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

ROBUST STATISTICS AND ECONOMETRICS

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Jan Ámos Víšek

Week 10

Content of lecture

1. At the beginning of any lecture let us repeat .....
2. Problems with orthogonality condition
   - Total least squares - from definition to robustification
   - Instrumental variables - from definition to robustification

Repeating LWS - definition and normal equations

We have defined the estimator $\hat{\beta}^{(LWS,n,w)}$

$$\arg\min_{\beta \in R^p} \sum_{i=1}^{n} w\left(\frac{k-1}{n}\right) r^2_{(k)}(\beta)$$

where $r^2_{(k)}(\beta)$ is the k-th order statistics among the squared residuals $r^2_i(\beta) = (Y_i - X'_i\beta)^2$, i. e.

$$r^2_{(1)}(\beta) \le r^2_{(2)}(\beta) \le \dots \le r^2_{(n)}(\beta).$$

Put

$$\pi(\beta, i) = k \in \{1, 2, \dots, n\} \qquad \text{if} \qquad r^2_i(\beta) = r^2_{(k)}(\beta).$$

By words:

$\pi(\beta, i)$ is the number of squared residuals which are not larger than the i-th squared residual.

Repeating LWS - definition and normal equations

So, once again we have defined the estimator $\hat{\beta}^{(LWS,n,w)}$

$$\underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w\left(\frac{k-1}{n}\right) r_{(k)}^2(\beta)$$

and we have put

$$\pi(\beta, i) = k \in \{1, 2, ..., n\} \quad \text{if} \quad r_i^2(\beta) = r_{(k)}^2(\beta), i.e.$$

So, we have

$$\underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w\left(\frac{\pi(\beta, i)-1}{n}\right) r_i^2(\beta).$$

Repeating LWS - normal equations

Then we have proved that the estimator $\hat{\beta}^{(LWS,n,w)}$
is one of the solutions of the *normal equations*

$$\sum_{i=1}^{n} w\left(\frac{\pi(\beta,i)-1}{n}\right) X_i(Y_i - X_i'\beta) = 0$$

where $\pi(\beta, i)$ is the rank of the i-th squared residual, i. e.

$$\pi(\beta,i) = k \in \{1,2,...,n\} \qquad \text{if} \qquad r_i^2(\beta) = r_{(k)}^2(\beta).$$

By words:
$\pi(\beta, i)$ is the number of squared residuals
which are not larger than the i-th squared residual.

Ranks of squared residuals and of absolute values of residuals

Rewrite from previous slide

$$\pi(\beta, i) = k \in \{1, 2, ..., n\} \quad \text{if} \quad r_i^2(\beta) = r_{(k)}^2(\beta).$$

Now, lets realize that

$$r_k^2(\beta) \le r_\ell^2(\beta) \quad \Leftrightarrow \quad |r_k(\beta)| \le |r_\ell(\beta)|.$$

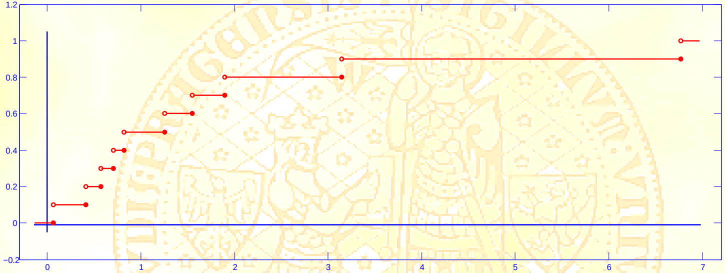
Denoting (a bit non-traditionally) $|r(\beta)|_{(i)}$ the i-th order statistic amog the absolute values of residuals, we have

$$\pi(\beta, i) = k \in \{1, 2, ..., n\} \quad \text{if} \quad r_i(\beta) = |r(\beta)|_{(k)}.$$

By words:

$\pi(\beta, i)$ is the number of residuals which absolute value is not larger than the i-th value among these absolute values.

Consider e. d. f. $F_n(.)$ of absolute values of residuals.



Fix $x_0 \in R$ and ask what is the value of $F_n(x_0)$?
It is, of course, the number of absolute values of residuals which are smaller than x_0 divided by n.
And what is now the value of $F_n(.)$ at $|r_\ell(\beta)|$?
It is again the number of absolute values of residuals which are smaller than $|r_\ell(\beta)|$ divided by n.

But it is just $\frac{\pi(\beta,\ell)-1}{n}$, as we have found on the previous slide !

Normal equations for LWS

We have proved that the estimator $\hat{\beta}^{(LWS,n,w)}$ is one of the solutions of the *normal equations*

$$\sum_{i=1}^{n} w\left(\frac{\pi(\beta,i)-1}{n}\right) X_i(Y_i - X_i'\beta) = 0.$$

We can substitute $\frac{\pi(\beta,i)-1}{n}$ by $F_n(|r_i(\beta)|)$, i. e.

$$\sum_{i=1}^{n} w\left(F_n(|r_i(\beta)|)\right) X_i(Y_i - X_i'\beta) = 0.$$

Comparing OLS and LWS - the definitions

The ordinary least squares

$$\hat{\beta}^{(OLS,n)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} r_i^2(\beta) = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} r_{(i)}^2(\beta)$$

The least weighted squares

$$\hat{\beta}^{(LWS,n,w)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} \boxed{w\left(\frac{i-1}{n}\right)} \; r_{(i)}^2(\beta)$$

Notice that robustification of the ordinary least squares
is accomplished just by an "implicit" weighting,
i. e. assigning the weights to the order statistics.

Comparing OLS and LWS - the normal equations

The ordinary least squares

$$\sum_{i=1}^{n} X_i(Y_i - X_i'\beta) = 0.$$

The least weighted squares

$$\sum_{i=1}^{n} \boxed{w\left(F_n(|r_i(\beta)|)\right)} X_i(Y_i - X_i'\beta) = 0.$$

Notice that robustification of the OLS normal equations is accomplished again just by an "implicit" weighting, i. e. including the weights $w\left(F_n(|r_i(\beta)|)\right)$.

