INSTITUTE OF ECONOMIC STUDIES, FACULTY OF SOCIAL SCIENCES
CHARLES UNIVERSITY IN PRAGUE (*established 1348*)

Charles University

The emblem of

Charles University in Prague, founded 1348, April 7,

- the foundation documents were symbolically
and evidently humbly handed over

by **Charles the IV.**,
the Czech King and the Holy Roman Emperor,
so probably the most powerfull man of those days,

to the representative of Higher Power,
to the Knight of God Army, Saint Venceslav.

ROBUST STATISTICS AND ECONOMETRICS

INSTITUTE OF ECONOMIC STUDIES
FACULTY OF SOCIAL SCIENCES
CHARLES UNIVERSITY IN PRAGUE

JAN ÁMOS VÍŠEK

Week 11

Content of lecture

Regression model

Basic (technical) conditions of “classical” framework

- Orthogonality condition $\boldsymbol{E}\{\varepsilon\,|X\} = 0$
- Sphericality condition $\boldsymbol{E}\{\varepsilon \cdot \varepsilon'\,|X\} = \sigma^2 \boldsymbol{I}$
- $\boldsymbol{E}X \cdot X' = Q$, Q - regular matrix

$\hat{\beta}^{(OLS,n)}$ is BLUE

It means that one-eyed is among the blinds the king!!

Regression model

$\hat{\beta}^{(OLS,n)}$ is BLUE

Restriction on the *family of linear estimators* is sometimes justified
by the *linearity of model* !?!?

Karl Weierstrass (1885): Űber die analytische Darstellbarkeit sogenannter
willkűrlicher Functionen einer reellen Verǻnderlichen.
Sitzungsberichte der Kőniglich Preussischen
Akademie der Wissenschaften zu Berlin, 1885 (II), 633 - 639, 789 - 805.

The linearity of model doesn't represent signifficant restriction,
while <u>*the restriction on linear estimators is drastic !!*</u>

These two linearities has <u>*no interrelations !!*</u>

Regression framework

Really important condition

Normality of disturbances, i. e. $\mathcal{L}(\varepsilon) = \mathcal{N}(0, \sigma^2 \boldsymbol{I}) \quad \Rightarrow \quad \hat{\beta}^{(OLS,n)}$ is BUE

It means - the restriction on linear estimators disappeared !!
The normality is to be checked or reached ?!?

Prigogine, I., I. Stengers (1984):
Order out of Chaos. Man's New Dialog with Nature.
Bantam Books, New York.

Prigogine, I. (1982): *Only an Illusion*.
The Tanner Lectures on Human Values, Jawaharlal Nehru University.

Regression framework

The task is:

1. Not only to estimate the unknown coefficients
2. but also to confirm the assumed “shape” of model.

Under various structural frameworks:

- Static versus dynamic model (i. e. cross-sectional versus panel data)
- Modification of basic model (e. g. quantitave versus qualitative model)
- Dummy, proxy, limited, unobservable, latent (e. g. fixed versus random effects)
- Etc.

Regression framework

The task is:

1. Not only to estimate the unknown coefficients
2. but also to confirm the assumed “shape” of model

Under various assumptional frameworks:

- Random versus deterministic explanatory variables
- Orthogonality held or broken
- I. i. d. versus something else
  (e. g. homoscedasticity versus heteroscedasticity, ARMA, etc.)
- Normality versus anything else (e. g. heavy tails).
- Etc.

Regression framework

The task is:

1. Not only to estimate the unknown coefficients
2. but also to confirm the assumed “shape” of model

Under various estimation doctrines:

- Classical versus modern, especially robust (i. e. OLS, TLS, Maximum likelihood, (Generalized) Moment Method, Minimal distance, etc.)
- Equivariance and invariance
- Etc.

Qualitative information - about explanatory and/or about response variables

Examples of qualitative explanatory varaibles

1. Man $\rightarrow \quad x_{ij} = 0$, woman $\rightarrow \quad x_{ij} = 1$,
2. employed $\rightarrow \quad x_{ij} = 0$, unemployed $\rightarrow \quad x_{ij} = 1$,
3. single $\rightarrow \quad x_{ij} = 0$, married $\rightarrow \quad x_{ij} = 1$, divorsed $\rightarrow \quad x_{ij} = 2$.

Examples of qualitative response varaible

1. Passed the exam $\rightarrow \quad x_{ij} = 0$, failed $\rightarrow \quad x_{ij} = 1$,
2. won an opportunity $\rightarrow \quad x_{ij} = 0$, lost an opportunity $\rightarrow \quad x_{ij} = 1$,
3. bad performance $\rightarrow \quad x_{ij} = 0$, good performance $\rightarrow \quad x_{ij} = 1$, excellent performance $\rightarrow \quad x_{ij} = 2$.

Qualitative information - about response variable

Problem(s) with qualitative response variable

$$Y_i = X_i' \beta^0 + \varepsilon_i = \sum_{j=1}^{p} X_{ij} \beta_j^0 + \varepsilon_i, \quad i = 1, 2, ..., n$$

$$\mathcal{L}(\varepsilon_1) = \mathcal{N}(0, \sigma^2),$$

i. e. response variable is implicitly assumed to be continuous !!

Basic trick:

$$P(Y_i = 1) = F(X_i' \beta^0) \quad \rightarrow \quad \pi_i = P(Y_i = 1) + \varepsilon_i = F(X_i' \beta^0) + \varepsilon_i$$

where F is a d. f. and - for the case

of repeated observations of the i-th case

$$\pi_i = \frac{\sum_{k=1} Y_{ik}}{n_i}.$$

Qualitative information - about response variable

Qualitative response variable

- situation with one observation for each case:

Basic trick:

$$P(Y_i = 1) = F(X_i'\beta^0) \quad \text{and} \quad P(Y_i = 0) = 1 - F(X_i'\beta^0),$$

i. e.

$$P(Y_i = y) = F^y(X_i'\beta^0) \cdot (1 - F(X_i'\beta^0))^{(1-y)}$$

Then

$$\hat{\beta}^{(n)} = \underset{\beta \in R^p}{\arg\max} \prod_{i=1}^{n} \left[F^{y_i}(X_i'\beta) \cdot (1 - F(X_i'\beta))^{(1-y_i)} \right]$$

Qualitative information - about one explanatory variable

Why we can have problems with it and what problems?

Problems with qualitative explanatory variables - remember example

1. Man $\rightarrow$ $x_{ij} = 0$, woman $\rightarrow$ $x_{ij} = 1$

Data consist of two (several) groups.

As we want to take into account it, the fact of belonging to a given group is signifficant for explanation of data.

We can disaggregate data and to create model for individual groups.

Why we want to create model simultaneouly for the pooled data?

(The different groups have common structure, features, etc.)

Qualitative information - about one explanatory variable

The different groups have common structure, features, etc.

$\Rightarrow$ they have the same slopes coefficents in model.

But then the solution is straightforward - which?

$$\begin{bmatrix} 1, & 0, & x_{1,2}, & \ldots, & x_{1,k} \\ 1, & 0, & x_{2,2}, & \ldots, & x_{2,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1, & 0, & x_{\ell,2}, & \ldots, & x_{\ell,k} \\ 0, & 1, & x_{\ell+1,2}, & \ldots, & x_{\ell+1,k} \\ 0, & 1, & x_{\ell+2,2}, & \ldots, & x_{\ell+2,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0, & 1, & x_{n,2}, & \ldots, & x_{n,k} \end{bmatrix}$$

Qualitative information - about one explanatory variable

The different groups have common structure, features, etc.

$\Rightarrow$ they have the same slopes coefficents in model.

So we have the model - with $x_{i,0} = 1$ or 0 and $x_{i,1} = 0$ or 1

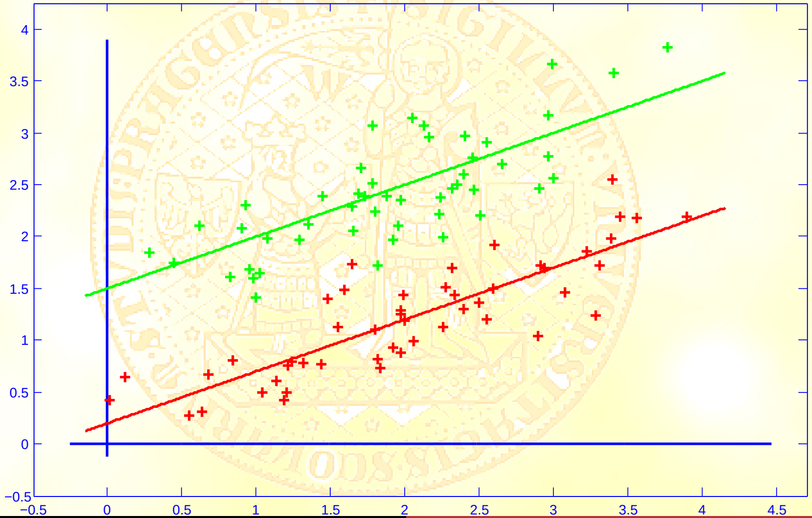
$$y_i = \beta_0^0 \cdot x_{i,0} + \beta_1^0 \cdot x_{i,1} + \beta_2^0 \cdot x_{i,2} + \ldots + \beta_k^0 \cdot x_{i,k},$$

with $\beta_0^0 \neq \beta_1^0$, i.e. with <u>two intercepts</u>.

What is a ("graphical") consequence for the model?

Draw a figure !!

