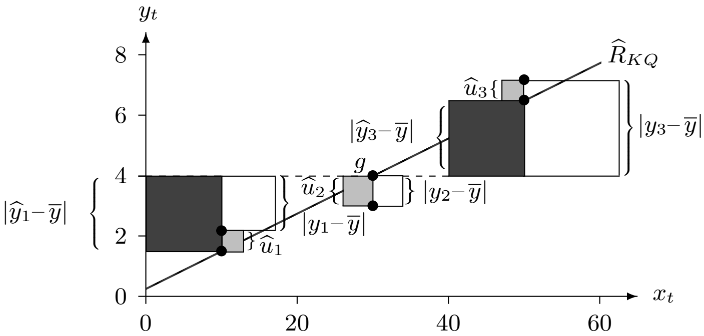
Variation of the endogenous variable

$$S_{yy} = S_{\hat{y}\hat{y}} + S_{\hat{u}\hat{u}}$$

$\bar{R} : \bar{y} = \hat{a} + \hat{\beta}\bar{x}$,
$g : (\bar{x}, \bar{y})$



We try to explain the variation in y_t through the variation in x_t. $\bar{R}$ is a bad choice since the variation in y_t is equal the variation in u_t, but not in x_t. R^{KQ} must go through g. We rotate the line such that $S_{\hat{u}\hat{u}}$



This shows that the entire variation S_{yy} can be broken down into two components: the explained variation $S_{\hat{y}\hat{y}}$ and the unexplained variation $S_{\hat{u}\hat{u}}$. Graphically, this means that the sum of dark grey and medium grey areas corresponds to the sum of the light grey areas. However, this connection does not apply to a single observation point!

We get the equation:

$$\sum (y_t - \bar{y})^2 = \sum (\hat{y}_t - \bar{y})^2 + \sum \hat{u}_t^2$$

Proof: Rewrite the normal equations:

$$\begin{aligned}\sum_{t=1}^{T} y_t &= \sum_{t=1}^{T} \hat{\alpha} + \hat{\beta} \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} x_t y_t &= \hat{\alpha} \sum_{t=1}^{T} x_t + \hat{\beta} \sum_{t=1}^{T} x_t^2\end{aligned}$$

or

$$\begin{aligned}\sum_{t=1}^{T} y_t - \sum_{t=1}^{T} \hat{\alpha} - \hat{\beta} \sum_{t=1}^{T} x_t &= 0 \\ \sum_{t=1}^{T} x_t y_t - \hat{\alpha} \sum_{t=1}^{T} x_t - \hat{\beta} \sum_{t=1}^{T} x_t^2 &= 0\end{aligned}$$

or

$$\begin{aligned}\sum_{t=1}^{T} \left(y_t - \hat{\alpha} - \hat{\beta} x_t \right) &= 0 \\ \sum_{t=1}^{T} x_t \left(y_t - \hat{\alpha} - \hat{\beta} x_t \right) &= 0.\end{aligned}$$

Because of $\hat{u}_t = y_t - \hat{\alpha} - \hat{\beta} x_t$ this equals

$$\begin{aligned}\sum_{t=1}^{T} \hat{u}_t &= 0 && (5)\\ \sum_{t=1}^{T} x_t \hat{u}_t &= 0. && (6)\end{aligned}$$

The decomposition is now easy to show:

$$\begin{aligned}\sum (y_t - \bar{y})^2 &= \sum (y_t - \hat{y}_t + \hat{y}_t - \bar{y})^2 \\ &= \sum \hat{u}_t^2 + \sum (\hat{y}_t - \bar{y})^2 \\ &\quad + \sum 2\hat{u}_t (\hat{y}_t - \bar{y}).\end{aligned}$$

The last term is

$$\begin{aligned}\sum 2\hat{u}_t (\hat{y}_t - \bar{y}) &= \sum 2\hat{u}_t \left(\hat{\alpha} + \hat{\beta} x_t - \bar{y}\right) \\ &= 2\hat{\alpha} \sum \hat{u}_t + 2\hat{\beta} \sum \hat{u}_t x_t - 2\bar{y} \sum \hat{u}_t.\end{aligned}$$

Due to (5) and (6) this term equals zero. Hence, the decomposition is proved.

Coefficient of determination (Bestimmtheitsmaß) R^2

$$R^2 = \frac{\text{‘explained variation”}}{\text{“total variation”}} = \frac{S_{yy} - S_{\hat{u}\hat{u}}}{S_{yy}} = \frac{S_{\hat{y}\hat{y}}}{S_{yy}}$$

Different way of computing R^2:

$$R^2 = \frac{\hat{\beta} S_{xy}}{S_{yy}} = \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

For our example with the three points: We already know that

$$\begin{aligned}S_{xx} &= \sum (x_t - \bar{x})^2 = 800 \\ S_{xy} &= \sum (x_t - \bar{x})(y_t - \bar{y}) = 100\end{aligned}$$

Now compute

$$S_{yy} = \sum (y_t - \bar{y})^2 = 14$$

Calculate

$$\begin{aligned}R^2 &= \frac{S_{xy}^2}{S_{xx} S_{yy}} \\ &= \frac{100^2}{800 \cdot 14} \approx 89.3\%\end{aligned}$$

or

$$\begin{aligned}R^2 &= \frac{S_{yy} - S_{\hat{u}\hat{u}}}{S_{yy}} \\ &= \frac{14 - 1.5}{14} \approx 89.3\%\end{aligned}$$

Note that in the simple linear regression model: $\widehat{corr}(x,y) = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \pm\sqrt{\frac{S_{xx}^2}{S_{xx}S_{yy}}} = \pm\sqrt{R^2}$

Note also that in models without an intercept the coefficient of determination could be negative.

4 Properties of the estimators

The estimates

$$\begin{aligned}\hat{\beta} &= S_{xy}/S_{xx} \\ \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x}\end{aligned}$$

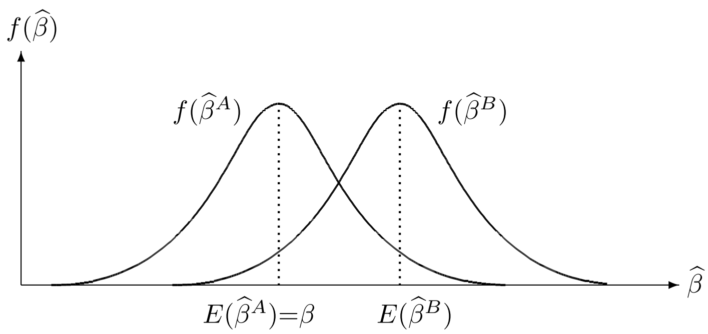
are random variables.

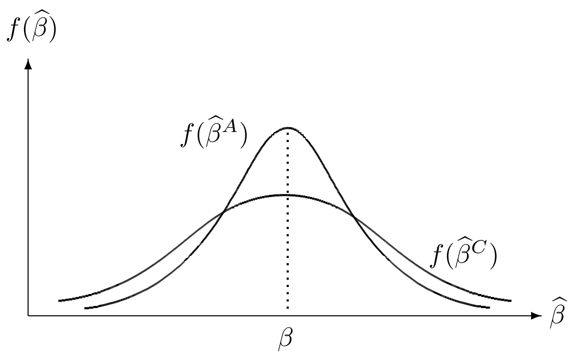
Thought experiment: repeated samples

Computer simulation experiment.R

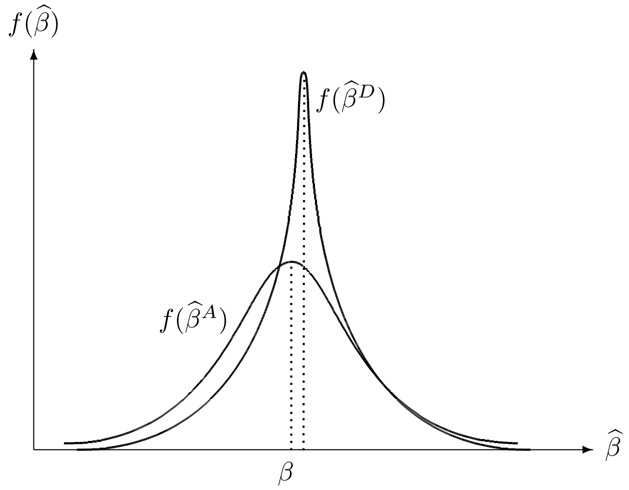
In order to compare different estimators we need criteria for „good" estimators:

- Unbiasedness: An estimator $\hat{\beta}^A$ is unbiased if given an infinite amount of repeated samples, the obtained estimates $\hat{\beta}^A$ are on average equal to the true value of β, i.e. $E(\hat{\beta}^A) = \beta$.
- Efficiency: An estimator $\hat{\beta}^A$ is efficient, if it is the one with the smallest variance $var(\hat{\beta}^A)$ **within** the class of unbiased estimators.





Efficiency is subordinated criteria. We focus on unbiased estimators first, then we compare these. What about a slightly biased estimator (of which we even can compute the bias) which has a much less variance?



This question cannot be answered in general, it depends on the focus of the researcher.

Furthermore, what about considering α and β jointly? We will come back to this issue in the multivariate case.

Expectation of the estimators

Unbiasedness: Under the a-, b- and c-assumptions (without b2 and b4)

$$\begin{aligned} E(\hat{\alpha}) &= \alpha \\ E(\hat{\beta}) &= \beta \end{aligned}$$

Proof: We will use the following equality at a later stage:

$$\sum (x_t - \bar{x})\, y_t = \sum (x_t - \bar{x})\,(y_t - \bar{y}),$$

because

$$\sum (x_t - \bar{x})\,(y_t - \bar{y}) = \sum x_t y_t - \bar{x} \sum y_t - \underbrace{\bar{y} \sum x_t}_{=T\bar{x}\bar{y}} + \underbrace{\sum \bar{x}\bar{y}}_{=T\bar{x}\bar{y}}$$

The expectation is easy to calculate. Since

$$\begin{aligned}
\hat{\beta} &= \frac{S_{xy}}{S_{xx}} \\
&= \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2} \\
&= \frac{\sum (x_t - \bar{x})(\alpha + \beta x_t + u_t - \alpha - \beta\bar{x} - \bar{u})}{\sum (x_t - \bar{x})^2} \\
&= \frac{\sum (x_t - \bar{x})(\beta (x_t - \bar{x}) + (u_t - \bar{u}))}{\sum (x_t - \bar{x})^2} \\
&= \beta + \frac{\sum (x_t - \bar{x})}{\sum (x_t - \bar{x})^2} (u_t - \bar{u}) \qquad (7)
\end{aligned}$$

we find

$$\begin{aligned}
E(\hat{\beta}) &= E\left(\beta + \frac{\sum (x_t - \bar{x})}{\sum (x_t - \bar{x})^2} (u_t - \bar{u})\right) \\
&= \beta + \frac{\sum (x_t - \bar{x})}{\sum (x_t - \bar{x})^2} E(u_t - \bar{u}) \\
&= \beta.
\end{aligned}$$

To find the expectation of $\hat{\alpha}$ we use the following equations

$$\begin{aligned}
\bar{y} &= \hat{\alpha} + \hat{\beta}\bar{x} \\
\bar{y} &= \alpha + \beta\bar{x} + \bar{u}.
\end{aligned}$$

Substract:

$$\begin{aligned}
0 &= \hat{\alpha} - \alpha + \left(\hat{\beta} - \beta\right)\bar{x} - \bar{u} \\
\hat{\alpha} &= \alpha - (\hat{\beta} - \beta)\bar{x} + \bar{u}. \qquad (8)
\end{aligned}$$

Thus,

$$\begin{aligned}
E(\hat{\alpha}) &= \alpha + E(\hat{\beta} - \beta)\bar{x} + E(\bar{u}) \\
&= \alpha.
\end{aligned}$$

Variance and covariance of the estimators

The variances and covariances are given by

$$\begin{aligned}
Cov(\hat{\alpha}, \hat{\beta}) &= -\sigma^2 (\bar{x}/S_{xx}) \\
Var(\hat{\alpha}) &= \sigma^2 \left(1/T + \bar{x}^2/S_{xx}\right) \\
Var(\hat{\beta}) &= \sigma^2/S_{xx}
\end{aligned}$$

The variance is smaller if (i) the variance of the error term is smaller and (ii) the larger the variation in x, i.e. S_{xx}. S_{xx} is higher the more observations we have, therefore the more observations the more efficient the estimators.

Proof:

$$
\begin{aligned}
\hat{\beta} &= \frac{S_{xy}}{S_{xx}} \\
&= \frac{\sum (x_t - \bar{x})\, y_t}{\sum (x_t - \bar{x})^2} \\
&= \frac{\sum (x_t - \bar{x})\,(\alpha + \beta x_t + u_t)}{\sum (x_t - \bar{x})^2} \\
&= \frac{\sum (x_t - \bar{x})\,(\alpha + \beta x_t)}{\sum (x_t - \bar{x})^2} + \frac{\sum (x_t - \bar{x})\, u_t}{\sum (x_t - \bar{x})^2}
\end{aligned}
$$

Hence

$$
\begin{aligned}
Var(\hat{\beta}) &= Var\left(\frac{\sum (x_t - \bar{x})\, u_t}{\sum (x_t - \bar{x})^2}\right) \\
&= \frac{Var\left(\sum (x_t - \bar{x})\, u_t\right)}{\left(\sum (x_t - \bar{x})^2\right)^2} \\
&= \frac{\sum Var\left((x_t - \bar{x})\, u_t\right)}{\left(\sum (x_t - \bar{x})^2\right)^2} \\
&= \frac{\sum (x_t - \bar{x})^2\, Var\,(u_t)}{\left(\sum (x_t - \bar{x})^2\right)^2} \\
&= Var\,(u_t)\, \frac{\sum (x_t - \bar{x})^2}{\left(\sum (x_t - \bar{x})^2\right)^2} \\
&= \frac{\sigma^2}{S_{xx}}.
\end{aligned}
$$

For the Variance of $\hat{\alpha}$ we make use of (8),

$$
\begin{aligned}
\hat{\alpha} &= \alpha - (\hat{\beta} - \beta)\bar{x} + \bar{u} \\
&= \alpha + \beta\bar{x} - \hat{\beta}\bar{x} + \bar{u}
\end{aligned}
$$

Dropping the constants, we get

$$
\begin{aligned}
Var\,(\hat{\alpha}) &= Var(\hat{\beta}\bar{x}) + Var(\bar{u}) - 2Cov(\hat{\beta}\bar{x}, \bar{u}) \\
&= \bar{x}^2 Var(\hat{\beta}) + Var(\bar{u}) - 2\bar{x}Cov(\hat{\beta}, \bar{u}).
\end{aligned}
$$

We will show shortly, that $Cov(\hat{\beta}, \bar{u}) = 0$, hence

$$
\begin{aligned}
Var\,(\hat{\alpha}) &= \bar{x}^2 Var(\hat{\beta}) + Var\left(\frac{1}{T}\sum u_t\right) \\
&= \bar{x}^2 \frac{\sigma^2}{S_{xx}} + \frac{1}{T^2}\sum \sigma^2 \\
&= \sigma^2 \left(1/T + \bar{x}^2/S_{xx}\right).
\end{aligned}
$$

To see, that the covariance vanishes, use (7),

$$\hat{\beta} = \beta + \frac{\sum (x_t - \bar{x})(u_t - \bar{u})}{\sum (x_t - \bar{x})^2}.$$

Since $E(\bar{u}) = 0$ the covariance is given by

$$\begin{aligned} Cov(\hat{\beta}, \bar{u}) &= E\left(\left(\hat{\beta} - \beta\right)\bar{u}\right) \\ &= E\left(\frac{\sum (x_t - \bar{x})(u_t - \bar{u})\,\bar{u}}{\sum (x_t - \bar{x})^2}\right) \\ &= \frac{\sum (x_t - \bar{x})\, E\left((u_t - \bar{u})\,\bar{u}\right)}{\sum (x_t - \bar{x})^2}. \end{aligned}$$

The expectation is

$$\begin{aligned} E\left((u_t - \bar{u})\,\bar{u}\right) &= E\left(u_t \bar{u}\right) - E\left(\bar{u}^2\right) \\ &= E\left(u_t \frac{1}{T}\sum_i u_i\right) - E\left(\frac{1}{T^2}\sum_i \sum_j u_i u_j\right) \\ &= \frac{1}{T}\sum_i E\left(u_t u_i\right) - \frac{1}{T^2}\sum_i \sum_j E\left(u_i u_j\right). \end{aligned}$$

Since $E(u_i u_j) = 0$ for $i \neq j$ and $E(u_i^2) = \sigma^2$ we find

$$\begin{aligned} E\left((u_t - \bar{u})\,\bar{u}\right) &= \frac{1}{T}\sum_i E\left(u_t u_i\right) - \frac{1}{T^2}\sum_i \sum_j E\left(u_i u_j\right) \\ &= \frac{1}{T}\sigma^2 - \frac{1}{T^2}\sum_i \sigma^2 \\ &= 0. \end{aligned}$$

This completes the derivation of the variance of $\hat{\alpha}$.

The covariance between $\hat{\alpha}$ and $\hat{\beta}$ is easy to derive. Using the definition of the covariance, unbiasedness and (8) we get

$$\begin{aligned} Cov(\hat{\alpha}, \hat{\beta}) &= E\left[(\hat{\alpha} - \alpha)\left(\hat{\beta} - \beta\right)\right] \\ &= E\left[\left(\bar{u} - \left(\hat{\beta} - \beta\right)\bar{x}\right)\left(\hat{\beta} - \beta\right)\right] \\ &= E\left[\bar{u}\left(\hat{\beta} - \beta\right)\right] - E\left[\left(\hat{\beta} - \beta\right)^2 \bar{x}\right] \\ &= -\bar{x} Var(\hat{\beta}) \\ &= -\frac{\bar{x}\sigma^2}{S_{xx}}. \end{aligned}$$

the first term vanishes because $Cov(\hat{\beta}, \bar{u}) = 0$ as has been shown above.

Distribution

We now know the expected values, variances and covariance of the estimators $\hat{\alpha}$ and $\hat{\beta}$. How are the estimators distributed? Which distribution family do they belong to?

Pre-consideration: Because of

$$y_t = \alpha + \beta x_t + u_t$$

the y_t are linear transformations (in fact just shifts) of the normally distributed error terms. Hence, y_t is also normally distributed. Due to the independence of the errors we get that y_t are also independent.

$E(y_t) = \alpha + \beta x_t$ and $var(y_t) = \sigma^2$. $y_t \sim NID(\alpha + \beta x_t, \sigma^2)$.

Because of

$$\begin{aligned} \hat{\beta} &= \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2} \\ &= \frac{\sum (x_t - \bar{x}) y_t}{\sum (x_t - \bar{x})^2} \\ &= \frac{\sum (x_t - \bar{x})}{\sum (x_t - \bar{x})^2} y_t \end{aligned}$$

we see that $\hat{\beta}$ is a linear transformation of the y_t. Linear transformations of (jointly) normally distributed random variables are also normally distributed. Thus, $\hat{\beta}$ is normally distributed.

A similar argument holds for $\hat{\alpha}$.

The normal distribution is uniquely identified by its expectation and variances.

The distribution of both estimators is thus:

$$\begin{aligned} \hat{\alpha} &\sim N\left(\alpha, \sigma^2 \left(\frac{1}{T} + \frac{\bar{x}^2}{S_{xx}}\right)\right) \\ \hat{\beta} &\sim N\left(\beta, \frac{\sigma^2}{S_{xx}}\right). \end{aligned}$$

BLUE-Property

LS estimator is a linear estimator, linear function in $\hat{\beta}$ and $\hat{\alpha}$. It is even the Best Linear Unbiased Estimator (without b4). With Normal distribution assumption (b4) we even get BUE, Best Unbiased Estimator (Cramer-Rao)

A linear estimator is a linear function of the y-values. Let $\tilde{\beta}$ denote an arbitrary linear unbiased estimator of β,

$$\tilde{\beta} = \sum_{t=1}^{T} c_t y_t.$$

Since the estimator is unbiased,

$$E(\tilde{\beta}) = E\left(\sum_{t=1}^{T} c_t y_t\right) = \beta,$$

i.e.

$$\begin{aligned}
\sum_{t=1}^{T} c_t E\left(y_t\right) &= \beta \\
\sum_{t=1}^{T} c_t\left(\alpha+\beta x_t\right) &= \beta \\
\alpha \sum_{t=1}^{T} c_t+\beta \sum_{t=1}^{T} c_t x_t &= \beta .
\end{aligned}$$

Therefore

$$\begin{aligned}
\sum_{t=1}^{T} c_t &= 0 \\
\sum_{t=1}^{T} c_t x_t &= 1 .
\end{aligned}$$

We now determine the values $c_1, \ldots, c_T$ that minimize the variance of the estimator $\tilde{\beta}$. The variance is

$$\begin{aligned}
Var(\tilde{\beta}) &= Var\left(\sum_{t=1}^{T} c_t y_t\right) \\
&= \sum_{t=1}^{T} c_t^2 Var\left(y_t\right) \\
&= \sigma^2 \sum_{t=1}^{T} c_t^2 .
\end{aligned}$$

This is a standard optimization problem with constraints. Obviously, σ^2 is immaterial and for simplicity we normalize it to $\sigma^2=1/2$. The method of Lagrange is

$$\frac{1}{2} \sum_{t=1}^{T} c_t^2 \longrightarrow \min_{c_1, \ldots, c_T}$$

subject to

$$\begin{aligned}
\sum_{t=1}^{T} c_t &= 0 \\
\sum_{t=1}^{T} c_t x_t &= 1 .
\end{aligned}$$

The Lagrange function is

$$L(c_1, \ldots, c_T, \lambda_1, \lambda_2)=\sum_{t=1}^{T} c_t^2-\lambda_1\left(\sum_{t=1}^{T} c_t\right)-\lambda_2\left(\left[\sum_{t=1}^{T} c_t x_t\right]-1\right) .$$

The derivatives are

$$\begin{aligned}
\frac{\partial L}{\partial c_t} &= c_t - \lambda_1 - \lambda_2 x_t, \quad t = 1, \ldots, T \\
\frac{\partial L}{\partial \lambda_1} &= -\sum_{t=1}^{T} c_t \\
\frac{\partial L}{\partial \lambda_2} &= \left(- \left[\sum_{t=1}^{T} c_t x_t \right] + 1 \right).
\end{aligned}$$

The first order conditions are

$$c_t - \lambda_1 - \lambda_2 x_t = 0, \quad t = 1, \ldots, T \tag{9}$$

$$\sum_{t=1}^{T} c_t = 0 \tag{10}$$

$$\sum_{t=1}^{T} c_t x_t = 1. \tag{11}$$

From (9) we find the optimal values

$$c_t = \lambda_1 + \lambda_2 x_t. \tag{12}$$

Next, we determine the Lagrange multipliers λ_1 and λ_2. Add all T equations (9) and use (10):

$$\underbrace{\sum_{t=1}^{T} c_t}_{=0} - T\lambda_1 - \lambda_2 \sum_{t=1}^{T} x_t = 0,$$

such that

$$T\lambda_1 + \lambda_2 \sum_{t=1}^{T} x_t = 0. \tag{13}$$

Now, multiply all T equations (9) with x_t and add them; using (11) we find

$$\underbrace{\sum_{t=1}^{T} c_t x_t}_{=1} - \lambda_1 \sum_{t=1}^{T} x_t - \lambda_2 \sum_{t=1}^{T} x_t^2 = 0,$$

such that

$$\lambda_1 \sum_{t=1}^{T} x_t + \lambda_2 \sum_{t=1}^{T} x_t^2 = 1. \tag{14}$$

The solution of the equation system (13) and (14) is

$$\begin{aligned}
\lambda_1 &= \frac{\begin{vmatrix} 0 & \sum_{t=1}^{T} x_t \\ 1 & \sum_{t=1}^{T} x_t^2 \end{vmatrix}}{\begin{vmatrix} T & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} x_t^2 \end{vmatrix}} = \frac{-\sum_{t=1}^{T} x_t}{T\sum_{t=1}^{T} x_t^2 - \left(\sum_{t=1}^{T} x_t\right)^2} \\
&= \frac{-T\bar{x}}{T\sum_{t=1}^{T} x_t^2 - (T\bar{x})^2} = \frac{-\bar{x}}{\sum_{t=1}^{T} x_t^2 - T\bar{x}^2} = \frac{-\bar{x}}{S_{xx}} \\
\lambda_2 &= \frac{\begin{vmatrix} T & 0 \\ \sum_{t=1}^{T} x_t & 1 \end{vmatrix}}{\begin{vmatrix} T & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} x_t^2 \end{vmatrix}} = \frac{T}{T\sum_{t=1}^{T} x_t^2 - \left(\sum_{t=1}^{T} x_t\right)^2} \\
&= \frac{1}{\sum_{t=1}^{T} x_t^2 - T\bar{x}^2} = \frac{1}{S_{xx}}.
\end{aligned}$$

Hence, we may rewrite (12) as

$$\begin{aligned}
c_t &= \lambda_1 + \lambda_2 x_t \\
&= \frac{-\bar{x}}{S_{xx}} + \frac{1}{S_{xx}} x_t \\
&= \frac{x_t - \bar{x}}{S_{xx}}.
\end{aligned}$$

Therefore, the optimal linear unbiased estimator is

$$\begin{aligned}
\hat{\beta} &= \sum_{t=1}^{T} c_t y_t \\
&= \sum_{t=1}^{T} \frac{x_t - \bar{x}}{S_{xx}} y_t \\
&= \frac{\sum_{t=1}^{T}(x_t - \bar{x}) y_t}{S_{xx}} \\
&= \frac{S_{xy}}{S_{xx}}.
\end{aligned}$$

5 Confidence intervals and interval estimation

Up to now, we only habe one numerical balue for $\hat{\alpha}$ and $\hat{\beta}$, i.e. a point estimator. How reliable is this? Again thought experiment: repeated samples, for fixed x we randomly draw new u and get a different y, recompute the point estimators. And lets do this for a very large number of times. All of our estimates should be close to the true value, but maybe some are not. Given those samples, let's quantify reliability. We already know that $\hat{\beta}$ is a random variable and

$$\hat{\beta} \sim N\left(\beta, \sigma^2/S_{xx}\right),$$

Instead of a point estimator $\hat{\beta}$ we now want an interval estimator

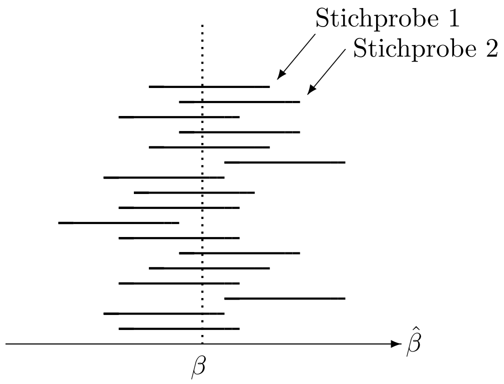
$$P(\beta - k \leq \hat{\beta} \leq \beta + k) = 1 - a.$$

That is a fixed ratio k of samples such that the interval should contain the true value in $1 - a\%$ of times.