We will need it a bit later.

An algorith for LWS

LWS - ALGORITHM

A

Find the plane through $p + 1$ randomly selected observations.

Evaluate squared residuals of all observations. Then sum up the products of the weights and of the order statistics of squared residuals and the sum denote $S(\hat{\beta}_{present})$.

Is $S(\hat{\beta}_{present})$ less than $S(\hat{\beta}_{past})$?

no → B

yes

Establish *new* $\hat{\beta}_{present}$ just applying *WLS* on the reordered observations (reoredered according to the squared residuals).

An algorith for LWS

LWS - ALGORITHM(*continued*)

B

Was ℓ-times found the same model with minimal value of $S(\beta)$?

no

Was already k-times repeated outer cycle ?

no → A

yes

yes

As $\hat{\beta}^{(LWS,n,w)}$ we will assume $\beta \in R^p$ for which the functional $S(\beta)$ attained - through just described iterations - minimal value.

PROS AND CONS OF LWS

"Inherited" from LTS:

$\sqrt{n}$-consistency (even under heteroscedasticity)

Scale- and affine-equivariance

Quick and reliable algorithm (implemented in MATLAB and R)

PROS AND CONS OF LWS (*continued*)

Moreover:

Breakdown point and efficiency adaptable not only to level but also to character of contamination

Diagnostic tools:

1. Significance of the individual explanatory variable
2. Durbin-Watson test, White test, Hausman test
3. Test of submodels

Modifications for nonstandard situations (e. g. instrumental variables, models with fixed and random effects, ridge regression, estimation with constraints)

Low sensitivity to the shift and deletion of observation(s)

Applicability for panel data

“Coping automatically” with heteroscedasticity of data - empirical experience

Pros and cons of LWS$_{(\textit{continued})}$

Still (more or less) lacking:

Determination of model

Problems with the broken orthogonality condition

The ordinary least squares

$$\hat{\beta}^{(OLS,n)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} (Y_i - X_i'\beta)^2 = (X'X)^{-1} X'Y$$

$$= \beta^0 + \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i \varepsilon_i.$$

IF ORTHOGONALITY CONDITION IS BROKEN:

Explanatory variables are correlated with disturbances.

$$\operatorname*{plim}_{n \to \infty} \frac{1}{n}\sum_{i=1}^{n} X_i \varepsilon_i = \mathbb{E}\{X_1 \varepsilon_1\} \neq 0,$$

$\hat{\beta}^{(OLS,n)}$ is biased and inconsistent.

How frequently does it happen ?

Problems with the broken orthogonality condition

Frequently given examples of situations when:

Explanatory variables are correlated with disturbances

General examples:

1. Measurement of explanatory variable with a random error,

The most important fact (in social sciences):
Disturbances contain some (part of) explanatory variables !!!
(typically correlated with the variables in model)

Specific examples:

1. Consumption always depends on the income of households,
2. infla

How to cope with it ?

A remedy for the broken orthogonality condition

Basically two possibilities

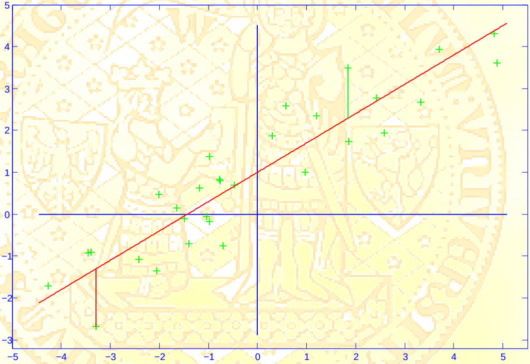
The total least squares
- mostly used in natural and technical sciences

The instrumental variables
- typically used in econometrics, generally in social sciences

Both methods are not robust !!!

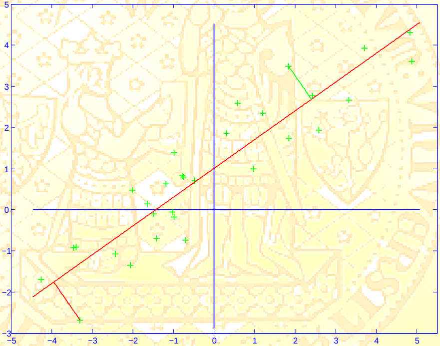
OLS

OLS minimizes the sum of squared residuals parallel to axe y.



TLS

The total least squares minimizes the sum of squared residuals orthogonal to regression plane.



Minimal modification of n equations to obtain a solution

An alternative approach to data processing

1. Assume a system of n equations with p unknown variables, $(p < n)$, say $Y = X\beta$, i. e. β is unknown.
2. The system is typically overdetermined and hence it has generally no solution.
3. We admit that we measured Y with an error e but we hope that e is small.
4. Define an extremal problem:

$$\hat{e} = \arg\min_{e \in R^n} \left\{ \|e\|^2 : Y - e = X\beta \text{ has (at least) one solution} \right\}.$$

(last row is rewritten on the next slide).

Minimal modification of n equations to obtain a solution

Solve the problem

$$\hat{e} = \arg\min_{e \in R^n} \left\{ \|e\|^2 : Y - e = X\beta \text{ has (at least) one solution} \right\}.$$

It is equivalent to:

$$\hat{e} = \arg\min_{e \in R^n} \left\{ \|Y - X\beta\|^2 : Y - e = X\beta \text{ has (at least) one solution} \right\}. \quad (1)$$

It appeared that whenever we have $\hat{\beta} \in R^p$ so that

$$\hat{\beta} = \arg\min_{\beta \in R^p} \left\{ \|Y - X\beta\|^2 \right\},$$

$\hat{e} = Y - X\hat{\beta}$ is the unique solution of (1)

with $\hat{\beta} = (X'X)^{-1} X'Y$ and hence $\hat{e} = \left(I\!I - X(X'X)^{-1} X' \right) Y$

(where we have assumed X to be of full rank

- and we'll assume it hereafter).