Two intercepts - dummy for two groups of observations



Qualitative information - about one explanatory variable

The previous model is equivalent to model with design matrix - notice the first column

$$\begin{bmatrix} 1, & 0, & x_{2,1}, & \ldots, & x_{k,1} \\ 1, & 0, & x_{2,2}, & \ldots, & x_{k,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1, & 0, & x_{2,\ell}, & \ldots, & x_{k,\ell} \\ 1, & 1, & x_{2,\ell+1}, & \ldots, & x_{k,\ell+1} \\ 1, & 1, & x_{2,\ell+2}, & \ldots, & x_{k,\ell+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1, & 1, & x_{2,n}, & \ldots, & x_{k,n} \end{bmatrix}$$

Now we have the model - with $x_{i,0} = 1$ for all i's and $x_{i,1} = 0$ or 1

$$y_i = \beta_0^0 \cdot x_{i,0} + \beta_1^0 \cdot x_{i,1} + \beta_2^0 \cdot x_{i,2} + \ldots + \beta_k^0 \cdot x_{i,k},$$

with $\beta_0^0 \neq \beta_1^0$, i.e. with two intercepts - the first intercept is the same as in the previous model, the second one is difference between the second and the first intercept from the previous model.

Qualitative information - about one explanatory variable

So, the models with design matrices

$$\begin{bmatrix} 1, & 0, & x_{2,1}, & \ldots, & x_{k,1} \\ 1, & 0, & x_{2,2}, & \ldots, & x_{k,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1, & 0, & x_{2,\ell}, & \ldots, & x_{k,\ell} \\ 1, & 1, & x_{2,\ell+1}, & \ldots, & x_{k,\ell+1} \\ 1, & 1, & x_{2,\ell+2}, & \ldots, & x_{k,\ell+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1, & 1, & x_{2,n}, & \ldots, & x_{k,n} \end{bmatrix} \text{ and } \begin{bmatrix} 1, & 0, & x_{1,2}, & \ldots, & x_{1,k} \\ 1, & 0, & x_{2,2}, & \ldots, & x_{2,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1, & 0, & x_{\ell,2}, & \ldots, & x_{\ell,k} \\ 0, & 1, & x_{\ell+1,2}, & \ldots, & x_{\ell+1,k} \\ 0, & 1, & x_{\ell+2,2}, & \ldots, & x_{\ell+2,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0, & 1, & x_{n,2}, & \ldots, & x_{n,k} \end{bmatrix}$$

are equivalent but not the same.

Qualitative information - about more explanatory variables

If there are no relations between (among) them, the generalization is straightforward.

$$\begin{bmatrix}
1, & 0, & 1, & 0, & x_{4,1}, & \ldots, & x_{k,1} \\
1, & 0, & 1, & 0, & x_{4,2}, & \ldots, & x_{k,2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots, & \vdots \\
1, & 0, & 1, & 0, & x_{4,\ell_1}, & \ldots, & x_{k,\ell_1} \\
1, & 0, & 0, & 1, & x_{4,\ell_1+1}, & \ldots, & x_{k,\ell_1+1} \\
1, & 0, & 0, & 1, & x_{4,\ell_1+2}, & \ldots, & x_{k,\ell_1+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots, & \vdots \\
1, & 0, & 0, & 1, & x_{4,\ell_2}, & \ldots, & x_{k,\ell_2} \\
0, & 1, & 1, & 0, & x_{4,\ell_2+1}, & \ldots, & x_{k,\ell_2+1} \\
0, & 1, & 1, & 0, & x_{4,\ell_2+2}, & \ldots, & x_{k,\ell_2+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots, & \vdots \\
0, & 1, & 1, & 0, & x_{4,\ell_2+\ell_1}, & \ldots, & x_{k,\ell_2+\ell_1} \\
0, & 1, & 0, & 1, & x_{4,\ell_2+\ell_1+1}, & \ldots, & x_{k,\ell_2+\ell_1+1} \\
0, & 1, & 0, & 1, & x_{4,\ell_2+\ell_1+2}, & \ldots, & x_{k,\ell_2+\ell_1+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots, & \vdots \\
0, & 1, & 0, & 1, & x_{4,n}, & \ldots, & x_{k,n}
\end{bmatrix} = \begin{bmatrix}
\text{the first block} \\
\text{of } \ell_1 \text{ rows} \\
\text{where the first as well as} \\
\text{the second dummy} = 1 \\
\\
\text{the second block,} \\
\text{of } \ell_2 - \ell_1 \text{ rows,} \\
\text{the first dummy} = 1 \\
\text{while the second one} = 0 \\
\\
\text{the third block} \\
\text{of } \ell_1 \text{ rows,} \\
\text{the first dummy} = 0 \\
\text{while the second one} = 1 \\
\\
\text{the fourth block} \\
\text{of } \ell_2 - \ell_1 \text{ rows,} \\
\text{where the first as well as} \\
\text{the second dummy} = 0
\end{bmatrix}$$

Qualitative information - about more explanatory variables

Let's think about the model once again:

$$\begin{bmatrix}
1, & 0, & 1, & 0, & x_{4,1}, & \ldots, & x_{k,1} \\
1, & 0, & 1, & 0, & x_{4,2}, & \ldots, & x_{k,2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots, & \vdots \\
1, & 0, & 1, & 0, & x_{4,\ell_1}, & \ldots, & x_{k,\ell_1} \\
1, & 0, & 0, & 1, & x_{4,\ell_1+1}, & \ldots, & x_{k,\ell_1+1} \\
1, & 0, & 0, & 1, & x_{4,\ell_1+2}, & \ldots, & x_{k,\ell_1+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots, & \vdots \\
1, & 0, & 0, & 1, & x_{4,\ell_2}, & \ldots, & x_{k,\ell_2} \\
0, & 1, & 1, & 0, & x_{4,\ell_2+1}, & \ldots, & x_{k,\ell_2+1} \\
0, & 1, & 1, & 0, & x_{4,\ell_2+2}, & \ldots, & x_{k,\ell_2+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots, & \vdots \\
0, & 1, & 1, & 0, & x_{4,\ell_2+\ell_1}, & \ldots, & x_{k,\ell_2+\ell_1} \\
0, & 1, & 0, & 1, & x_{4,\ell_2+\ell_1+1}, & \ldots, & x_{k,\ell_2+\ell_1+1} \\
0, & 1, & 0, & 1, & x_{4,\ell_2+\ell_1+2}, & \ldots, & x_{k,\ell_2+\ell_1+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots, & \vdots \\
0, & 1, & 0, & 1, & x_{4,n}, & \ldots, & x_{k,n}
\end{bmatrix}$$

Plese realize, the first and the second columns are for the first dummy variable.

The third and the four columns are, for the second dummy variable.

So, the dummy variables are separated.

Qualitative information - about more explanatory variables

Similarly as above we can find the equivalence of the matrices

$$\begin{bmatrix}
1, & 0, & 1, & 0, & x_{4,1}, & \ldots, & x_{k,1} \\
1, & 0, & 1, & 0, & x_{4,2}, & \ldots, & x_{k,2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1, & 0, & 1, & 0, & x_{4,\ell_1}, & \ldots, & x_{k,\ell_1} \\
1, & 0, & 0, & 1, & x_{4,\ell_1+1}, & \ldots, & x_{k,\ell_1+1} \\
1, & 0, & 0, & 1, & x_{4,\ell_1+2}, & \ldots, & x_{k,\ell_1+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1, & 0, & 0, & 1, & x_{4,\ell_2}, & \ldots, & x_{k,\ell_2} \\
0, & 1, & 1, & 0, & x_{4,\ell_2+1}, & \ldots, & x_{k,\ell_2+1} \\
0, & 1, & 1, & 0, & x_{4,\ell_2+2}, & \ldots, & x_{k,\ell_2+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0, & 1, & 1, & 0, & x_{4,\ell_2+\ell_1}, & \ldots, & x_{k,\ell_2+\ell_1} \\
0, & 1, & 0, & 1, & x_{4,\ell_2+\ell_1+1}, & \ldots, & x_{k,\ell_2+\ell_1+1} \\
0, & 1, & 0, & 1, & x_{4,\ell_2+\ell_1+2}, & \ldots, & x_{k,\ell_2+\ell_1+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0, & 1, & 0, & 1, & x_{4,n}, & \ldots, & x_{k,n}
\end{bmatrix}
\text{ equiv. }
\begin{bmatrix}
1, & 0, & 0, & 0, & x_{4,1}, & \ldots, & x_{k,1} \\
1, & 0, & 0, & 0, & x_{4,2}, & \ldots, & x_{k,2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1, & 0, & 0, & 0, & x_{4,\ell_1}, & \ldots, & x_{k,\ell_1} \\
0, & 1, & 0, & 0, & x_{4,\ell_1+1}, & \ldots, & x_{k,\ell_1+1} \\
0, & 1, & 0, & 0, & x_{4,\ell_1+2}, & \ldots, & x_{k,\ell_1+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0, & 1, & 0, & 0, & x_{4,\ell_2}, & \ldots, & x_{k,\ell_2} \\
0, & 0, & 1, & 0, & x_{4,\ell_2+1}, & \ldots, & x_{k,\ell_2+1} \\
0, & 0, & 1, & 0, & x_{4,\ell_2+2}, & \ldots, & x_{k,\ell_2+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0, & 0, & 1, & 0, & x_{4,\ell_2+\ell_1}, & \ldots, & x_{k,\ell_2+\ell_1} \\
0, & 0, & 0, & 1, & x_{4,\ell_2+\ell_1+1}, & \ldots, & x_{k,\ell_2+\ell_1+1} \\
0, & 0, & 0, & 1, & x_{4,\ell_2+\ell_1+2}, & \ldots, & x_{k,\ell_2+\ell_1+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0, & 0, & 0, & 1, & x_{4,n}, & \ldots, & x_{k,n}
\end{bmatrix}$$

In the left model we estimate 4 unknown intercepts, in the right also,
hence they are mutually uniquely determined and models are equivalent.