Both inequalities within the probability can be transformed to

$$\begin{aligned} P(\beta - k \leq \hat{\beta} \leq \beta + k) &= 1 - a \\ P(-\hat{\beta} - k \leq -\beta \leq -\hat{\beta} + k) &= 1 - a \\ P(\hat{\beta} - k \leq \beta \leq \hat{\beta} + k) &= 1 - a. \end{aligned}$$

The interval $[\hat{\beta} - k\ ;\ \hat{\beta} + k]$ is called $(1 - a)$-confidence interval (Konfidenzintervall). a is called confidence level.



How to choose k?

Case 1: Confidence interval when σ^2 is known

- Step 1: Standardization of $\hat{\beta}$

$$\begin{aligned} se(\hat{\beta}) &= \sqrt{\sigma^2 / S_{xx}} \\ z &= \frac{\hat{\beta} - E(\hat{\beta})}{se(\hat{\beta})} \\ &= \frac{\hat{\beta} - \beta}{se(\hat{\beta})} \sim N\left(0, 1\right) \end{aligned}$$

- The random variable $z = (\hat{\beta} - \beta)/se(\hat{\beta})$ is a standardized normally distributed variable. Furthermore, it is a pivot (Pivot), i.e. its distribution does not depend on unknown parameters. It is a normed indicator for the deviations from the estimates from the true value.
  Side note: why go this way? Because we know everything about N(0,1) (it is a pivot), as we do not know the exact position (E()) and shape (var) of $\hat{\beta}$.

- Step 2: Find the $(1-\alpha/2)$-quantile $z_{a/2}$

$$P(-z_{a/2} \leq z \leq z_{a/2}) = 1 - a$$

Why $\alpha/2$: because we focus on a symmetric interval!

- Step 3: Substitute z by $(\hat{\beta} - \beta)/se(\hat{\beta})$

$$P\left(-z_{a/2} \leq \frac{\hat{\beta}-\beta}{se(\hat{\beta})} \leq z_{a/2}\right) = 1 - a$$

- Rewriting yields the $(1-a)$-interval

$$\begin{aligned}
P\left(-z_{a/2} \leq z \leq z_{a/2}\right) &= 1-a \\
P\left(-z_{a/2} \leq \frac{\hat{\beta}-\beta}{se(\hat{\beta})} \leq z_{a/2}\right) &= 1-a \\
P\left(-\hat{\beta} - z_{a/2} \cdot se(\hat{\beta}) \leq -\beta \leq -\hat{\beta} + z_{a/2} \cdot se(\hat{\beta})\right) &= 1-a \\
P\left(\hat{\beta} + z_{a/2} \cdot se(\hat{\beta}) \geq \beta \geq \hat{\beta} - z_{a/2} \cdot se(\hat{\beta})\right) &= 1-a \\
P\left(\hat{\beta} - z_{a/2} \cdot se(\hat{\beta}) \leq \beta \leq \hat{\beta} + z_{a/2} \cdot se(\hat{\beta})\right) &= 1-a.
\end{aligned}$$

Hence:

$$\left[\hat{\beta} - z_{a/2} \cdot se(\hat{\beta}) \ ; \ \hat{\beta} + z_{a/2} \cdot se(\hat{\beta})\right]$$

- Numerical example: The 0.975-quantile $z_{0.975}$ can be found in statistical tables or by using built-in computer functions, e.g. the R-command `qnorm(0.975): 1.959964`. Assume that $\sigma^2 = 2$. The standard error of $\hat{\beta}$ is

$$se(\hat{\beta}) = \sqrt{\sigma^2/S_{xx}} = \sqrt{2/800} = 0.05$$

Because of $\hat{\beta} = 0.125$ the interval estimator is

$$[0.125 - 1.96 \cdot 0.05; 0.125 + 1.96 \cdot 0.05] = [0.027; 0.223]\,.$$

Case 2: Confidence interval when σ^2 is unknown

- Step 1: Consistent and unbiased estimation of σ^2 and $se(\hat{\beta})$:

$$\hat{\sigma}^2 = \frac{1}{T-2}\sum_{t=1}^{T}\hat{u}_t^2$$
$$\widehat{se}(\hat{\beta}) = \sqrt{\hat{\sigma}^2/S_{xx}}$$

(we postpone the proofs). We divide by the degrees of freedom $T-2$, as we have T observations but two parameters which we need to estimate first. Again: if we had only two observations ($T = 2$), then the parameters were fixed by these two observations (the regression line goes through these two points). But then we could not say anything about the deviation of the errors or residuals!. Independent of the random draws we would always get the same regression line (only two points), residuals would always be zero. With two observations there is no „free information“ to estimate the variance of the errors. We need more than two observations to start making inference for the variance of the error term. So with three observations we would have $T - 2 = 3 - 2 = 1$ degree of freedom.

- Step 2: Standardization of $\hat{\beta}$

$$t = \frac{\hat{\beta} - E(\hat{\beta})}{\widehat{se}(\hat{\beta})} = \frac{\hat{\beta} - \beta}{\widehat{se}(\hat{\beta})} \sim t_{(T-2)}.$$

  The random variable $t = (\hat{\beta} - \beta)/\widehat{se}(\hat{\beta})$ is a pivot. Side note: Why is this a t-distributed value?

- Step 3: Find the $(1 - \alpha/2)$-quantile $t_{a/2}$

$$P(-t_{a/2} \leq t \leq t_{a/2}) = 1 - a.$$

- Step 4: Substitute and solve for β,

$$P(\hat{\beta} - t_{a/2} \cdot \widehat{se}(\hat{\beta}) \leq \beta \leq \hat{\beta} + t_{a/2} \cdot \widehat{se}(\hat{\beta})) = 1 - a.$$

  The interval estimator is

$$\left[\hat{\beta} - t_{a/2} \cdot \widehat{se}(\hat{\beta});\ \hat{\beta} + t_{a/2} \cdot \widehat{se}(\hat{\beta})\right]$$

- Numerical example: The estimated variance of the error terms (standard error of regression) is

$$\begin{aligned} \hat{\sigma}^2 &= S_{\hat{u}\hat{u}} / (T-2) \\ &= 1.5/1 \\ &= 1.5 \end{aligned}$$

  The estimated standard error is

$$\begin{aligned} \widehat{se}(\hat{\beta}) &= \sqrt{\hat{\sigma}^2 / S_{xx}} \\ &= \sqrt{1.5/800} \\ &= 0.0433, \end{aligned}$$

  and for α we find

$$\begin{aligned} \widehat{se}(\hat{\alpha}) &= \sqrt{1.5 \cdot (1/3 + 30^2/800)} = 1.4790 \\ t_{a/2} &= 12.706 \\ \hat{\alpha} &= 0.25 \\ \text{intervall} &= [-18,5424\ ;\ 19,0424] \end{aligned}$$

Interval estimator for intercept α:

$$\left[\hat{\alpha} - t_{a/2} \cdot \widehat{se}(\hat{\alpha})\ ;\ \hat{\alpha} + t_{a/2} \cdot \widehat{se}(\hat{\alpha})\right]$$

where

$$\widehat{se}(\hat{\alpha}) = \sqrt{\widehat{\sigma}^2(1/T + \overline{x}^2/S_{xx})}.$$

Some terminology: The standard error (Standardfehler) is $se(\hat{\beta})$;
the estimated standard error is $\widehat{se}(\hat{\beta})$
Usually, both $se(\hat{\beta})$ and $\widehat{se}(\hat{\beta})$ are called standard error (Standardfehler)
Interpretation of interval estimators? The interval means that for a portion of 1− of the repeated samples, the calculated interval covers the true value. If one were to pick out one of the infinite number of samples by chance, then the probability would be that a random sample that covers the true value is exactly $1 - a$. The equation does not say that the true parameter is within the interval estimator determined by our actual sample with a probability of $1 - a$! The actually calculated interval estimator is an expression of a random variable „interval estimator". A statement of probability is only permitted in connection with possible outcomes of a random variable, but not in connection with the actual outcome of this random variable. A simple example may illustrate this: Consider the random variable „dice number". The probability to observe the possible outcome 4 is 1/6. Once the dice has been thrown, it is 4 or it is not 4, but it is by no means 4 with probability 1/6. This thought is also transferable to the random variable „interval estimator". Once the interval estimator of our actual sample has been determined (actual outcome of the random variable „interval estimator"), it covers the fixed value or it does not cover it, but it does not cover it with a certain probability (e. g. 95%). interval.R

The actual calculated interval estimator can only be used as an **indicator** for the expected shape of the interval estimators determined in repeated samples.

6 Hypotheses Tests

How can we test hypotheses about the regression coefficients (usually about the slope β)?

Null hypothesis H_0 and alternative hypothesis H_1 (There are one-sided and two-sided tests):

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

We already know that

$$\hat{\beta} \sim N(\beta, \sigma^2/S_{xx})$$

If the null hypothesis $H_0 : \beta = q$ is true, then β can be substituted by q

$$\hat{\beta} \sim N(q, \sigma^2/S_{xx})$$

Graphical representation: If H_0 is true, then $\hat{\beta}$ lies in the interval $[q - k; q + k]$ with (high) probability $1 - a$. If $\hat{\beta}$ lies outside the interval, that is evidence against the null hypothesis. ttest.R (I)

Analytical approach is slightly different: we don't look at $\hat{\beta}$ directly, but standardize it. ttest.R (II)

- Step 1: fix the significance level a and set

$$H_0 : \beta = q$$
$$H_1 : \beta \neq q$$

- Step 2: Estimate the standard error $se(\hat{\beta})$ with

$$\widehat{se}(\hat{\beta}) = \sqrt{\hat{\sigma}^2/S_{xx}}$$

where

$$\hat{\sigma}^2 = S_{\hat{u}\hat{u}}/(T-2)$$

- Step 3: Compute the pivotal t-test statistic

$$t = \frac{\hat{\beta} - q}{\widehat{se}(\hat{\beta})}$$

If H_0 is true, then $t \sim t_{(T-2)}$.

- Step 4: Find the critical value $t_{a/2}$

$$P(-t_{a/2} \leq t \leq t_{a/2}) = 1 - a$$

that is the $(1 - a/2)$-quantile of the t-distribution with $T - 2$ degrees of freedom

- Step 5: Compare t and $t_{a/2}$. If t is outside of $[-t_{a/2}, t_{a/2}]$, i.e. if $|t| > t_{a/2}$, then reject H_0

Example: Our three observations.

Step 1: $H_0 : \beta = 0.7$ and $H_1 : \beta \neq 0.7$, the significance level should be $a = 5\%$

Step 2: $T = 3$, $S_{\hat{u}\hat{u}} = 1,5$ and $S_{xx} = 800$. So

$$\hat{\sigma}^2 = S_{\hat{u}\hat{u}}/\,(T-2) = 1.5/1 = 1.5$$

and

$$\widehat{se}(\hat{\beta}) = \sqrt{\hat{\sigma}^2/S_{xx}} = \sqrt{1.5/800} = 0.0433$$

Step 3: The test statistic is

$$t = \frac{\hat{\beta} - q}{\widehat{se}(\hat{\beta})} = \frac{0.125 - 0.7}{0.0433} = -13.279$$

Step 4: the critical value is

$$t_{a/2} = 12.7062$$

Step 5: Since the test statistic -13.279 is outside $[-12.7062\ ;\ 12.7062]$ we reject $H_0 : \beta = 0.7$

Connections between hypothesis testing and confidence intervals: Under the (two-sided) null hypothesis H_0

$$P(q - t_{a/2} \cdot \widehat{se}(\hat{\beta}) \leq \hat{\beta} \leq q + t_{a/2} \cdot \widehat{se}(\hat{\beta})) = 1 - a$$

The $(1 - a)$-confidence interval is

$$[\hat{\beta} - t_{a/2} \cdot \widehat{se}(\hat{\beta}); \hat{\beta} + t_{a/2} \cdot \widehat{se}(\hat{\beta})]$$

Conclusion: If q is outside the confidence interval, H_0 is rejected.

Example: The 0.95-confidence interval for β is $[-0.4252; 0.6752]$; since $q = 0.7$ is outside the interval, $H_0 : \beta = 0.7$ is rejected at the 5% level.

One-sided tests, e.g. right-sided null hypothesis:

$$H_0 : \beta \leq q$$
$$H_1 : \beta > q$$

The basic idea remains the same: If $\hat{\beta}$ is „much larger“ than q, reject H_0.

Step 4 changes to: Find the critical value t_a

$$P(t \leq t_a) = 1 - a$$

For left-sided null hypotheses, the steps 1, 2 and 3 are the same; the critical value is t_{1-a} with $P(t < t_a) = a$

Step 5: Compare t_a and t; reject H_0, if $t > t_a$ For left-sided null hypotheses, H_0 is rejected if t is less than the critical value, $t < t_{1-a}$

Example: We suspect that there is a positive impact of the billing amount on the tip.

Step 1: $H_0 : \beta \leq 0$ and $H_1 : \beta > 0$. The significance level is set to 5%.

Step 2: We have already calculated $\widehat{se}(\hat{\beta}) = 0.0433$.

Step 3: The t-value is

$$t = \frac{\hat{\beta} - q}{\widehat{se}(\hat{\beta})} = \frac{0.125 - 0}{0.0433} = 2.8868$$

Step 4: According to Tabelle T.2 $t_a = 6.3138$; or in R: `qt(0.95,df=1): 6.313752.`

Step 5: Since $t = 2.8868 < t_a = 6.3138$, we *cannot* reject H_0. The data are compatible with the hypothesis that $\beta \leq 0$. The positive estimate $\hat{\beta} = 0.125$ could have happened just by chance.

p-value:

The p-value is

- the probability that the test statistic (a random variable) is greater than the realized test statistic
- the smallest significance level for which the null hypothesis is just rejected

Traditional approach: Reject the null hypothesis if the test statistic is inside the critical region, e.g. if $t > t_a$

Alternative approach: Comparison of probabilities; reject the null hypothesis if the p-value is less than the significance level a

Plot for p-Value (right-sided,left-sided,two-sided) pwert.R

Analytical computation of the p value for the right-sided test: the p-value is

$$\begin{aligned} P(T > t) &= 1 - P(T \leq t) \\ &= 1 - F_{t_{T-2}}(t), \end{aligned}$$

where $F_{t_{T-2}}$ is the distribution function of the t_{T-2}-distribution

Similar for the p value of left-sided tests:

$$\begin{aligned} P(T < t) &= P(T \le t) \\ &= F_{t_{T-2}}(t). \end{aligned}$$

For two-sided tests

$$\begin{aligned} P(|T| > t) &= P(T < -t) + P(T > t) \\ &= 2P(T > t) \quad \text{(due to symmetry)} \\ &= 2(1 - P(T \le t)) \\ &= 2\left(1 - F_{t_{T-2}}(t)\right). \end{aligned}$$

Example: Consider the one-sided null and alternative hypotheses

$$\begin{aligned} H_0 &: \beta \le 0 \\ H_1 &: \beta > 0. \end{aligned}$$

We choose our significance level at 5%. We already know: $\hat{\beta} = 0.125$ and $\widehat{se}(\hat{\beta}) = 0.0433$; thus the t-statistic is 2.8868. Standard software returns $p = 10.6\%$. Since $10.6\% > 5\%$, we *cannot* reject H_0.

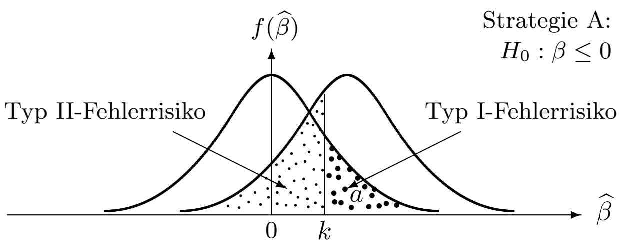
Interpretation of p-values? Advocates of p values find that you can see by how much the null is rejected.

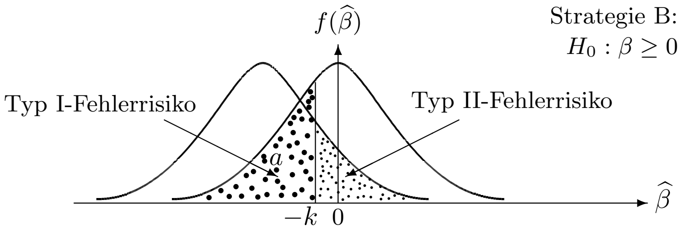
How to choose the null and alternative hypothesis? E.g. for the tip example: one-sided test, we want to state that there is a positive relationship of the bill amount on tips $\beta > 0$.

- There are basically two strategies:
  - State the opposite of the conjecture as the null hypothesis and try to reject it (this is the dominant strategy), $H_0 : \beta \le 0$
  - State the conjecture as the null hypothesis and show that it cannot be rejected $H_0 : \beta > 0$
- There is an important asymmetry between rejection and non-rejection

	H_0 is not rejected	H_0 is rejected
H_0 is true	no error	Type I error
H_0 is false	Type II error	no error

Type I errors are less or equal to a. If the null is rejected, then we are quite sure that it is not true. A lower a increases, however, the Type II error.





Because the test is constructed in such a way that an error of the first type only occurs with a small probability, we interpret a rejection of H_0 as a "statistical underpinning" of H_1.

Attention: Statements about the probability that H_0 or H_1 is correct are nonsense even after testing. If we cannot reject H_0, it does not mean that H_0 is correct but rather there is not enough statistical evidence that the difference between $\hat{\beta}$ and q is large enough.

Maximum-Likelihood-Estimation

Main idea: Find those parameter values that maximize the probability (or likelihood) of observing the actually observed data

Notation:

$$\begin{aligned} \theta &: \text{ vector of parameters, e.g. } \theta = (\alpha, \beta, \sigma^2) \\ L(\theta) &: \text{ likelihood (conditional on data)} \\ \ln L(\theta) &: \text{ log-likelihood} \end{aligned}$$

Maximum-Likelihood-Estimator

$$\hat{\theta} = \arg\max \ln L(\theta) = \arg\min(-\ln L(\theta))$$

We already know that for all $t = 1, \ldots, T$

$$y_t \sim NID(\alpha + \beta x_t, \sigma^2),$$

Hence, the density of each y_t

$$f_{y_t}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(y - \alpha - \beta x_t)^2}{\sigma^2}\right).$$

Due to independence we get

$$\begin{aligned} L(\alpha, \beta, \sigma^2) &= f_{y_1,\ldots,y_T}(y_1, \ldots, y_T) = \prod_{t=1}^{T} f_{y_t}(y_t) \\ \ln L(\alpha, \beta, \sigma^2) &= \ln f_{y_1,\ldots,y_T}(y_1, \ldots, y_T) = \sum_{t=1}^{T} \ln f_{y_t}(y_t) \end{aligned}$$

Maximizing

$$\begin{aligned}
\ln L(\alpha, \beta, \sigma^2) &= \ln f_{y_1,\ldots,y_T}(y_1, \ldots, y_T) \\
&= \sum_{t=1}^{T} \ln \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(y_t - \alpha - \beta x_t)^2}{\sigma^2} \right) \right] \\
&= \sum_{t=1}^{T} \left[\ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2} \frac{(y_t - \alpha - \beta x_t)^2}{\sigma^2} \right] \\
&= \sum_{t=1}^{T} \left[-\frac{1}{2} \ln\left(2\pi\sigma^2\right) - \frac{1}{2} \frac{(y_t - \alpha - \beta x_t)^2}{\sigma^2} \right] \\
&= -\frac{T}{2} \ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_t - \alpha - \beta x_t)^2 .
\end{aligned}$$

w.r.t. α, β, σ^2.

For any given value of σ^2 the loglikelihood is obviously maximized if the sum of squared residuals (SSR) is minimized. Thus, the ML estimators of α and β are identical to the OLS estimators,

$$\begin{aligned}
\hat{\beta} &= \frac{S_{xy}}{S_{xx}} \\
\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x}.
\end{aligned}$$

Differentiating with respect to σ^2 yields the first order condition,

$$\frac{\partial \ln f_{y_1,\ldots,y_T}}{\partial \sigma^2} = -\frac{T}{2}\frac{1}{\sigma^2} + \frac{SSR}{2}\frac{1}{\sigma^4} = 0$$

Solve the first order condition

$$\hat{\sigma}^2_{ML} = \frac{SSR}{T} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2$$

which is different from the OLS estimator

$$\hat{\sigma}^2_{OLS} = \frac{1}{T-2}\sum_{t=1}^{T} \hat{u}_t^2$$

7 Forecasting

Conditional forecast: the value of the exogenous variable is known and non-stochastic: x_0

Point forecast of the endogenous variable is:

$$\hat{y}_0 = \hat{\alpha} + \hat{\beta} x_0$$

The true value of y_0 is usually not $\hat{y}_0$ but

$$y_0 = \alpha + \beta x_0 + u_0$$

The forecasting error is

$$\begin{aligned}\hat{y}_0 - y_0 &= \hat{\alpha} + \hat{\beta} x_0 - (\alpha + \beta x_0 + u_0) \\ &= (\hat{\alpha} - \alpha) + (\hat{\beta} - \beta) x_0 - u_0\end{aligned}$$

There are two error sources:

1. The error term u_0 will not vanish, in general.
2. The parameter estimates $\hat{\alpha}$ and $\hat{\beta}$ will deviate from the true values α and β.

Properties of the point forecast

Expected forecasting error:

$$E(\hat{y}_0 - y_0) = E(\hat{\alpha} - \alpha) + E(\hat{\beta} - \beta) x_0 - E(u_0) = 0$$

Variance of the forecasting error:

$$Var\left(\hat{y}_0 - y_0\right) = Var(\hat{\alpha}) + x_0^2 Var(\hat{\beta}) + \sigma^2 + 2x_0 Cov(\hat{\alpha}, \hat{\beta}).$$

we already derived the variances and covariances, inserting:

$$\begin{aligned}Var\left(\hat{y}_0 - y_0\right) &= \sigma^2 \left(\frac{1}{T} + \frac{\bar{x}^2}{S_{xx}}\right) + x_0^2 \frac{\sigma^2}{S_{xx}} + \sigma^2 - 2x_0 \frac{\bar{x}\sigma^2}{S_{xx}} \\ &= \sigma^2 \left(\frac{1}{T} + \frac{\bar{x}^2}{S_{xx}} + \frac{x_0^2}{S_{xx}} + 1 - \frac{2x_0\bar{x}}{S_{xx}}\right) \\ &= \sigma^2 \left(1 + \frac{1}{T} + \frac{\left(x_0 - \bar{x}\right)^2}{S_{xx}}\right).\end{aligned}$$

Estimated variance of the forecasting error:

$$\widehat{Var}(\hat{y}_0 - y_0) = \hat{\sigma}^2 \left[1 + 1/T + \left(x_0 - \overline{x}\right)^2 / S_{xx}\right]$$

Interval forecast

- Step 1: Estimation of $se(\hat{y}_0 - y_0)$
- Step 2: Standardization of $(\hat{y}_0 - y_0)$

$$t = \frac{(\hat{y}_0 - y_0) - \overbrace{E(\hat{y}_0 - y_0)}^{=0}}{\widehat{se}(\hat{y}_0 - y_0)} = \frac{\hat{y}_0 - y_0}{\widehat{se}(\hat{y}_0 - y_0)} \sim t_{T-2}$$

- Step 3: Find the $t_{a/2}$-value (from statistical tables or using statistical computer software)
- Step 4: With large probability $1 - \alpha$, the random variable t will be inside the interval $[-t_{a/2}; t_{a/2}]$

$$P\left(-t_{a/2} \leq \frac{\hat{y}_0 - y_0}{\widehat{se}(\hat{y}_0 - y_0)} \leq t_{a/2}\right) = 1 - a$$

  Solving for y_0 yields the interval forecast prognose.R

$$\left[\hat{y}_0 - t_{a/2} \cdot \widehat{se}(\hat{y}_0 - y_0);\ \hat{y}_0 + t_{a/2} \cdot \widehat{se}(\hat{y}_0 - y_0)\right]$$

Example: Suppose $x_0 = 20$. Insert it into the estimated model, to get the point forecast

$$\begin{aligned} \widehat{y}_0 &= 0.25 + 0.125 \cdot x_0 \\ &= 2.75. \end{aligned}$$

The expected forecasting error is 0. We know already that $\hat{\sigma}^2 = 1.5$, $S_{xx} = 800$ and $\bar{x} = 30$. Thus, the estimated variance of the forecast error is

$$\begin{aligned} \widehat{Var}(\hat{y}_0 - y_0) &= 1.5\left[1 + 1/3 + (20 - 30)^2 / 800\right] \\ &= 2.1875. \end{aligned}$$

The estimated standard deviation of the forecast error is $\widehat{se}(\hat{y}_0 - y_0) = \sqrt{2.1875} = 1.4790$.

The critical value is the 0.975-quantile of the t-distribution with $T - 2 = 1$ degrees of freedom, i.e. 12.706. Hence, 0.95-interval forecast is

$$\begin{aligned} &\left[\hat{y}_0 \pm t_{a/2} \cdot \widehat{se}(\hat{y}_0 - y_0)\right] \\ = \; &2.75 \pm 12.706 \cdot 1.4790 \\ = \; &[-16.0422;\ 21.5422]. \end{aligned}$$

II. Multiple linear regression model

Until today we only considered a *single* exogenous variable, but in most empirical problems we face many exogenous variables

Many of the results from the simple linear regression model can be transferred to the multiple case

Important tool: matrix algebra
(main diagonal, transpose, addition, scalar multiplication, inner product, matrix multiplication, idem potent, determinant, rank, inverse, trace, definit matrices, semidefinite matrices)

Example: Estimation of a production function for barley Conduct an experiment where the barley output (Gerste, g_t) is observed for different combinations of phosphate (p_t) and nitrogen (n_t) There are $T = 30$ different combinations, see table Slide gerstebsp.R

Functional Specification (A-Annahmen): The economic (agro-economic) model formalizes the connection between the barley output (g) and the fertilizers (p and n)

$$g = f(p, n),$$

Possible function form

$$g = \alpha + \beta_1 p + \beta_2 n.$$

A more realistic functional form

$$g = Ap^{\beta_1} n^{\beta_2}$$

where A, β_1 and β_2 are constant parameters. Taking logs on both sides, we get

$$\ln g = \ln A + \beta_1 \ln p + \beta_2 \ln n.$$

Define $\alpha = \ln A$, $y = \ln g$, $x_1 = \ln p$ and $x_2 = \ln n$, then the econometric model (for $t = 1, \ldots, T$) is

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

In general for K exogenous variables

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_K x_{kt} + u_t$$

or

$$\begin{aligned} y_1 &= \alpha + \beta_1 x_{11} + \beta_2 x_{21} + ... + \beta_K x_{K1} + u_1 \\ y_2 &= \alpha + \beta_1 x_{12} + \beta_2 x_{22} + ... + \beta_K x_{K2} + u_2 \\ &\vdots \\ y_T &= \alpha + \beta_1 x_{1T} + \beta_2 x_{2T} + ... + \beta_K x_{KT} + u_T \end{aligned}$$

Matrix notation: Define

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}; \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{K1} \\ 1 & x_{12} & \ldots & x_{K2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1T} & \ldots & x_{KT} \end{bmatrix}; \quad \beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}; \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}$$

Then the model equations are

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{K1} \\ 1 & x_{12} & \dots & x_{K2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1T} & \dots & x_{KT} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}$$

or compactly:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}.$$

8 Specification

A-Assumptions:

A1: No relevant exogenous variable is omitted from the econometric model, and all exogenous variables included in the model are relevant

A2: The true functional dependence between $\mathbf{X}$ and $\mathbf{y}$ is linear

A3: The parameters β are constant for all T observations $(\mathbf{x}_t, y_t)$

B-Assumptions:

The B-assumptions are the same as in the simple linear model, i.e. $E(u_t) = 0$, $Var(u_t) = \sigma^2$, $Cov(u_t, u_s) = 0$ für $t \neq s$ and normality.