Recalling technicalities

In what follows we will assume
- similarly as in the basic course of econometrics:

1. The design matrix X is of full rank,
   i. e. the columns are linearly independent
   - or in other words - $X'X$ is regular.

2. The matrix $[Y, X]$ is also of full rank,
   i. e. Y is not a linear combination of the columns of design matrix,
   - or in other words - $[Y, X]' [Y, X]$ is regular.

Minimal modifications of n equations to obtain a solution

An alternative approach to data processing

1. Assume a system of n equations with p unknown variables, ($p < n$), say $Y = X\beta$, i. e. β is unknown.
2. The system is typically overdetermined and hence it has generally no solution.

The first two points of framework are the same - - too much equations for p unknown quantities !

Minimal modifications of n equations to obtain a solution

An alternative approach to data processing

1. We admit that we measured:
   - Y with an error e

   and

   - X with an error $\hat{X} - X$

   but we hope that $\|e\|$ and $\left\|\hat{X} - X\right\|$ are small.

2. Define an extremal problem:

$$\left(\hat{e}, \hat{X}\right) = \underset{e \in R^n,\ \tilde{X}\ \text{matrix of type}\ R^n \times R^p}{\arg\min} \left\{ \|e\|^2 + \left\|\tilde{X} - X\right\|^2 : \right.$$
$$\left. Y - e = \tilde{X}\beta \text{ has (at least) one solution} \right\}.$$

The solution is not so simple as in the previous case !

Total least squares - TLS

$$\hat{\beta}^{TLS,n} = \underset{\beta \in R^p,\ \tilde{X}\ matrix\ of\ type\ R^n \times R^p}{\arg\min} \sum_{i=1}^{n} r_i^2(\beta, \tilde{X})$$

where $r_i^2(\beta, \tilde{X}) = \left(Y_i - \tilde{X}_i'\beta\right)^2 + \left\|X_i - \tilde{X}_i\right\|^2$ are consistent

but not robust.

Evaluating “ordinary” total least squares

Denoting $\hat{e}_i(\beta, \hat{X}) = Y_i - \hat{X}_i'\beta$,

$$\hat{e}(\beta, \hat{X}) = \begin{bmatrix} \hat{e}_1(\beta, \hat{X}) \\ \hat{e}_2(\beta, \hat{X}) \\ \vdots \\ \hat{e}_n(\beta, \hat{X}) \end{bmatrix}, \quad \text{and} \quad \hat{X} = \begin{bmatrix} \hat{X}_1' \\ \hat{X}_2' \\ \vdots \\ \hat{X}_n' \end{bmatrix},$$

we look for a pair $\left(\hat{\beta}, \hat{X}\right)$ which solves

$$Y = \hat{X}\hat{\beta} + \hat{e}(\hat{\beta}, \hat{X}) \tag{2}$$

with $S(\hat{\beta}, \hat{X}) = \sum_{i=1}^{n} \left\{ \left(Y_i - \hat{X}_i'\beta\right)^2 + \left\| X_i - \hat{X}_i \right\|^2 \right\}$ is to be minimal.

Evaluating “ordinary” total least squares

(2) can be written as

$$\hat{e}(\hat{\beta}, \hat{X}) - Y + \hat{X}\hat{\beta} = 0.$$

Denoting $\hat{Y} = Y - \hat{e}(\hat{\beta}, \hat{X})$, we have

$$-\hat{Y} + \hat{X}\hat{\beta} = 0$$

or

$$\left[\ \hat{Y}, \hat{X}\ \right] \cdot \begin{bmatrix} -1 \\ \hat{\beta} \end{bmatrix} = 0$$

with $\hat{e}(\hat{\beta}, \hat{X}) = Y - \hat{Y}$ and $\sum_{i=1}^{n} \left(\hat{e}_i^2(\hat{\beta}, \hat{X}) + \left\| X_i - \hat{X}_i \right\|^2 \right)$ minimal.

(The last two rows are rewritten on the next slide.)

Making preparatory steps for finding TLS

$$\left[\ \hat{Y}, \hat{X}\ \right] \cdot \begin{bmatrix} -1 \\ \hat{\beta} \end{bmatrix} = 0$$

with $\hat{e}(\hat{\beta}, \hat{X}) = Y - \hat{Y}$ and $\sum_{i=1}^{n} \left(\hat{e}_i^2(\hat{\beta}, \hat{X}) + \left\| X_i - \hat{X}_i \right\|^2 \right)$ minimal.

It means:

1. We look for $\left[\hat{Y}, \hat{X}\right]$ having the difference in norm from $[Y, X]$ as small as possible.
2. We want to find a vector $\xi \perp \left[\hat{Y}, \hat{X}\right]$.

To be able to find $\xi \perp \left[\hat{Y}, \hat{X}\right]$,

we have to have $rank(\left[\hat{Y}, \hat{X}\right]) < rank([Y, X])$

as $[Y, X]$ is of full rank $p + 1$).

Making preparatory steps for finding TLS

We will need singular decomposition of matrix $[Y, X]$.

Let us consider the matrix $[Y, X]' \cdot [Y, X]$ and denote

$$q_1, q_2, ..., q_{p+1} \quad \text{and} \quad s_1^2 \geq s_2^2 \geq ... \geq s_{p+1}^2 > 0$$

its eigenvectors and eigenvalues, i. e.

$$[Y, X]' \cdot [Y, X] \cdot q_i = s_i^2 \cdot q_i$$

(recall - they are real and positive).

Put $Q = [q_1, q_2, ..., q_{p+1}]$, then $Q'Q = QQ' = I\!\!I$ = (commutative).

Put $S = diag\{s_1, s_2, ..., s_{p+1}\}$ and (in matrix form) we have

$$[Y, X]' \cdot [Y, X]\, Q = Q \cdot S \cdot S$$

remember it, we will need it.