Qualitative information - about more explanatory variables

But the right matrix from the previous slide is as follows:

$$\begin{bmatrix} 1, & 0, & 0, & 0, & x_{4,1}, & \ldots, & x_{k,1} \\ 1, & 0, & 0, & 0, & x_{4,2}, & \ldots, & x_{k,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1, & 0, & 0, & 0, & x_{4,\ell_1}, & \ldots, & x_{k,\ell_1} \\ 0, & 1, & 0, & 0, & x_{4,\ell_1+1}, & \ldots, & x_{k,\ell_1+1} \\ 0, & 1, & 0, & 0, & x_{4,\ell_1+2}, & \ldots, & x_{k,\ell_1+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0, & 1, & 0, & 0, & x_{4,\ell_2}, & \ldots, & x_{k,\ell_2} \\ 0, & 0, & 1, & 0, & x_{4,\ell_2+1}, & \ldots, & x_{k,\ell_2+1} \\ 0, & 0, & 1, & 0, & x_{4,\ell_2+2}, & \ldots, & x_{k,\ell_2+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0, & 0, & 1, & 0, & x_{4,\ell_2+\ell_1}, & \ldots, & x_{k,\ell_2+\ell_1} \\ 0, & 0, & 0, & 1, & x_{4,\ell_2+\ell_1+1}, & \ldots, & x_{k,\ell_2+\ell_1+1} \\ 0, & 0, & 0, & 1, & x_{4,\ell_2+\ell_1+2}, & \ldots, & x_{k,\ell_2+\ell_1+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0, & 0, & 0, & 1, & x_{4,n}, & \ldots, & x_{k,n} \end{bmatrix} = \begin{bmatrix} \text{the first block of } \ell_1 \text{ rows,} \\ \text{the first as well as} \\ \text{the second dummy} = 1. \\ \text{the second block} \\ \text{of } \ell_2 - \ell_1 \text{ rows,} \\ \text{the first dummy} = 1 \\ \text{while the second one} = 0 \\ \\ \text{the third block of } \ell_1 \text{ rows,} \\ \text{the first dummy} = 0 \\ \text{while the second one} = 1. \\ \\ \text{The fourth block} \\ \text{of } \ell_2 - \ell_1 \text{ rows,} \\ \text{the first dummy} = 0, \\ \text{the second one} = 0. \end{bmatrix}$$

But then the first 4 columns represent the all possible subgroups of data, generated by correponding two properties (features) for which we introduced dummies.

Qualitative information - about more explanatory variables

In other words,

$$
\begin{bmatrix}
1, & 0, & 0, & 0, & x_{4,1}, & \ldots, & x_{k,1} \\
1, & 0, & 0, & 0, & x_{4,2}, & \ldots, & x_{k,2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1, & 0, & 0, & 0, & x_{4,\ell_1}, & \ldots, & x_{k,\ell_1} \\
0, & 1, & 0, & 0, & x_{4,\ell_1+1}, & \ldots, & x_{k,\ell_1+1} \\
0, & 1, & 0, & 0, & x_{4,\ell_1+2}, & \ldots, & x_{k,\ell_1+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0, & 1, & 0, & 0, & x_{4,\ell_2}, & \ldots, & x_{k,\ell_2} \\
0, & 0, & 1, & 0, & x_{4,\ell_2+1}, & \ldots, & x_{k,\ell_2+1} \\
0, & 0, & 1, & 0, & x_{4,\ell_2+2}, & \ldots, & x_{k,\ell_2+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0, & 0, & 1, & 0, & x_{4,\ell_2+\ell_1}, & \ldots, & x_{k,\ell_2+\ell_1} \\
0, & 0, & 0, & 1, & x_{4,\ell_2+\ell_1+1}, & \ldots, & x_{k,\ell_2+\ell_1+1} \\
0, & 0, & 0, & 1, & x_{4,\ell_2+\ell_1+2}, & \ldots, & x_{k,\ell_2+\ell_1+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0, & 0, & 0, & 1, & x_{4,n}, & \ldots, & x_{k,n}
\end{bmatrix}
=
\begin{bmatrix}
\text{The first block} \\
\text{of } \ell_1 \text{ rows,} \\
\text{represents the situation when} \\
\text{both properties have value} = 1. \\
\\
\text{The second block} \\
\text{of } \ell_2 - \ell_1 \text{ rows,} \\
\text{represents the situation} \\
\text{when the first property has} \\
\text{value} = 1 \text{ while the second} = 0. \\
\\
\text{The third block} \\
\text{of } \ell_1 \text{ rows,} \\
\text{represents the situation when} \\
\text{the first property has value} = 0 \\
\text{while the second one} = 1. \\
\\
\text{The fourth block} \\
\text{of } \ell_2 - \ell_1 \text{ rows,} \\
\text{represents the situation when} \\
\text{both properties have value} = 0.
\end{bmatrix}
$$

It means that this matrix is for the model involving the interactions of dummies.

Notations and setup

Regression model

$$Y_{it} = X_{it}'\beta^0 + u_i + e_{it}, \quad i = 1, 2, ..., n, \;\; t = 1, 2, ..., T$$

X_{it}'s - p-dimensional random vector (explanatory variables)
u_i's - effects e_{it}'s - disturbances

DEFINITION

$cov(X_{itj}, u_i) = 0$	$\rightarrow$	the random effects model
otherwise	$\rightarrow$	the fixed effects model

Estimating the model with random effects - the classical approach

$$\text{Random effects} \;\equiv\; cov\,(X_{itj}, u_i) = 0$$

$$\Rightarrow \quad Y_{it} = X_{it}'\beta^0 + u_i + e_{it}, \quad \rightarrow \quad Y_{it} = X_{it}'\beta^0 + v_{it}$$

$$\mathbb{E} v_{it} = 0, \quad \operatorname{cov}(X_{itj}, v_{it}) = 0 \quad \mathbb{E}\,[v_{ks}, v_{is}] = 0,$$

$$\text{but } \mathbb{E}\,[v_{it}, v_{is}] = \sigma_u^2 \text{ and } \mathbb{E}\left[v_{it}^2\right] = \sigma_e^2 + \sigma_u^2$$

$\Rightarrow$ OLS are inefficient, hence ...

Transforming data $\quad \tilde{Y}_{it} = Y_{it} - \lambda \bar{Y}_i, \quad \tilde{X}_{it} = X_{it} - \lambda \bar{X}_i$

with $\lambda = 1 - \sigma_e^2 \left(\sigma_e^2 + T \cdot \sigma_u^2\right)^{-1}$,

(where $\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$ and $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$) and we can apply the OLS on the transformed data, efficiently - if we know λ.

Estimating λ by $\hat{\lambda} = 1 - \hat{\sigma}_e^2 \left(\hat{\sigma}_e^2 + T \cdot \hat{\sigma}_u^2\right)^{-1}$, etc.

(In numerical study below denoted as $\hat{\beta}^{RE}$).

Estimating the model with fixed effects - the classical approach

$$\boxed{\text{Fixed effects} \equiv \quad cov\left(X_{itj}, u_i\right) \neq 0}$$

$$\Rightarrow \qquad Y_{it} = X_{it}'\beta^0 + u_i + e_{it}, \qquad \rightarrow \qquad Y_{it} = X_{it}'\beta^0 + v_{it}$$

$$I\!\!E v_{it} = 0, \quad \mathrm{cov}\left(X_{itj}, v_{it}\right) \neq 0$$

$\Rightarrow$ OLS are inconsistent, hence ...

Transforming data

$$\tilde{Y}_{it} = Y_{it} - \bar{Y}_i, \qquad \tilde{X}_{it} = X_{it} - \bar{X}_i,$$

$\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$ and $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$, we rid of the effects u_i's and we can apply the OLS on the transformed data, consistently.

Notice that the transformed data don't include intercept, if original included it.

(In numerical study below denoted as $\hat{\beta}^{FE}$).

Hausman test

Hausman, J. (1978): Specification test in econometrics.
Econometrica, 46, 1978, 1251 - 1271.

Denote $\hat{\beta}^{(RE,n,T)}$ - the efficient estimator for the case when we have assumed that the effects are not correlated with explanatory variables.

Further denote $\hat{\beta}^{(FE,n,T)}$ - that (nearly) efficient estimator for the case when we have assumed that effects are correlated with explanatory variables and put

$$q_{(n,T)} = \hat{\beta}^{(RE,n,T)} - \hat{\beta}^{(FE,n,T)}.$$

(Discuss magnitude of $q_{(n,T)}$.)

Finally, denote $V_{(n,T)}$ the covariance matrix of $q_{(n,T)}$.

Hausman test

Hausman, J. (1978): Specification test in econometrics.
Econometrica, 46, 1978, 1251 - 1271.

Then, under normality of disturbances and effects and their independence from explanatory variables

$$\mathcal{L}\left\{ q'_{(n,T)} V^{-1}_{(n,T)} q_{(n,T)} \right\}$$

$$= \mathcal{L}\left\{ \left[\hat{\beta}^{(RE,n,T)} - \hat{\beta}^{(FE,n,T)} \right]' V^{-1}_{(n,T)} \left[\hat{\beta}^{(RE,n,T)} - \hat{\beta}^{(FE,n,T)} \right] \right\} = \chi^2(p)$$

where p is number of explanatory variables and moreover

$$V_{(n,T)} = \mathrm{cov}(q_{(n,T)}) = \mathrm{cov}(\hat{\beta}^{(RE,n,T)}) - \mathrm{cov}(\hat{\beta}^{(FE,n,T)})$$

(remember the idea of test for the next numerical study).