B1 to B4 in matrix notation:

$$\mathbf{u} \sim N\left(\mathbf{0}, \sigma^2 \mathbf{I}_T\right)$$

C-Assumptions:

C1: The exogenous variables $x_{1t}, \dots, x_{Kt}$ are not stochastic, but can be controlled as in an experimental situation

C2: No perfect multicollinearity: The are no parameter values γ_0, γ_1, γ_2, $\dots, \gamma_K$ (with at least one $\gamma_k \neq 0$), such that for all $t = 1, \dots, T$:

$$\gamma_0 + \gamma_1 x_{1t} + \gamma_2 x_{2t} + \dots + \gamma_K x_{Kt} = 0$$

In matrix notation:

$$rank(\mathbf{X}) = K + 1$$

(Implication: $T \geq K + 1$)

Perfect multicollinearity with two regressors: If C2 is violated, there are, γ_0, γ_1, γ_2 (not all 0), such that

$$\gamma_0 + \gamma_1 x_{1t} + \gamma_2 x_{2t} = 0$$

for all $t = 1, \dots, T$. Then

$$x_{2t} = -\left(\gamma_0/\gamma_2\right) - \left(\gamma_1/\gamma_2\right) x_{1t} = \delta_0 + \delta_1 x_{1t}$$

with $\delta_0 = -\left(\gamma_0/\gamma_2\right)$ and $\delta_1 = -\left(\gamma_1/\gamma_2\right)$ Hence, there are not really two regressors, since

$$\begin{aligned} y_t &= \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \\ &= \underbrace{\left(\alpha + \beta_2 \delta_0\right)}_{=\alpha'} + \underbrace{\left(\beta_1 + \beta_2 \delta_1\right)}_{=\beta'} x_{1t} + u_t \end{aligned}$$

9 Point estimation

The econometric model is:

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\beta + \mathbf{u}\\ y_t &= \alpha + \beta_1 x_{1t} + \ldots + \beta_K x_{Kt} + u_t \text{ for } t = 1, \ldots, T\end{aligned}$$

The estimated model is:

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{X}\hat{\beta}\\ \hat{y}_t &= \hat{\alpha} + \hat{\beta}_1 x_{1t} + \ldots + \hat{\beta}_K x_{Kt} \text{ for } t = 1, \ldots, T\end{aligned}$$

Define the residuals

$$\begin{aligned}\hat{\mathbf{u}} &= \mathbf{y} - \hat{\mathbf{y}}\\ \hat{u}_t &= y_t - \hat{y}_t \text{ for } t = 1, \ldots, T\end{aligned}$$

How can we find an estimator $\hat{\beta}$ in the multiple regression model?

The sum of squared residuals is

$$\begin{aligned}S_{\hat{u}\hat{u}} &= \hat{\mathbf{u}}'\hat{\mathbf{u}}\\ &= \sum \hat{u}_t^2\end{aligned}$$

Due to

$$\begin{aligned}\hat{\mathbf{u}} &= \mathbf{y} - \mathbf{X}\hat{\beta}\\ &= y_t - \hat{\alpha} - \hat{\beta}_1 x_{1t} - \ldots - \hat{\beta}_K x_{Kt}\end{aligned}$$

we have

$$\begin{aligned}S_{\hat{u}\hat{u}} &= \left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)' \left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)\\ &= \sum \left(y_t - \hat{\alpha} - \hat{\beta}_1 x_{1t} - \ldots - \hat{\beta}_K x_{Kt}\right)^2\end{aligned}$$

First order conditions

$$\frac{\partial S_{\hat{u}\hat{u}}}{\partial \hat{\beta}} = \begin{bmatrix} \partial S_{\hat{u}\hat{u}}/\partial \hat{\alpha} \\ \partial S_{\hat{u}\hat{u}}/\partial \hat{\beta}_1 \\ \vdots \\ \partial S_{\hat{u}\hat{u}}/\partial \hat{\beta}_K \end{bmatrix} = \mathbf{0}$$

Vector of derivatives

$$\begin{aligned}\frac{\partial S_{\hat{u}\hat{u}}}{\partial \hat{\beta}} &= \frac{\partial}{\partial \hat{\beta}} \left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)' \left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)\\ &= \frac{\partial}{\partial \hat{\beta}} \mathbf{y}'\mathbf{y} - \frac{\partial}{\partial \hat{\beta}} 2\mathbf{y}'\mathbf{X}\hat{\beta} + \frac{\partial}{\partial \hat{\beta}} \hat{\beta}\mathbf{X}'\mathbf{X}\hat{\beta}\\ &= -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\beta}\end{aligned}$$

See J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics und Econometrics, rev. ed., John Wiley & Sons: Chichester, 1999, or: Phoebus J. Dhrymes, Mathematics for Econometrics, 3rd ed., Springer: New York, 2000.

Rules:

$$\frac{\partial \beta' A}{\partial \beta} = \frac{\partial A' \beta}{\partial \beta'} = A$$

or $\frac{\partial A\beta}{\partial \beta'}$ upper number of rows must be equal to lower number of columns. For quadratic functions:

$$\frac{\partial \beta' A \beta}{\partial \beta} = A\beta + A'\beta = 2A\beta$$

as $A = (X'X)^{-1'} = (X'X)^{-1}$ is symmetric.

Solving the first order conditions yields the normal equations

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$$

and thus

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

The terms are:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} T & \sum x_{1t} & \dots & \sum x_{Kt} \\ \sum x_{1t} & \sum x_{1t}^2 & \dots & \sum x_{1t}x_{Kt} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_{Kt} & \sum x_{Kt}x_{1t} & \dots & \sum x_{Kt}^2 \end{bmatrix}, \quad \mathbf{X}'\mathbf{y} = \begin{bmatrix} \sum y_t \\ \sum x_{1t}y_t \\ \vdots \\ \sum x_{Kt}y_t \end{bmatrix}$$

Second-order derivatives:

$$\frac{\partial^2 S}{\partial \beta \partial \beta'} = 2X'X$$

This matrix is positive definite, as for all vectors $a \neq 0$: $a'X'Xa > 0$. Hence we found the minimum.

Meaning of the estimators $\hat{\alpha}, \hat{\beta}_1$ and $\hat{\beta}_2$:

Formal meaning

$$\frac{\partial \hat{y}_t}{\partial x_{1t}} = \hat{\beta}_1 \text{ and } \frac{\partial \hat{y}_t}{\partial x_{2t}} = \hat{\beta}_2$$

Meaning of $\hat{\alpha}$: for $x_{1t} = x_{2t} = 0$:

$$\begin{aligned} \ln \hat{g}_t &= \hat{\alpha} = 0.9543 \\ \hat{g}_t &= e^{0.9543} = 2.5969 \end{aligned}$$

Meaning of $\hat{\beta}_1$ and $\hat{\beta}_2$ in the barley-output-model:

$$\hat{\beta}_1 = \frac{\partial \hat{y}_t}{\partial x_{1t}} = \frac{\partial (\ln \hat{g}_t)}{\partial (\ln p_t)}$$

Because of

$$\frac{\partial \ln \hat{g}_t}{\partial \hat{g}_t} = \frac{1}{\hat{g}_t} \quad \text{und} \quad \frac{\partial \ln p_t}{\partial p_t} = \frac{1}{p_t}$$

we have

$$\hat{\beta}_1 = \frac{\partial \hat{g}_t / \hat{g}_t}{\partial p_t / p_t}$$

i.e. $\hat{\beta}_1$ is the estimated elasticity of the barley output with respect to the phosphate fertilizer

Example: From the $T = 30$ observations compute (rounded to two decimals)

$$\begin{aligned}
T &= 30 \\
\sum x_{1t} &= 96.77 \\
\sum x_{2t} &= 129.72 \\
\sum y_t &= 120.42 \\
\sum x_{1t}^2 &= 312.39 \\
\sum x_{1t} x_{2t} &= 418.46 \\
\sum x_{1t} y_t &= 388.57 \\
\sum x_{2t}^2 &= 564.63 \\
\sum x_{2t} y_t &= 521.66
\end{aligned}$$

Compute

$$\begin{aligned}
\hat{\beta} &= \begin{bmatrix} 30 & 96.77 & 129.72 \\ 96.77 & 312.39 & 418.46 \\ 129.72 & 418.46 & 564.63 \end{bmatrix}^{-1} \begin{bmatrix} 120.42 \\ 388.57 \\ 521.66 \end{bmatrix} \\
&= \begin{bmatrix} 0.9543 \\ 0.5965 \\ 0.2626 \end{bmatrix} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}
\end{aligned}$$

The estimated model is

$$\hat{y}_t = 0.9543 + 0.5965 \cdot x_{1t} + 0.2626 \cdot x_{2t}$$

Present the computations also in R using matrix notation.

The coefficient of determination R^2

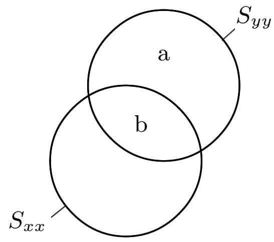
The total variation of y can be decomposed in the same way as in the simple linear model,

$$\underbrace{S_{yy}}_{\text{„total variation"}} = \underbrace{S_{\hat{y}\hat{y}}}_{\text{"explained variation"}} + \underbrace{S_{\hat{u}\hat{u}}}_{\text{"unexplained variation"}}$$

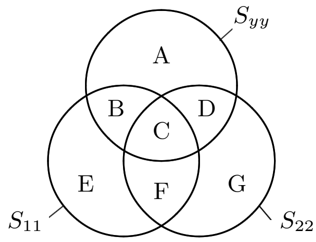
The coefficient of determination is defined as

$$\begin{aligned}
R^2 &= \frac{\text{"explained variation"}}{\text{"total variation"}} \\
&= \frac{S_{\hat{y}\hat{y}}}{S_{yy}} \\
&= \frac{S_{yy} - S_{\hat{u}\hat{u}}}{S_{yy}}
\end{aligned}$$

Graphical illustration for the simple and multiple linear regression model using Venn-diagrams:



(a)



(b)

$$R^2 = \frac{A+B+C}{A+B+C+D}.$$

Note that R^2 always gets larger if we include variables (even if t-statistic is very small). Hence for model comparison we often rely on the corrected coefficient of determination which is defined as

$$\begin{aligned}\overline{R}^2 &= 1 - \frac{S_{\hat{u}\hat{u}}/(T-K-1)}{S_{yy}/(T-1)} \\ &= 1 - \left(1 - R^2\right)\frac{T-1}{T-K-1}\end{aligned}$$

Note that in the denominator is the unbiased estimator for $\hat{\sigma}^2$ and in the nominator unbiased estimator for the variance of y.

Properties of the OLS estimators $\hat{\beta}$

The estimator $\hat{\beta}$ is a random vector. The expectation vector is

$$E(\hat{\beta}) = \beta,$$

i.e. the estimator is unbiased. Due to

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

and

$$y = X\beta + u$$

we get

$$\begin{aligned}\hat{\beta} &= \left(X'X\right)^{-1} X'\left(X\beta + u\right) \\ &= \left(X'X\right)^{-1} X'X\beta + \left(X'X\right)^{-1} X'u \\ &= \beta + \left(X'X\right)^{-1} X'u.\end{aligned}$$

Hence

$$\begin{aligned}E(\hat{\beta}) &= E\left(\beta + \left(X'X\right)^{-1} X'u\right) \\ &= \beta + \left(X'X\right)^{-1} X'E\left(u\right) \\ &= \beta.\end{aligned}$$

The covariance matrix of $\hat{\beta}$ is

$$\begin{aligned}
Cov\left(\hat{\beta}\right) &= E\left(\left(\hat{\beta}-E(\hat{\beta})\right)\left(\hat{\beta}-E(\hat{\beta})\right)'\right) \\
&= E\left(\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)'\right) \\
&= E\left(\left(X'X\right)^{-1}X'uu'X\left(X'X\right)^{-1}\right) \\
&= \left(X'X\right)^{-1}X'E\left(uu'\right)X\left(X'X\right)^{-1} \\
&= \left(X'X\right)^{-1}X'Cov\left(u\right)X\left(X'X\right)^{-1} \\
&= \left(X'X\right)^{-1}X'\sigma^2 IX\left(X'X\right)^{-1} \\
&= \sigma^2\left(X'X\right)^{-1}X'X\left(X'X\right)^{-1} \\
&= \sigma^2\left(X'X\right)^{-1}.
\end{aligned}$$

The variances $Var(\hat{\alpha})$ and $Var(\hat{\beta})$ and the covariance $Cov(\hat{\alpha},\hat{\beta})$, derived for the simple linear regression model are simply special cases of this general result.

BE CAREFUL: Do not mistake with $cov(u) = (Euu') = \sigma^2 I$!

Special case: Covariance matrix in the two regressor model:

$$\begin{aligned}
Var(\hat{\beta}_1) &= \frac{\sigma^2}{S_{11}\left(1-R^2_{1\cdot 2}\right)} \\
Var(\hat{\beta}_2) &= \frac{\sigma^2}{S_{22}\left(1-R^2_{1\cdot 2}\right)} \\
Var\left(\hat{\alpha}\right) &= \sigma^2/T + \bar{x}_1^2 Var(\hat{\beta}_1) \\
&\quad +2\bar{x}_1\bar{x}_2 Cov(\hat{\beta}_1,\hat{\beta}_2) + \bar{x}_2^2 Var(\hat{\beta}_2) \\
Cov(\hat{\beta}_1,\hat{\beta}_2) &= \frac{-\sigma^2 R^2_{1\cdot 2}}{S_{12}\left(1-R^2_{1\cdot 2}\right)}
\end{aligned}$$

with

$$R^2_{1\cdot 2} = \frac{S_{12}^2}{S_{11}S_{22}}.$$

Gauss-Markov-Theorem

The estimator $\hat{\beta}$ is linear in $\mathbf{y}$, since

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y \\
&= Dy
\end{aligned}$$

with $D = (X'X)^{-1}X'$. We will now show that $\hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$ is not only unbiased but also efficient.

Let $\check{\beta}$ be another linear unbiased estimator of β,

$$\begin{aligned}
\check{\beta} &= Cy \\
&= C(X\beta+u) \\
&= CX\beta + Cu.
\end{aligned}$$

Unbiasedness requires

$$\begin{aligned} E(\check{\beta}) &= E(CX\beta + Cu) \\ &= CX\beta \\ &= \beta, \end{aligned}$$

i.e. the matrix C must fulfill $CX = I$. Hence

$$\begin{aligned} \check{\beta} &= CX\beta + Cu \\ &= \beta + Cu \end{aligned}$$

or

$$\check{\beta} - \beta = Cu.$$

The covariance matrix of $\check{\beta}$ is

$$\begin{aligned} V(\check{\beta}) &= E(Cu.u'C') \\ &= \sigma^2 CC'. \end{aligned}$$

Now, compare the covariance matrices $V(\hat{\beta})$ and $V(\check{\beta})$ (using $CX = I$):

$$\begin{aligned} V(\check{\beta}) - V(\hat{\beta}) &= \sigma^2 CC' - \sigma^2 (X'X)^{-1} \\ &= \sigma^2 \left(CC' - (X'X)^{-1} \right) \\ &= \sigma^2 \left(C - (X'X)^{-1}X' \right) \left(C - (X'X)^{-1}X' \right)', \end{aligned}$$

as can be easily verified. The dimension of the matrix

$$A = C - (X'X)^{-1}X'$$

is $(K+1) \times T$. The matrix AA' is positive semi-definite, since for all $(K+1)$-vectors a:

$$a'A \underbrace{A'a}_{=b} = b'b = \sum_{t=1}^{T} b_t^2 \geq 0$$

Hence the OLS estimator is best linear unbiased estimator.

Distribution of y and $\hat{\beta}$

From $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$ und $\mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_T)$ we conclude that $\mathbf{y}$ is multivariate normally distributed. EExpectation vector and covariance matrix of endogenous variable $\mathbf{y}$:

$$\begin{aligned} E(\mathbf{y}) &= E(\mathbf{X}\beta + \mathbf{u}) = \mathbf{X}\beta \\ \mathbf{V}(\mathbf{y}) &= \mathbf{V}(\mathbf{X}\beta + \mathbf{u}) = \mathbf{V}(\mathbf{u}) = \sigma^2 \mathbf{I}_T \end{aligned}$$

Thus $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I}_T)$.

From $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ we conclude that the estimator $\hat{\beta}$ also has a multivariate normal distribution.

Expectation vector and covariance matrix are already known

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

Estimation of the error term variance

Problem: The error term variance σ^2 is unknown. The covariance matrix $\mathbf{V}(\hat{\beta})$ cannot be computed without σ^2

Solution: We need an estimator for σ^2. An estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{S_{\hat{u}\hat{u}}}{T-K-1}$$

This estimator is unbiased. Define

$$M = I - X(X'X)^{-1}X'.$$

The matrix is symmetric and idempotent, i.e. $M = M'$ and $MM = M$. Often, M is called residual maker matrix, since

$$\hat{u} = Mu = My.$$

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

Orthogonal to X:

$$X'\hat{u} = X'y - X'X(X'X)^{-1}X'y = X'y - X'y = 0$$

Hence:

$$\hat{u} = M(X\beta + u) = Mu$$

For the following derivations we need some rules of calculus for the trace of a matrix. The trace of a quadratic $(n \times n)$-Matrix A is the sum of its diagonal elements:

$$tr(A) = \sum_{i=1}^{n} a_{ii}.$$

For suitable Matrices A and B and a scalar λ):

$$\begin{aligned} tr(A+B) &= tr(A) + tr(B) \\ tr(\lambda A) &= \lambda tr(A) \\ tr(A') &= tr(A) \\ tr(AB) &= tr(BA). \end{aligned}$$

Back to the proof: The sum of squared residuals is

$$\begin{aligned} \hat{u}'\hat{u} &= u'M'Mu \\ &= u'Mu \\ &= tr(u'Mu) \\ &= tr(uu'M). \end{aligned}$$

Hence

$$\begin{aligned} E(\hat{u}'\hat{u}) &= E\left(tr(uu'M)\right) \\ &= tr(E(uu'M)) \\ &= tr(E(uu')M) \\ &= tr(\sigma^2 I_T M) \\ &= \sigma^2 tr(M) \\ &= \sigma^2 \left[tr(I_T) - tr(X(X'X)^{-1}X')\right] \\ &= \sigma^2(T-K-1). \end{aligned}$$

Thus

$$E(\hat{\sigma}^2) = \sigma^2.$$

Interval estimation

Interval estimation of a single component $\hat{\beta}_k$ of $\hat{\beta}$,

$$P\left(\hat{\beta}_k - c \leq \beta_k \leq \hat{\beta}_k + c\right) = 1 - a$$

We know that

$$\hat{\beta}_k \sim N(\beta_k, Var(\hat{\beta}_k)),$$

where $Var(\hat{\beta}_k)$ is the $(k+1)$ diagonal element of $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Problem: σ^2 and $Var(\hat{\beta}_k)$ are unknown and must be estimated first

Interval estimation of β_k:

- Step 1: Estimation of σ^2 by $\hat{\sigma}^2$ and $se(\hat{\beta}_k) = \sqrt{Var(\hat{\beta}_k)}$ by

$$\widehat{se}(\hat{\beta}_k) = \sqrt{\widehat{Var}(\hat{\beta}_k)}.$$

- Step 2: Standardization of $\hat{\beta}_k$

$$t = \frac{\hat{\beta}_k - E(\hat{\beta}_k)}{\widehat{se}(\hat{\beta}_k)} = \frac{\hat{\beta}_k - \beta_k}{\widehat{se}(\hat{\beta}_k)} \sim t_{(T-K-1)}$$

- Step 3: Find the $t_{a/2}$-value
- Step 4: The $(1-\alpha)$-interval estimator is

$$\left[\hat{\beta}_k - t_{a/2} \cdot \widehat{se}(\hat{\beta}_k)\ ;\ \hat{\beta}_k + t_{a/2} \cdot \widehat{se}(\hat{\beta}_k)\right]$$

Interval estimation of linear combinations of $\hat{\beta}$

Let $\mathbf{r}$ be an arbitrary $(K+1)$-Vektor. How can we find a confidence interval of $\mathbf{r}'\beta$?

Fertilizer example: $\mathbf{r} = [0, 1, 1]'$, then $\mathbf{r}'\beta = \beta_1 + \beta_2$ (economies von scale?)

The point estimator of $\mathbf{r}'\beta$ is $\mathbf{r}'\hat{\beta}$. The variance of $\mathbf{r}'\hat{\beta}$ ist $\mathbf{r}'\mathbf{V}(\hat{\beta})\mathbf{r} = \sigma^2\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}$.

The confidence interval for $\mathbf{r}'\beta$ is

$$\left[\mathbf{r}'\hat{\beta} - t_{a/2} \cdot \hat{\sigma}\sqrt{\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}}\ ;\ \mathbf{r}'\hat{\beta} + t_{a/2} \cdot \hat{\sigma}\sqrt{\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}}\right]$$

Special case of a single component

$$\beta_k = \mathbf{r}'\beta$$

for

$$\mathbf{r} = [0, \ldots, 0, 1, 0, \ldots, 0]'$$

where the 1 is located at the k^{th} position

10 Hypothesis tests: t-test

There are tests of a single linear combination (t-tests) and tests of multiple linear combinations (F-tests)

t-Test

Testing a single linear combination of parameters (two-sided):

Remember: In the simple linear regression case

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

In the multiple linear model the null and alternative hypotheses are

$$\begin{aligned} H_0 &: r_0\alpha + r_1\beta_1 + \ldots + r_K\beta_K = q \\ H_1 &: r_0\alpha + r_1\beta_1 + \ldots + r_K\beta_K \neq q \end{aligned}$$

or

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

where

$$\mathbf{r} = [r_0, r_1, \ldots, r_K]'$$

The test procedure:

1. Set up H_0 and H_1 and fix the significance level a
2. Estimate $se(\mathbf{r}'\hat{\beta})$
3. Compute the t-statistic
4. Find the critical value $t_{1-a/2}$
5. Test decision: Compare $t_{1-a/2}$ and t. Reject H_0, if $|t| > t_{1-a/2}$ ist.

Left-sided t-test

$$\begin{aligned} H_0 &: \mathbf{r}'\beta \geq q \\ H_1 &: \mathbf{r}'\beta < q \end{aligned}$$

Right-sided t-test

$$\begin{aligned} H_0 &: \mathbf{r}'\beta \leq q \\ H_1 &: \mathbf{r}'\beta > q \end{aligned}$$

The critical values are lower quantiles of the t-distribution for the left-sided test and upper quantiles for the right-sided test

F-Test

Simultaneous test of two or more linear combinations (restrictions) Null hypothesis and alternative hypothesis

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

Examples:

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \ldots = \beta_K = 0 \\ H_0 &: \beta_1 = \beta_2 = \ldots = \beta_K \\ H_0 &: \beta_1 + \ldots + \beta_k = 1 \text{ und } \beta_1 = 2\beta_2 \\ H_0 &: \beta_1 = 5 \text{ und } \beta_2 = \ldots = \beta_K = 0 \end{aligned}$$

Basic idea of the F-test: Compare the restricted and the unrestricted model

Sum of squared residuals of the econometric model and the model under the null hypothesis

$$\begin{aligned} S_{\hat{u}\hat{u}} &= \hat{\mathbf{u}}'\hat{\mathbf{u}} = \sum_{t=1}^{T} \hat{u}_t^2 \\ S_{\hat{u}^0\hat{u}^0} &= \hat{\mathbf{u}}^{0\prime}\hat{\mathbf{u}}^0 = \sum_{t=1}^{T} \left(\hat{u}_t^0\right)^2 \end{aligned}$$

where $\hat{\mathbf{u}}^0$ are the residuals if the model is estimated under the restrictions of the null hypothesis

Example: Null hypothesis

$$y_t = \alpha + 0 \cdot x_{1t} + \ldots + 0 \cdot x_{Kt} + u_t = \alpha + u_t$$

Obviously, $S^0_{\widehat{u}\widehat{u}} \geq S_{\widehat{u}\widehat{u}}$; the null hypothesis is likely to be false if $S^0_{\widehat{u}\widehat{u}}$ is "much larger" than $S_{\widehat{u}\widehat{u}}$

The test statistic is

$$F = \frac{\left(S^0_{\widehat{u}\widehat{u}} - S_{\widehat{u}\widehat{u}}\right) / L}{S_{\widehat{u}\widehat{u}} / (T - K - 1)}$$

where L is the number of restrictions in H_0 If the null hypothesis is true, then

$$F \sim F_{(L,T-K-1)}.$$

The five steps of the F-test

1. Set up H_0 and H_1 and choose the significance level a
2. Calculate $S_{\widehat{u}\widehat{u}}$ and $S^0_{\widehat{u}\widehat{u}}$ (more on the computation of $S^0_{\widehat{u}\widehat{u}}$ later)
3. Compute the F-test statistic
4. Find the critical value F_a, i.e. the upper a-quantile of the $F_{L,T-K-1}$-distribution
5. Reject H_0 if $F > F_a$

For $L = 1$ the F-test is identical to a two-sided t-test

Careful: A combination of t-tests is *not* the same as a single F-test The decisions of t-tests and an F-test can be contradicting Distinction between individual t-tests and a simultaneous F tFtests.R

Der restringierte KQ-Schätzer:

Estimate β subject to the restrictions $\mathbf{R}\beta = \mathbf{q}$ given in the null hypothesis. Optimization under constraints: Minimize

$$S_{\hat{u}^0\hat{u}^0}(\mathbf{b}) = (\mathbf{y} - \mathbf{X}\mathbf{b})' (\mathbf{y} - \mathbf{X}\mathbf{b})$$

with respect to $\mathbf{b}$ subject to $\mathbf{R}\mathbf{b} - \mathbf{q} = \mathbf{0}$. A standard Lagrange approach yields

$$\begin{aligned} \mathcal{L}(\mathbf{b}, \lambda) &= (\mathbf{y} - \mathbf{X}\mathbf{b})' (\mathbf{y} - \mathbf{X}\mathbf{b}) + \lambda' (\mathbf{R}\mathbf{b} - \mathbf{q}) \\ &= \mathbf{y}'\mathbf{y} - \mathbf{2}\mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} + \lambda'\mathbf{R}\mathbf{b} - \lambda'\mathbf{q} \end{aligned}$$

where λ is a vector (having length L) of Lagrange multipliers. The first order conditions are

$$\frac{\partial \mathcal{L}}{\partial \mathbf{b}} = (-\mathbf{2}\mathbf{y}'\mathbf{X})' + 2\mathbf{X}'\mathbf{X}\mathbf{b} + (\lambda'\mathbf{R})' = \mathbf{0} \tag{15}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \mathbf{R}\mathbf{b} - \mathbf{q} = \mathbf{0}. \tag{16}$$

The vector satisfying these conditions is the restricted least squares estimator, $\hat{\beta}^{RLS}$. Rewriting (15) ergibt

$$\begin{aligned} \mathbf{X}'\mathbf{X}\hat{\beta}^{RLS} &= \mathbf{X}'\mathbf{y} - \mathbf{R}'\lambda/2 \\ \hat{\beta}^{RLS} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'\lambda/2 \\ &= \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'\lambda/2. \end{aligned} \tag{17}$$

This implies

$$\mathbf{R}\hat{\beta}^{RLS} = \mathbf{R}\hat{\beta} - \mathbf{R} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'\lambda/2,$$

but since $\mathbf{R}\hat{\beta}^{RLS} = \mathbf{q}$s the restriction enforced in (16) we can also write

$$\mathbf{R} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'\lambda/2 = \mathbf{R}\hat{\beta} - \mathbf{q}.$$

The $(L \times L)$-Matrix $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ is invertible and thus

$$\lambda/2 = \left(\mathbf{R} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'\right)^{-1} \left(\mathbf{R}\hat{\beta} - \mathbf{q}\right).$$

Substituting the Lagrange multipliers in (17) results in

$$\begin{aligned} \hat{\beta}^{RLS} &= \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'\lambda/2 \\ &= \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \left(\mathbf{R} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'\right)^{-1} \left(\mathbf{R}\hat{\beta} - \mathbf{q}\right). \end{aligned} \tag{18}$$

Residuals of the restricted model: $\hat{\mathbf{u}}^0 = \mathbf{y} - \mathbf{X}\hat{\beta}^{RLS}$.