Making preparatory steps for finding TLS

Let's recall that q_i's can be selected so that they create orthonormal base of R^{p+1}.

Recall that $Q = [q_1, q_2, ..., q_p]$ and $Q'Q = QQ' = \mathbb{I}$ = (commutative).
Put

$$u_i = s_i^{-1} \cdot [Y, X] \cdot q_i, \quad U = [u_1, u_2, ..., u_{p+1}]$$

and recall that

$$S = diag\{s_1, s_2, ..., s_{p+1}\}.$$

Then

$$U = [Y, X] \cdot Q \cdot S^{-1}.$$

Making preparatory steps for finding TLS

From previous slide $U = [Y, X] \cdot Q \cdot S^{-1}$.

Hence $U \cdot S = [Y, X] \cdot Q$ and $U \cdot S \cdot Q' = [Y, X]$.

It is called singular decomposition

$$[Y, X] = U \cdot S \cdot Q' = \sum_{i=1}^{p+1} s_i \cdot u_i \cdot q_i',$$

(notice that $[Y, X]$ is combination of u_i's).

moreover

$$U'U = S^{-1} \cdot Q' [Y, X]' \cdot [Y, X] \cdot Q \cdot S^{-1}$$

$$= S^{-1} \cdot Q' \cdot Q \cdot S = I$$

(remeber it !!).

Making preparatory steps for finding TLS

Let A be a matrix of type $\ell \times k$. Then

$$\|A\|^2 = trace\,(A'A)\,.$$

Proof: Recalling that $\|A\|^2 = \sum_{j=1}^{k} \sum_{r=1}^{\ell} A_{r,j}^2$, we have

$$trace\,(A'A) = \sum_{j=1}^{k} (A'A)_{j,j} = \sum_{j=1}^{k} \left[\sum_{r=1}^{\ell} A_{r,j}^2 \right].$$

Q.E.D.

Similarly as for $[Y,X] = U \cdot S \cdot Q' = \sum_{i=1}^{p+1} s_i \cdot u_i \cdot q_i'$, we can look for

$$\left[\hat{Y},\hat{X}\right] = \sum_{i=1}^{p+1} \lambda_i s_i \cdot u_i \cdot q_i' = U \cdot \Lambda \cdot S \cdot Q'$$

(notice summation up to $p+1$)

for some (unknown) numbers λ_i's and we have put

$$\Lambda = diag\{\lambda_1, \lambda_2, ..., \lambda_{p+1}\}.$$

Making preparatory steps for finding TLS

Then
$$[Y, X] - \left[\hat{Y}, \hat{X}\right] = U \cdot (I - \Lambda) \cdot S \cdot Q'$$

with $S' = S$, $(I - \Lambda)' = (I - \Lambda)$ and

$$S' \cdot (I - \Lambda)' \cdot (I - \Lambda) \cdot S = (I - \Lambda)^2 \cdot S^2$$
$$= diag\{(1-\lambda_1)^2, (1-\lambda_2)^2, ..., (1-\lambda_{p+1})^2\} \cdot diag\left\{s_1^2, s_2^2, ..., s_{p+1}^2\right\}.$$

It gives

$$\left\{[Y, X] - \left[\hat{Y}, \hat{X}\right]\right\}' \left\{[Y, X] - \left[\hat{Y}, \hat{X}\right]\right\} = Q \cdot S \cdot (I - \Lambda) \cdot \underbrace{U'U}_{I} \cdot (I - \Lambda) \cdot S \cdot Q'$$
$$= Q \cdot (I - \Lambda)^2 \cdot S^2 \cdot Q'.$$

Making preparatory steps for finding TLS

Finally

$$\left\| [Y, X] - \left[\hat{Y}, \hat{X}\right] \right\|^2 = trace\left(Q \cdot (\mathbb{I} - \Lambda)^2 \cdot S^2 \cdot Q'\right)$$

$$trace\left(\underbrace{Q'Q}_{\mathbb{I}} \cdot (\mathbb{I} - \Lambda)^2 \cdot S^2\right) = trace\left((\mathbb{I} - \Lambda)^2 \cdot S^2\right)$$

which is evidently minimal for $\lambda_i = 1, i = 1, 2, ..., p$ and $\lambda_{p+1} = 0$.

Then $\left\| [Y, X] - \left[\hat{Y}, \hat{X}\right] \right\|^2 = s^2_{p+1}$.

Establishing the solution for TLS

So, we have for $I\!\!I - \Lambda = diag\{1, 1, ..., 1, 0\}$ and

$$\left[\hat{Y}, \hat{X}\right] = U \cdot (I\!\!I - \Lambda) \cdot S \cdot Q' = \sum_{i=1}^{p} s_i \cdot u_i \cdot q_i'.$$

Then we have

$$\left[\hat{Y}, \hat{X}\right] \cdot q_{p+1} = \sum_{i=1}^{p} s_i \cdot u_i \cdot q_i' \cdot q_{p+1} = 0$$

and hence $\begin{bmatrix} -1 \\ \hat{\beta} \end{bmatrix} = \gamma \cdot q_{p+1}$.