Assumptions (the classical ones for model with effects)

$$Y_{it} = X_{it}'\beta^0 + u_i + e_{it}, \quad i = 1, 2, ..., n, \;\; t = 1, 2, ..., T$$

- $\left\{ \{X_{it}\}_{t=1}^T \right\}_{i=1}^\infty$ T-tuples of i. i. d. p-dimensional r.v.'s, $\exists q \in [0, 1] \;:\; \mathbb{E}\,\|X_{11}\|^{2q} < \infty$,
- $\{u_i\}_{i=1}^\infty$ i. i. d. r. v.'s, $\mathbb{E}\,(u_i) = 0, \;\; \mathrm{var}\,(u_i) = \sigma_u^2$,
- $\left\{ \{e_{it}\}_{t=1}^T \right\}_{i=1}^\infty$ T-tuples of independent r.v.'s, $F_{e_{it}}(r) = F_e(\sigma_{it} \cdot r)$, $\mathbb{E}\,(e_{it}) = 0, \;\; \mathrm{var}\,(e_{it}) = \sigma_{it}^2$,

---

- e_{it}'s independent from X_{it}'s,
- u_i's independent from e_{it}'s

but u_i's need not be independent from X_{it}'s

Assumption (non-classical)

Residuals $\forall \beta \in R \quad \to \quad r_{it}(\beta) = Y_{it} - X_{it}'\beta$

D. f. of the absolute values of residual

$$\forall \beta \in R \quad \to \quad F_\beta^{(it)}(r) = P\left(|r_{it}(\beta)| < r\right)$$

IDENTIFICATION CONDITION

For any $n \in N$ there is the only solution of

$$\left(\beta - \beta^0\right)' E\left\{\sum_{i=1}^{n}\sum_{t=1}^{T}\left[w\left(F_\beta^{(it)}(|r_{it}(\beta)|)\right)X_{it}\left(e_{it} - X_{it}'\left(\beta - \beta^0\right)\right)\right]\right\} = 0,$$

namely β^0.

(Notice, we have left Euclidean metric - hence identification condition.)

Estimating the model with random effects - recalling the classical approach

$$\boxed{\text{Random effects} \equiv \quad cov\,(X_{itj}, u_i) = 0}$$

$$\Rightarrow \qquad Y_{it} = X_{it}'\beta^0 + u_i + e_{it}, \qquad \to \qquad Y_{it} = X_{it}'\beta^0 + v_{it}$$

$$\mathbb{E}v_{it} = 0, \quad \operatorname{cov}(X_{itj}, v_{it}) = 0 \quad \mathbb{E}\,[v_{it}, v_{is}] = \sigma_u^2$$

$\Rightarrow$ OLS are inefficient, hence ...

Transforming data $\quad \tilde{Y}_{it} = Y_{it} - \lambda\bar{Y}_i, \quad \tilde{X}_{it} = X_{it} - \lambda\bar{X}_i$

with $\lambda = 1 - \sigma_e^2\left(\sigma_e^2 + T\cdot\sigma_u^2\right)^{-1}$,

(where $\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$ and $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$)

and we can apply the OLS on the transformed data, efficiently

- if we know λ.

Estimating the model with random effects - recalling the classical approach

$$\text{Random effects} \;\equiv\; cov\,(X_{itj}, u_i) = 0$$

Estimating λ by $\hat{\lambda} = 1 - \hat{\sigma}_e^2 \left(\hat{\sigma}_e^2 + T \cdot \hat{\sigma}_u^2\right)^{-1}$, etc.

Transforming data

$$\tilde{Y}_{it} = Y_{it} - \hat{\lambda}\bar{Y}_i, \quad \tilde{X}_{it} = X_{it} - \hat{\lambda}\bar{X}_i$$

and we can apply the OLS on the transformed data,

now already efficiently

(In numerical study below denoted as $\hat{\beta}^{RE}$).

Robustifying the classical approach

$$\boxed{\text{Random effects} \equiv \; cov\,(X_{itj}, u_i) = 0}$$

$$\Rightarrow \quad Y_{it} = X_{it}'\beta^0 + u_i + e_{it}, \quad \rightarrow \quad Y_{it} = X_{it}'\beta^0 + v_{it}$$

$$I\!\!E v_{it} = 0, \quad \mathrm{cov}\,(X_{itj}, v_{it}) = 0 \quad I\!\!E\,[v_{it}, v_{is}] = \sigma_u^2$$

$\Rightarrow$ $\boxed{\text{LWS}}$ are (probably) inefficient, hence ...

(the differences against the previous slide - red and in the box)

Transforming data

$$\tilde{Y}_{it} = Y_{it} - \lambda \; \boxed{\bar{Y}_i^{LWS}} \; , \quad \tilde{X}_{it} = X_{it} - \lambda \; \boxed{\bar{X}_i^{LWS}}$$

$$\text{with } \lambda = 1 - \sigma_e^2 \left(\sigma_e^2 + T \cdot \sigma_u^2\right)^{-1}$$

and we can apply $\boxed{\text{LWS}}$ on the transformed data, again unknown λ, ...

Robustifying the classical approach

Random effects $\equiv \quad cov\left(X_{itj}, u_i\right) = 0$

Put $r_{Y,it}(y) = Y_{it} - y$ and denote $r^2_{(Y,it)}(y)$ the t-th order statistic among the squared residuals $r^2_{Y,i1}(y), r^2_{Y,i2}(y), ..., r^2_{Y,iT}(y)$. Then

$$\bar{Y}_i^{LWS} = \arg\min_{y \in R} \sum_{t=1}^{T} w\left(\frac{i-1}{n}\right) r^2_{(Y,it)}(y).$$

Similarly, put $r_{X,itj}(x) = X_{itj} - x$ and denote $r^2_{(X,itj)}(x)$ the t-th order statistic among the squared residuals $r^2_{X,i1j}(x), r^2_{X,i2j}(x), ..., r^2_{X,iTj}(x),$. Then

$$\bar{X}_{ij}^{LWS} = \arg\min_{x \in R} \sum_{t=1}^{T} w\left(\frac{i-1}{n}\right) r^2_{(X,itj)}(x)$$

and put $\quad \bar{X}_i^{LWS} = \left(\bar{X}_{i1}^{LWS}, \bar{X}_{i2}^{LWS}, ..., \bar{X}_{ip}^{LWS}\right)'.$

Recalling estimation of σ_e^2 and σ_u^2

$$\boxed{\text{Random effects} \equiv cov\left(X_{itj}, u_i\right) = 0}$$

Put $r_{it}(\hat{\beta}^{(OLS,n)}) = Y_{it} - X_{it}'\hat{\beta}^{(OLS,n)}$. Then

$$\hat{\sigma}_v^2 = \frac{1}{n \cdot T - 1} \sum_{i=1}^{n} \sum_{t=1}^{T} r_{it}^2(\hat{\beta}^{(OLS,n)}),$$

$$\hat{\sigma}_u^2 = \frac{1}{(n \cdot T - 1)^2} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{s \neq t} r_{it}(\hat{\beta}^{(OLS,n)}) r_{is}(\hat{\beta}^{(OLS,n)})$$

and

$$\hat{\sigma}_e^2 = \hat{\sigma}_v^2 - \hat{\sigma}_u^2$$

Robustifying the classical approach

Random effects $\equiv$ $cov\,(X_{itj}, u_i) = 0$

Put $r_{it}(\hat{\beta}^{(LWS,n,w)}) = Y_{it} - X_{it}'\hat{\beta}^{(LWS,n,w)}$ and denote $r^2_{(k)}(\hat{\beta}^{(LWS,n,w)})$ the k-th order statistic among all squared residuals $r^2_{it}(\hat{\beta}^{(LWS,n,w)}),\ \ i = 1, 2, .., n$, $t = 1, 2, ..., T$. Then

$$\hat{\sigma}^2_{(LWS,w,v)} = \frac{1}{n \cdot T - 1} \sum_{k=1}^{n \cdot T} w\left(\frac{k-1}{n \cdot T}\right) r^2_{(k)}(\hat{\beta}^{(LWS,n,w)}),$$

$$\hat{\sigma}^2_{(LWS,w,u)} = \frac{1}{(n \cdot T - 1)^2} \sum_{k=1}^{n \cdot T} \sum_{k \neq j} w\left(\frac{k-1}{n \cdot T}\right) w\left(\frac{j-1}{n \cdot T}\right) r_{[k]}(\hat{\beta}^{(LWS,n,w)}) r_{[j]}(\hat{\beta}^{(LWS,n,w)})$$

where $r_{[k]}(\hat{\beta}^{(LWS,n,w)})$ is the residual corresponding to $r^2_{(k)}(\hat{\beta}^{(LWS,n,w)})$ and

$$\hat{\sigma}^2_{(LWS,w,e)} = \hat{\sigma}^2_{(LWS,w,v)} - \hat{\sigma}^2_{(LWS,w,u)}.$$

Robustifying the classical approach

$$\boxed{\text{Random effects} \;\equiv\; cov\,(X_{itj}, u_i) = 0}$$

Estimating λ by $\hat{\lambda} = 1 - \boxed{\hat{\sigma}^2_{(LWS,w,e)}} \cdot \left(\boxed{\hat{\sigma}^2_{(LWS,w,e)}} + T \cdot \boxed{\hat{\sigma}^2_{(LWS,w,u)}} \right)^{-1}$,
etc.

Transforming data

$$\tilde{Y}_{it} = Y_{it} - \hat{\lambda}\,\boxed{\bar{Y}_i^{LWS}}\,, \qquad \tilde{X}_{it} = X_{it} - \hat{\lambda}\,\boxed{\bar{X}_i^{LWS}}$$

$$\text{with } \hat{\lambda} = 1 - \hat{\sigma}^2_{(LWS,w,e)} \left(\hat{\sigma}^2_{(LWS,w,e)} + T \cdot \hat{\sigma}^2_{(LWS,w,u)} \right)^{-1}$$

and we can apply $\boxed{\text{LWS}}$ on the transformed data, now already efficiently.

(In numerical study below denoted as $\underline{\hat{\beta}^{RWE}}$).