An alternative way to write (or compute) the F-statistic

$$F = \frac{\left(S^0_{\widehat{u}\widehat{u}} - S_{\widehat{u}\widehat{u}}\right) / L}{S_{\widehat{u}\widehat{u}} / (T - K - 1)}$$

is

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

The denominator is obviously identical. To see the equality of the numerator use

$$\begin{aligned}\hat{\mathbf{u}}^0 &= \mathbf{y}-\mathbf{X}\hat{\beta}^{RLS}\\ &= \mathbf{y}-\mathbf{X}\hat{\beta}+\mathbf{X}\hat{\beta}-\mathbf{X}\hat{\beta}^{RLS}\\ &= \mathbf{y}-\mathbf{X}\hat{\beta}+\mathbf{X}\left(\hat{\beta}-\hat{\beta}^{RLS}\right)\\ &= \hat{\mathbf{u}}+\mathbf{X}\left(\hat{\beta}-\hat{\beta}^{RLS}\right).\end{aligned}$$

Hence

$$\begin{aligned}S^0_{\hat{u}\hat{u}} &= \hat{\mathbf{u}}^{0\prime}\hat{\mathbf{u}}^0\\ &= \left(\hat{\mathbf{u}}+\mathbf{X}\left(\hat{\beta}-\hat{\beta}^{RLS}\right)\right)'\left(\hat{\mathbf{u}}+\mathbf{X}\left(\hat{\beta}-\hat{\beta}^{RLS}\right)\right)\\ &= \hat{\mathbf{u}}'\hat{\mathbf{u}}+\mathbf{2}\hat{\mathbf{u}}'\mathbf{X}\left(\hat{\beta}-\hat{\beta}^{RLS}\right)+\left(\hat{\beta}-\hat{\beta}^{RLS}\right)'\mathbf{X}'\mathbf{X}\left(\hat{\beta}-\hat{\beta}^{RLS}\right)\\ &= \hat{\mathbf{u}}'\hat{\mathbf{u}}+\left(\hat{\beta}-\hat{\beta}^{RLS}\right)'\mathbf{X}'\mathbf{X}\left(\hat{\beta}-\hat{\beta}^{RLS}\right)\end{aligned}$$

since $\hat{\mathbf{u}}'\mathbf{X}=\mathbf{0}$ (these are the normal equations), so

$$S^0_{\hat{u}\hat{u}}-S_{\hat{u}\hat{u}}=\left(\hat{\beta}-\hat{\beta}^{RLS}\right)'\mathbf{X}'\mathbf{X}\left(\hat{\beta}-\hat{\beta}^{RLS}\right).$$

From (18) we have

$$\hat{\beta}-\hat{\beta}^{RLS}=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\left(\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\hat{\beta}-\mathbf{q}\right).$$

Inserting these terms it is easy to see that

$$S^0_{\hat{u}\hat{u}}-S_{\hat{u}\hat{u}}=\left(\mathbf{R}\hat{\beta}-\mathbf{q}\right)'\left[\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{R}\hat{\beta}-\mathbf{q}\right).$$

Note the similarity to the t-test statistic

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

Maximum likelihood estimation

Repetition: If $\mathbf{X}$ is a K-dimensional random vector with multivariate normal distribution $N(\mu,\mathbf{\Sigma})$ then its joint density is

$$f_{\mathbf{X}}\left(\mathbf{x}\right)=\left(2\pi\right)^{-K/2}\left(\det\mathbf{\Sigma}\right)^{-1/2}\exp\left(-\frac{1}{2}\left(\mathbf{x}-\mu\right)'\mathbf{\Sigma}^{-1}\left(\mathbf{x}-\mu\right)\right).$$

Multiple linear regression model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u} \quad \text{mit } \mathbf{u} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$$

Distribution of the endogenous variables:

$$\mathbf{y} \sim N\left(\mathbf{X}\beta, \sigma^2\mathbf{I}\right).$$

Joint density of $\mathbf{y}$

$$\begin{aligned} & f_{\mathbf{y}}(\mathbf{y}) \\ = \quad & (2\pi)^{-\frac{T}{2}} \left(\det \sigma^2\mathbf{I}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{X}\beta)' \left(\sigma^2\mathbf{I}\right)^{-1}(\mathbf{y}-\mathbf{X}\beta)\right) \\ = \quad & (2\pi)^{-T/2} \left(\sigma^{2T}\right)^{-1/2} \exp\left(-\frac{(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)}{2\sigma^2}\right). \end{aligned}$$

Log-likelihood function

$$\ln L\left(\beta, \sigma^2\right) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)}{2\sigma^2}.$$

First order condition for a maximum

$$\begin{bmatrix} \frac{\partial \ln L}{\partial \beta} \\ \\ \frac{\partial \ln L}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{X}'(\mathbf{y}-\mathbf{X}\beta)}{\sigma^2} \\ \\ \frac{-T}{2\sigma^2} + \frac{(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}$$

Solution of the FOCs

$$\frac{\mathbf{X}'(\mathbf{y}-\mathbf{X}'\beta)}{\sigma^2} = \mathbf{0}$$

is obviously

$$\hat{\beta}_{ML} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

The ML estimator is identical to the OLS estimator. The second part of the FOCs concerns the error term variance. Inserting $\hat{\beta}_{ML}$ for β yields

$$\begin{aligned} \frac{-T}{2\hat{\sigma}^2_{ML}} + \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{2\hat{\sigma}^4_{ML}} &= 0 \\ -T + \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\hat{\sigma}^2_{ML}} &= 0 \\ \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\hat{\sigma}^2_{ML}} &= T \\ \hat{\sigma}^2_{ML} &= \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}. \end{aligned}$$

The ML estimator of β is identical to the OLS estimator, the ML estimator of σ^2 is different and thus biased (but asymptotically unbiased)

The classical tests (LR, Wald, LM)

Illustration of the basic test ideas. threetests.R classtest.R

Generalization to multiple restrictions:

$$\begin{aligned} H_0 &: \mathbf{g}(\beta) = \mathbf{0} \\ H_1 &: \mathbf{g}(\beta) \neq \mathbf{0} \end{aligned}$$

where β is the coefficient vector of a multiple linear regression model ; and $\mathbf{g}$ is a (possibly nonlinear) vector-valued function. Special Case: Test of L linear restrictions:

$$\mathbf{g}(\beta) = \mathbf{R}\beta - \mathbf{q}.$$

Wald-Test

Idea: If $\mathbf{g}(\hat{\beta}_{ML})$ is significantly different from $\mathbf{0}$, reject H_0

Test statistic (for multiple restrictions)

$$W = \mathbf{g}\left(\hat{\beta}_{ML}\right)' \left[\widehat{Cov}\left(\mathbf{g}\left(\hat{\beta}_{ML}\right)\right)\right]^{-1} \mathbf{g}\left(\hat{\beta}_{ML}\right) \xrightarrow{d} U \sim \chi^2_L,$$

if the null hypothesis is true

Wald test statistic for L linear restrictions (write $\hat{\beta}$ for $\hat{\beta}_{ML}$ as they are the same):

$$W = \mathbf{g}\left(\hat{\beta}\right)' \left[\widehat{Cov}\left(\mathbf{g}\left(\hat{\beta}\right)\right)\right]^{-1} \mathbf{g}\left(\hat{\beta}\right)$$

where $\mathbf{g}(\beta) = \mathbf{R}\beta - \mathbf{q}$. The covariance matrix of $\mathbf{g}(\hat{\beta})$ is

$$\begin{aligned} Cov(\mathbf{g}(\hat{\beta})) &= Cov(\mathbf{R}\hat{\beta} - \mathbf{q}) \\ &= \mathbf{R}Cov(\hat{\beta})\mathbf{R}' \\ &= \sigma^2 \mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}' \end{aligned}$$

and

$$\widehat{Cov}(\mathbf{g}(\hat{\beta})) = \hat{\sigma}^2_{ML}\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'.$$

Hence, the Wald statistic is

$$\begin{aligned} W &= \left(\mathbf{R}\hat{\beta} - \mathbf{q}\right)' \left[\hat{\sigma}^2_{ML}\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right]^{-1} \left(\mathbf{R}\hat{\beta} - \mathbf{q}\right) \\ &= \frac{\left(\mathbf{R}\hat{\beta} - \mathbf{q}\right)' \left[\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right]^{-1} \left(\mathbf{R}\hat{\beta} - \mathbf{q}\right)}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/T}. \end{aligned}$$

This term is almost identical to L times the F-statistic

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

Likelihood-ratio-Test (LR)

Idea: If the maximal likelihood under the restrictions $L(\hat{\beta}_R, \hat{\sigma}^2_R)$ is significantly lower than the maximal likelihood without restrictions $L(\hat{\beta}_{ML}, \hat{\sigma}^2_{ML})$,then reject H_0 ab.

Test statistic

$$LR = 2\left(\ln L\left(\hat{\beta}_{ML}, \hat{\sigma}^2_{ML}\right) - \ln L\left(\hat{\beta}_R, \hat{\sigma}^2_R\right)\right) \xrightarrow{d} U \sim \chi^2_L,$$

if the null hypothesis is true

The restricted estimators are (without proof)

$$\begin{aligned} \hat{\beta}_R &= \hat{\beta}^{RLS} \\ \hat{\sigma}^2_R &= \hat{\mathbf{u}}^{0\prime}\hat{\mathbf{u}}^0/T. \end{aligned}$$

The log-likelihood function evaluated at the two points is,

$$\begin{aligned} \ln L(\hat{\beta}_{ML}, \hat{\sigma}^2_{ML}) &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\hat{\sigma}^2_{ML} - \frac{\left(\mathbf{y}-\mathbf{X}\hat{\beta}_{ML}\right)'\left(\mathbf{y}-\mathbf{X}\hat{\beta}_{ML}\right)}{2\hat{\sigma}^2_{ML}} \\ &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}\right) - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{2\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}\right)} \\ &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}\right) - \frac{T}{2} \\ \ln L(\hat{\beta}_R, \hat{\sigma}^2_R) &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\hat{\sigma}^2_R - \frac{\left(\mathbf{y}-\mathbf{X}\hat{\beta}_R\right)'\left(\mathbf{y}-\mathbf{X}\hat{\beta}_R\right)}{2\hat{\sigma}^2_R}. \\ &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\left(\frac{\hat{\mathbf{u}}^{0\prime}\hat{\mathbf{u}}^0}{T}\right) - \frac{\hat{\mathbf{u}}^{0\prime}\hat{\mathbf{u}}^0}{2\left(\frac{\hat{\mathbf{u}}^{0\prime}\hat{\mathbf{u}}^0}{T}\right)} \\ &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\left(\frac{\hat{\mathbf{u}}^{0\prime}\hat{\mathbf{u}}^0}{T}\right) - \frac{T}{2} \end{aligned}$$

The difference times 2 is

$$\begin{aligned} LR &= T\left(\ln\left(\frac{\hat{\mathbf{u}}^{0\prime}\hat{\mathbf{u}}^0}{T}\right) - \ln\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}\right)\right) \\ &= T\ln\left(\frac{\hat{\mathbf{u}}^{0\prime}\hat{\mathbf{u}}^0}{\hat{\mathbf{u}}'\hat{\mathbf{u}}}\right) \end{aligned}$$

Since $\ln x \approx x - 1$ for values near unity we can approximate the test statistic by

$$\begin{aligned} LR &\approx T\left(\frac{\hat{\mathbf{u}}^{0\prime}\hat{\mathbf{u}}^0}{\hat{\mathbf{u}}'\hat{\mathbf{u}}} - 1\right) \\ &= \frac{\hat{\mathbf{u}}^{0\prime}\hat{\mathbf{u}}^0 - \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/T}. \end{aligned}$$

which is close to L times the usual F-statistic

$$F = \frac{\left(\hat{\mathbf{u}}^{0\prime}\hat{\mathbf{u}}^0 - \hat{\mathbf{u}}'\hat{\mathbf{u}}\right)/L}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(T-K-1)}.$$

Lagrange-Multiplier-Test (LM)

Idea: If the slope of the log-likelihood function, evaluated at the restricted estimator, $\partial \ln L(\hat{\beta}_R)/\partial\beta$ is significantly different from $\mathbf{0}$, reject H_0

For test statistic we make use of the information matrix

$$I(\theta) = -E\left(\frac{\partial^2 \ln L(\theta)}{\partial\theta\partial\theta'}\right)$$

where $\theta = (\beta, \sigma)'$.

The vector of first-order derivatives is

$$\frac{\partial \ln L(\theta)}{\partial\theta} = \begin{bmatrix} \frac{\partial \ln L(\beta,\sigma)}{\partial\beta} \\ \frac{\partial \ln L(\beta,\sigma)}{\partial\sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{X}'(\mathbf{y}-\mathbf{X}\beta)}{\sigma^2} \\ \frac{-T}{2\sigma^2} + \frac{(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}$$

The matrix of Second-order derivatives is

$$\frac{\partial^2 \ln L(\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} -\frac{X'X}{\sigma^2} & -\frac{X'(y-X\beta)}{\sigma^4} \\ -\frac{(y-X\beta)'X}{\sigma^4} & \frac{T}{2\sigma^4} - \frac{(y-X\beta)'(y-X\beta)}{\sigma^6} \end{bmatrix}.$$

Due to $E(y - X\beta) = 0$ and $E((y - X\beta)'(y - X\beta)) = T\sigma^2$ we get

$$I(\theta) = \begin{bmatrix} \frac{X'X}{\sigma^2} & 0 \\ 0 & \frac{T}{2\sigma^4} \end{bmatrix}$$

and

$$I(\theta)^{-1} = \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{T} \end{bmatrix}$$

Test statistic:

$$LM = \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial\theta}\right)' \left[I(\hat{\theta}_R)\right]^{-1} \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial\theta}\right) \xrightarrow{d} U \sim \chi^2_L,$$

under the null hypothesis H_0.

Evaluating the score at $\hat{\theta}_R = (\hat{\beta}_R, \hat{\sigma}^2_R)'$, we get

$$\begin{aligned} \frac{\partial \ln L(\hat{\theta}_R)}{\partial\theta} &= \begin{bmatrix} \frac{\partial \ln L(\hat{\beta}_R,\hat{\sigma}_R)}{\partial\beta} \\ \frac{\partial \ln L(\hat{\beta}_R,\hat{\sigma}_R)}{\partial\sigma^2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{X'(y-X\hat{\beta}_R)}{\hat{\sigma}^2_R} \\ \frac{-T}{2\hat{\sigma}^2_R} + \frac{(y-X\hat{\beta}_R)'(y-X\hat{\beta}_R)}{2\hat{\sigma}^4_R} \end{bmatrix}. \end{aligned}$$

However,

$$(y - X\hat{\beta}_R)'(y - X\hat{\beta}_R) = T\hat{\sigma}^2_R,$$

such that the lower component of the score vector is 0.

Therefore, the simplified test statistic is given by

$$\begin{aligned} LM &= \left(\frac{\partial \ln L(\hat\beta_R)}{\partial \beta}\right)' \hat\sigma_R^2 (X'X)^{-1} \left(\frac{\partial \ln L(\hat\beta_R)}{\partial \beta}\right) \\ &= \frac{(y - X\hat\beta_R)'X(X'X)^{-1}X'(y - X\hat\beta_R)}{\hat\sigma_R^2} \end{aligned}$$

Because of

$$\hat\beta_R = \hat\beta - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat\beta - q)$$

we can rewrite $\hat u^0 = y - X\hat\beta_R$ to

$$\begin{aligned} \hat u^0 &= y - X\hat\beta_R \\ &= y - X\left(\hat\beta - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat\beta - q)\right) \\ &= y - X\hat\beta + X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat\beta - q) \\ &= \hat u + X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat\beta - q). \end{aligned}$$

Taking into account that $X'\hat u = 0$ (normal equations) we rewrite the numerator as

$$\begin{aligned} \hat u'^0 X(X'X)^{-1}X'\hat u^0 &= (R\hat\beta - q)'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'X \\ &\qquad (X'X)^{-1}X'X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat\beta - q) \\ &= (R\hat\beta - q)'[R(X'X)^{-1}R']^{-1}(R\hat\beta - q), \end{aligned}$$

such that

$$LM = \frac{(R\hat\beta - q)'[R(X'X)^{-1}R']^{-1}(R\hat\beta - q)}{\hat u'^0 \hat u^0 / T}.$$

Now we clearly see the similarity to the F-Test and the two other classical tests.

11 Forecasting

The approach is similar to forecasting in the simple linear regression Let $\mathbf{x}_0 = [1, x_{10}, x_{20}, \ldots, x_{K0}]'$ denote the vector of exogenous variables

Point forecast

$$\hat y_0 = \mathbf{x}_0'\hat\beta.$$

Variance of the forecast error

$$Var\left(\hat y_0 - y_0\right) = \sigma^2 \left(1 + \mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0\right)$$

12 Presentation of the results

In the literature, the results of regression analyses are often presented as follows

$$\begin{array}{cccccc} \hat{y} = & \hat{\alpha} & + & \hat{\beta}_1 x_1 & + \ldots + & \hat{\beta}_K x_K \\ & (\widehat{se}(\hat{\alpha})) & & (\widehat{se}(\hat{\beta}_1)) & & (\widehat{se}(\hat{\beta}_K)) \end{array}$$

Sometimes you find t-values in the parentheses, i.e. the values of the test statistics for the tests $H_0 : \beta_k = 0$ vs $H_1 : \beta_k \neq 0$

Often, R^2 and $\hat{\sigma}$ and the value of the test statistic of the F test

$$H_0 : \beta_1 = \ldots = \beta_K = 0 \quad \text{vs} \quad H_1 : \text{not } H_0$$

are reported additionally

Fertilizer example:

$$\begin{array}{cccccc} \hat{y} = & 0.95432 & + & 0.59652 x_1 & + & 0.26255 x_2 \\ & (0.46943) & & (0.13788) & & (0.03400) \end{array}$$

Additional results:

$$\begin{aligned} R^2 &= 0.743 \\ \hat{\sigma}^2 &= 0.00425 \\ \hat{\sigma} &= 0.0652 \end{aligned}$$

Test statistics:

$$\begin{array}{lcl} H_0 : \beta_1 = 0 & \longrightarrow & 4.326 \\ H_0 : \beta_2 = 0 & \longrightarrow & 7.723 \\ H_0 : \beta_1 = \beta_2 = 0 & \longrightarrow & 38.98 \end{array}$$

Examples of computer output (R, Stata, SPSS, matlab, Excel) Slides 22-30

13 Omitted or irrelevant variables

Assumption A1: No relevant exogenous variable is omitted from the econometric model, and all exogenous variables included in the model are relevant

What happens if relevant variables are missing? What happens if there are irrelevant variables included in the model?

Example: Wage structure in a firm with 20 employees; what are the determinants of the wage y_t? Slide 31

Data: Education x_{1t}; age x_{2t}; firm tenure x_{3t}

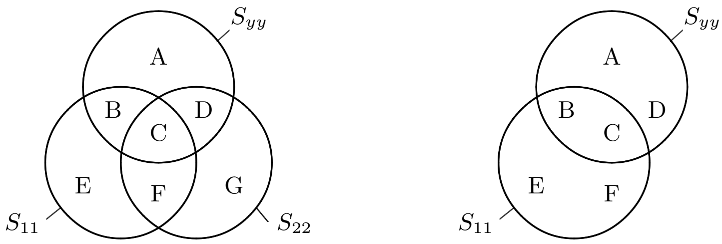
Three potential models (M2 is the true model):

$$\begin{array}{lrcl} (M1) & y_t & = & \alpha + \beta x_{1t} + u_t' \\ (M2) & y_t & = & \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \\ (M3) & y_t & = & \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t'' \end{array}$$

Estimation results for the three variables. Slide 32

Omitted variables

Graphical representation: What happens if we omit the relevant variable 2?



(a) (b)

Consider the model M1, where the relevant variable x_2 is missing.

$$(M1) \qquad y_t = \alpha + \beta x_{1t} + u_t'$$

The error terms

$$\begin{array}{rcl} u_t' & = & \beta_2 x_{2t} + u_t \\ E(u_t') & = & E(\beta_2 x_{2t} + u_t) \\ & = & \beta_2 x_{2t} + E(u_t) \\ & = & \beta_2 x_{2t} + 0 \\ & \neq & 0 \end{array}$$

If a relevant exogenous variable is omitted, assumption B1 is violated!

Consequence for point estimation:

$$
\begin{aligned}
E(\hat{\beta}_1') &= E\left(\frac{S_{1y}}{S_{11}}\right) \\
&= E\left(\frac{\sum_t (x_{1t}-\bar{x}_1)(y_t-\bar{y})}{\sum_t (x_{1t}-\bar{x}_1)^2}\right) \\
&= E\left(\frac{\sum_t (x_{1t}-\bar{x}_1)y_t}{\sum_t (x_{1t}-\bar{x}_1)^2}\right) \\
&= E\left(\frac{\sum_t (x_{1t}-\bar{x}_1)(\alpha+\beta_1 x_{1t}+\beta_2 x_{2t}+u_t)}{\sum_t (x_{1t}-\bar{x}_1)^2}\right) \\
&= E\left(\frac{\beta_1 S_{11}+\beta_2 S_{12}+\sum_t (x_{1t}-\bar{x}_1)u_t}{S_{11}}\right) \\
&= \beta_1+\beta_2\frac{S_{12}}{S_{11}}+\frac{\sum_t (x_{1t}-\bar{x}_1)E(u_t)}{S_{11}} \\
&= \beta_1+\beta_2\frac{S_{12}}{S_{11}}
\end{aligned}
$$

meaning that $\hat{\beta}_1$ is biased. The direction of the bias is dependent on the sign of β_2 and S_{12} ab.

Consequence for interval estimation:

$$[\hat{\beta}_1' - t_{a/2}\cdot\widehat{se}(\hat{\beta}_1')\ ;\ \hat{\beta}_1' + t_{a/2}\cdot\widehat{se}(\hat{\beta}_1')]$$

meaning it is not correctly centered. Further

$$se(\hat{\beta}_1') = \sqrt{var(\hat{\beta}_1')}$$

with

$$var(\hat{\beta}_1') = \frac{\sigma^2}{S_{11}}.$$

The estimator

$$\hat{\sigma}^2 = \frac{S_{\widehat{u}'\widehat{u}'}}{T-2}$$

is biased. The unbiased estimator is

$$\hat{\sigma}^2 = \frac{S_{\widehat{u}\widehat{u}}}{T-3}$$

Conclusion: The coverage probability of the confidence intervals is not $1-\alpha$

Hypothesis tests are also biased: The probability of an error of the first kind does not equal the significance level

If a relevant exogenous variable is omitted, then

- the point estimators are biased and inconsistent (we show that later)
- the interval estimators and hypothesis tests are no longer valid

In matrix notation: True model is $y_t = \alpha+\beta_1 x_{1t}+\beta_2 x_{2t}+u_t$ or in matrix notation

$$y = X_a\beta_a + X_2\beta_2 + u$$

where

$$X_a = \begin{pmatrix} 1 & x_{11} \\ \vdots & \vdots \\ 1 & x_{1T} \end{pmatrix}, \qquad X_2 = \begin{pmatrix} x_{21} \\ \vdots \\ x_{2T} \end{pmatrix}, \quad \beta_a = \begin{pmatrix} \alpha \ \beta_1 \end{pmatrix}$$

If we omit x_{2t} we estimate

$$\begin{aligned} \hat{\beta}_a &= (X_a' X_a)^{-1} X_a' y \\ &= (X_a' X_a)^{-1} X_a' (X_a \beta_a + X_2 \beta_2 + u) \\ &= (X_a' X_a)^{-1} X_a' X_a \beta_a + (X_a' X_a)^{-1} X_a' X_2 \beta_2 + (X_a' X_a)^{-1} X_a' u \end{aligned}$$

Taking the expectation and noting that $E(u) = 0$ we get

$$E(\hat{\beta}_a) = \beta_a + (X_a' X_a)^{-1} X_a' X_2 \beta_2$$

Hence the bias is equal to

$$bias = E(\hat{\beta}_a) - \beta_a = (X_a' X_a)^{-1} X_a' X_2 \beta_2$$

Note that if $X_a' X_2 = 0$ there is no bias. So looking into $X'X$ is a good indicator. That is $X = (X_a X_2)$, $X'X = \begin{pmatrix} X_a' X_a & X_a' X_2 \\ X_2' X_a & X_2' X_2 \end{pmatrix}$

Irrelevant variables

The error term in the misspecified model M3 is

$$u_t'' = u_t - \beta_3 x_{3t}$$

and since $\beta_3 = 0$:

$$u_t'' = u_t$$

Consequently,

$$\begin{aligned} E(\widehat{\alpha}_1'') &= \alpha \\ E(\hat{\beta}_1'') &= \beta_1 \\ E(\hat{\beta}_2'') &= \beta_2 \\ E(\hat{\beta}_3'') &= \beta_3 = 0. \end{aligned}$$

The variances of the estimators are

$$\begin{aligned} Var(\hat{\beta}_1) &= \frac{\sigma^2}{S_{11}\,(1 - R_{1\cdot 2}^2)} \\ Var(\hat{\beta}_1'') &= \frac{\sigma^2}{S_{11}\,(1 - R_{1\cdot 2}^2 - R_{1\cdot 3}^2)} \end{aligned}$$

The estimated error term variance is

$$\widehat{\sigma}^2 = \frac{S_{\widehat{u}''\widehat{u}''}}{T - 4}$$

Conclusion: Omitted relevant variables are a serious problem, redundant variables are not (but they inflate the standard errors)

Diagnosis

How can we find the correct model?