Then by normalization we have $\gamma = \frac{1}{s_{p+1}}$ and

$$\begin{bmatrix} -1 \\ \hat{\beta} \end{bmatrix} = \frac{1}{s_{p+1}} q_{p+1}, \qquad \left[\hat{Y}, \hat{X}\right] \begin{bmatrix} -1 \\ \hat{\beta} \end{bmatrix} = 0$$

and

$$S(\hat{\beta}, \hat{X}) = \sum_{i=1}^{n} \hat{e}_i^2(\hat{\beta}, \hat{X}) = \sum_{i=1}^{n} \left\{ \left(Y_i - \hat{X}_i' \hat{\beta} \right)^2 + \left\| X_i - \hat{X}_i \right\|^2 \right\} = s_{p+1}^2.$$

Robustifying *TLS* - generalized *M*-estimators - William H. Jeffrys

Instead of

$$\hat{\beta}^{TLS,n} = \underset{\beta \in R^p,\ \tilde{X}\ matrix\ of\ type\ R^n \times R^p}{\arg\min} \sum_{i=1}^{n} r_i^2(\beta, \tilde{X})$$

where $r_i^2(\beta, \tilde{X}) = \left(Y_i - \tilde{X}_i' \beta\right)^2 + \left\| X_i - \tilde{X}_i \right\|^2$

put

$$\hat{\beta}^{MTLS,n} = \underset{\beta \in R^p,\ \tilde{X}\ matrix\ of\ type\ R^n \times R^p}{\arg\min} \sum_{i=1}^{n} \rho\left(\frac{r_i^2(\beta, \tilde{X})}{s_n}\right).$$

Jeffrys, W. H. (1990): Robust estimation when more than one variable per equation of condition has error. *Biometrika 77, 597 - 607.*

Robustifying *TLS* - Total least weighted squares (*TLWS*)

Instead of

$$\hat{\beta}^{TLS,n} = \underset{\beta \in R^p,\ \tilde{X}\ matrix\ of\ type\ R^n \times R^p}{\arg\min} \sum_{i=1}^{n} r_i^2(\beta, \tilde{X})$$

where $r_i^2(\beta, \tilde{X}) = \left(Y_i - \tilde{X}_i'\beta\right)^2 + \left\| X_i - \tilde{X}_i \right\|^2$

put

$$\hat{\beta}^{TLWS,n} = \underset{\beta \in R^p,\ \tilde{X}\ matrix\ of\ type\ R^n \times R^p}{\arg\min} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta, \tilde{X})$$

$$= \underset{\beta \in R^p,\ \tilde{X}\ matrix\ of\ type\ R^n \times R^p}{\arg\min} \sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right) r_i^2(\beta, \tilde{X}).$$

($F_{\beta}^{(n)}(|r_j(\beta)|)$ is again e.d.f. of $|r_j(\beta)|$'s)

Robustifying *TLS* - preparing algorithm

$\hat{\beta}^{(TLWS,n,w)}$ minimizes the functional

$$\tilde{S}(\beta,\tilde{X}) = \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) \cdot r^2_{(i)}(\beta,\tilde{X}).$$

Let $F^{(n)}_{\beta}(|r_j(\beta)|)$ be e.d.f. of $|r_i(\beta,\tilde{X})| = \sqrt{\left(Y_i - \tilde{X}_i'\beta\right)^2 + \left\|X_i - \tilde{X}_i\right\|^2}$
and put

$$\tilde{Y}_i = \sqrt{w\left(F^{(n)}_{\beta}(|r_j(\beta)|)\right)} \cdot Y_i \quad \text{and} \quad \tilde{X}_i = \sqrt{w\left(F^{(n)}_{\beta}(|r_j(\beta)|)\right)} \cdot X_i$$

Robustifying *TLS* - algorithm

A Find the plane through $p+1$ randomly selected observations.

↓

Evaluate $r_i^2(\beta, \tilde{X})$ of all observations, find their ranks and transform data to $\left(\tilde{Y}, \tilde{X}\right)$. Evaluate $\tilde{S}(\hat{\beta}_{present}, \tilde{X}_{present})$.

↓

Is $\tilde{S}(\hat{\beta}_{present})$ less than $\tilde{S}(\hat{\beta}_{past})$?

no → **B**

↓ yes

Establish *new* estimate of β^0 employing data $\left(\tilde{Y}, \tilde{X}\right)$ and the algorithm for "Ordinary" TLS and consider it instead of previous estimate.

(loop back to "Evaluate")

Robustifying *TLS* - algorithm

B

Was ℓ-times found the same model with minimal value of $\tilde{S}(\beta, \tilde{X})$?

no ↓

Was already k-times repeated outer cycle ? → no → A

yes ↓ (also yes from the first question)

As $\hat{\beta}^{(TLWS,n,w)}$ we will assume $\beta \in R^p$ for which the functional $\tilde{S}(\beta, \tilde{X})$ attained - through just described iterations - minimal value.

Let's recall once again: Problems with the broken orthogonality condition

The ordinary least squares

$$\hat{\beta}^{(OLS,n)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} (Y_i - X_i'\beta)^2 = (X'X)^{-1} X'Y$$

$$= \beta^0 + \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i \varepsilon_i.$$

IF ORTHOGONALITY CONDITION IS BROKEN:

Explanatory variables are correlated with disturbances.

$$\operatorname{plim}_{n \to \infty} \frac{1}{n}\sum_{i=1}^{n} X_i \varepsilon_i = \mathbb{E}\{X_1 \varepsilon_1\} \neq 0,$$

$\hat{\beta}^{(OLS,n)}$ is biased and inconsistent.

Method of the instrumental variables - heuristics

Method of *the least squares* is the solution of normal equations

$$\sum_{j=1}^{n} X_j \left(Y_j - X_j' \beta \right) = 0.$$

Let's look for *"substitutes" (instruments)* for X_j, say Z_j,
which will be "near" to X_j, but

$$\operatorname*{plim}_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} Z_j \varepsilon_j = \mathbb{E} Z_1 \varepsilon_1 = 0.$$

Then the solution of normal equations

$$\sum_{j=1}^{n} Z_j \left(Y_j - X_j' \beta \right) = 0$$

will be called the estimate by means of
the method of instrumental variables.

Method of the instrumental variables

The vector equation

$$\sum_{i=1}^{n} Z_i \left(Y_i - X_i' \beta \right) = 0$$

can be rewritten into the matrix form

$$Z \left(Y - X\beta \right) = 0.$$

It yields immediately the formula for estimator

$$\hat{\beta}^{(IV,n)} = \left(Z'X \right)^{-1} Z'Y.$$

Plugging in the formula for the regression model $Y = X\beta^0 + e$, we obtain

$$\hat{\beta}^{(IV,n)} = \beta^0 + \left(Z'X \right)^{-1} Z'e.$$

Method of the instrumental variables

So, the estimate by means of the *Instrumental Variables*

$$\hat{\beta}^{(IV,n)} = \beta^0 + \left(\frac{1}{n}\sum_{i=1}^n Z_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n Z_i \varepsilon_t,$$

→ *unbiased* and *consistent,* not *robust*.