Estimating the model with fixed effects - recalling the classical approach

$$\boxed{\text{Fixed effects} \quad \equiv \quad cov\,(X_{itj}, u_i) \neq 0}$$

$$\Rightarrow \qquad Y_{it} = X_{it}'\beta^0 + u_i + e_{it}, \qquad \rightarrow \qquad Y_{it} = X_{it}'\beta^0 + v_{it}$$

$$\mathbb{E}v_{it} = 0, \quad \mathrm{cov}\,(X_{itj}, v_{it}) \neq 0$$

$\Rightarrow$ OLS are inconsistent, hence ...

Transforming data

$$\tilde{Y}_{it} = Y_{it} - \bar{Y}_i, \quad \tilde{X}_{it} = X_{it} - \bar{X}_i,$$

$\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$ and $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$, <u>we rid of the effects u_i's</u>

and we can apply the OLS on the transformed data, consistently.

Notice that the transformed data don't include intercept.

(In numerical study below denoted as $\hat{\beta}^{FE}$).

Robustifying the classical approach

$$\text{Fixed effects} \equiv cov\,(X_{itj}, u_i) \neq 0$$

$$\Rightarrow \quad Y_{it} = X_{it}'\beta^0 + u_i + e_{it}, \quad \rightarrow \quad Y_{it} = X_{it}'\beta^0 + v_{it}$$

$$\mathbb{E}v_{it} = 0, \quad \mathrm{cov}\,(X_{itj}, v_{it}) \neq 0$$

$\Rightarrow$ $\boxed{LWS}$ are inconsistent, hence ...

(the differences against the previous slide - again red and in the box)

Transforming data

$$\tilde{Y}_{it} = Y_{it} - \boxed{\bar{Y}_i^{LWS}}, \quad \tilde{X}_{it} = X_{it} - \boxed{\bar{X}_i^{LWS}},$$

we rid of u_i's (approximately)

and we can apply the $\boxed{LWS}$ on the transformed data.

(In numerical study below denoted as $\underline{\hat{\beta}^{FWE}}$).

Numerical study

***The framework*:**

- 500 data sets, each of them containing:
  - 50 cases (i. e. $n = 50$ of observed objects),
  - observed for 20 time periods (i. e. $T = 20$),

    $\Rightarrow$ each data set has 1 000 rows.
  - Each coordinate of explanatory vector correlated with fixed effect on the level $\frac{1}{\sqrt{2}}$.
- 7 levels (0.25% - 15%) and several types of contamination
  - * outliers - randomly selected observations $\rightarrow \; Y_i = -2 * Y_i$,
  - * leverage points - selected observations on the outskirts $\rightarrow \; \tilde{X}_i = 10 \cdot X_i$ and $Y_i = -\tilde{X}_i' \cdot \beta^0 + e_i$

    - see http://samba.fsv.cuni.cz/~visek/Oxford*2013/

Numerical study

***The framework** (continued)*:

- The optimal weight function used for LWS.
- Exhibited are

$$\hat{\beta}_j^{(index)} = \frac{1}{500} \sum_{k=1}^{500} \hat{\beta}_j^{(index,k)}$$

and

$$\widehat{\mathrm{MSE}} \left(\hat{\beta}_j^{(index)} \right) = \frac{1}{500} \sum_{k=1}^{500} \left[\hat{\beta}_j^{(index,k)} - \hat{\beta}_j^{0} \right]^2 .$$

- The empirical distributuion function of Hausman test is also given

- notice value on the x-axe.

All else will be clear from the context.

Estimating the model with random effects - patterns of numerical study

TABLE 1

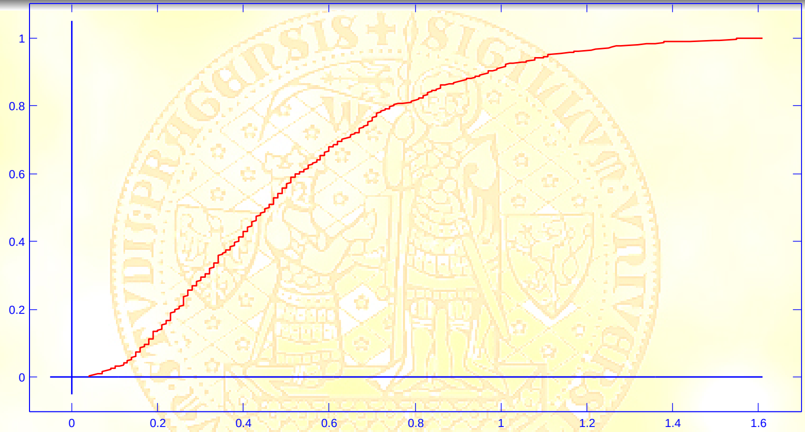
True coeffs β^0	1	-2	3	-4	5
!!! These coefficients were used in the whole numerical study. !!!					

The disturbances are homoscedastic, both the disturbances and the effects are independent from explanatory variables. Data were without contamination.

Variances of the disturbances and effects were both equal to 1.

$\hat{\beta}^{OLS}_{(\mathrm{MSE}(\hat{\beta}^{OLS}))}$	$1.00_{(0.201)}$	$-2.00_{(0.185)}$	$3.00_{(0.204)}$	$-4.00_{(0.215)}$	$5.00_{(0.192)}$
$\hat{\beta}^{FE}_{(\mathrm{MSE}(\hat{\beta}^{FE}))}$	$1.00_{(0.101)}$	$-2.00_{(0.108)}$	$3.00_{(0.106)}$	$-4.00_{(0.106)}$	$5.00_{(0.101)}$
$\hat{\beta}^{RE}_{(\mathrm{MSE}(\hat{\beta}^{RE}))}$	$1.00_{(0.100)}$	$-2.00_{(0.107)}$	$3.00_{(0.106)}$	$-4.00_{(0.106)}$	$5.00_{(0.100)}$
$\hat{\beta}^{LWS}_{(\mathrm{MSE}(\hat{\beta}^{LWS}))}$	$1.00_{(0.210)}$	$-2.00_{(0.192)}$	$3.00_{(0.209)}$	$-4.00_{(0.221)}$	$5.00_{(0.198)}$
$\hat{\beta}^{FWE}_{(\mathrm{MSE}(\hat{\beta}^{FWE}))}$	$1.00_{(0.105)}$	$-2.00_{(0.111)}$	$3.00_{(0.113)}$	$-4.00_{(0.108)}$	$5.00_{(0.103)}$
$\hat{\beta}^{RWE}_{(\mathrm{MSE}(\hat{\beta}^{RWE}))}$	$1.00_{(0.106)}$	$-2.00_{(0.111)}$	$3.00_{(0.112)}$	$-4.00_{(0.109)}$	$5.00_{(0.103)}$

Hausman test



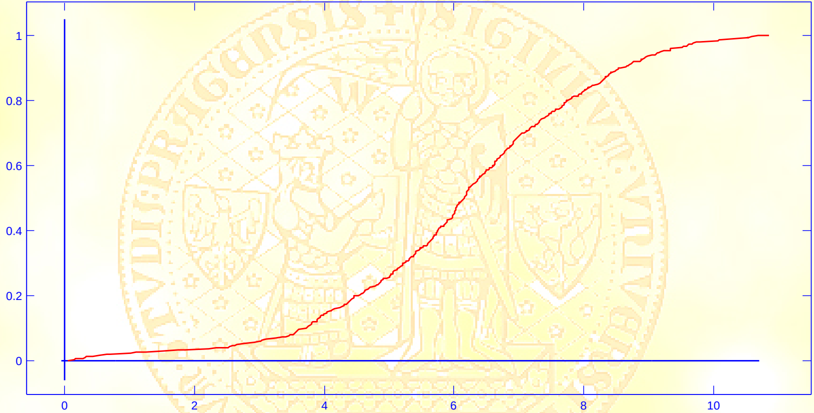
The d. f.'s of test statistic of Hausman test (values of test statistics multiplied by 100). The details of framework are given in the head of previous table.

Estimating the model with fixed effects - patterns of numerical study

TABLE 2

True coeffs β^0	**1**	**-2**	**3**	**-4**	**5**
The disturbances are homoscedastic, the disturbances are independent while the effects are correlated with explanatory variables. Data were without contamination.					
Variances of the disturbances and effects were both equal to 1.					
$\hat\beta^{OLS}_{(\mathrm{var}(\hat\beta^{OLS}))}$	$1.03_{(0.208)}$	$-1.93_{(0.576)}$	$3.10_{(1.134)}$	$-3.87_{(1.884)}$	$5.17_{(2.919)}$
$\hat\beta^{FE}_{(\mathrm{MSE}(\hat\beta^{FE}))}$	$1.00_{(0.105)}$	$\boxed{-2.00_{(0.103)}}$	$3.00_{(0.102)}$	$\boxed{-4.00_{(0.102)}}$	$5.00_{(0.096)}$
$\hat\beta^{RE}_{(\mathrm{MSE}(\hat\beta^{RE}))}$	$1.03_{(0.207)}$	$-1.93_{(0.572)}$	$3.10_{(1.122)}$	$-3.87_{(1.865)}$	$5.17_{(2.888)}$
$\hat\beta^{LWS}_{(\mathrm{MSE}(\hat\beta^{LWS}))}$	$1.03_{(0.211)}$	$-1.93_{(0.578)}$	$3.10_{(1.128)}$	$-3.87_{(1.882)}$	$5.17_{(2.929)}$
$\hat\beta^{FWE}_{(\mathrm{MSE}(\hat\beta^{FWE}))}$	$1.00_{(0.109)}$	$-2.00_{(0.102)}$	$3.00_{(0.103)}$	$-4.00_{(0.104)}$	$5.00_{(0.096)}$
$\hat\beta^{RWE}_{(\mathrm{MSE}(\hat\beta^{RWE}))}$	$1.03_{(0.208)}$	$-1.93_{(0.567)}$	$3.10_{(1.104)}$	$-3.87_{(1.838)}$	$5.17_{(2.865)}$

Hausman test



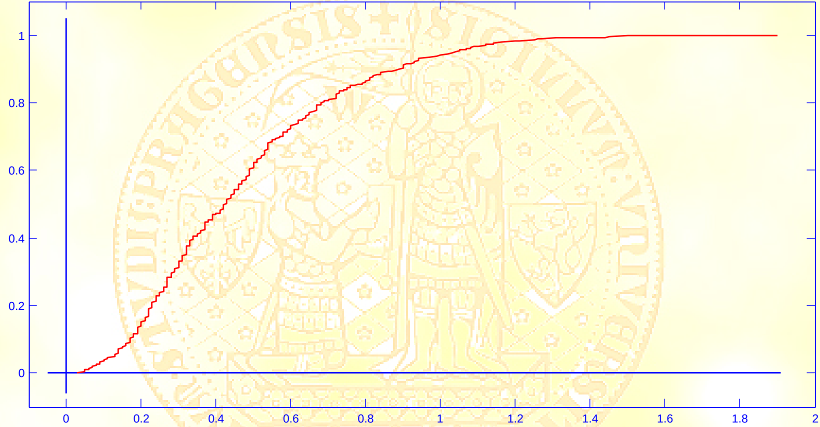
The d. f.'s of test statistic of Hausman test (values of test statistics multiplied by 100). The details of framework are given in the head of previous table.