The coeffcient of determination R^2 does not help select a model

Adjusted R^2

$$\begin{aligned}\overline{R}^2 &= 1 - \frac{S_{\widehat{u}\widehat{u}}/(T-K-1)}{S_{yy}/(T-1)} \\ &= 1 - \left(1 - R^2\right)\frac{T-1}{T-K-1}\end{aligned}$$

Further model selection criteria: trade-off between biasedness and inefficiency. Another important criteria is the Akaike information criterion (AIC)

$$AIC = \ln\left(\frac{S_{\widehat{u}\widehat{u}}}{T}\right) + \frac{2(K+1)}{T}.$$

The smaller AIC the better the model

We can also use t-test for single variables to select the correct model or an F-test for multiple variables

14 Functional form

Assumption A2: The true functional dependence between $\mathbf{X}$ and $\mathbf{y}$ is linear

Milk example: Milk production m depends on amount of concentrated feed f. Slide 32

scatterplot shows that linear relationship is infeasible. milch.R

A misspecified model returns useless results

Some nonlinear dependencies:

$$\begin{aligned}\text{Semi-logarithmisch} &: m_t = \alpha + \beta \ln f_t + u_t \\ \text{Invers} &: m_t = \alpha + \beta\,(1/f_t) + u_t \\ \text{Exponential} &: \ln m_t = \alpha + \beta f_t + u_t \\ \text{Logarithmisch} &: \ln m_t = \alpha + \beta \ln f_t + u_t \\ \text{Quadratisch} &: m_t = \alpha + \beta_1 f_t + \beta_2 f_t^2 + u_t\end{aligned}$$

Approach I: Estimation of a nonlinear regression

$$y_t = g(x_t) + u_t$$

with criterion function

$$\sum_{t=1}^{T}\left(y_t - g(x_t)\right)^2.$$

Optimization by numerical methods

Approach II: Linearization of the model; then linear regression

$$\begin{aligned}y_t &= \alpha + \beta x_t + u_t \\ y_t &= \ln m_t \\ x_t &= \ln f_t\end{aligned}$$

Diagnosis: Regression Specification Error Test (RESET)

Higher order Taylor approximation

$$y_t = f(x_t) = \alpha + \beta_1 x_t + \beta_2 x_t^2 + \beta_3 x_t^3 + \dots$$

Test: Are the higher orders (jointly) significant? F-test of

$$H_0 : \beta_2 = \beta_3 = \dots = 0.$$

If we reject H_0 then there is evidence for a nonlinear relationship.

Problem: What happens if there are many exogenous variables?

Basic idea of the RESET Test: $\widehat{y}_t^2$, $\widehat{y}_t^3$, ... are included as additional exogenous variables,

$$y_t = \alpha + \beta_1 x_t + \gamma_2 \widehat{y}_t^2 + \gamma_3 \widehat{y}_t^3 + u_t$$

If γ_2 and/or γ_3 are significant, then there are nonlinearities

The null hypothesis is

$$H_0 : \gamma_2 = \gamma_3 = 0$$

(maybe even higher orders). The test is a common F test and implemented in many statistical software packages

RESET in the linear model:

1. Estimate the linear model and calculate $S_{\widehat{u}\widehat{u}}$ and the fitted $\widehat{y}_t$
2. Add L powers of $\hat{y}_t$ to the linear model

   $$y_t = \alpha + \beta_1 x_t + \gamma_2 \hat{y}_t^2 + \gamma_3 \hat{y}_t^3 + u_t$$

   Estimate the extended model and calculate the sum of squared residuals $S^*_{\widehat{u}\widehat{u}}$.

   The null hypothesis is $H_0 : \gamma_2 = \gamma_3 = 0$. Compute the F-test statistic

   $$F_{(L,T-K^*-1)} = \frac{(S_{\widehat{u}\widehat{u}} - S^*_{\widehat{u}\widehat{u}})/L}{S^*_{\widehat{u}\widehat{u}}/(T - K^* - 1)},$$

   where K^* is the number of exogenous variables in the extended model
3. If $F > F_a$ ist (significance level a, degress of freedom L and $T - K^* - 1$) then H_0 is rejected and the linear model is discarded

Milk example reset.R

Qualitative exogenous variables

Assumption A3: The parameters β are constant for all T observations $(\mathbf{x}_t, y_t)$

Example: The wage y_t depends on education x_{1t} and age x_{2t},

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t.$$

The wage equations for males and females might be different:

$$\begin{aligned} y_t &= \alpha_M + \beta_{M1} x_{1t} + \beta_{M2} x_{2t} + u_t \\ y_t &= \alpha_F + \beta_{F1} x_{1t} + \beta_{F2} x_{2t} + u_t \end{aligned}$$

What happens if the difference is neglected? qualitative.R

Dummy variable

$$D_t = \begin{cases} 0 & \text{wenn männlich} \\ 1 & \text{wenn weiblich} \end{cases}$$

Extended model

$$y_t = \alpha + D_t\gamma + \beta_1 x_{1t} + \delta_1 D_t x_{1t} + \beta_2 x_{2t} + \delta_2 D_t x_{2t} + u_t.$$

Model for men ($D_t = 0$):

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t.$$

Model for women ($D_t = 1$):

$$y_t = (\alpha + \gamma) + (\beta_1 + \delta_1)\, x_{1t} + (\beta_2 + \delta_2)\, x_{2t} + u_t$$

If the qualitative variable has more than two values, we need more than one dummy variable.
Example: Religion (protestant, catholic, other)

$$\begin{aligned} D_{Pt} &= \begin{cases} 0 & \text{for other} \\ 1 & \text{for protestant} \\ 0 & \text{for catholic} \end{cases} \\ D_{Ct} &= \begin{cases} 0 & \text{for other} \\ 0 & \text{for protestant} \\ 1 & \text{for catholic} \end{cases} \end{aligned}$$

Meaning of the coefficients; testing structural stability

Estimation of the model:

Use the ordinary t- or F-tests to detect differences in the coefficients, e.g.

$$H_0 : \gamma = \delta_1 = \delta_2 = 0$$

Very often, the model includes only a level effect, i.e.

$$y_t = \alpha + \gamma D_t + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

Then use a t-test for γ

Estimation of the wage equation model

$$y_t = \alpha + D_t\gamma + \beta_1 x_{1t} + \delta_1 D_t x_{1t} + \beta_2 x_{2t} + \delta_2 D_t x_{2t} + u_t$$

Compare with separat estimation of the two models: wages.R

$$\begin{aligned} y_t &= \alpha_M + \beta_{M1} x_{1t} + \beta_{M2} x_{2t} + u_t && \text{für Männer} \\ y_t &= \alpha_F + \beta_{F1} x_{1t} + \beta_{F2} x_{2t} + u_t && \text{für Frauen} \end{aligned}$$

The point estimates and the sum of squared residuals are identical (why?) The standard errors differ (why?)

For simplicity we only consider one exogenous variable

$$y_t = \alpha + \gamma D_t + \beta x_t + \delta D_t x_t + u_t.$$

Order the observations such that $D_t = 0$ for $t = 1, \ldots, T_1$ and $D_t = 1$ for $t = T_1 + 1, \ldots, T$.

The joint estimation minimizes (with respect to $\alpha, \beta, \gamma, \delta$)

$$S(\alpha, \beta, \gamma, \delta) = \sum_{t=1}^{T_1} (y_t - \alpha - \beta x_t)^2 + \sum_{t=T_1+1}^{T} (y_t - (\alpha + \gamma) - (\beta + \delta) x_t)^2.$$

The first order conditions for the joint estimation are

$$\begin{aligned}
\frac{\partial S}{\partial \alpha} &= -\sum_{t=1}^{T_1} (y_t - \alpha - \beta x_t) - \sum_{t=T_1+1}^{T} (y_t - (\alpha + \gamma) - (\beta + \delta) x_t) = 0 \\
\frac{\partial S}{\partial \beta} &= -\sum_{t=1}^{T_1} (y_t - \alpha - \beta x_t) x_t - \sum_{t=T_1+1}^{T} (y_t - (\alpha + \gamma) - (\beta + \delta) x_t) x_t = 0 \\
\frac{\partial S}{\partial \gamma} &= -\sum_{t=T_1+1}^{T} (y_t - (\alpha + \gamma) - (\beta + \delta) x_t) = 0 \\
\frac{\partial S}{\partial \delta} &= -\sum_{t=T_1+1}^{T} (y_t - (\alpha + \gamma) - (\beta + \delta) x_t) x_t = 0.
\end{aligned}$$

ence, the point estimates in the joint estimation are identical to those of the separat estimations

If the point estimates are identical, then so are the residuals; and if the residuals are identical, then so are the sums of squared residuals

As to the standard errors, in the joint model we estimate

$$\hat{\sigma}^2 = S_{\hat{u}\hat{u}} / (T - 4)$$

while in the separat estimations we estimate

$$\begin{aligned}
\hat{\sigma}_0^2 &= S_{\hat{u}\hat{u}}^0 / (T_1 - 2) \\
\hat{\sigma}_1^2 &= S_{\hat{u}\hat{u}}^1 / ((T - T_1) - 2)
\end{aligned}$$

Remarks:

- What happens if the dummy variables are not 0/1-coded but 1/2-coded?
- Consider the model

$$y_t = \alpha + \gamma D_{1t} + \delta D_{2t} + \beta x_t + u_t$$

  where

$$\begin{aligned}
D_{1t} &= \begin{cases} 0 & \text{for males} \\ 1 & \text{for females} \end{cases} \\
D_{2t} &= \begin{cases} 0 & \text{for German citizenship} \\ 1 & \text{else} \end{cases}
\end{aligned}$$

- Interaction terms

17 Heteroskedastizität

Annahme B2: $Var(u_t) = \sigma^2$ für alle $t = 1, \ldots, T$.

Graphische Darstellung der Beispieldaten [rentexample.R].

Welche Eigenschaften hat der OLS-Schätzer $\hat{\beta}$, wenn die Annahme verletzt ist?

Der Schätzer $\hat{\beta} = (X'X)^{-1}X'y$ ist weiterhin unverzerrt, denn

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

Also ist

$$\begin{aligned}
E(\hat{\beta}) &= E(\beta + (X'X)^{-1}X'u) \\
&= \beta + (X'X)^{-1}X'E(u) \\
&= \beta.
\end{aligned}$$

An keiner Stelle wurde die Annahme B2 verwendet. Das Ergebnis gilt also auch, wenn B2 verletzt ist.

Über die Kovarianzmatrix von $\hat{\beta}$ lässt sich im allgemeinen Fall recht wenig sagen. Es gilt

$$\begin{aligned}
Cov(\hat{\beta}) &= E((\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))') \\
&= E((\hat{\beta} - \beta)(\hat{\beta} - \beta)') \\
&= E((X'X)^{-1}X'uu'X(X'X)^{-1}) \\
&= (X'X)^{-1}X'E(uu')X(X'X)^{-1} \\
&= (X'X)^{-1}X'Cov(u)X(X'X)^{-1}.
\end{aligned}$$

Falls $Cov(u) = \sigma^2 I$ (also im Fall von Homoskedastizität) kann man diesen Ausdruck weiter vereinfachen. Im Fall der Heteroskedastizität jedoch nicht. Nur unter bestimmten Annahmen über die genaue Form der Heteroskedastizität wären weitere Umformungen möglich.

Die Normalität von $\hat{\beta}$ gilt weiterhin, denn $\hat{\beta}$ ist immer noch eine lineare Funktion der normalverteilten endogenen Variable.

Wir betrachten nun das einfache lineare Modell

$$y_t = \alpha + \beta x_t + u_t$$

und machen nun eine (restriktive und willkürliche) Annahme über die Varianz der Störterme:

$$\sigma_t^2 = \sigma^2 x_t.$$

Wir gehen also davon aus, dass die Störterme umso stärker streuen, je weiter die Beobachtungen vom Zentrum entfernt sind.

Das Modell wird nun wie folgt transformiert:

$$\begin{aligned}
\frac{y_t}{\sqrt{x_t}} &= \alpha \frac{1}{\sqrt{x_t}} + \beta \frac{x_t}{\sqrt{x_t}} + \underbrace{\frac{u_t}{\sqrt{x_t}}}_{\text{error term}} \\
y_t^* &= \alpha z_t^* + \beta x_t^* + u_t^*.
\end{aligned}$$

Der neue Fehlerterm

$$u_t^* = \frac{u_t}{\sqrt{x_t}}.$$

hat folgende Eigenschaften:

$$\begin{aligned}
E(u_t^*) &= E\left(\frac{u_t}{\sqrt{x_t}}\right) \\
&= \frac{E(u_t)}{\sqrt{x_t}} \\
&= 0,
\end{aligned}$$

$$\begin{aligned}
Var(u_t^*) &= Var\left(\frac{u_t}{\sqrt{x_t}}\right) \\
&= \frac{1}{x_t} Var(u_t) \\
&= \frac{1}{x_t} \sigma^2 x_t \\
&= \sigma^2
\end{aligned}$$

und

$$\begin{aligned}
Cov(u_s^*, u_t^*) &= Cov\left(\frac{u_s}{\sqrt{x_s}}, \frac{u_t}{\sqrt{x_t}}\right) \\
&= \frac{1}{\sqrt{x_s}\sqrt{x_t}} Cov(u_s, u_t) \\
&= 0.
\end{aligned}$$

Außerdem sind die transformierten Störterme normalverteilt, da sie lineare Transformationen der u_t sind.

Das transformierte Modell erfüllt also alle A-, B- und C-Annahmen!

KQ-Schätzung des transformierten Modells:

$$\begin{aligned}
\hat{\alpha}^* &= \frac{S_{z^* y^*}}{S_{z^* z^*}} \\
\hat{\beta}^* &= \frac{S_{x^* y^*}}{S_{x^* x^*}} \\
&= \frac{\sum \left(x_t^* - \overline{x}^*\right)\left(y_t^* - \overline{y}\right)}{\sum \left(x_t^* - \overline{x}^*\right)^2} \\
&= \frac{\sum \frac{1}{x_t} \left(x_t - \overline{x}\right)\left(y_t - \overline{y}\right)}{\sum \frac{1}{x_t} \left(x_t - \overline{x}\right)^2}
\end{aligned}$$

Die üblichen Schätzer

$$\begin{aligned}
\hat{\beta} &= \frac{\sum \left(x_t - \overline{x}\right)\left(y_t - \overline{y}\right)}{\sum \left(x_t - \overline{x}\right)^2} \\
\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x}
\end{aligned}$$

sind anders und daher ineffizient (Gauß-Markov-Theorem),

Die Störtermvarianz

$$Var(u_t^*) = \sigma^2$$

kann man unverzerrt schätzen durch

$$\hat{\sigma}^2 = \frac{S_{\hat{u}^*\hat{u}^*}}{T-2}.$$

Aus $\sigma_t^2 = \sigma^2 x_t$ folgt, dass

$$\hat{\sigma}_t^2 = \hat{\sigma}^2 \cdot x_t$$

ein unverzerrter Schätzer von $Var(u_t)$ ist.

Herleitung der Varianz von $\hat{\beta}$: Betrachte das einfache lineare Modell

$$y_t = \alpha + \beta x_t + u_t$$

mit $Var(u_t) = \sigma^2 x_t$. Der KQ-Schätzer von β ist

$$\begin{aligned}
\hat{\beta} &= \frac{S_{xy}}{S_{xx}} \\
&= \frac{\sum (x_t - \bar{x})\, y_t}{\sum (x_t - \bar{x})^2} \\
&= \frac{\sum (x_t - \bar{x})(\alpha + \beta x_t + u_t)}{\sum (x_t - \bar{x})^2} \\
&= \frac{\sum (x_t - \bar{x})(\alpha + \beta x_t)}{\sum (x_t - \bar{x})^2} + \frac{\sum (x_t - \bar{x})\, u_t}{\sum (x_t - \bar{x})^2}.
\end{aligned}$$

Also ist

$$\begin{aligned}
Var(\hat{\beta}) &= Var\left(\frac{\sum (x_t - \bar{x})\, u_t}{\sum (x_t - \bar{x})^2}\right) \\
&= \frac{1}{S_{xx}^2} Var\left(\sum (x_t - \bar{x})\, u_t\right) \\
&= \frac{1}{S_{xx}^2} \sum Var\left((x_t - \bar{x})\, u_t\right) \\
&= \frac{1}{S_{xx}^2} \sum (x_t - \bar{x})^2\, Var(u_t) \\
&= \frac{\sum (x_t - \bar{x})^2 \sigma_t^2}{S_{xx}^2}.
\end{aligned}$$

Achtung: Die üblichen Formeln

$$Var(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$$

und

$$\hat{\sigma}^2 = \frac{S_{\hat{u}\hat{u}}}{T-2}$$

sind unter Heteroskedastizität falsch!

Wie kann man aufdecken, ob Heteroskedastizität ein Problem sein könnte? Wir behandeln zwei Testverfahren, nämlich den Goldfeld-Quandt-Test und den White-Test. Wir starten mit dem Goldfeld-Quandt-Test.

Goldfeld-Quandt-Test

- Schritt 1: Ordne die Beobachtungen aufsteigend nach den x_t-Werten (oder nach einer anderen "Heteroskedastizitätsquelle")
- Schritt 2: Definiere zwei Gruppen:
  - T_1 Beobachtungen mit niedrigen x_t-Werten;
  - T_2 Beobachtungen mit hohen x_t-Werten.

  Oft wählt man so, dass $T_1 + T_2 = T$.
- Schritt 3: Wir nehmen an, dass $\sigma_2^2 > \sigma_1^2$; also

$$\begin{aligned} H_0 &: \sigma_2^2 = \sigma_1^2 \\ H_1 &: \sigma_2^2 > \sigma_1^2 \end{aligned}$$

- Schritt 4: Separate KQ-Schätzung für beide Gruppen; berechne $S^1_{\hat{u}\hat{u}}$ und $S^2_{\hat{u}\hat{u}}$.
- Schritt 5: Goldfeld und Quandt (1972) zeigen, dass unter H_0

$$F = \frac{S^2_{\hat{u}\hat{u}}/(T_2 - K - 1)}{S^1_{\hat{u}\hat{u}}/(T_1 - K - 1)}$$

  einer $F_{(T_2-K-1,T_1-K-1)}$-Verteilung folgt.
- Schritt 6: Vergleiche F mit dem kritischen Wert F_a. Wenn $F > F_a$, lehne H_0 ab.

Numerische Illustration: rentexample.R

1. Ordne die Beobachtungen nach der x_t-Variable.
2. Gruppe Z: Zentrum($T_Z = 5$); Gruppe P: Peripherie ($T_P = 7$)
3. Nullhypothese: $H_0 : \sigma_P^2 \leq \sigma_Z^2$
4. Summe der quadrierten Residuen

$$S^Z_{\hat{u}\hat{u}} = 0.246$$

   und

$$S^P_{\hat{u}\hat{u}} = 4.666$$

5. Also gilt

$$F = \frac{4.666/5}{0.246/3} = 11.4$$

6. Auf dem Niveau $a = 5\%$ ist der kritische Wert 9.01. Die Nullhypothese wird also verworfen. Die Daten sprechen für Heteroskedastizität.

White test

Der Goldfeld-Quandt-Test ist dann gut anwendbar, wenn man eine einzelne Variable als „Treiber“ für die Heteroskedastizität identifizieren kann. Wenn man keine konkrete Vorstellung von der Form der Heteroskedastizität hat, hilft der White-Test weiter. Wir betrachten das lineare Modell mit zwei exogenen Variablen,

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

Der White-Test ist ein LM-Test und wird wie folgt durchgeführt.

- Schritt 1: H_0: Homoskedastizität gegen H_1: Heteroskedastizität.
- Schritt 2: Berechne die KQ-Residuen $\hat{u}_t$
- Schritt 3: Schätze die Hilfsregression
$$\hat{u}_t^2 = \gamma_0 + \gamma_1 x_{1t} + \gamma_2 x_{2t} + \gamma_3 x_{1t}^2 + \gamma_4 x_{2t}^2 + \gamma_5 x_{1t} x_{2t} + v_t.$$
- Schritt 4: Unter H_0 gilt
$$R^2 \cdot T \sim \chi_r^2,$$
wobei r die Anzahl der Steigungskoeffizienten in der Hilfsregression ist (in diesem Beispiel also $r = 5$).
- Wenn die Teststatistik $T \cdot R^2$ größer ist als der kritische Werte der χ_r^2-Verteilung, wird H_0 verworfen.

Eine Ablehnung erfolgt, wenn die quadrierten Residuen zumindest teilweise durch die exogenen Variablen erklärt werden können.

Illustration [rentexample.R]

Was ist zu tun, wenn Heteroskedastizität vorliegt?

- Möglichkeit 1: Passe die Schätzmethode an $\longrightarrow$ VKQ oder geschätzte VKQ
- Möglichkeit 2: Benutze weiterhin die KQ-Methode, aber passe die Berechnung der Standardfehler an $\longrightarrow$ Whites heteroskedastizitätskonsistente Kovarianzmatrixschätzung

Verallgemeinerte Kleinste-Quadrate-Methode (VKQ)

Die verallgemeinerte Kleinste-Quadrate-Methode wird auch oft generalised least squares method (GLS) genannt. Ausgangspunkt ist das Regressionsmodell

$$y = X\beta + u,$$

wobei wir nun davon ausgehen, dass die Kovarianzmatrix der Störterme nicht $Cov(u) = \sigma^2 I$ ist, sondern

$$Cov(u) = \sigma^2 \Omega.$$

Beispiel: Wenn $\sigma_t^2 = \sigma^2 x_{kt}$, dann ist

$$\Omega = \begin{bmatrix} x_{k1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & x_{kT} \end{bmatrix}$$

Transformation des Modells: Da Ω positiv definit ist, gibt es eine $(T \times T)$-Matrix P mit

$$P'P = \Omega^{-1}.$$

Beispiel: Wenn

$$\Omega = \begin{bmatrix} x_{k1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & x_{kT} \end{bmatrix},$$

dann ist

$$P = \begin{bmatrix} 1/\sqrt{x_{k1}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1/\sqrt{x_{kT}} \end{bmatrix}.$$

Aus $P'P = \Omega^{-1}$ folgt, dass

$$\begin{aligned} P'P &= \Omega^{-1} \\ P\Omega P'PP^{-1} &= P\Omega\Omega^{-1}P^{-1} \\ P\Omega P' &= I. \end{aligned}$$

Multipliziert man P von links an $y = X\beta + u$, so ergibt sich

$$\begin{aligned} Py &= PX\beta + Pu \\ y^* &= X^*\beta + u^*. \end{aligned}$$

Eigenschaften des transformierten Störtermvektors: Der Erwartungswertvektor ist

$$\begin{aligned} E(u^*) &= E(Pu) \\ &= PE(u) \\ &= 0 \end{aligned}$$

und die Kovarianzmatrix ist

$$\begin{aligned} V(u^*) &= E(u^*u^{*\prime}) \\ &= E(Puu'P') \\ &= PE(uu')P' \\ &= P\sigma^2\Omega P' \\ &= \sigma^2 I. \end{aligned}$$

Außerdem ist der Vektor u^* normalverteilt, da er eine lineare Transformation von normalverteilten Größen ist (nämlich Pu).

Folgerung: Das transformierte Modell erfüllt alle A-, B- und C-Annahmen!

Der GLS-Schätzer ist der OLS-Schätzer des transformierten Modells,

$$
\begin{aligned}
\hat{\beta}^{GLS} &= (X^{*\prime} X^*)^{-1} X^{*\prime} y^* \\
&= (X' P' P X)^{-1} X' P' P y \\
&= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y.
\end{aligned}
$$

Die Kovarianzmatrix des GLS-Schätzers ist

$$
\begin{aligned}
V(\hat{\beta}^{GLS}) &= \sigma^2 (X^{*\prime} X^*)^{-1} \\
&= \sigma^2 \left(X' P' P X\right)^{-1} \\
&= \sigma^2 \left(X' \Omega^{-1} X\right)^{-1}.
\end{aligned}
$$

Vergleich mit der gewöhnlichen Kovarianzmatrix (im OLS-Fall)

$$
\begin{aligned}
V(\hat{\beta}) &= E\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'\right] \\
&= E\left[\left(X'X\right)^{-1} X'u \left(\left(X'X\right)^{-1} X'u\right)'\right] \\
&= E\left[\left(X'X\right)^{-1} X'uu'X' \left(X'X\right)^{-1}\right] \\
&= \left(X'X\right)^{-1} X' E\left[uu'\right] X' \left(X'X\right)^{-1} \\
&= \sigma^2 \left(X'X\right)^{-1} X' \Omega X' \left(X'X\right)^{-1}.
\end{aligned}
$$

Man schätzt σ^2 durch

$$
\hat{\sigma}^2 = \frac{\hat{u}^{*\prime} \hat{u}^*}{T - K - 1} = \frac{\hat{u}' \Omega^{-1} \hat{u}}{T - K - 1}.
$$

Würde man die Heteroskedastizität ignorieren, würde man mit

$$
\begin{aligned}
V(\hat{\beta}) &= \sigma^2 (X'X)^{-1} \\
\hat{\sigma}^2 &= \frac{\hat{u}'\hat{u}}{T - K - 1}
\end{aligned}
$$

rechnen. In diesem Fall wären die Intervallschätzer und die Hypothesentests nicht korrekt.