Robustification is straightforward !

$$\sum_{i=1}^n w\left(F_\beta^{(n)}(|r_i(\beta)|)\right) Z_i \left(Y_i - X_i'\beta\right) = 0$$

Method of the instrumental variables

Definition

The estimate by means of the *Instrumental weighted variables* $\hat{\beta}^{(IWV,n,w)}$ is any solution of *normal equations*

$$\sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right) Z_i\left(Y_i - X_i'\beta\right) = 0.$$

Víšek, J. Á. (2004): Robustifying instrumental variables.
Proc. COMPSTAT'2004,Physica-Verlag/Springer, 1947 - 1954.

Instrumental weighted variables (IWV) - asymptotic theory

C1 *Random variables*

- $\{(X_i', Z_i', \varepsilon_i)'\}_{i=1}^{\infty} \subset R^{2p+1}$ is a sequence of i.i.d. r.v.'s,
- $\forall \ (i \in N) \quad Z_i$ a ε_i are independent,
- distribution function $F_{X,Z}(x, z)$ is absolutly continuous,
- $\mathbb{E}\left\{w\left(F_{\beta^0}(|\varepsilon_1|)\right) Z_1 X_1'\right\}$ a $\mathbb{E}\left\{Z_1 Z_1'\right\}$ is pozitive definite,
- $\exists \ (q > 1) \ : \quad \mathbb{E}\left\{\|Z_1\| \cdot \|X_1\|\right\}^q < \infty.$

Instrumental weighted variables - asymptotic theory

C2 *Weight function is as follows:*

- $w(\alpha) : [0,1] \rightarrow [0,1], w(0) = 1,$
- absolutely continuous, non-increasing,
- $\exists$ almost everywhere the derivative $w'(\alpha) > -L, \; L \in R^+$.

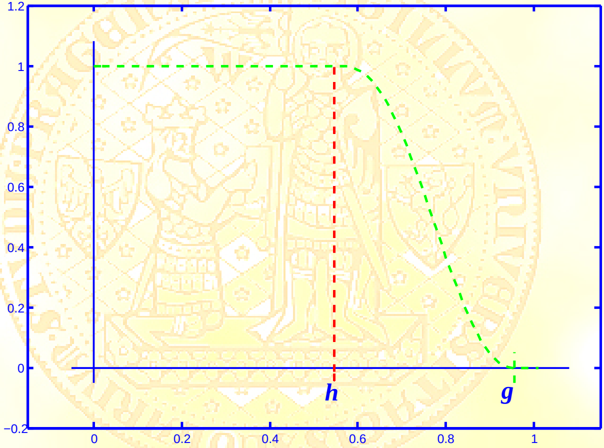
Denote $F_\beta(r) = P(|Y_1 - X_1'\beta| < r)$.

C3 *Identifiability of β^0*

- $\exists$ the only solution of equation

$$\beta' \mathbb{E}\left[w\left(F_\beta(|r_1(\beta)|)\right) Z_1 \left(e_1 - X_1'\beta\right)\right] = 0.$$

General shape of weight function



Instrumental weighted variables - asymptotic theory

Enlarging the notations:

$F_{\beta' ZX' \beta}(v)$ - d. f. of r. v. $\beta' Z_1 X_1' \beta$

$$\gamma_{\lambda,a} = \sup_{\|\beta\|=\lambda} F_{\beta' ZX' \beta}(a)$$

and

$$\tau_\lambda = - \inf_{\|\beta\| \le \lambda} \beta' \mathbb{E}\left[Z_1 X_1' \cdot I\{\beta' Z_1 X_1' \beta < 0\}\right] \beta.$$

C4 *Quality of approximation of X_i by Z_i*

- $\exists \left(a > 0, b \in (0,1) \text{ a } \lambda > 0 \right) \quad a \cdot (b - \gamma_{\lambda,a}) \cdot w(b) > \tau_\lambda$

Instrumental weighted variables - consistency

Theorem

Let **C1, C2, C3** and **C4** hold. Then

$$\hat{\beta}^{(IWV,n,w)} \xrightarrow{p} \beta^0 \quad \textit{for} \;\; T \to \infty.$$

Víšek, J. Á. (2009): Consistency of the instrumental weighted variables.
Annals of the Institute of Statistical Mathematics, (2009) 61, 543 - 578.

Víšek, J. Á. (2011):
Consistency of the instrumental weighted variables under heteroscedasticity.
Kybernetika 47 , 179-206.

Instrumental weighted variables - asymptotic theory

NC1 *Random variables*

- ✓ $\{(X_i', Z_i', \varepsilon_i)'\}_{i=1}^{\infty} \subset R^{2p+1}$ - i.i.d. r.v.'s,
- ✓ $\forall \ (t \in N) \quad Z_i$ and ε_i - independent,
- ✓ D.f. $F_{X,Z}(x, z)$ - absolutely continuous,
- ✓ $I\!\!E \left\{ w \left(F_{\beta^0}(|\varepsilon_1|) \right) Z_1 X_1' \right\}$ and $I\!\!E \left\{ Z_1 Z_1' \right\}$ - pozitive definite,
- ✓ $\exists \ (q > 1) \ : \quad I\!\!E \left\{ \|Z_1\| \cdot \|X_1\| \right\}^q < \infty$,
- density $f_{e|X}(r|X_1 = x)$ is uniformly in x Lipschitz of the first order,
- $|f_e'(r)| < U < \infty$.

