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 12

Content of lecture

1. Disqualifying classical regression analysis by ignoring ...
   - ... broken orthogonality condition
   - ... collinearity
2. Significance of individual explanatory variable
   - ... for LWS-estimation
   - ... simulations of p-values

Ignoring the orthogonality condition - breaking the consistency of $\hat{\beta}^{(OLS,n)}$

We have discussed it in details in the tenth lecture.

We have robustified the *Instrumental Variables* by means of *implicit weighting*, i. e. we have defined

Instrumental Weighted Variables (IWV) as a solutions of the normal equations:

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

But there are still at least two questions:

1. Firstly, can we robustify *Instrumental Variables* by some other way, employing some other idea of robustification of the classical econometric methods?
2. Secondly (and more importantly), how we learn that we should use IWV?

Answering the first question

We can define *Instrumental M-estimators*, say $\hat{\beta}^{(IM,n)}$, as a solution of:

$$\boxed{\sum_{i=1}^{n} Z_i \psi\left(\frac{Y_i - X_i'\beta}{\hat{\sigma}_n}\right) = 0}$$

where Z_i's have to be selected so that $\mathbb{E} Z_1 \psi(e_1 \sigma_{e_1}^{-1}) = 0$
and $|\mathbb{E} Z_{1j}(X_{1j} - \mathbb{E} X_{1j})|$ as large as possible for all $j = 1, 2, ..., p$.

Answering the second question

Let's make some preparation steps for the Hausman test
Under hypothesis of orthogonality:

$$\mathcal{L}_{\mathrm{as}}\left(\sqrt{n}(\hat{\beta}^{(IM,n)} - \hat{\beta}^{(M,n)})\right) = \mathcal{N}(0, C^H)$$

with

$$C^H = \mathbb{E}\psi^2(e_1\sigma_{e_1}^{-1})\left[Q_Z^{-1}\mathbb{E}\left\{Z_1 Z_i'\right\}[Q_Z^{-1}]' - Q^{-1}\mathbb{E}^{-1}\psi'(e_1\sigma_{e_1}^{-1})\right].$$

(Notice that the second term in parentheses is $\hat{\beta}^{(M,n)}$)
- this is a preparatory considerations for *Hausman test*.)

Answering the second question

Let's continue in the preparation steps for the Hausman test:
Under alternative:

$$\mathcal{L}_{\text{as}}(\sqrt{n}\left(\hat{\beta}^{(IM,n)} - \hat{\beta}^{(M,n)} - Q^{-1}\mathbb{E}\left[X_1\psi(e_1\sigma_{e_1}^{-1})\right]\right)) = \mathcal{N}(0, C^A)$$

with

$$\begin{aligned} C^A = Q^{-1}\Big[\mathbb{E}\left\{\psi^2(e_1\sigma_{e_1}^{-1})X_1X_1'\right\} \\ -\mathbb{E}\left\{\psi(e_1\sigma_{e_1}^{-1})X_1\right\}\mathbb{E}\left\{\psi(e_1\sigma_{e_1}^{-1})X_1'\right\}\Big]Q^{-1} \\ -Q^{-1}\mathbb{E}\{X_1Z_1'\psi^2(e_1\sigma_{e_1}^{-1})\}[Q_Z^{-1}]' \\ -Q_Z^{-1}\mathbb{E}\{Z_1X_1'\psi^2(e_1\sigma_{e_1}^{-1})\}[Q^{-1}]' \\ +\mathbb{E}\left\{\psi^2(e_1\sigma_{e_1}^{-1})\right\}Q_Z^{-1}\mathbb{E}\left\{Z_1Z_1'\right\}[Q_Z^{-1}]'. \end{aligned}$$

Under hypothesis we can show that $C^A \longrightarrow C^H$.

How we learn that the orthogonality condition is broken?

The classical econometrics offered the *Hausman test*.

Put

$$q = \sqrt{n}\left(\hat{\beta}^{(IV,n)} - \hat{\beta}^{(LS,n)}\right), \qquad \hat{X} = Z\left(Z'Z\right)^{-1} Z'X$$

$$V = \left(\frac{1}{n}\hat{X}'\hat{X}\right)^{-1} - \left(\frac{1}{n}X'X\right)^{-1} \quad \text{and} \quad \lambda = \mathbb{E}\left\{X_1'e_1\right\} V^{-1} \mathbb{E}\left\{X_1 e_1\right\}.$$

Then

$$\boxed{\mathcal{L}_{\text{as}}\left(\frac{q'V^{-1}q}{s^2}\right) = \chi^2(p,\lambda).}$$

(Notice that under the hypothesis χ^2 is central,
while under the alternative we have some parameter λ of noncentrality.)

Hausman specification test for *M*-estimators

Theorem: Under hypothesis & alternative:

Víšek, J. Á. (1998): Robust specification test.
Proceedings of Prague Stochastics'98 (eds. Hušková, M. & Lachout, P.), Union of Czechoslovak Mathematicians and Physicists, 581 - 586.

Víšek, J. Á. (1998): Robust instrumental variables and specification test.
PRASTAN 2000, Proceedings of conference "Mathematical statistics, numerical mathematics and their application" (eds. Kalina, M., Minárová, M., Nánásiová, O.), 133 - 164.

(I believe that for LWS the same result can be derived - but an exact proof is still to be written.)