Was kann man tun, wenn Ω unbekannt ist?

Beispiel:

$$
W = \sigma^2 \Omega = \begin{bmatrix}
\sigma_I^2 & 0 & \dots & \dots & \dots & 0 \\
0 & \ddots & & & & \vdots \\
\vdots & & \sigma_I^2 & & & \vdots \\
\vdots & & & \sigma_{II}^2 & & \vdots \\
\vdots & & & & \ddots & 0 \\
0 & \dots & \dots & \dots & 0 & \sigma_{II}^2
\end{bmatrix}.
$$

Ansatz: Feasible Generalized Least Squares (FGLS), geschätzte verallgemeinerte Kleinste-Quadrate (GVKQ).

Man geht in zwei Stufen vor. Zuerst schätzt man die unbekannten Größen in $W = \sigma^2\Omega$. Anschließend ist der FGLS-Schätzer

$$\hat{\beta}^{FGLS} = (X'\hat{W}^{-1}X)^{-1}X'\hat{W}^{-1}y.$$

Die geschätzte Kovarianzmatrix ist dann

$$\hat{V}(\hat{\beta}^{FGLS}) = (X'\hat{W}^{-1}X)^{-1}.$$

Was kann man tun, wenn man überhaupt keine Informationen über die Form der Heteroskedastizität hat?

Whites heteroskedastizitätskonsistenter Kovarianzmatrixschätzer

Davidson und MacKinnon, chap. 5.5. Das ökonometrische Modell lautet

$$y = X\beta + u.$$

Die Kovarianzmatrix sei $V(u) = W$ mit

$$W = diag(\sigma_1^2, \dots, \sigma_T^2).$$

Der OLS-Schätzer

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

hat die Kovarianzmatrix

$$Cov(\hat{\beta}) = (X'X)^{-1}X'WX'(X'X)^{-1}.$$

Eine konsistente Schätzung von W ist nicht möglich. White (1980) hat jedoch gezeigt, dass eine konsistente Schätzung von

$$\begin{aligned}\Sigma &= \frac{1}{T}X'WX \\ &= \frac{1}{T}\sum_{t=1}^{T}\sigma_t^2 x_t x_t'\end{aligned}$$

möglich ist, wobei x_t die t-te Zeile der Matrix X ist (als Spaltenvektor geschrieben).

Ein konsistenter Schätzer von Σ ist

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 x_t x_t'.$$

Die geschätzte Kovarianzmatrix ist

$$\hat{V}(\hat{\beta}) = (X'X)^{-1}X'\hat{W}X(X'X)^{-1}$$

mit

$$\hat{W} = \begin{bmatrix} \hat{u}_1^2 & & \\ & \ddots & \\ & & \hat{u}_T^2 \end{bmatrix}$$

Wegen seiner Form nennt man diesen Schätzer auch Sandwich-Schätzer.

Illustration [rentexample.R]

18 Autokorrelation

Annahme B3: Die Störterme sind unkorreliert,

$$Cov(u_t, u_s) = 0$$

für alle $t \neq s$.

Beispiel: [waterfilter.R]: Die Nachfragefunktion nach Wasserfiltern sei

$$y_t = \alpha + \beta x_t + u_t,$$

mit y_t verkaufter Menge, x_t Preis für die Monate Januar 2001 bis Dezember 2002.

Wir nehmen an, dass die Autokorrelation folgende Form hat:

$$u_t = \rho u_{t-1} + e_t$$

mit $-1 < \rho < 1$. Dabei sei

$$e_t \sim NID(0, \sigma_e^2)$$

Eigenschaften der Störterme u_t:

Der Erwartungswert (von dem wir annehmen, dass er zeitinvariant ist) beträgt

$$\begin{aligned}
E(u_t) &= E(\rho u_{t-1} + e_t) \\
&= \rho E(u_{t-1}) \\
E(u_t) &= 0.
\end{aligned}$$

Die Varianz (von der wir ebenfalls annehmen, dass sie zeitinvariant ist) beträgt

$$\begin{aligned}
Var(u_t) &= Var(\rho u_{t-1} + e_t) \\
&= \rho^2 Var(u_{t-1}) + \sigma_e^2 \\
Var(u_t) &= \sigma_e^2 / \left(1 - \rho^2\right)
\end{aligned}$$

und die Kovarianz (der Ordnung 1) ist

$$\begin{aligned}
Cov(u_t, u_{t-1}) &= E(u_t u_{t-1}) \\
&= E\left((\rho u_{t-1} + e_t)\, u_{t-1}\right) \\
&= \rho E(u_{t-1}^2) + E(e_t u_{t-1}) \\
&= \rho \cdot Var(u_t) \\
&= \rho \sigma_e^2 / \left(1 - \rho^2\right)
\end{aligned}$$

Für höhere Ordnung ergibt sich rekursiv

$$Cov(u_t, u_{t-j}) = \rho^j \left(\frac{\sigma_e^2}{1 - \rho^2}\right).$$

Folgerung: B1, B2 und B4 sind weiterhin erfüllt. Nur B3 ist verletzt!

Transformation des Modells:

Betrachte

$$y_t = \alpha + \beta x_t + u_t$$

mit

$$u_t = \rho u_{t-1} + e_t$$

und $e_t \sim UN(0, \sigma_e^2)$. Addiert man

$$-\rho y_{t-1} = -\rho\alpha - \rho\beta x_{t-1} - \rho u_{t-1},$$

so erhält man

$$\begin{aligned} y_t - \rho y_{t-1} &= \alpha - \rho\alpha + \beta x_t - \rho\beta x_{t-1} + u_t - \rho u_{t-1} \\ &= \alpha\,(1-\rho) + \beta\,(x_t - \rho x_{t-1}) + e_t \\ y_t^* &= \alpha^* + \beta x_t^* + e_t \end{aligned}$$

Der Störterm des transformierten Modells erfüllt alle Modellannahmen (insbesondere B1 bis B4). Allerding muss man für die Transformation den Wert von ρ kennen.

Da die Schätzer für α und β des transformierten Modells anders aussehen als die KQ-Schätzer, sind die KQ-Schätzer ineffizient. Außerdem sind die üblichen KQ-Formeln

$$Var(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$$

und

$$\hat{\sigma}^2 = \frac{S_{\hat{u}\hat{u}}}{T-2}$$

nicht korrekt. Die Folgen sind also identisch zum Fall der Heteroskedastizität.

Diagnose von Autokorrelation:

Einfachster Weg: Graph der Residuen $\hat{u}_t$ über die Zeit oder Streudiagramm der Paare $(\hat{u}_{t-1}, \hat{u}_t)$. Beispiel

Schätzung von ρ: Wegen $u_t = \rho u_{t-1} + e_t$ kann man ρ durch die Regression

$$\widehat{u}_t = \rho\widehat{u}_{t-1} + e_t^*$$

schätzen. Der KQ-Schätzer lautet

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=2}^{T} \widehat{u}_{t-1}^2}.$$

Achtung: Wegen der Zweistufigkeit ist der gewöhnliche t-Test nicht gültig.

Durbin-Watson-Test

Schritt 1: Aufstellen der Hypothesen

$$\begin{aligned} H_0 &: \rho \leq 0 \\ H_1 &: \rho > 0 \end{aligned}$$

Schritt 2: Berechne die Durbin-Watson-Teststatistik

$$d = \frac{\sum_{t=2}^{T} (\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=1}^{T} \widehat{u}_t^2}$$

Umschreiben der Teststatistik ergibt

$$\begin{aligned} d &= \frac{\sum_{t=2}^{T} \left(\widehat{u}_t - \widehat{u}_{t-1}\right)^2}{\sum_{t=1}^{T} \widehat{u}_t^2} \\ &= \frac{\sum_{t=2}^{T} \left(\widehat{u}_t^2 - 2\widehat{u}_t \widehat{u}_{t-1} + \widehat{u}_{t-1}^2\right)}{\sum_{t=1}^{T} \widehat{u}_t^2} \\ &= \frac{\sum_{t=2}^{T} \widehat{u}_t^2}{\sum_{t=1}^{T} \widehat{u}_t^2} - 2\frac{\sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=1}^{T} \widehat{u}_t^2} + \frac{\sum_{t=2}^{T} \widehat{u}_{t-1}^2}{\sum_{t=1}^{T} \widehat{u}_t^2} \end{aligned}$$

Also gilt

$$\frac{\sum_{t=2}^{T} \widehat{u}_t^2}{\sum_{t=1}^{T} \widehat{u}_t^2} \approx 1 \quad \text{und} \quad \frac{\sum_{t=2}^{T} \widehat{u}_{t-1}^2}{\sum_{t=1}^{T} \widehat{u}_t^2} \approx 1.$$

Außerdem ist

$$\frac{\sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=1}^{T} \widehat{u}_t^2} \approx \frac{\sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=2}^{T} \widehat{u}_{t-1}^2} = \hat{\rho}.$$

Folglich ist

$$d \approx 2\,(1 - \hat{\rho})\,.$$

Schritt 3: Finde den kritischen Wert d_a (mit Hilfe von Computersoftware). Wenn $d < d_a$, wird H_0 verworfen.

Problem: Der kritische Wert d_a hängt von X ab. Wenn die Software den tatsächlichen kritischen Wert d_a nicht bestimmen kann, gibt es Tabellen für eine Abschätzung des kritischen Werts und eine Obergrenze d_a^H und eine Untergrenze d_a^L für d_a.

Schritt 4: Vergleiche die Teststatistik d mit d_a^L und d_a^H.

Entscheidungregeln:

- Wenn $d < d_{0,05}^L$, lehne $H_0 : \rho \leq 0$ ab.
- Wenn $d > d_{0,05}^H$, lehne $H_0 : \rho \leq 0$ nicht ab.
- Wenn $d_{0,05}^L \leq d \leq d_{0,05}^H$, muss die Entscheidung offen bleiben.

Nachteile des Durbin-Watson-Tests:

- In manchen Fällen ist keine Entscheidung möglich.
- Eine verzögerte endogene Variable ist nicht erlaubt (dazu später mehr).
- Der Test ist nur auf $AR(1)$-Prozesse zugeschnitten.

Es gibt eine Reihe von alternativen Tests auf Autokorrelation, die auch in vielen Software-Programmen verfügbar sind.

GLS und Autokorrelation

Im Regressionsmodell

$$y = X\beta + u$$

hat die Kovarianzmatrix der Störterme bei Autokorrelation die Form $V(u) = \sigma^2\Omega$ mit

$$\mathbf{\Omega} = \begin{bmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & 1 \end{bmatrix}.$$

Es gibt eine Matrix P, die $P'P = \Omega^{-1}$ erfüllt. Mit dieser Matrix wird das Modell transformiert. Man kann nachprüfen, dass

$$P = \frac{1}{\sqrt{1-\rho^2}} \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -\rho & 1 \end{bmatrix}.$$

Der GLS-Schätzer ist der gleiche wie im Fall der Heteroskedastizität,

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

und die Kovarianzmatrix lautet

$$V(\hat{\beta}^{GLS}) = \sigma^2(X'\Omega^{-1}X)^{-1}.$$

Als Schätzer für die Varianz der Störterme (bzw. die Skalierung der Kovarianzmatrix der Störterme) dient

$$\hat{\sigma}^2 = \frac{\hat{u}'\Omega^{-1}\hat{u}}{T-K-1}.$$

Achtung: GLS ist nicht möglich, weil ρ (und damit P) unbekannt ist.

Hildreth-Lu-Ansatz: Definiere ein feines Gitter für ρ über das Intervall $[-1;1]$; wähle das ρ, für das $\hat{\sigma}^2(1-\rho^2) = \hat{\sigma}_e^2$ minimal ist.

Cochrane-Orcutt-Ansatz: Schätze $\hat{\rho}$ aus den OLS-Residuen. Führe GLS mit $\hat{\rho}$ durch. Iteriere bis die Schätzung konvergiert.

Es ist auch möglich, eine Kovarianzmatrixschätzung durchzuführen, die sowohl heteroskedastizitätskonsistent als auch autokorrelationskonsistent ist (Newey and West, Econometrica 1987). Die Vorgehensweise ist ähnlich zum HC-Schätzer nach White.

19 Nicht normalverteilt Störterme

Annahme B4: Die Störterme sind normalverteilt.

Wir haben diese Annahme gebraucht, um die Normalverteilt der Schätzer $\hat{\beta}$ herzuleiten, um die t-Verteilung der Teststatistik des t-Tests herzuleiten und um die F-Verteilung der Teststatistik des F-Tests herzuleiten.

Erinnerung: Der Schätzer $\hat{\beta}$ ist ein linearer Schätzer,

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= Cy\end{aligned}$$

Für eine einzelne Komponente von $\hat{\beta}$ bedeutet das,

$$\hat{\beta}_k = \sum_{t=1}^{T} c_{kt} y_t.$$

Die Zufallsvariablen $y_1, \ldots, y_T$ sind stochastisch unabhängig. Folglich ist $\hat{\beta}_k$ die Summe von unabhängigen (aber nicht unbedingt identisch verteilten) Zufallsvariablen.

Zentraler Grenzwertsatz: Die Summe von vielen i.i.d. Zufallsvariablen ist approximativ normalverteilt. Der zentrale Grenzwertsatz gilt (unter bestimmten Bedingungen) auch, wenn die Summanden nicht identisch verteilt sind.

Ferner gilt: Der Zufallsvektor $\hat{\beta}$ ist approximativ multivariat normalverteilt,

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

Für die approximative Normalität müssen einige (schwache) Regularitätsbedingungen erfüllt sein. Die Normalität kann unter Umständen nicht gelten (aber gewöhnlich tut sie es).

Simulation [b4.R]: Trinkgeldbeispiel (aus dem letzten Semester):

$$y_t = 0.5 + 0.1 \cdot x_t + u_t$$

erfülle alle A-, B-, C-Annahmen außer B4. Wir nehmen an, dass die Störterme die Dichtefunktion

$$f_{u_t}(u) = \exp(-(u+1))$$

haben.

Wegen

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

gilt für eine einzelne Komponente $\hat{\beta}_k$

$$\frac{\hat{\beta}_k - \beta_k}{SE(\hat{\beta}_k)} \xrightarrow{d} U \sim N(0,1)$$

für $k = 1, \ldots, K$.

Folglich sind Konfidenzintervalle für t-Tests asymptotisch gültig (man benutzt die Quantile der $N(0,1)$ anstelle der t-Verteilung). Auch die F-Tests sind asymptotisch gültig (sie konvergieren gegen die χ^2-Verteilung).

Stochastische Konvergenz

Definition: Die reelle Folge $\{a_n\}_{n\in\mathbb{N}}$ konvergiert gegen ihren Limes a, wenn für jedes (beliebig kleine) $\varepsilon > 0$ ein $N(\varepsilon)$ existiert, so dass $|a_n - a| < \varepsilon$ für alle $n \geq N(\varepsilon)$.

Notation: $\lim_{n\to\infty} a_n = a$ oder $a_n \to a$.

Beispiele:

$$\lim_{n\to\infty} 1/n = 0$$
$$\lim_{n\to\infty} \left[(n^2+n+6)/(3n^2-2n+2)\right] = 1/3$$

Beispiel: Graph der konvergenten Folge $(n^2+n+6)/(3n^2-2n+2)$

Fragen:

- Wie kann man die Idee von Konvergenz auf Folgen von Zufallsvariablen übertragen?
- Was ist eine Folge von Zufallsvariablen?
- Welche Folgen von Zufallsvariablen treten in der Ökonometrie typischerweise auf?

Definition: Seien $X_1, X_2, \ldots$ Zufallsvariablen

$$X_i : \Omega \to \mathbb{R}.$$

Wir nennen $X_1, X_2, \ldots$ eine Folge von Zufallsvariablen. $X_1, X_2, \ldots$ sind (abzählbar unendlich viele) multivariate Zufallsvariable.

Formal handelt es sich um eine Folge von Funktionen (*nicht* von reellen Zahlen).

Definition: Die Folge $X_1, X_2, \ldots$ konvergiert *fast sicher (almost surely)* gegen eine Zufallsvariable X, wenn

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

Notation:

$$X_n \overset{f.s.}{\to} X$$
$$X_n \overset{a.s.}{\to} X$$

Diese Art von Konvergenz spielt in der Ökonometrie meist keine große Rolle.

Definition: Die Folge $X_1, X_2, \ldots$ konvergiert *nach Wahrscheinlichkeit (in probability)* gegen eine Zufallsvariable X, wenn

$$\lim_{n\to\infty} P\left(|X_n - X| < \varepsilon\right) = 1.$$

Notation:

$$X_n \overset{p}{\to} X$$
$$\text{plim } X_n = X$$

Die Art von Konvergenz spielt in der Ökonometrie eine sehr wichtige Rolle.

Spezialfall: Konvergenz nach Wahrscheinlichkeit gegen eine Konstante a liegt vor wenn,

$$\lim_{n\to\infty} P\left(|X_n - a| < \varepsilon\right) = 1.$$

Notation:

$$\begin{aligned} X_n &\xrightarrow{p} a \\ \operatorname{plim} X_n &= a \end{aligned}$$

Dies ist der Fall, den wir in der Ökonometrie am häufigsten benötigen.

Definition: Die Folge $X_1, X_2, \ldots$ (mit den Verteilungsfunktionen $F_1, F_2, \ldots$) konvergiert *nach Verteilung (in distribution, in law)* gegen eine Zufallsvariable X (mit Verteilungsfunktion F), wenn

$$\lim_{n\to\infty} F_n(x) = F(x)$$

für alle $x \in \mathbb{R}$, an denen $F(x)$ stetig ist.

Notation:

$$X_n \xrightarrow{d} X$$

Es gibt noch weitere Definitionen von stochastischer Konvergenz, die wir aber nicht betrachten.

Beziehungen zwischen den Arten von Konvergenz.

$$X_n \xrightarrow{f.s.} X \quad \Rightarrow \quad X_n \xrightarrow{p} X \quad \Rightarrow \quad X_n \xrightarrow{d} X.$$

Grenzwertsätze

Es gibt zwei wichtige Klassen von Grenzwertsätzen, nämlich zum einen die Gesetze der großen Zahlen (laws of large numbers) und zum anderen die zentralen Grenzwertsätze (central limit theorems).

Sei $X_1, X_2, \ldots$ eine Folge von Zufallsvariablen. Definiere eine neue Folge $\bar{X}_1, \bar{X}_2, \ldots$ mit

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Starkes Gesetz der großen Zahl: Sei $X_1, X_2, \ldots$ eine Folge von Zufallsvariablen mit $\mu_i = E(X_i) < \infty$ und $Var(X_i) < \infty$ für $i = 1, 2, \ldots$ Wenn $\sum_{k=1}^{\infty} Var(X_k)/k^2 < \infty$, dann gilt

$$P\left(\lim_{n\to\infty}\left(\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right) = 0\right) = 1.$$

Spezialfall: i.i.d.-Folge, $\bar{X}_n \xrightarrow{f.s.} \mu$.

Schwaches Gesetz der großen Zahl (Chebyshev): Sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvariablen mit $\mu_i = E(X_i) < \infty$ und $Var(X_i) < c < \infty$. Dann ist

$$\lim_{n\to\infty} P\left(\left|\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| < \varepsilon\right) = 1.$$

Spezialfall: i.i.d.-Folge, plim $\bar{X}_n = \mu$.

Schwaches Gesetz der großen Zahl (Khinchin): Sei $X_1, X_2, \ldots$ eine Folge von i.i.d. Zufallsvariablen mit $E(X_i) = \mu$. Dann gilt:

$$\lim_{n\to\infty} P\left(\left|\bar{X}_n - \mu\right| < \varepsilon\right) = 1.$$

Es gibt auch Gesetze der großen Zahl für stochastische Prozesse, z.B. für Martingaldifferenzenfolgen. Die Gesetze der großen Zahl können leicht auf den multivariaten Fall übertragen werden.

Zentraler Grenzwertsatz:

Sei $X_1, X_2, \ldots$ eine Folge von Zufallsvariablen. Betrachte die Folge der standardisierten kumulativen Summen

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} \qquad \text{mit} \quad S_n = \sum_{i=1}^{n} X_i$$

Wie ist Z_n für $n \to \infty$ verteilt? Wir machen nur wenige Annahmen über die Verteilung der X_is.

ZGS (Lindeberg-Levy):

Sei $X_1, X_2, \ldots$ eine Folge von i.i.d. Zufallsvariablen mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2 < \infty$. Mit $F_n(z) = P(Z_n \leq z)$ bezeichnen wir die Verteilungsfunktion von Z_n. Dann gilt

$$\lim_{n\to\infty} F_n(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du.$$

Konvergenz nach Verteilung: $Z_n \overset{d}{\to} Z \sim N(0,1)$.

ZGS (Liapunov):

Sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvariablen mit $E(X_i) = \mu_i$, $Var(X_i) = \sigma_i^2 < \infty$, und $E(|X_i|^{2+\delta}) < \infty$ für (beliebig kleines) $\delta > 0$. Definiere $c_n = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$. Wenn

$$\lim_{n\to\infty} \left(\frac{1}{c_n^{2+\delta}} \sum_{i=1}^{n} E\left(|X_i - \mu_i|\right)^{2+\delta}\right) = 0,$$

dann gilt $Z_n \overset{d}{\to} Z \sim N(0,1)$.

Bedeutung der Annahme: keine einzelne Zufallsvariable dominiert die Summe. Jedes $(X_i - \mu_i)/\sigma_i$ liefert nur einen kleinen Anteil zu der Summe $(S_n - E(S_n))/c_n$.

Häufige Notation (im i.i.d. Fall):

$$S_n \overset{appr}{\sim} N(n\mu, n\sigma^2)$$
$$\bar{X}_n \overset{appr}{\sim} N(\mu, \sigma^2/n)$$

Wir können die Summe als normalverteilt betrachten (wenn n groß genug ist).

Nützliche Rechenregeln für stochastische Konvergenz: Wenn $\text{plim } X_n = a$ und $\text{plim } Y_n = b$, dann gilt

$$\begin{aligned} \text{plim } (X_n \pm Y_n) &= a \pm b \\ \text{plim } (X_n Y_n) &= ab \\ \text{plim } \left(\frac{X_n}{Y_n}\right) &= \frac{a}{b}, \text{ wenn } b \neq 0 \end{aligned}$$

Wenn eine Funktion g an der Stelle a stetig ist, dann gilt

$$\text{plim } g(X_n) = g(a).$$

Wenn $Y_n \xrightarrow{d} Z$ und h ist eine stetige Funktion, dann

$$h(Y_n) \xrightarrow{d} h(Z).$$

Cramérs Theorem: Wenn $X_n \xrightarrow{p} a$ und $Y_n \xrightarrow{d} Z$, dann

$$\begin{aligned} X_n + Y_n &\xrightarrow{d} a + Z \\ X_n Y_n &\xrightarrow{d} aZ \end{aligned}$$

Cramérs Theorem ist sehr nützlich, wenn es unbekannte Parameter in einer asymptotischen Verteilung gibt, die man konsistent schätzen kann (dazu später mehr).

Beispiel für Cramérs Theorem:

Sei $X_1, \ldots, X_n$ eine Zufallsstichprobe aus X; wir wissen, dass

$$\begin{aligned} S_n^{*2} &= \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \xrightarrow{p} \sigma^2 \\ S_n^2 &= \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \xrightarrow{p} \sigma^2 \end{aligned}$$

Also

$$\frac{\sigma}{S_n^*} \xrightarrow{p} 1 \quad \text{und} \quad \frac{\sigma}{S_n} \xrightarrow{p} 1.$$

Laut ZGS

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} Z \sim N(0,1).$$

Wegen

$$\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \cdot \frac{\sigma}{S_n}$$

und $\sigma / S_n \xrightarrow{p} 1$ ergibt sich

$$\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \xrightarrow{d} Z \cdot 1 = Z \sim N\left(0, 1\right).$$

Für $\sqrt{n}(\bar{X}_n - \mu)/S_n^*$ geht man analog vor.

20 Stochastische exogene Variable

Annahme C1: Die Matrix X ist nicht stochastisch.

Was passiert, wenn X (zumindest einige Elemente) stochastisch ist? Wir unterscheiden drei Fälle:

1. X und u sind stochastisch unabhängig.
2. Kontemporäre Unkorreliertheit: $Cov(x_{kt}, u_t) = 0$ für alle t, k
3. X und u sind kontemporär korreliert.

Exkurs zu bedingten Erwartungen und Erwartungswerten

Sei (X, Y) gemeinsam stetig verteilt mit der Dichtefunktion $f_{X,Y}(x, y)$. Die Randdichten sind

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx. \end{aligned}$$

Die bedingte Dichte von X gegeben $Y = y$ ist

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Der bedingte Erwartungswert von X gegeben $Y = y$ ist

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y=y}(x)\, dx.$$

Die bedingte Erwartung von X gegeben Y,

$$E(X|Y),$$

ist eine Zufallsvariable mit der Realisation $E(X|Y = y)$ wenn $Y = y$ ist.

Der bedingte Erwartungswert $E(X|Y = y)$ ist eine reelle Zahl (für gegebenes y). Die bedingte Erwartung $E(X|Y)$ ist eine Zufallsvariable.