Instrumental weighted variables - $\sqrt{n}$-consistency

NC2 *Weight function*

- ✓ $w(\alpha) : [0,1] \rightarrow [0,1], w(0) = 1$,
- ✓ absolutely continuous, non-increasing,
- ✓ $\exists$ derivative $w'(\alpha) > -L, \; L \in R^+$,
- $w'(\alpha)$ is Lipschitz of the first order.

Theorem

Let **NC1, NC2, C3** ana **C4** hold. Then

$$\forall(\epsilon > 0) \;\; \exists(K_\epsilon \in R, n_\epsilon \in N) \;\; \forall(n > n_\epsilon)$$

$$P\left(\left\|\sqrt{n}\left(\hat{\beta}^{(IWV,n,w)} - \beta^0\right)\right\| < K_\epsilon\right) > 1 - \epsilon.$$

Víšek, J. Á. (2010):
Weak $\sqrt{n}$-consistency of the least weighted squares under heteroscedasticity.
Acta Universitatis Carolinae, Mathematica et Physica, 2/51, 71 - 82.

Instrumental weighted variables - asymptotic representation

Denote by $g(r)$ density of r.v. e_1^2.

AC1 Density of error term is that

- $\forall (a \in R^+) \;\; \exists \;\; (\Delta(a) > 0) \qquad \inf_{r \in (0, a+\Delta(a))} g(r) > L_g > 0$
- $\exists (s > 1) : \; E|e_1|^{2s} < \infty$

Víšek, J. Á. (2011): Asymptotic representation of the instrumental weighted variables:
Part I - deriving the formula for the asymptotic representation.
Part II - numerical study.
Bulletin of the Czech Econometric Society 21 (32), 1 - 47 and 48 - 72.

C3, C4 and **AC1** hold. Then

$$\sqrt{n}\left(\hat{\beta}^{(IWV,n,w)} - \beta^0\right) = Q^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w\left(F_{\beta^0}(|e_i|)\right) \cdot Z_i e_i + o_p(1)$$

for $n \to \infty$.

Numerical study

General framework:

1. Data generated by

$$Y_i = \beta_0^0 + \beta_1^0 \cdot X_{i1} + \beta_2^0 \cdot X_{i2} + \beta_3^0 \cdot X_{i3} + \beta_4^0 \cdot X_{i4} + \varepsilon_i$$

2. 1000 sets with 100 observations - referred empirical mean values and empirical mean square errors,

i. e.

$$\mathbb{E}_{emp}\hat{\beta}_j = \frac{1}{1000}\sum_{k=1}^{1000}\hat{\beta}_j^{(k)} \qquad \mathrm{MSE}_{emp}\left(\hat{\beta}_j\right) = \frac{1}{1000}\sum_{k=1}^{1000}\left(\hat{\beta}_j^{(k)} - \beta_j^0\right)^2 .$$

3. 50% autocorrelation of explanatory variables.
4. Instruments - lagged values of explanatory variables.
5. Outliers - randomly selected observations $\rightarrow \quad Y_i = -2 * Y_i$.
6. Leverage points - selected observations on the outskirts $\rightarrow \quad \tilde{X}_i = 10 \cdot X_i \qquad$ and $Y_i = -\tilde{X}_i' \cdot \beta^0 + e_i$.

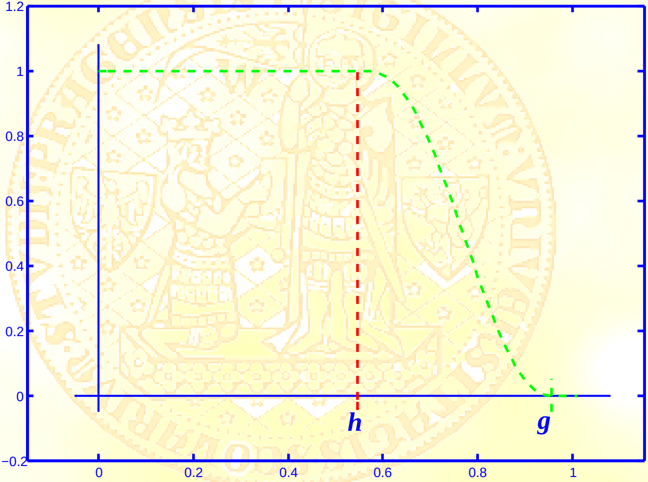
Numerical study

True coeffs β^0	6.3	-0.9	-5.2	6.8	3.1
Heteroscedastic disturbances, independent from explanatory variables					
Outliers (10%) & leverage points (2%)					
$\hat{\beta}^{OLS}_{(\mathrm{MSE}(\hat{\beta}^{OLS}))}$	$-6.79_{(40.7)}$	$1.13_{(54.8)}$	$5.75_{(44.0)}$	$-7.33_{(40.0)}$	$-3.71_{(49.0)}$
$\hat{\beta}^{IV}_{(\mathrm{MSE}(\hat{\beta}^{IV}))}$	$-6.52_{(136)}$	$0.60_{(116)}$	$5.53_{(145)}$	$-7.21_{(124)}$	$-3.61_{(136)}$
$\hat{\beta}^{LWS}_{(\mathrm{MSE}(\hat{\beta}^{LWS}))}$	$6.25_{(0.03)}$	$-0.89_{(0.02)}$	$-5.16_{(0.03)}$	$6.76_{(0.03)}$	$3.08_{(0.02)}$
$\hat{\beta}^{IWV}_{(\mathrm{MSE}(\hat{\beta}^{IWV}))}$	$6.26_{(0.28)}$	$-0.88_{(0.35)}$	$-5.17_{(0.29)}$	$6.73_{(0.55)}$	$3.11_{(1.17)}$