Estimating the model with random effects - patterns of numerical study

TABLE 3

True coeffs β^0	1	-2	3	-4	5
The disturbances are homoscedastic, both the disturbances and the effects are independent from explanatory variables. Variances of the disturbances and effects were both equal to 1. Data were contaminated by outliers on the level 0.5%					
$\hat\beta^{OLS}_{(\mathrm{var}(\hat\beta^{OLS}))}$	$0.95_{(2.008)}$	$-1.91_{(2.596)}$	$2.86_{(4.047)}$	$-3.80_{(6.353)}$	$4.76_{(8.082)}$
$\hat\beta^{FE}_{(\mathrm{MSE}(\hat\beta^{FE}))}$	$0.95_{(1.933)}$	$-1.91_{(2.533)}$	$2.86_{(3.886)}$	$-3.80_{(6.315)}$	$4.76_{(8.182)}$
$\hat\beta^{RE}_{(\mathrm{MSE}(\hat\beta^{RE}))}$	$0.95_{(1.928)}$	$\boxed{-1.91_{(2.522)}}$	$2.86_{(3.899)}$	$\boxed{-3.80_{(6.312)}}$	$4.76_{(8.122)}$
$\hat\beta^{LWS}_{(\mathrm{MSE}(\hat\beta^{LWS}))}$	$1.00_{(0.209)}$	$-2.00_{(0.254)}$	$3.00_{(0.208)}$	$-4.00_{(0.231)}$	$5.00_{(0.256)}$
$\hat\beta^{FWE}_{(\mathrm{MSE}(\hat\beta^{FWE}))}$	$1.00_{(0.125)}$	$-2.00_{(0.146)}$	$2.99_{(0.130)}$	$-3.99_{(0.134)}$	$4.99_{(0.134)}$
$\hat\beta^{RWE}_{(\mathrm{MSE}(\hat\beta^{RWE}))}$	$1.00_{(0.123)}$	$-2.00_{(0.143)}$	$2.99_{(0.126)}$	$-3.99_{(0.127)}$	$4.99_{(0.131)}$

Hausman test



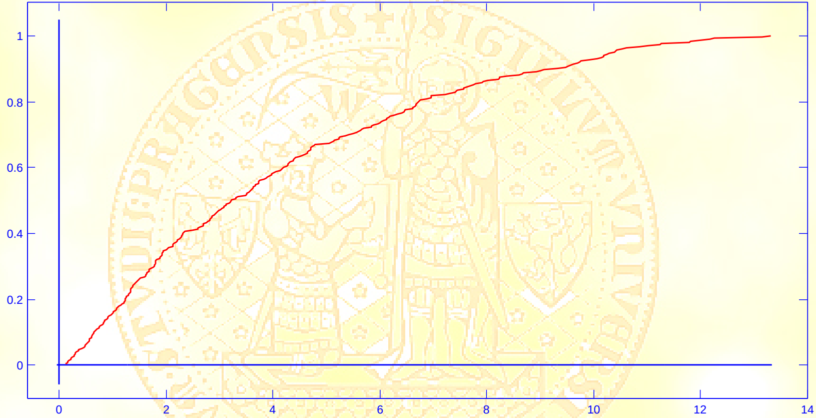
The d. f.'s of test statistic of Hausman test (values of test statistics multiplied by 100).
The details of framework are given in the head of previous table.

Estimating the model with random effects - patterns of numerical study

TABLE 4

True coeffs β^0	1	-2	3	-4	5
The disturbances are homoscedastic, both the disturbances and the effects are independent from explanatory variables. Variances of the disturbances and effects were both equal to 1. Data were contaminated by leverage points on the level 0.5%.					
$\hat{\beta}^{OLS}$ $(\mathrm{var}(\hat{\beta}^{OLS}))$	$-0.07_{(460.675)}$	$0.09_{(788.855)}$	$-0.07_{(1362.958)}$	$0.41_{(2460.109)}$	$-0.31_{(3432.956)}$
$\hat{\beta}^{FE}$ $(\mathrm{MSE}(\hat{\beta}^{FE}))$	$-0.06_{(447.410)}$	$0.07_{(772.587)}$	$-0.04_{(1330.554)}$	$0.37_{(2411.864)}$	$-0.27_{(3371.086)}$
$\hat{\beta}^{RE}$ $(\mathrm{MSE}(\hat{\beta}^{RE}))$	$-0.07_{(460.598)}$	$\boxed{0.09_{(789.090)}}$	$-0.07_{(1362.957)}$	$\boxed{0.41_{(2460.699)}}$	$-0.31_{(3433.437)}$
$\hat{\beta}^{LWS}$ $(\mathrm{MSE}(\hat{\beta}^{LWS}))$	$1.00_{(0.210)}$	$-2.00_{(0.254)}$	$3.00_{(0.215)}$	$-4.00_{(0.237)}$	$5.00_{(0.258)}$
$\hat{\beta}^{FWE}$ $(\mathrm{MSE}(\hat{\beta}^{FWE}))$	$1.00_{(0.126)}$	$-1.99_{(0.163)}$	$2.99_{(0.141)}$	$-3.99_{(0.148)}$	$4.98_{(0.159)}$
$\hat{\beta}^{RWE}$ $(\mathrm{MSE}(\hat{\beta}^{RWE}))$	$1.00_{(0.126)}$	$-2.00_{(0.160)}$	$2.99_{(0.141)}$	$-3.99_{(0.142)}$	$4.99_{(0.146)}$

Hausman test



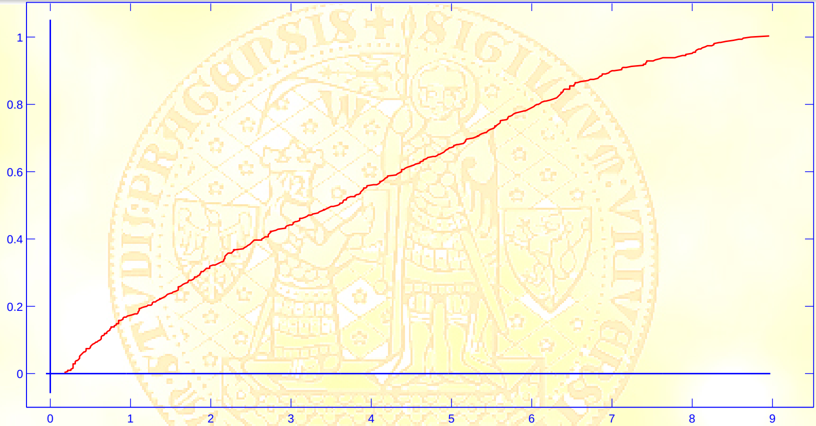
The d. f.'s of test statistic of Hausman test (values of test statistics multiplied by 100). The details of framework are given in the head of previous table.

Estimating the model with fixed effects - patterns of numerical study

TABLE 5

True coeffs β^0	1	-2	3	-4	5
The disturbances are independent while the effects are correlated with explanatory variables. Variances of the disturbances and effects were both equal to 1. Data were contaminated by outliers on the level 0.5%.					
$\hat{\beta}^{OLS}_{(\text{var}(\hat{\beta}^{OLS}))}$	$1.00_{(1.766)}$	$-1.96_{(1.542)}$	$3.00_{(2.956)}$	$-3.90_{(2.243)}$	$5.01_{(5.792)}$
$\hat{\beta}^{FE}_{(\text{MSE}(\hat{\beta}^{FE}))}$	$0.95_{(1.463)}$	$\boxed{-2.04_{(1.434)}}$	$2.87_{(4.652)}$	$\boxed{-4.07_{(2.188)}}$	$4.80_{(10.039)}$
$\hat{\beta}^{RE}_{(\text{MSE}(\hat{\beta}^{RE}))}$	$0.99_{(1.240)}$	$-1.96_{(1.171)}$	$2.99_{(2.817)}$	$-3.92_{(1.706)}$	$4.99_{(5.764)}$
$\hat{\beta}^{LWS}_{(\text{MSE}(\hat{\beta}^{LWS}))}$	$1.03_{(0.237)}$	$-1.93_{(0.572)}$	$3.10_{(1.128)}$	$-3.86_{(1.919)}$	$5.17_{(2.929)}$
$\hat{\beta}^{FWE}_{(\text{MSE}(\hat{\beta}^{FWE}))}$	$1.00_{(0.129)}$	$-2.00_{(0.127)}$	$2.99_{(0.155)}$	$-4.00_{(0.147)}$	$4.99_{(0.163)}$
$\hat{\beta}^{RWE}_{(\text{MSE}(\hat{\beta}^{RWE}))}$	$1.02_{(0.201)}$	$-1.95_{(0.415)}$	$3.07_{(0.768)}$	$-3.90_{(1.311)}$	$5.12_{(1.914)}$

Hausman test



The d. f.'s of test statistic of Hausman test (values of test statistics multiplied by 100).
The details of framework are given in the head of previous table.