Nützliche Rechenregeln für bedingte Erwartungen:

1. Gesetz der iterierten Erwartung (law of iterated expectations):
$$E(E(X|Y)) = E(X)$$
2. Unabhängigkeit: Wenn X und Y unabhängig sind, dann gilt
$$E(X|Y) = E(X)$$
3. Linearität: Für $a_1, a_2 \in \mathbb{R}$, gilt
$$E(a_1X_1 + a_2X_2|Y) = a_1E(X_1|Y) + a_2E(X_2|Y)$$
4. Bedingte Zufallsvariablen können wir Konstanten behandelt werden,
$$E(f(X)g(Y)|Y) = g(Y)E(f(X)|Y)$$

Stochastische exogene Variable, Fall 1

Wir betrachten das Modell $y = X\beta + u$ mit X und u stochastisch unabhängig.
Die Schätzer $\hat{\beta}$ und $\hat{\sigma}^2$ sind weiterhin unverzerrt,

$$\begin{aligned}
\hat{\beta} &= \beta + (X'X)^{-1}X'u \\
E(\hat{\beta}) &= \beta + E\left((X'X)^{-1}X'\right)E(u) \\
&= \beta.
\end{aligned}$$

Die Varianz der Störterme wird geschätzt durch

$$\begin{aligned}
\hat{\sigma}^2 &= \hat{u}'\hat{u}/(T-K-1) \\
E(\hat{\sigma}^2) &= E\left[E\left(\hat{\sigma}^2|X\right)\right] \\
&= E\left[E\left(\hat{u}'\hat{u}/(T-K-1)|X\right)\right] \\
&= E\left[\frac{1}{T-K-1}E\left(\hat{u}'\hat{u}|X\right)\right] \\
&= E\left[\frac{1}{T-K-1}\sigma^2\left(T-K-1\right)\right] \\
&= \sigma^2.
\end{aligned}$$

Die Kovarianzmatrix von $\hat{\beta}$ ist

$$\begin{aligned}
V(\hat{\beta}) &= E\left(\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)'\right) \\
&= E\left((X'X)^{-1}X'uu'X(X'X)^{-1}\right) \\
&= E\left[E\left((X'X)^{-1}X'uu'X(X'X)^{-1}|X\right)\right] \\
&= E\left[(X'X)^{-1}X'E(uu'|X)X(X'X)^{-1}\right] \\
&= E\left[(X'X)^{-1}X'\sigma^2 IX(X'X)^{-1}\right] \\
&= \sigma^2 E\left[(X'X)^{-1}\right].
\end{aligned}$$

Ein unverzerrter Schätzer dieser Kovarianzmatrix ist

$$\begin{aligned}
\hat{V}(\hat{\beta}) &= \hat{\sigma}^2(X'X)^{-1} \\
E(\hat{V}(\hat{\beta})) &= E\left(\hat{\sigma}^2(X'X)^{-1}\right) \\
&= E\left[E\left(\hat{\sigma}^2(X'X)^{-1}|X\right)\right] \\
&= E\left[E\left(\hat{\sigma}^2|X\right)(X'X)^{-1}\right] \\
&= E\left[\sigma^2(X'X)^{-1}\right] \\
&= \sigma^2 E\left[(X'X)^{-1}\right] \\
&= V(\hat{\beta}).
\end{aligned}$$

Der Schätzer $\hat{\beta}$ ist konsistent, denn

$$\begin{aligned}
\operatorname{plim} \hat{\beta} &= \operatorname{plim} \left[\beta + (X'X)^{-1} X'u\right] \\
&= \operatorname{plim} (\beta) + \operatorname{plim} \left((X'X)^{-1} X'u\right) \\
&= \beta + \operatorname{plim} \left(\left(\frac{X'X}{T}\right)^{-1} \frac{X'u}{T}\right)
\end{aligned}$$

Wir brauchen nun zusätzliche Annahmen. Wir nehmen an, dass

$$\begin{aligned}
\operatorname{plim} \left(\frac{X'X}{T}\right) &= Q_{XX} \\
\lim_{T\to\infty} E\left(\frac{X'X}{T}\right) &= Q_{XX}
\end{aligned}$$

Mit diesen Annahmen werden Trends ausgeschlossen (und zwar sowohl deterministische Trends als auch stochastische Trends).

Unter dieser Annahme gilt

$$\operatorname{plim} \left(\frac{X'u}{T}\right) = 0.$$

Für den Beweis der Konvergenz nach Wahrscheinlichkeit einer Folge von Zufallsvariablen gegen eine Konstante a reicht es aus zu zeigen, dass der Erwartungswert gegen a und die Varianz gegen 0 konvergieren (diese Bedingung ist nicht notwendig, aber hinreichend).

Wegen der Unabhängigkeit gilt

$$E(X'u/T) = E(X)'E(u)/T = 0.$$

Außerdem ist

$$\begin{aligned}
V\left(X'u/T\right) &= \frac{1}{T^2} E\left(X'uu'X\right) \\
&= \frac{1}{T^2} E\left(E\left(X'uu'X|X\right)\right) \\
&= \frac{1}{T^2} E\left(X'E\left(uu'|X\right)X\right) \\
&= \frac{1}{T^2} E\left(X'\sigma^2 IX\right) \\
&= \frac{1}{T}\sigma^2 E\left(\frac{X'X}{T}\right),
\end{aligned}$$

also

$$\lim_{T\to\infty} V\left(X'u/T\right) = 0.$$

Damit ist die Konvergenz nach Wahrscheinlichkeit von $X'u/T$ gegen 0 bewiesen. Also gilt

$$\begin{aligned}
\operatorname{plim}\left(\left(\frac{X'X}{T}\right)^{-1}\frac{X'u}{T}\right) &= \operatorname{plim}\left(\left(\frac{X'X}{T}\right)^{-1}\right)\operatorname{plim}\left(\frac{X'u}{T}\right)\\
&= \left(\operatorname{plim}\left(\frac{X'X}{T}\right)\right)^{-1}\operatorname{plim}\left(\frac{X'u}{T}\right)\\
&= Q_{XX}^{-1}\cdot 0\\
&= 0.
\end{aligned}$$

Daraus folgt, dass $\hat{\beta}$ ein konsistenter Schätzer für β ist.

Wir beweisen nun die Konsistenz von $\hat{\sigma}^2 = \hat{u}'\hat{u}/(T-K-1)$. Asymptotisch spielt es keine Rolle ob wir durch $T-K-1$ oder T dividieren,

$$\begin{aligned}
\operatorname{plim}\hat{\sigma}^2 &= \operatorname{plim}\left(\hat{u}'\hat{u}/\left(T-K-1\right)\right)\\
&= \operatorname{plim}\left(\hat{u}'\hat{u}/T\right).
\end{aligned}$$

Aus dem letzten Semester kennen wir bereits die sogenannte Residuenmacher-Matrix (oder Dach-Matrix) $M = I - X(X'X)^{-1}X'$.

$$\begin{aligned}
\hat{u}'\hat{u} &= u'Mu\\
&= u'u - u'X\left(X'X\right)^{-1}X'u.
\end{aligned}$$

Also

$$\begin{aligned}
\operatorname{plim}\left(\hat{u}'\hat{u}/T\right) &= \operatorname{plim}\left(u'u/T\right) - \operatorname{plim}\left(u'X/T\right)\operatorname{plim}\left(X'X/T\right)^{-1}\operatorname{plim}\left(X'u/T\right)\\
&= \sigma^2 - 0\cdot Q_{XX}^{-1}\cdot 0\\
&= \sigma^2.
\end{aligned}$$

Ein konsistenter Schätzer für $\sigma^2 Q_{XX}^{-1}$ ist

$$\begin{aligned}
\operatorname{plim}\left(\hat{\sigma}^2\left(\frac{X'X}{T}\right)^{-1}\right) &= \operatorname{plim}\left(\hat{\sigma}^2\right)\operatorname{plim}\left(\frac{X'X}{T}\right)^{-1}\\
&= \sigma^2 Q_{XX}^{-1}
\end{aligned}$$

Da asymptotisch

$$\sqrt{T}\left(\hat{\beta}-\beta\right) \sim N\left(0, \sigma^2 Q_{XX}^{-1}\right),$$

können wir die Verteilung von $\hat{\beta}$ approximieren durch

$$\hat{\beta} \sim N\left(\beta, \sigma^2 Q_{XX}^{-1}/T\right).$$

Ein geeigneter Schätzer für $\sigma_u^2 Q_{XX}^{-1}/T$ ist

$$\begin{aligned}
\hat{V}(\hat{\beta}) &= \hat{\sigma}^2\left(X'X/T\right)^{-1}/T\\
&= \hat{\sigma}^2\left(X'X\right)^{-1}.
\end{aligned}$$

Also ändert sich letztlich nichts! Wenn X und u stochastisch unabhängig sind, können wir alle Methoden weiterhin anwenden.

Stochastische exogene Variable, Fall 2

Im Fall 2 sind der Störtermvektor und die Matrix der exogenen Variablen kontemporär unkorreliert, aber sie könnten über die Zeit hinweg korreliert sein. Typischer Fall: Verzögerte endogene Variable als Regressor.

Die Unverzerrtheit geht in diesem Fall verloren, aber die Konsistenz und die asymptotische Normalverteilung bleiben weiterhin erhalten.

Um die Konsistenz zu beweisen, geht man praktisch genauso vor wie in Fall 1. Zu zeigen ist insbesondere, dass

$$\text{plim}\left(\frac{X'u}{T}\right) = 0$$

auch gilt, wenn X und u nicht unabhängig sind. Betrachte den Zufallsvektor $X'u/T$ elementweise,

$$X'u/T = \begin{bmatrix} \frac{1}{T}\sum u_t \\ \frac{1}{T}\sum x_{1t}u_t \\ \vdots \\ \frac{1}{T}\sum x_{Kt}u_t \end{bmatrix}.$$

Wenn ein Gesetz der großen Zahl gilt (und das nehmen wir einfach an), dann ist

$$\text{plim}\ \frac{1}{T}\sum x_{kt}u_t = E(x_k u).$$

Wegen der kontemporären Unkorreliertheit ist $E(x_k u) = E(x_k)E(u) = 0$. Damit ergibt sich das Resultat unmittelbar.

Schlussfolgerung: Wenn kontemporäre Korrelation vorliegt, kommt es kaum zu Problemen, wenn die Stichprobe nur groß genug ist.

Stochastische exogene Variable, Fall 3

In diesem Fall sind der Störterm und die exogenen Variablen auch kontemporär miteinander korreliert.

Graphisches Beispiel: Streudiagramm mit korrelierten Störtermen.

Wodurch kann es zu kontemporärer Korrelation kommen?

Fehler in den Variablen (Errors-in-variables). Beispiel:

$$y_t = \alpha + \beta x_t^* + e_t,$$

wobei x_t^* nicht direkt beobachtet werden kann. Stattdessen beobachten wir eine verrauschte Version der exogenen Variablen. Die Annahme C1 ist also verletzt.

$$x_t = x_t^* + v_t.$$

Annahme: $E(e_t v_t) = 0$ für alle s, t; beide Störterme sollen B1, B2, B3 erfüllen.

Das beobachtete Modell ist

$$\begin{aligned} y_t &= \alpha + \beta\left(x_t - v_t\right) + e_t \\ &= \alpha + \beta x_t + \underbrace{\left(e_t - \beta v_t\right)}_{=u_t} \end{aligned}$$

Eigenschaften von u_t:

B1

$$E(u_t) = 0$$

B2

$$V(u_t) = \sigma_e^2 + \beta^2 \sigma_v^2$$

B3

$$\begin{aligned} Cov(u_t, u_s) &= E(u_t u_s) \\ &= E\left(\left(e_t - \beta v_t\right)\left(e_s - \beta v_s\right)\right) \\ &= 0, \end{aligned}$$

aber die Störterme sind kontemporär korreliert mit der beobachteten exogenen Variable,

$$\begin{aligned} Cov(x_t, u_t) &= E\left(x_t u_t\right) \\ &= E\left(\left(x_t^* + v_t\right)\left(e_t - \beta v_t\right)\right) \\ &= E\left(x_t^* e_t - \beta x_t^* v_t + v_t e_t - \beta v_t^2\right) \\ &= -\beta \sigma_v^2 \\ &\neq 0. \end{aligned}$$

Simultanes Gleichungssystem: Die exogene Variable ist im Rahmen eines umfassenden ökonomischen Modells selbst endogen (das behandeln wir später noch ausführlicher).

Einfaches Beispiel: Makroökonomische Konsumfunktion

$$c_t = \alpha + \beta y_t + u_t$$

und

$$y_t = c_t + i_t.$$

Dann gilt

$$\begin{aligned} Cov\left(y_t, u_t\right) &= E\left(y_t u_t\right) \\ &= E\left(\left(c_t + i_t\right) u_t\right) \\ &= E\left(\left(\alpha + \beta y_t + u_t + i_t\right) u_t\right) \\ &= E\left(\alpha u_t + \beta u_t y_t + u_t^2 + u_t i_t\right) \\ &= \beta E(y_t u_t) + \sigma_u^2 \\ E\left(y_t u_t\right) &= \sigma_u^2 / \left(1 - \beta\right) \\ &\neq 0. \end{aligned}$$

Also kommt es zu kontemporärer Korrelation zwischen Störterm und endogener Variable.

Was sind die Folgen kontemporärer Korrelation für den OLS-Schätzer $\hat{\beta}$?

Der Schätzer ist verzerrt,

$$E(\hat{\beta}) \neq \beta.$$

Schlimmer noch: Der Schätzer ist inkonsistent. Sei

$$\text{plim}\left(\frac{X'u}{T}\right) =: q,$$

dann ist

$$\begin{aligned}
\text{plim}\ \hat{\beta} &= \beta + \text{plim}\left(\left(\frac{X'X}{T}\right)^{-1}\frac{X'u}{T}\right)\\
&= \beta + \text{plim}\left(\frac{X'X}{T}\right)^{-1}\text{plim}\left(\frac{X'u}{T}\right)\\
&= \beta + Q_{XX}^{-1}q\\
&\neq \beta.
\end{aligned}$$

Folglich hilft auch ein großer Stichprobenumfang nicht! Inkonsistenz eines Schätzer ist ein gravierendes Problem. Wie kann man das Problem lösen?

Instrumentvariablen (IV-Schätzung)

Wir betrachten das übliche Modell

$$y = X\beta + u$$

mit kontemporärer Korrelation zwischen X und u.

Als Instrumentvariable (oder Instrument) bezeichnet man eine Zufallsvariable Z, die kontemporär unkorreliert ist mit u, aber kontemporär korreliert ist mit X.

Sei Z eine $(T \times (L+1))$-Matrix von Instrumenten und

$$P = Z(Z'Z)^{-1}Z'.$$

Die Matrix P is symmetrisch und idempotent, d.h. $P'P = PP = P$. Es muss gelten $L \geq K$ (oft ist $L = K$).

Wir transformieren das Modell zu

$$Py = PX\beta + Pu.$$

Der KQ-Schätzer von β im transformierten Modell,

$$\begin{aligned}
\hat{\beta}^{IV} &= (X'P'PX)^{-1}X'P'Py\\
&= (X'PX)^{-1}X'Py
\end{aligned}$$

heißt IV-Schätzer.

Im Spezialfall $L = K$ gilt

$$\begin{aligned}
\hat{\beta}^{IV} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y\\
&= (Z'X)^{-1}(Z'Z)(X'Z)^{-1}X'Z(Z'Z)^{-1}Z'y\\
&= (Z'X)^{-1}Z'y
\end{aligned}$$

und in der einfachen linearen Regression ($L = K = 1$)

$$\hat{\beta}^{IV} = \frac{\sum (z_t - \bar{z})(y_t - \bar{y})}{\sum (z_t - \bar{z})(x_t - \bar{x})}.$$

Folgende Annahmen treffen wir über Z:

- Konvergenz:
$$\operatorname{plim} \frac{Z'Z}{T} = \lim_{T \to \infty} E\left(\frac{Z'Z}{T}\right) = Q_{ZZ}$$
mit Q_{ZZ} positiv definit.
- Asymptotische Korrelation mit exogenen Variablen
$$\operatorname{plim} \frac{Z'X}{T} = Q_{ZX}$$
mit
$$\operatorname{rang}(Q_{ZX}) = K + 1.$$
- Asymptotische Unkorreliertheit mit den Störtermen
$$\operatorname{plim} \frac{Z'u}{T} = \lim_{T \to \infty} E\left(\frac{Z'u}{T}\right) = 0$$

Unter diesen Annahmen ist der IV-Schätzer konsistent, aber nicht erwartungstreu. Um die Konsistenz zu zeigen, formen wir zuerst den Schätzer etwas um,

$$\begin{aligned}\hat{\beta}^{IV} &= (X'PX)^{-1}X'Py \\ &= \beta + (X'PX)^{-1}X'Pu.\end{aligned}$$

Also gilt

$$\begin{aligned}\operatorname{plim} \hat{\beta}^{IV} &= \beta + \operatorname{plim}\left[(X'PX)^{-1}X'Pu\right] \\ &= \beta + \operatorname{plim}\left[(X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'u\right] \\ &= \beta + \operatorname{plim}\left[\left(\frac{X'Z}{T}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'X}{T}\right)^{-1}\frac{X'Z}{T}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'u}{T}\right] \\ &= \beta + (Q'_{ZX}Q_{ZZ}^{-1}Q_{ZX})^{-1}Q'_{ZX}Q_{ZZ}^{-1}0 \\ &= \beta.\end{aligned}$$

Für den Beweis wird sowohl die asymptotische Unkorreliertheit von X und u als auch die asymptotische Korreliertheit von X und Z benötigt. Wenn die Instrumente schwach sind (d.h., $Q_{ZX} = 0$), gilt die Herleitung nicht mehr.

Der IV-Schätzer ist asymptotisch normalverteilt. Aus

$$\hat{\beta}^{IV} - \beta = (X'PX)^{-1}X'Pu$$

folgt

$$\sqrt{T}(\hat{\beta}^{IV} - \beta) = \left(\frac{X'PX}{T}\right)^{-1} \cdot T^{-1/2}X'Pu.$$

Der Term auf der rechten Seite kann wie folgt umgeschrieben werden,

$$\begin{aligned} T^{-1/2}X'Pu &= T^{-1/2}X'Z(Z'Z)^{-1}Z'u \\ &= \left(\frac{X'Z}{T}\right)\left(\frac{Z'Z}{T}\right)^{-1} T^{-1/2}Z'u. \end{aligned}$$

Für $T \to \infty$ konvergieren die Terme $\left(\frac{X'Z}{T}\right)$ und $\left(\frac{Z'Z}{T}\right)^{-1}$ nach Wahrscheinlichkeit gegen Q_{XZ} und Q_{ZZ}^{-1}. Der Faktor $T^{-1/2}Z'u$ kann elementweise geschrieben werden als:

$$T^{-1/2}Z'u = \begin{bmatrix} \sqrt{T}\frac{1}{T}\sum_{t=1}^{T} u_t \\ \sqrt{T}\frac{1}{T}\sum_{t=1}^{T} Z_{1t}u_t \\ \sqrt{T}\frac{1}{T}\sum_{t=1}^{T} Z_{2t}u_t \\ \vdots \\ \sqrt{T}\frac{1}{T}\sum_{t=1}^{T} Z_{Lt}u_t \end{bmatrix}.$$

Nach dem zentralen Grenzwertsatz und weil $E(Z_{kt}u_t) = 0$ ist, ist dieser Vektor asymptotisch normalverteilt mit Erwartungswertvektor 0 und Kovarianzmatrix

$$\sigma^2 \text{plim}\frac{Z'Z}{T} = \sigma^2 Q_{ZZ}.$$

Folglich gilt

$$\begin{aligned} T^{-1/2}X'Pu &= \left(\frac{X'Z}{T}\right)\left(\frac{Z'Z}{T}\right)^{-1} T^{-1/2}Z'u \\ &\to N\left(0, Q_{XZ}Q_{ZZ}^{-1}\sigma^2 Q_{ZZ}Q_{ZZ}^{-1}Q'_{XZ}\right) \\ &= N\left(0, \sigma^2 Q_{XZ}Q_{ZZ}^{-1}Q'_{XZ}\right). \end{aligned}$$

und schließlich

$$\begin{aligned} \sqrt{T}(\hat{\beta}^{IV} - \beta) &= \left(\frac{X'PX}{T}\right)^{-1} \cdot T^{-1/2}X'Pu \\ &= \left(\frac{X'Z}{T}\left(\frac{Z'Z}{T}\right)^{-1}\frac{Z'X}{T}\right)^{-1} \cdot T^{-1/2}X'Pu \\ &\to N\left(0, \left(Q_{X'Z}Q_{ZZ}^{-1}Q'_{XZ}\right)^{-1}\sigma^2 Q_{XZ}Q_{ZZ}^{-1}Q'_{XZ}\left(Q_{X'Z}Q_{ZZ}^{-1}Q'_{XZ}\right)^{-1}\right) \\ &= N\left(0, \sigma^2\left(Q_{X'Z}Q_{ZZ}^{-1}Q'_{XZ}\right)^{-1}\right). \end{aligned}$$

Ein Schätzer für die Kovarianzmatrix

$$V(\hat{\beta}^{IV}) = \sigma^2\left(Q_{X'Z}Q_{ZZ}^{-1}Q'_{XZ}\right)^{-1}/T$$

ist

$$
\begin{aligned}
\hat{V}(\hat{\beta}^{IV}) &= \hat{\sigma}^2 \left(\left(\frac{X'Z}{T}\right) \left(\frac{Z'Z}{T}\right)^{-1} \left(\frac{Z'X}{T}\right) \right)^{-1} /T \\
&= \hat{\sigma}^2 \left(X'Z\,(Z'Z)^{-1}\,Z'X \right)^{-1} \\
&= \hat{\sigma}^2\,(X'PX)^{-1}
\end{aligned}
$$

mit

$$
\begin{aligned}
\hat{\sigma}^2 &= \hat{u}^{IV\prime}\hat{u}^{IV}/(T-K-1) \\
\hat{u}^{IV} &= y - X\hat{\beta}^{IV}.
\end{aligned}
$$

Hausman test (Durbin-Wu-Hausman test)

Die Hypothesen lauten

$$
\begin{aligned}
H_0 &: \operatorname{plim} \frac{X'u}{T} = 0 \\
H_1 &: \operatorname{plim} \frac{X'u}{T} \neq 0.
\end{aligned}
$$

Testidee: Unter H_0 sind sowohl der OLS- als auch der IV-Schätzer konsistent, aber unter H_1 ist nur der IV-Schätzer konsistent. Lehne H_0 ab, wenn $\hat{\beta}^{IV}$ „zu weit” von $\hat{\beta}$ abweicht.

Die Teststatistik ist

$$
(\hat{\beta}^{IV} - \hat{\beta})' \left[\hat{V}(\hat{\beta}^{IV}) - \hat{V}(\hat{\beta}) \right]^{-1} (\hat{\beta}^{IV} - \hat{\beta}).
$$

Unter H_0 folgt die Teststatistik einer $\chi^2_{K^*}$-Verteilung, wobei K^* die Zahl der Spalten in Z ist, die nicht in X enthalten sind (also die “echten Instrumente”).

Multikollinearität

Perfekte Multikollinearität beruht praktisch immer auf einem Denkfehler bei der Variablendefinition, z.B. bei Dummy-Variablen oder Anteilen.

Bei imperfekter Multikollinearität lassen sich die Einflüsse der einzelnen exogenen Variablen auf die endogene Variable schlecht voneinander trennen. Ein wirkliches Problem liegt aber nicht vor, da alle Modellannahmen weiterhin erfüllt sind.

21 Dynamische Modelle

Notation und Begriffe:

- Stochastischer Prozess: $x_1, \ldots, x_T$
- Momentfunktionen: $E(x_t)$, $Var(x_t)$, $Cov(x_t, x_{t+\tau})$

- (Schwache) Stationarität

$$\begin{aligned} E(x_t) &= \mu \\ Var(x_t) &= \sigma_x^2 \\ Cov(x_t, x_{t+\tau}) &= \gamma_\tau \end{aligned}$$

- Integrationsordnung des Prozesses, $I(d)$

Das einfachste dynamische Modell hat verzögerte exogene Variable

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \ldots + \beta_K x_{t-K} + v_t$$

Aus den Parametern ergeben sich der kurzfristige und der langfristige Multiplikator. Im ungestörten langfristigen Gleichgewicht gilt

$$\begin{aligned} y_{t-1} = \tilde{y} &= \alpha + \beta_0 \tilde{x} + \ldots + \beta_K \tilde{x} \\ &= \alpha + \tilde{x} \sum_{k=0}^{K} \beta_k. \end{aligned}$$

Eine Erhöhung von $\tilde{x}$ um eine Einheit in Periode t ergibt

$$\begin{aligned} y_t &= \alpha + \beta_0 (\tilde{x} + 1) + \beta_1 \tilde{x} + \ldots + \beta_K \tilde{x} \\ &= \alpha + \tilde{x} \sum_{k=0}^{K} \beta_k + \beta_0 \\ &= y_{t-1} + \beta_0. \end{aligned}$$

Daher wird β_0 kurzfristiger Multiplikator genannt.

Im neuen ungestörten Gleichgewicht ist langfristig

$$\begin{aligned} y_{t+K} &= \alpha + \beta_0 (\tilde{x} + 1) + \beta_1 (\tilde{x} + 1) + \ldots + \beta_K (\tilde{x} + 1) \\ &= \alpha + \tilde{x} \sum_{k=0}^{K} \beta_k + \sum_{k=0}^{K} \beta_k \\ &= y_{t-1} + \sum_{k=0}^{K} \beta_k. \end{aligned}$$

Daher heißt $\sum_{k=0}^{K} \beta_k$ langfristiger Multiplikator.

Bei der Schätzung dieses dynamischen Modells kommt es zu mehreren Problemen:

- es gibt viele Parameter,
- Multikollinearität,
- eine präzise Schätzung der einzelnen Komponenten β_k ist nicht möglich.

Anmerkung: Die Varianz des langfristigen Multiplikators kann selbst dann klein sein, wenn alle Komponenten $\hat{\beta}_k$ eine große Varianz haben, denn

$$Var\left(\sum_{k=1}^{K}\hat{\beta}_k\right) = \sum_{k=1}^{K} Var\left(\hat{\beta}_k\right) + 2\sum_{j=0}^{K}\sum_{k=0}^{j-1} Cov\left(\hat{\beta}_j, \hat{\beta}_k\right).$$

Wenn die Kovarianzen negativ sind, bleibt die Varianz der Summe klein. Eine Umparametrisierung des Modells hilft bei der Schätzung des langfristigen Multiplikators. Wir schreiben:

$$\begin{aligned}
y_t &= \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \ldots + \beta_K x_{t-K} + v_t \\
&= \alpha + \beta_0 x_t \underline{-\beta_0 x_{t-1} + \beta_0 x_{t-1}} + \beta_1 x_{t-1} \\
&\quad + \ldots + \beta_K x_{t-K} + v_t \\
&= \alpha + \beta_0 \Delta x_t + \beta_0 x_{t-1} + \beta_1 x_{t-1} \\
&\quad + \ldots + \beta_K x_{t-K} + v_t \\
&= \alpha + \beta_0 \Delta x_t + (\beta_0 + \beta_1)\, x_{t-1} \underline{- (\beta_0 + \beta_1)\, x_{t-2} + (\beta_0 + \beta_1)\, x_{t-2}} + \beta_2 x_{t-2} \\
&\quad + \ldots + \beta_K x_{t-K} + v_t \\
&= \alpha + \beta_0 \Delta x_t + (\beta_0 + \beta_1)\, \Delta x_{t-1} + (\beta_0 + \beta_1 + \beta_2)\, x_{t-2} + \beta_3 x_{t-3} \\
&\quad + \ldots + \beta_K x_{t-K} + v_t \\
y_t &= \alpha + \delta_0 \Delta x_t + \delta_1 \Delta x_{t-1} + \delta_2 \Delta x_{t-2} + \ldots + \delta_{K-1} \Delta x_{t-k+1} + \delta_K x_{t-K} + u_t,
\end{aligned}$$

wobei

$$\begin{aligned}
\delta_0 &= \beta_0 \\
\delta_1 &= \beta_0 + \beta_1 \\
\delta_2 &= \beta_0 + \beta_1 + \beta_2 \\
&\vdots \\
\delta_K &= \beta_0 + \ldots + \beta_K.
\end{aligned}$$

In dieser Schreibweise erhält man den Standardfehler des langfristigen Multiplikators sofort.