Numerical study

True coeffs β^0	3.5	-1.1	8.4	5.2	9.8
Heteroscedastic disturbances, correlated (50%) with explanatory variables					
Outliers (20%) & leverage points (4%)					
$\hat{\beta}^{OLS}_{(\mathrm{MSE}(\hat{\beta}^{OLS}))}$	$3.06_{(17.7)}$	$-0.78_{(18.0)}$	$7.65_{(18.0)}$	$4.73_{(17.8)}$	$8.89_{(19.7)}$
$\hat{\beta}^{IV}_{(\mathrm{MSE}(\hat{\beta}^{IV}))}$	$3.21_{(93.8)}$	$-0.75_{(78.9)}$	$8.04_{(84.4)}$	$4.58_{(94.7)}$	$8.65_{(85.1)}$
$\hat{\beta}^{LWS}_{(\mathrm{MSE}(\hat{\beta}^{LWS}))}$	$3.99_{(0.04)}$	$-0.59_{(0.04)}$	$8.88_{(0.04)}$	$5.69_{(0.04)}$	$10.3_{(0.04)}$
$\hat{\beta}^{IWV}_{(\mathrm{MSE}(\hat{\beta}^{IWV}))}$	$3.45_{(0.65)}$	$-1.13_{(0.72)}$	$8.33_{(0.60)}$	$5.17_{(0.56)}$	$9.72_{(0.59)}$

Intuitively expected optimal weight function



Looking for minimal (cumulative) bias (over *h* and *g*)

$$E_{emp}\text{bias}\left(\hat{\beta}_j\right) = \frac{1}{1000}\sum_{j=0}^{4}\sum_{k=1}^{1000}\left|\hat{\beta}_j^{(k)} - \beta_j^{(true)}\right|$$

From 0 up to 76, 77, ..., 83 the weights are equal to one,
starting from 79, 80, ..., 87 up to 100 the weights are equal to zero
inbetween linearly decreasing.

$g \downarrow$ $h \rightarrow$	76	77	78	79	80	81	82	83
87	0.1452	0.1467	0.1514	0.1550	0.1648	0.1729	0.1949	0.2178
86	0.1330	0.1361	0.1378	0.1469	0.1496	0.1586	0.1724	0.1966
85	0.1280	0.1281	0.1321	0.1342	0.1348	0.1416	0.1589	0.1777
84	0.1223	0.1263	0.1240	0.1269	0.1297	0.1347	0.1501	0.1655
83	0.1224	0.1223	0.1239	0.1245	0.1269	0.1267	0.1387	-
82	0.1244	0.1225	**0.1208**	0.1222	0.1239	0.1264	-	-
81	0.1297	0.1279	0.1259	0.1265	0.1228	-	-	-
80	0.1310	0.1320	0.1271	0.1288	-	-	-	-
79	0.1330	0.1328	0.1292	-	-	-	-	

Looking for minimal (cumulative) MSE (over h and g)

$$\text{MSE}_{emp}^{(cumm)}\left(\hat{\beta}_j\right) = \frac{1}{1000}\sum_{j=0}^{4}\sum_{k=1}^{1000}\left(\hat{\beta}_j^{(k)} - \beta_j^0\right)^2$$

From 0 up to 76, 77, ..., 83 the weights are equal to one,
starting from 79, 80, ..., 87 up to 100 the weights are equal to zero
inbetween linearly decreasing.

$g \downarrow$ $h \rightarrow$	76	77	78	79	80	81	82	83
87	0.0338	0.0350	0.0373	0.0407	0.0452	0.0512	0.0638	0.0822
86	0.0282	0.0297	0.0305	0.0349	0.0375	0.0411	0.0504	0.0680
85	0.0258	0.0262	0.0274	0.0288	0.0307	0.0350	0.0423	0.0544
84	0.0237	0.0249	0.0258	0.0252	0.0272	0.0298	0.0375	0.0471
83	0.0241	0.0247	0.0248	0.0244	0.0256	0.0265	0.0308	-
82	0.0247	0.0238	**0.0232**	0.0239	0.0242	0.0252	-	-
81	0.0264	0.0257	0.0251	0.0251	0.0237	-	-	-
80	0.0267	0.0272	0.0265	0.0261	-	-	-	-
79	0.0280	0.0280	0.0275	-	-	-	-	-

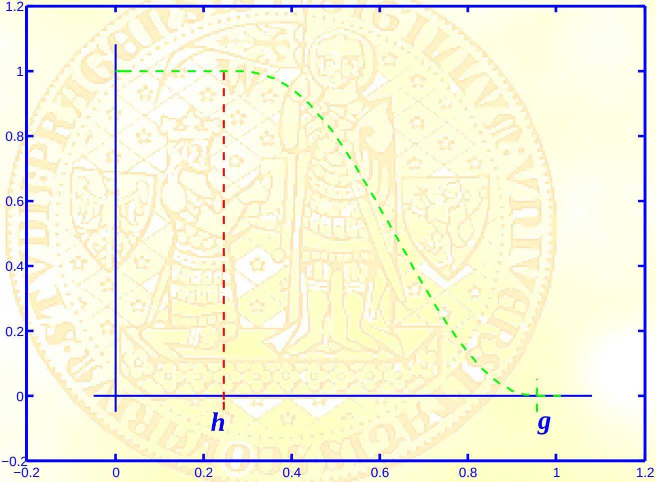
Looking for minimal (cumulative) bias over all *h*'s and over all *g*'s

Outliers (20%) & leverage points (4%)

Minima of cumulative absolute bias for all *h*'s and all *g*'s Corresponding cumulative MSE (on the next line)					
h	24	22	20	27	28
g	85	83	85	85	84
Bias	0.1067	0.1067	0.1070	0.1076	0.1076
MSE	0.0186	0.0188	0.0188	0.0186	0.0188

Minima of cumulative absolute bias for all *h*'s and all *g*'s Corresponding cumulative MSE (on the next line)					
h	23	25	21	21	30
g	83	84	85	84	83
Bias	0.1082	0.1082	0.1082	0.1084	0.1085
MSE	0.0191	0.0192	0.0191	0.0194	0.0190

Really optimal weight function



Conclusion of numerical study

It is worthwhile to use (and optimize) weights,
better than to delete only just some observations !

The differences in efficiency were small
even under a rather high level of contamination !

THANKS FOR ATTENTION