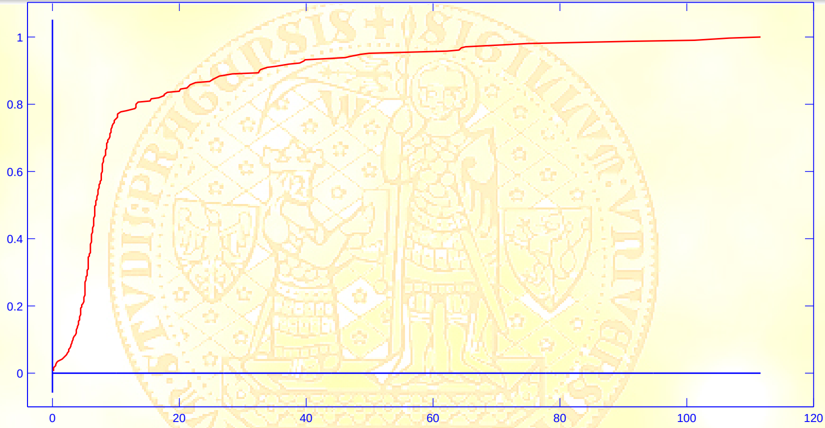
Estimating the model with fixed effects - patterns of numerical study

TABLE 6

The disturbances are homoscedastic, the disturbances are independent while the effects are correlated with explanatory variables.
Variances of the disturbances and effects were both equal to 1.
Data were contaminated by leverage points on the level 0.5%.

$\hat{\beta}^{OLS}$ $(\mathrm{var}(\hat{\beta}^{OLS}))$	$0.48_{(290.714)}$	$-1.36_{(305.095)}$	$1.30_{(846.638)}$	$-2.78_{(636.709)}$	$2.13_{(1844.263)}$
$\hat{\beta}^{FE}$ $(\mathrm{MSE}(\hat{\beta}^{FE}))$	$0.38_{(158.625)}$	$\boxed{-1.57_{(214.387)}}$	$1.11_{(706.569)}$	$\boxed{-3.11_{(731.946)}}$	$1.87_{(1727.656)}$
$\hat{\beta}^{RE}$ $(\mathrm{MSE}(\hat{\beta}^{RE}))$	$0.37_{(170.499)}$	$-1.60_{(182.087)}$	$1.03_{(814.620)}$	$-3.24_{(520.855)}$	$1.67_{(2118.376)}$
$\hat{\beta}^{LWS}$ $(\mathrm{MSE}(\hat{\beta}^{LWS}))$	$1.03_{(0.245)}$	$-1.93_{(0.593)}$	$3.10_{(1.127)}$	$-3.87_{(1.918)}$	$5.17_{(2.986)}$
$\hat{\beta}^{FWE}$ $(\mathrm{MSE}(\hat{\beta}^{FWE}))$	$0.95_{(2.478)}$	$-2.08_{(8.707)}$	$2.85_{(20.521)}$	$-4.18_{(33.784)}$	$4.75_{(57.394)}$
$\hat{\beta}^{RWE}$ $(\mathrm{MSE}(\hat{\beta}^{RWE}))$	$1.01_{(1.693)}$	$-2.02_{(6.141)}$	$2.95_{(14.617)}$	$-4.05_{(23.262)}$	$4.93_{(37.868)}$

Hausman test



The d. f.'s of test statistic of Hausman test (values of test statistics multiplied by 100). The details of framework are given in the head of previous table.

Sensitivity study

The classical formula for OLS:

$$\hat{\beta}^{(OLS,n)} - \hat{\beta}^{(OLS,n-1,\ell)} = \left\{[X^{(n-1,\ell)}]' X^{(n-1,\ell)}\right\}^{-1} X_\ell \left(Y_\ell - X_\ell' \hat{\beta}^{(OLS,n)}\right)$$

Proof: First of all let's recall that $\sum_{i=1}^n X_i X_i' = X'X$ and then consider the difference of normal equations

$$\sum_{i=1}^n X_i \left(Y_i - X_i' \hat{\beta}^{(OLS,n)}\right) = 0 \quad \text{and} \quad \sum_{i=1, i\neq\ell}^n X_i \left(Y_i - X_i' \hat{\beta}^{(OLS,n-1,\ell)}\right) = 0.$$

We have

$$X_\ell \left(Y_\ell - X_\ell' \hat{\beta}^{(OLS,n)}\right) = \sum_{i=1, i\neq\ell}^n X_i X_i' \left(\hat{\beta}^{(OLS,n)} - \hat{\beta}^{(OLS,n-1,\ell)}\right)$$

i. e.

$$X_\ell \left(Y_\ell - X_\ell' \hat{\beta}^{(OLS,n)}\right) = [X^{(n-1,\ell)}]' X^{(n-1,\ell)} \cdot \left(\hat{\beta}^{(OLS,n)} - \hat{\beta}^{(OLS,n-1,\ell)}\right). \quad \text{Q.E.D.}$$

For many examples of other diagnostic tools see e. g. :

Draper, N. R., H. Smith (1966): *Applied Regression Analysis.* New York: J.Wiley & Sons, 1st edition.

Chatterjee, S., A. S. Hadi (1988): *Sensitivity Analysis in Linear Regression.* New York: J. Wiley & Sons.

Zvára, K. (1989): *Regresní analýza (Regression Analysis – in Czech) Prague: Academia.*

We can look for:

An asymptotic representation of the difference

$$\hat{\beta}^{(*,n)} - \hat{\beta}^{(*,n-1,\ell)}. \tag{1}$$

Definition

If the norm of (1) is - uniformly in ℓ - low, we speak about the *low subsample sensitivity*.

M-etimators with absolutely continuous ψ

$$n\left(\hat{\beta}^{(M,n)} - \hat{\beta}^{(M,n-1,\ell)}\right) = \hat{\sigma}_n \mathbb{E}_{F_\varepsilon}^{-1}\left\{\psi'\left(\frac{\varepsilon_1}{\hat{\sigma}_n}\right)\right\} Q^{-1} X_\ell \psi\left(\left[Y_\ell - X_\ell'\hat{\beta}^{(M,n)}\right]\hat{\sigma}_n^{-1}\right)$$

$$+ o_p(1) \quad \text{as } n \to \infty,$$

$$Q = \mathbb{E} X_1 X_1'$$

Víšek, J. Á. (1996): Sensitivity analysis of M-estimates.
Ann. Inst. of Statist. Mathematics, 48, 469-495.

(The paper contains also results for the discontinuous ψ, see the next slide.)

M-etimators with discontinuous ψ

$$n\left(\hat{\beta}^{(L_1,n)} - \hat{\beta}^{(L_1,n-1,\ell)}\right) = \frac{1}{2} f^{-1}(0) Q^{-1} X_\ell \psi_m \left(Y_\ell - X_\ell^T, \hat{\beta}^{(L_1,n)}\right) + \mathcal{R}_n$$

where

$$\mathcal{R} =_{\mathcal{D}} \frac{1}{2} f^{-1}(0) Q^{-1} \left[W_n^{(1)} - W_n^{(2)}\right] + o_p(1)$$

with

$$W_n^{(j)} = \left(W(\sum_{i=1}^{n} \tau_{i1}^{(j)}), W(\sum_{i=1}^{n} \tau_{i2}^{(j)}), ..., W(\sum_{i=1}^{n} \tau_{ip}^{(j)}) \right)', \quad j = 1 \text{ and } 2$$

for some stopping times $\tau_{ik}^{(j)}$ and $W(s)$ a Wiener process.

But there is a snag !! (See the next slide)

M-etimators with discontinuous ψ

$$n\left(\hat{\beta}^{(L_1,n)} - \hat{\beta}^{(L_1,n-1,\ell)}\right) = \frac{1}{2} f^{-1}(0) Q^{-1} X_\ell \psi_m \left(Y_\ell - X_\ell^T, \hat{\beta}^{(L_1,n)}\right) + \mathcal{R}_n$$

where

$$\mathcal{R} =_{\mathcal{D}} \frac{1}{2} f^{-1}(0) Q^{-1} \left[W_n^{(1)} - W_n^{(2)}\right] + o_p(1) = \mathcal{O}_p(1)$$

with

$$W_n^{(j)} = \left(W(\sum_{i=1}^{n} \tau_{i1}^{(j)}), W(\sum_{i=1}^{n} \tau_{i2}^{(j)}), ..., W(\sum_{i=1}^{n} \tau_{ip}^{(j)}) \right)', \quad j = 1 \text{ and } 2$$

for some stopping times $\tau_{ik}^{(j)}$ and $W(s)$ a Wiener process.

(There are also results for *set-subsample sensitivity*, see the next slides.)