Um die Zahl der zu schätzenden Parameter zu verringern, kann man eine funktionale Form für den Verlauf von $\beta_0, \beta_1, \ldots, \beta_K$ annehmen. Typische Annahmen sind:

- Polynom-Lags (Almon lags)
- Geometrische Lags (Koyck lags)

Polynom-Lags

Wir gehen davon aus, dass β_k polynomial in k ist, z.B. eine quadratische Funktion

$$\beta_k = \eta_0 + \eta_1 k + \eta_2 k^2$$

für $k = 0, \dots, K$. Es gibt nun weniger als K Parameter, denn

$$\begin{aligned}
y_t &= \alpha + \sum_{k=0}^{K} \beta_k x_{t-k} + v_t \\
&= \alpha + \sum_{k=0}^{K} \left(\eta_0 + \eta_1 k + \eta_2 k^2\right) x_{t-k} + v_t \\
&= \alpha + \eta_0 \sum_{k=0}^{K} x_{t-k} + \eta_1 \sum_{k=0}^{K} k x_{t-k} \\
&\quad + \eta_2 \sum_{k=0}^{K} k^2 x_{t-k} + v_t \\
&= \alpha + \eta_0 x_{1t}^* + \eta_1 x_{2t}^* + \eta_2 x_{3t}^* + v_t.
\end{aligned}$$

Die Gültigkeit der Restriktionen kann man wie üblich testen.

Geometrische Lags

Die β_k hängen nun wie folgt von k ab:

$$\beta_k = \beta_0 \lambda^k$$

mit $0 < \lambda < 1$. Es ist üblich, dass man $K = \infty$ setzt,

$$\begin{aligned}
y_t &= \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ldots + v_t \\
&= \alpha + \beta_0 x_t + \beta_0 \lambda x_{t-1} + \beta_0 \lambda^2 x_{t-2} + \ldots + v_t.
\end{aligned}$$

Der kurzfristige Multiplikator beträgt β_0. Der langfristige Multiplikator ist

$$\begin{aligned}
\sum_{k=0}^{\infty} \beta_k &= \beta_0 \sum_{k=0}^{\infty} \lambda^k \\
&= \beta_0 \frac{1}{1-\lambda}.
\end{aligned}$$

Koyck transformation:

$$y_t = \alpha + \beta_0 x_t + \beta_0 \lambda x_{t-1} + \beta_0 \lambda^2 x_{t-2} + \ldots + v_t$$

minus

$$\lambda y_{t-1} = \lambda \alpha + \beta_0 \lambda x_{t-1} + \beta_0 \lambda^2 x_{t-2} + \ldots + \lambda v_{t-1}$$

ergibt

$$\begin{aligned}
y_t - \lambda y_{t-1} &= (\alpha - \lambda \alpha) + \beta_0 x_t + (v_t - \lambda v_{t-1}) \\
y_t &= \alpha_0 + \beta_0 x_t + \lambda y_{t-1} + u_t.
\end{aligned}$$

Eine OLS-Schätzung ist problematisch, weil B3 und C1 verletzt sind.

Modell mit rationaler Lag-Verteilung

$$
\begin{aligned}
y_t = \alpha_0 + \beta_0 x_t + \mu_1 x_{t-1} + \ldots + \mu_K x_{t-K} \\
+ \lambda_1 y_{t-1} + \ldots + \lambda_M y_{t-M} + u_t.
\end{aligned}
$$

Spezialfall $K = M = 1$

$$y_t = \alpha_0 + \beta_0 x_t + \mu x_{t-1} + \lambda y_{t-1} + u_t$$

Aus

$$y_t = \alpha_0 + \beta_0 x_t + \mu x_{t-1} + \lambda y_{t-1} + u_t$$

folgt

$$y_t - \lambda y_{t-1} = \alpha_0 + \beta_0 x_t + \mu x_{t-1} + u_t$$

Langfristiges (ungestörtes) Gleichgewicht

$$y^* = \frac{\alpha_0}{1-\lambda} + \frac{\beta_0 + \mu}{1-\lambda} x^*.$$

Fehlerkorrekturmodell

Das Fehlerkorrekturmodell kann wie folgt hergeleitet werden. Subtrahiere y_{t-1} von

$$y_t = \alpha_0 + \beta_0 x_t + \mu x_{t-1} + \lambda y_{t-1} + u_t.$$

Es gilt

$$
\begin{aligned}
y_t - y_{t-1} &= \alpha_0 + \beta_0 x_t + \mu x_{t-1} + \lambda y_{t-1} - y_{t-1} + u_t \\
\Delta y_t &= \alpha_0 + \beta_0 x_t + \mu x_{t-1} - (1-\lambda)\, y_{t-1} + u_t \\
&= \alpha_0 + \beta_0 x_t \underline{- \beta_0 x_{t-1} + \beta_0 x_{t-1}} + \mu x_{t-1} - (1-\lambda)\, y_{t-1} + u_t \\
&= \alpha_0 + \beta_0 \Delta x_t + (\beta_0 + \mu)\, x_{t-1} - (1-\lambda)\, y_{t-1} + u_t \\
&= \beta_0 \Delta x_t + \left[-(1-\lambda)\, y_{t-1} + \alpha_0 + (\beta_0 + \mu)\, x_{t-1}\right] + u_t \\
&= \beta_0 \Delta x_t - (1-\lambda) \left[y_{t-1} - \frac{\alpha_0}{1-\lambda} - \frac{\beta_0 + \mu}{1-\lambda} x_{t-1} \right] + u_t.
\end{aligned}
$$

Fehlerkorrekturdarstellung

$$\Delta y_t = \beta_0 \Delta x_t - (1-\lambda)\, e_{t-1} + u_t$$

mit dem Ungleichgewichtsterm

$$e_{t-1} = y_{t-1} - \frac{\alpha_0}{1-\lambda} + \frac{\beta_0 + \mu}{1-\lambda} x_{t-1}.$$

Wenn x_t und y_t beide $I(1)$ sind und wenn e_{t-1} außerdem $I(0)$ ist, dann heißen x_t und y_t kointegriert.

Schätzung des Fehlerkorrekturmodells:

1. Bestimme die Integrationsordnung von x_t und y_t.

2. Schätze per OLS

$$y_{t-1} = \frac{\alpha_0}{1-\lambda} - \frac{\beta_0 + \mu}{1-\lambda} x_{t-1} + e_{t-1}$$

und berechne die Residuen $\hat{e}_{t-1}$.

3. Bestimme die Integrationsordnung von $\hat{e}_{t-1}$.

4. Wenn Kointegration vorliegt, schätze

$$\Delta y_t = \beta_0 \Delta x_t - (1-\lambda)\,\hat{e}_{t-1} + u_t$$

22 Simultane Gleichungssysteme

Ein einfaches Beispiel: In einer Pharma-Firma betrachten wir die Werbeausgaben w_t, die verkaufte Menge a_t, den Produktpreis p_t und den Werbepreis (pro Seite) q_t.

Modellgleichungen:

$$\begin{aligned} a_t &= \alpha + \beta_1 w_t + \beta_2 p_t + u_t \\ w_t &= \gamma + \delta_1 a_t + \delta_2 q_t + v_t \end{aligned}$$

Die Fehlerterme sollen alle B-Annahmen erfüllen. Außerdem nehmen wir an, dass $Cov(u_t, v_t) = \sigma_{uv}$ und $Cov(u_s, v_t) = 0$ für $s \neq t$. In der ersten Gleichung sind u_t und w_t korreliert! Folglich sind die OLS-Schätzer inkonsistent.

Strukturelle Form versus reduzierte Form:

Aus der strukturellen Form

$$\begin{aligned} a_t &= \alpha + \beta_1 w_t + \beta_2 p_t + u_t \\ w_t &= \gamma + \delta_1 a_t + \delta_2 q_t + v_t \end{aligned}$$

leiten wir die reduzierte Form ab: Das System

$$\begin{aligned} a_t &= \alpha + \beta_1 w_t + \beta_2 p_t + u_t \\ w_t &= \gamma + \delta_1 a_t + \delta_2 q_t + v_t \end{aligned}$$

löst man nach den endogenen Variablen a_t und w_t auf. Die erste Gleichung ist

$$\begin{aligned} a_t &= \alpha + \beta_1 w_t + \beta_2 p_t + u_t \\ &= \alpha + \beta_1 (\gamma + \delta_1 a_t + \delta_2 q_t + v_t) + \beta_2 p_t + u_t \\ &= \alpha + \beta_1 \gamma + \beta_1 \delta_1 a_t + \beta_1 \delta_2 q_t + \beta_1 v_t + \beta_2 p_t + u_t \\ a_t (1 - \beta_1 \delta_1) &= \alpha + \beta_1 \gamma + \beta_1 \delta_2 q_t + \beta_2 p_t + u_t + \beta_1 v_t \\ a_t &= \frac{\alpha + \beta_1 \gamma}{1 - \beta_1 \delta_1} + \frac{\beta_2}{1 - \beta_1 \delta_1} p_t + \frac{\beta_1 \delta_2}{1 - \beta_1 \delta_1} q_t + \frac{u_t + \beta_1 v_t}{1 - \beta_1 \delta_1} \end{aligned}$$

Analog ergibt sich für die zweite Gleichung

$$w_t = \frac{\gamma + \delta_1 \alpha}{1 - \beta_1 \delta_1} + \frac{\beta_2 \delta_1}{1 - \beta_1 \delta_1} p_t + \frac{\delta_2}{1 - \beta_1 \delta_1} q_t + \frac{\delta_1 u_t + v_t}{1 - \beta_1 \delta_1}.$$

Wir definieren eine neue Notation für die Parameter,

$$\begin{aligned} a_t &= \pi_1 + \pi_2 p_t + \pi_3 q_t + u_t^* \\ w_t &= \pi_4 + \pi_5 p_t + \pi_6 q_t + v_t^* \end{aligned}$$

mit

$$\begin{aligned} \pi_1 &= \frac{\alpha + \beta_1 \gamma}{1 - \beta_1 \delta_1} \\ \pi_2 &= \frac{\beta_2}{1 - \beta_1 \delta_1} \\ \pi_3 &= \frac{\beta_1 \delta_2}{1 - \beta_1 \delta_1} \\ \pi_4 &= \frac{\gamma + \delta_1 \alpha}{1 - \beta_1 \delta_1} \\ \pi_5 &= \frac{\beta_2 \delta_1}{1 - \beta_1 \delta_1} \\ \pi_6 &= \frac{\delta_2}{1 - \beta_1 \delta_1}. \end{aligned}$$

In der reduzierten Form stehen alle endogenen Variablen auf der linken Seite und alle exogenen Variablen auf der rechten Seite.

Die Gleichungen der reduzierten Form kann man mit OLS schätzen.

Die Schätzer $\hat{\pi}_1, \ldots, \hat{\pi}_6$ können zurück transformiert werden in die Schätzer $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}, \hat{\delta}_1, \hat{\delta}_2$, und zwar

$$\begin{aligned} \beta_1 &= \frac{\pi_3}{\pi_6} \\ \delta_1 &= \frac{\pi_5}{\pi_2}. \end{aligned}$$

Umschreiben ergibt

$$\begin{aligned} \alpha &= \pi_1 - \frac{\pi_3 \pi_4}{\pi_6} \\ \gamma &= \pi_4 - \frac{\pi_1 \pi_5}{\pi_2} \\ \beta_2 &= \pi_2 - \frac{\pi_3 \pi_5}{\pi_6} \\ \delta_2 &= \pi_6 - \frac{\pi_3 \pi_5}{\pi_2}. \end{aligned}$$

Die Schätzer $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}, \hat{\delta}_1, \hat{\delta}_2$ sind konsistent.

Manchmal lassen sich die strukturellen Parameter nicht aus den reduzierten Parametern herleiten (Identifikationsproblem), z.B. hat die kleinere strukturelle Form

$$\begin{aligned} a_t &= \alpha + \beta_1 w_t + \beta_2 p_t + u_t \\ w_t &= \gamma + \delta_1 a_t + v_t \end{aligned}$$

die reduzierte Form

$$\begin{aligned} a_t &= \pi_1 + \pi_2 p_t + u_t^* \\ w_t &= \pi_3 + \pi_4 p_t + v_t^*. \end{aligned}$$

Nun gibt es fünf strukturelle Parameter, aber nur vier reduzierte Parametern. Eine eindeutige Rücktransformation ist also unmöglich.

Manchmal gibt es auch mehr reduzierte Parameter als strukturelle Parameter.

Abzählkriterium: Sei

$$\begin{aligned} \dot{K} &= \text{Anzahl der exogenen Variablen im Gesamtmodell} \\ K^* &= \text{Anzahl der exogenen Variablen in der betrachteten Gleichung} \\ M^* &= \text{Anzahl der endogenen Variablen in der betrachteten Gleichung.} \end{aligned}$$

Eine Gleichung ist

$$\begin{array}{rcl} \text{unteridentifiziert, wenn } M^* - 1 & > & \dot{K} - K^* \\ \text{exakt identifiziert, wenn } M^* - 1 & = & \dot{K} - K^* \\ \text{überidentifiziert, wenn } M^* - 1 & < & \dot{K} - K^*. \end{array}$$

$M^* - 1$ ist die Anzahl der endogenen Regressoren (auf der rechten Seite); und $\dot{K} - K^*$ ist die Anzahl der exogenen Variablen in den anderen Gleichungen. Man braucht also mindestens eine exogene Variable aus den anderen Gleichungen für jeden endogenen Regressor.

Schätzung von exakt oder überidentifizierten Gleichungen

Die zweistufige Methode der kleinsten Quadrate (2SLS) hat folgende Grundidee: Generiere Instrumentvariablen aus der reduzierten Form.

Beispiel für 2SLS: Die zweite Gleichung von

$$\begin{aligned} a_t &= \alpha + \beta_1 w_t + \beta_2 p_t + u_t \\ w_t &= \gamma + \delta_1 a_t + \delta_2 q_t + v_t \end{aligned}$$

soll geschätzt werden.

Erste Stufe: Schätze

$$a_t = \pi_1 + \pi_2 p_t + \pi_3 q_t + u_t^*$$

per OLS und berechne $\hat{a}_t = \hat{\pi}_1 + \hat{\pi}_2 p_t + \hat{\pi}_3 q_t$.

Zweite Stufe: OLS-Schätzung von

$$w_t = \gamma + \delta_1 \hat{a}_t + \delta_2 q_t + v_t$$

Die 2SLS-Schätzer (IV-Schätzer) sind konsistent, aber die Standardfehler aus der zweiten Stufe müssen noch angepasst werden.

Die Schätzer haben eine einfache asymptotische Verteilung, aber die Verteilung der Schätzer in endlichen Stichproben ist kompliziert.

Matrixnotation für interdependente Gleichungssysteme

Allgemeine Notation: Sei M die Anzahl der Gleichungen im System. Die endogenen Variablen werden in der $(T \times M)$-Matrix

$$Y = [y_1 \quad y_2 \quad \ldots \quad y_M]$$

zusammengefasst. Die exogenen Variablen (incl. Achsenabschnitt) befinden sich in der $(T \times \dot{K})$-Matrix

$$X = [x_0 \quad x_1 \quad \ldots \quad x_K].$$

Die m-te Gleichung ist

$$\begin{aligned} y_m &= \alpha_m x_0 + \beta_{1m} x_1 + \beta_{2m} x_2 + \ldots + \beta_{Km} x_K \\ &\quad + \gamma_{1m} y_1 + \ldots + \gamma_{m-1m} y_{m-1} + \gamma_{m+1m} y_{m+1} + \ldots + \gamma_{Mm} y_M \\ &\quad + u_m. \end{aligned}$$

Setzt man $\gamma_{mm} = -1$, dann lässt sich die m-te Gleichung kompakt schreiben als

$$\gamma_{1m} y_1 + \ldots + \gamma_{Mm} y_M + \alpha_m x_0 + \beta_{1m} x_1 + \ldots + \beta_{Km} x_K + u_m = 0.$$

Wir definieren die Parametervektoren

$$\begin{aligned} \gamma_m &= (\gamma_{1m}, \gamma_{2m}, \ldots, \gamma_{Mm})' \\ \beta_m &= (\alpha_m, \beta_{1m}, \beta_{2m}, \ldots, \beta_{Km})'. \end{aligned}$$

Mit ihrer Hilfe kann man das komplette Gleichungssystem wie folgt schreiben

$$\begin{aligned} Y\gamma_1 + X\beta_1 + u_1 &= 0 \\ Y\gamma_2 + X\beta_2 + u_2 &= 0 \\ &\vdots \\ Y\gamma_M + X\beta_M + u_M &= 0 \end{aligned}$$

oder

$$Y\Gamma + XB + U = 0,$$

wobei (beachte die Dimensionen!)

$$\begin{aligned} \Gamma &= [\gamma_1 \quad \ldots \quad \gamma_M] \\ B &= [\beta_1 \quad \ldots \quad \beta_K] \\ U &= [u_1 \quad \ldots \quad u_M]. \end{aligned}$$

Die Störterme u_m, $m = 1, \ldots, M$, sollen alle B-Annahmen erfüllen. Es ist jedoch erlaubt, dass die Störterme verschiedener Gleichungen miteinander korreliert sind.

Annahmen:

$$\begin{aligned} E(u_m u_m') &= \sigma_m^2 I_T \qquad \text{für } m = 1, \ldots, M \\ E(u_m u_n') &= \sigma_{mn} I_T \qquad \text{für } m \neq n \end{aligned}$$

(Gibt es eine kompakte Notation für diese Annahmen für die Matrix U?)

Die reduzierte Form (alle endogenen Variablen stehen auf der linken Seite, alle exogenen Variablen rechts) lässt sich wie folgt bestimmen: Aus

$$Y\Gamma + XB + U = 0$$

folgt

$$Y\Gamma\Gamma^{-1} + XB\Gamma^{-1} + U\Gamma^{-1} = 0$$

oder

$$Y = X\Pi + V$$

mit $\Pi = -B\Gamma^{-1}$ und $V = -U\Gamma^{-1}$.

Die strukturellen Parameter in Γ und B können identifiziert werden, wenn sich ihre Werte aus den Parametern Π bestimmen lassen.

Anzahl der Parameter:

$$\begin{aligned} \Pi: &\quad \dot{K}M \\ \Gamma: &\quad M^2 - M \\ B: &\quad \dot{K}M \end{aligned}$$

Folglich brauchen wir mindestens $M^2 - M$ geeignete Restriktionen in Γ und/oder B.

Identifikation simultaner Gleichungssysteme

Im folgenden gehen wir davon aus, dass alle Restriktionen Null-Restriktionen sind. Das Modell ist

$$Y\Gamma + XB + U = 0.$$

Ohne Beschränkung der Allgemeinheit betrachten wir nur die erste Gleichung. Ferner seien die übrigen Gleichungen so sortiert, dass die $M_1^* - 1$ endogenen und K_1^* exogenen variablen der ersten Gleichung zuerst kommen (Y_1 oder X_1), danach folgen die $M - M_1^*$ endogenen und $\dot{K} - K_1^*$ exogenen Variablen, die nicht in der ersten Gleichung vorkommen.

$$[y \quad Y_1 \quad Y_2] \begin{bmatrix} -1 & \Gamma_{02} \\ \gamma_1 & \Gamma_{12} \\ 0 & \Gamma_{22} \end{bmatrix} + [X_1 \quad X_2] \begin{bmatrix} \beta_1 & B_{12} \\ 0 & B_{22} \end{bmatrix} + [u_1 \quad U_2] = 0.$$

Die Matrizen Γ und B seien analog partitioniert.

In dieser Notation lautet die erste Gleichung

$$-y + Y_1\gamma_1 + X_1\beta_1 + u_1 = 0$$

oder

$$y = Y_1\gamma_1 + X_1\beta_1 + u_1 = 0.$$

Das Endogenitätsproblem kann man durch eine IV-Schätzung lösen. Die $M_1^* - 1$ endogenen Variablen Y_1 müssen durch Instrumente ersetzt werden. Es gibt $\dot{K} - K_1^*$ Instrumente in X_2. Das Ordnungs- oder Abzählkriterium ist also

$$\dot{K} - K_1^* \geq M_1^* - 1.$$

Diese Bedingung ist jedoch nur eine notwendige Bedingung für die Identifikation. Die hinreichende Bedingung lautet, dass die Instrumente in X_2 tatsächlich valide für die IV-Schätzung sind.

Der erste 2SLS-Schritt ergibt sich aus der reduzierten Form

$$Y = X\left(B\Gamma^{-1}\right) + V$$

oder

$$[y \quad Y_1 \quad Y_2] = [X_1 \quad X_2]\left(\begin{bmatrix} \beta_1 & B_{12} \\ 0 & B_{22} \end{bmatrix}\Gamma^{-1}\right) + V.$$

Diese Gleichung kann man auch schreiben als

$$[y \quad Y_1 \quad Y_2] = [X_1 \quad X_2]\begin{bmatrix} \pi_1 & \Pi_{11} & \Pi_{12} \\ \pi_2 & \Pi_{21} & \Pi_{22} \end{bmatrix} + V.$$

Um sicherzustellen, dass alle Elemente von Y_1 mit X_2 korreliert sind, muss die zugehörige Partition der Matrix $B\Gamma^{-1}$ (d.h. Π_{21}) vollen Rang haben. Man spricht daher auch vom Rang-Kriterium. Man kann zeigen, dass diese Bedingung äquivalent ist zu

$$Rang\left(\begin{bmatrix} \Gamma_{22} \\ B_{22} \end{bmatrix}\right) = M - 1.$$

Schätzung interdependenter Gleichungssysteme

Reduzierte Form:

$$\begin{aligned} y_1 &= X\pi_1 + v_1 \\ &\vdots \\ y_M &= X\pi_M + v_M \end{aligned}$$

OLS-Schätzung der Gleichung m:

$$\hat{\pi}_m = (X'X)^{-1}X'y_m$$

OLS-Schätzung aller Gleichungen:

$$\hat{\Pi} = (X'X)^{-1}X'Y$$

Indirekte Kleinste-Quadrate-Methode: Wenn Gleichung m exakt identifiziert ist, kann man die strukturellen Parameter eindeutig aus der Matrix $\hat{\Pi}$ bestimmen.

Wenn die Gleichung m exakt oder überidentifiziert ist, kann man 2SLS verwenden. Dazu ordnet und partitioniert man die Matrizen, so dass

$$\left[y_m \quad \bar{Y}_m \quad \check{Y}_m\right]\begin{bmatrix} -1 \\ \bar{\gamma}_m \\ 0 \end{bmatrix} + X\beta_m + u_m = 0,$$

wobei $\bar{Y}_m$ die endogenen Regressoren aus Gleichung m enthält und $\check{Y}_m$ die nicht in Gleichung m enthaltenen endogenen variablen enthält.

In dieser Notation kann man Gleichung m so schreiben:

$$\begin{aligned} y_m &= \bar{Y}_m \bar{\gamma}_m + X\beta_m + u_m \\ &= \begin{bmatrix} \bar{Y}_m & X \end{bmatrix} \begin{bmatrix} \bar{\gamma}_m \\ \beta_m \end{bmatrix} + u_m \end{aligned}$$

Erster Schritt der 2SLS-Schätzung: Schätze

$$\hat{\Pi} = (X'X)^{-1}X'Y$$

bzw.

$$\begin{bmatrix} \hat{\pi}_m & \hat{\bar{\Pi}}_m & \hat{\check{\Pi}}_m \end{bmatrix} = (X'X)^{-1}X' \begin{bmatrix} y_m & \bar{Y}_m & \check{Y}_m \end{bmatrix}$$

Die Instrumente für die endogenen Regressoren sind

$$\hat{\bar{Y}}_m = X\hat{\bar{\Pi}}_m.$$

Zweiter Schritt: Ersetze die endogenen Regressoren $\bar{Y}_m$ in

$$y_m = \begin{bmatrix} \bar{Y}_m & X \end{bmatrix} \begin{bmatrix} \bar{\gamma}_m \\ \beta_m \end{bmatrix} + u_m$$

durch die Instrumente $\hat{\bar{Y}}_m$.

Der 2SLS-Schätzer lautet

$$\begin{bmatrix} \hat{\bar{\gamma}}_m^{ZSKQ} \\ \hat{\beta}_m^{ZSKQ} \end{bmatrix} = \left[\begin{bmatrix} \hat{\bar{Y}}_m & X \end{bmatrix}' \begin{bmatrix} \hat{\bar{Y}}_m & X \end{bmatrix} \right]^{-1} \begin{bmatrix} \hat{\bar{Y}}_m & X \end{bmatrix}' y_m$$

Die zugehörige Kovarianzmatrix des Schätzers

$$\begin{bmatrix} \hat{\bar{\gamma}}_m^{ZSKQ} \\ \hat{\beta}_m^{ZSKQ} \end{bmatrix}$$

ist

$$\hat{\sigma}^2 \left[\begin{bmatrix} \hat{\bar{Y}}_m & X \end{bmatrix}' \begin{bmatrix} \hat{\bar{Y}}_m & X \end{bmatrix} \right]^{-1}$$

mit

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \left(y_m - \begin{bmatrix} \bar{Y}_m & X \end{bmatrix} \begin{bmatrix} \hat{\bar{\gamma}}_m^{ZSKQ} \\ \hat{\beta}_m^{ZSKQ} \end{bmatrix} \right)^2$$

und NICHT

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \left(y_m - \begin{bmatrix} \hat{\bar{Y}}_m & X \end{bmatrix} \begin{bmatrix} \hat{\bar{\gamma}}_m^{ZSKQ} \\ \hat{\beta}_m^{ZSKQ} \end{bmatrix} \right)^2.$$

Weitere Themen in Ökonometrie

- Univariate Zeitreihenanalyse

- Multivariate Zeitreihenanalyse
- Paneldaten-Ökonometrie
- Bayesianische Ökonometrie und MCMC
- Qualitative abhängige Variable (Logit, Probit, Tobit, geordnetes Probit, multinomiales Probit, ...)
- Fortgeschrittene Schätzmethoden (GMM, MSM, Bootstrap, ...)
- Durationsanalyse
- Zähldaten; Stichprobenverfahren; fehlende Werte
- ...