M-etimators with absolutely continuous ψ

$$n\left(\hat{\beta}^{(M,n,I_{k_n})} - \hat{\beta}^{(M,n)}\right)$$

$$= -\gamma^{-1} Q^{-1} \sum_{i \in I_{k_n}} g'\left(X_i, \hat{\beta}^{(n,I_{k_n})}\right) \psi\left(\left[Y_i - g(X_i, \hat{\beta}^{(M,n,I_{k_n})})\right] \hat{\sigma}_n^{-1}\right) + o_p(1)$$

as $n \to \infty$,

$$\gamma = \sigma^{-1} \mathbb{E}_F \, \psi'\left(e_1 \cdot \sigma^{-1}\right) + \sum_{k=1}^{s_1} f(r_{1,k}\sigma)\left[\psi(r_{1,k}+) - \psi(r_{1,k}-)\right],$$

$$Q = \mathbb{E} X_1 X_1'$$

Víšek, J. Á. (2002): Sensitivity analysis of *M*-estimates of nonlinear regression model: Influence of data subsets. *Ann. Inst. of Statist. Mathematics, 54, 261 - 290.*

M-etimators with discontinuous ψ

$$n\left(\hat{\beta}^{(M,n,I_{k_n})} - \hat{\beta}^{(M,n)}\right) = -\gamma^{-1} Q^{-1} \left\{ \sum_{i \in I_{k_n}} X_i \cdot \psi\left(\left[Y_i - X_i' \hat{\beta}^{(M,n)}\right] \hat{\sigma}_n^{-1}\right) + \mathcal{R}_n \right\}$$

$$+o_p(1) \text{ as } n \to \infty$$

where

$$(\mathcal{R}_n)_j =_{\mathcal{D}} W_j\left(\sum_{i=1}^{n} \tau_{ijn}(\sqrt{n}(\hat{\beta}^{(n)} - \beta^0), n(\hat{\beta}^{(n,I_{k_n})} - \hat{\beta}^{(n)}), \sqrt{n}(\log\hat{\sigma}_n - \log\sigma))\right)$$

with

$$\max_{1 \le j \le p} \sup_{\|t\| \cdot \|u\| \cdot |v| < M} \left| W_j\left(\sum_{i=1}^{n} \tau_{ijn}(t, u, v)\right)\right| = \mathcal{O}_p(1) \qquad \text{as } n \to \infty,$$

for some stopping times $\tau_{ik}^{(j)}$ and $W(s)$ as a Wiener process.

We have met already with:

Hettmansperger, T. P., S. J. Sheather (1992):
A Cautionary Note on the Method of Least Median Squares.
The American Statistician 46, 79–83.

Engine Knock Data $(n = 16, p = 4, h = 11)$

c	x_1	x_2	x_3	x_4	y
1	13.3	13.9	31	697	84.4
2	13.3	14.1	30	697	84.1
3	13.4	15.2	32	700	88.4
4	12.7	13.8	31	669	84.2
⋮	⋮	⋮	⋮	⋮	⋮
14	12.7	16.1	35	649	93.0
15	12.9	15.1	36	721	93.3
16	12.7	15.9	37	696	93.1

x_1 is spark timing $\quad$ x_2 air/fuel ratio
x_3 intake temperature $\quad$ x_4 exhaust temperature
y engine knock number

Results of analysis of Engine Knock Data

(Point 3 appeared to be the most influential.)

L_1	Intercept	Air/Fuel	Intake
Full data	31.84	2.471	0.594
Data without point 3	34.10	1.500	0.950

Huber ψ with tunning constant 1.2	Intercept	Air/Fuel	Intake
Full data	31.97	1.785	0.896
Data without point 3	32.71	1.639	0.937

Hampel ψ with tunning constant 1.2	Intercept	Air/Fuel	Intake
Full data	27.58	2.096	0.885
Data without point 3	28.49	1.934	0.929

The estimates are rather *subsample stable*.

Results of analysis of Health Club Data

Rousseeuw, P. J., A. M. Leroy (1987):
Robust Regression and Outlier Detection.
New York: J.Wiley & Sons.

(Point 20 appeared to be the most influential.)

L_1	Intercept	Weight	Pulse	Strength	$\frac{1}{4}$ mile
Full data	-57.03	1.090	-0.928	-0.317	4.853
Data without point 20	8.69	0.806	-2.238	-0.365	5.958

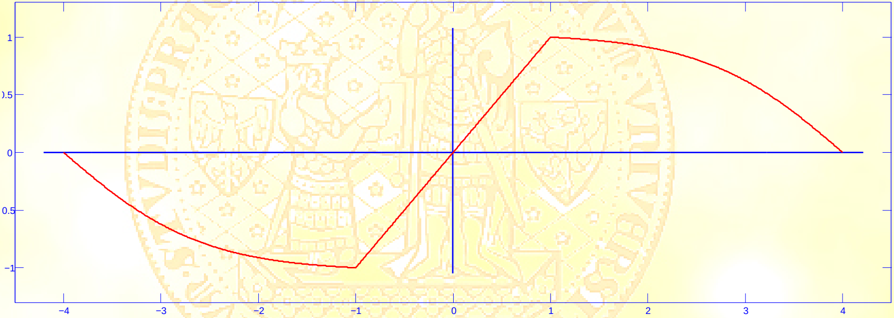
Huber ψ with t.c. 1.2	Intercept	Weight	Pulse	Strength	$\frac{1}{4}$ mile
Full data	8.06	1.303	-0.777	-0.538	3.969
Data without point 20	12.18	1.273	-0.868	-0.531	4.048

Hampel ψ with t.c. 1.2	Intercept	Weight	Pulse	Strength	$\frac{1}{4}$ mile
Full data	12.49	1.316	-0.873	-0.553	4.004
Data without point 20	12.82	1.298	-0.849	-0.549	4.019

The estimates are again rather *subsample stable*.

Example of searching for an optimal *M*-estimator of location.

Probably the most famous redescending ψ-function - Hample's one



Results of analysis of U. S. Crime Data

Rousseeuw, P. J., A. M. Leroy (1987):
Robust Regression and Outlier Detection.
New York: J.Wiley *&* Sons.

(Point 3 appeared to be the most influential.)

L_1	Intercept	Age	Education	Police	Income
Full data	450.3	0.426	-0.018	-2.096	-0.795
Data without point 3	389.5	0.507	0.250	-1.818	-0.946

Huber ψ with t.c 1.2	Intercept	Age	Education	Police	Income
Full data	406.8	0.476	0.241	-2.073	-0.819
Data without point 3	404.6	0.472	0.248	-2.066	-0.811

Hampel ψ with t.c. 1.2	Intercept	Age	Education	Police	Income
Full data	403.1	0.477	0.281	-2.120	-0.781
Data without point 3	399.9	0.471	0.292	-2.107	-0.773

The estimates are again rather *subsample stable*.

Preliminary conclusion

Stability of M-estimators on subsamples requires continuous ψ's.

Recalling:

$$\hat{\beta}^{(LTS,n,h)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{h} r_{(i)}^2(\beta)$$

Hampel, F. R. et al. (1986):
Robust Statistics – The Approach Based on Influence Functions.
New York: J.Wiley & Son.

Main result

Sensitivity study of LTS

$$n\left(\hat{\beta}^{(LTS,n,h)} - \hat{\beta}^{(LTS,n-1,\ell,h)}\right)$$

$$= Q_n^{-1}\left[(1-\alpha_0) - 2\cdot u_{\alpha_0}\cdot f(u_{\alpha_0}) + \mathcal{R}_n\right]^{-1} \times$$

$$\times\; X_\ell\left(Y_\ell - X_\ell'\hat{\beta}^{(LTS,n,h)}\right) I\left\{r_\ell^2(\hat{\beta}^{(LTS,n,h)}) \le r_{(h:n)}^2(\hat{\beta}^{(LTS,n,h)})\right\}$$

$$+ o_p(1) \text{ as } n \to \infty.$$

Víšek, J. Á. (2006): The least trimmed squares. Sensitivity study. *Proc. of the Prague Stochastics 2006, 728-738.*

Technicalities

$$n\left(\hat{\beta}^{(LTS,n,h)} - \hat{\beta}^{(LTS,n-1,\ell,h)}\right)$$

$$= Q_n^{-1}\left[(1-\alpha_0) - 2\cdot u_{\alpha_0}\cdot f(u_{\alpha_0}) + \mathcal{R}_n\right]^{-1} \times$$

$$\times\ X_\ell\left(Y_\ell - X_\ell'\hat{\beta}^{(LTS,n,h)}\right) I\left\{r_\ell^2(\hat{\beta}^{(LTS,n,h)}) \le r_{(h:n)}^2(\hat{\beta}^{(LTS,n,h)})\right\}$$

$$+o_p(1) \text{ as } n \to \infty.$$

$$\xi_i = I\left\{r_i^2(\hat{\beta}^{(LTS,n-1,\ell,h)}) \le r_{(h:n-1,\ell)}^2(\hat{\beta}^{(LTS,n-1,\ell,h)})\right\} - I\left\{r_i^2(\hat{\beta}^{(LTS,n,h)}) \le r_{(h:n)}^2(\hat{\beta}^{(LTS,n,h)})\right\}$$

and

$$\mathcal{R}_n = u_{\alpha_0}\sum_{i=1}^{n} sign(e_i)X_i\left(\xi_i - \boldsymbol{E}\xi_i\right)$$

where u_{α_0} is (two-sided) α_0-quantile.

Sensitivity study of LWS

$$n\left(\hat{\beta}^{(LWS,n,w)} - \hat{\beta}^{(LWS,n-1,\ell,w)}\right)$$

$$= \left[\mathbb{E}\left\{w(F(|e_1|))X_1X_1'\right\}\right]^{-1} w(F(|Y_\ell - X_\ell'\hat{\beta}^{(LWS,n,h)}|))\,X_\ell\left(Y_\ell - X_\ell'\hat{\beta}^{(LWS,n,w)}\right)$$

$$+o_p(1) \text{ as } n \to \infty.$$

(in draft)

Conjecture of the sensitivity of IWV

$$n\left(\hat{\beta}^{(IWV,n,w)} - \hat{\beta}^{(IWV,n-1,\ell,w)}\right)$$

$$= \left[\mathbb{E}\left\{ w(F(|e_1|))Z_1X_1'\right\}\right]^{-1} w(F(|Y_\ell - X_\ell'\hat{\beta}^{(IWV,n,h)}|))\, Z_\ell \left(Y_\ell - X_\ell'\hat{\beta}^{(IWV,n,w)}\right)$$

$$+o_p(1) \text{ as } n \to \infty.$$

THANKS FOR ATTENTION