Recognizing the collinearity and estimating the model

Condition number

$$\eta = \frac{\max_{1 \le i \le p} \sqrt{\lambda_i}}{\min_{1 \le i \le p} \sqrt{\lambda_i}}$$

Hoerl, A. E., R. W. Kennard (1970):
Ridge regression: Biased estimation for nonorthogonal problems.
Technometrics 12, 55 - 68.

Hoerl, A. E., R. W. Kennard (1970):
Ridge regression: Application to nonorthogonal problems.
Technometrics 12, 69 - 82.

$$\hat{\beta}^{(R,n)} = \left(X'X + \delta \cdot \mathbb{I}\right)^{-1} X'Y,$$

Biased - but $\delta \neq 0$ is not the sufficient explanation ... .

Generally, $(X'X + \delta \cdot \mathbb{I})^{-1} X'Y$ can't be equal to $(X'X)^{-1} X'Y$
because when $X'X$ is regular matrix, also $X'X + \delta \cdot \mathbb{I}$ is regular.

Estimating the model under collinearity

An alternative estimator:

The (ordinary) least squares with constraints

$$\hat{\beta}^{(OLS,n,C)} = \arg\min_{\beta \in R^p} \left\{ \sum_{i=1}^{n} (Y_i - X_i' \beta)^2 ; \quad C \cdot \beta + \kappa = 0 \right\}$$

They can be employed at least in two situations:

1. Depressing the influence of the collinearity

and

2. improving the combinations of forecasts.

A few words about the second possibility

Combining forecasts to reach an improvement

Bates, J. M., C. W. J. Granger (1969): The combination of forecasts.
Operational Research Quarterly, 20, 451-468.

Clemen, R. T. (1986):
Linear constraints and efficiency of combined forecasts.
Journal of Forecasting, 6, 31 - 38.

$$C = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 1 \end{bmatrix}, \quad \kappa = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

i. e.

$$\hat{\beta}_1^{(LS,C,n)} = 0 \qquad \sum_{j=2}^{p} \hat{\beta}_j^{(LS,C,n)} = 1.$$

Combining forecasts to reach an improvement

Hendry, D. F., M. P. Clements (2004): Pooling of forecasts.
Econometrics Journal 7, is 1, 1 - 31.

Araújo, M. B., M. New (2007):
Ensemble forecasting of species distributions.
Trends in Ecology & Evolution, 22, 42-47.

(both having more than 150 references)

Jore, A. S., J. Mitchell, S. P. Vahey (2010):
Combining forecast densities from VARs with uncertain instabilities.
Special Issue of J. of Applied Econometrics 25, 621-634.

Wonga, K. K. F., H. Song, S. F. Witta and D. C. Wua (2007):
Tourism forecasting: To combine or not to combine?
Tourism Management 28, 1068-1078.

Estimating the model under collinearity

We are looking for:

$$\hat{\beta}^{(OLS,n,C)} = \underset{\beta \in R^p}{\arg\min} \left\{ \sum_{i=1}^{n} (Y_i - X_i'\beta)^2 ; \;\; C \cdot \beta + \kappa = 0 \right\}.$$

If κ (can be assumed) random, we can write (for some $\delta \neq 0$)

$$\begin{bmatrix} Y \\ 0 \end{bmatrix} = \begin{bmatrix} X \\ \delta \cdot C \end{bmatrix} \cdot \beta + \begin{bmatrix} \varepsilon \\ \delta \cdot \kappa \end{bmatrix}$$

$$\hat{\beta}^{(OLS,n,C)} = \left\{ \begin{bmatrix} X \\ \delta \cdot C \end{bmatrix}' \cdot \begin{bmatrix} X \\ \delta \cdot C \end{bmatrix} \right\}^{-1} \cdot \begin{bmatrix} X \\ \delta \cdot C \end{bmatrix}' \cdot \begin{bmatrix} Y \\ 0 \end{bmatrix}$$

$$\hat{\beta}^{(OLS,n,C)} = \left[X'X + \delta^2 \cdot C'C \right]^{-1} \cdot X'Y.$$

(To keep a possibility to follow the explanation, a part of slide is rewritten on the next slide.)

Estimating the model under collinearity

We are looking for:

$$\hat{\beta}^{(OLS,n,C)} = \underset{\beta \in R^p}{\arg\min} \left\{ \sum_{i=1}^{n} (Y_i - X_i'\beta)^2 ; \;\; C \cdot \beta + \kappa = 0 \right\} \qquad (1)$$

and it gives

$$\hat{\beta}^{(OLS,n,C)} = \left[X'X + \delta^2 \cdot C'C\right]^{-1} \cdot X'Y. \qquad (2)$$

So it seems that $\hat{\beta}^{(OLS,n,C)}$ is a special case of $\hat{\beta}^{(R,n)}$

$= (X'X + \delta \cdot I\!I)^{-1} X'Y$, or vice versa.

Why is this conclusion wrong ?

Also the derivation of $\hat{\beta}^{(OLS,n,C)}$ needs some attention !!

The answer on the second sentence is straightforward:

Notice that (1) doesn't depend on δ while (2) does.

(As κ is not too much specified, the solution of (1) is also solution for

$$\hat{\beta}^{(OLS,n,C)} = \underset{\beta \in R^p}{\arg\min} \left\{ \sum_{i=1}^{n} (Y_i - X_i'\beta)^2 ; \;\; \delta \cdot C \cdot \beta + \tilde{\kappa} = 0 \right\}$$

for some $\tilde{\kappa}$ and it affects of course $\hat{\beta}^{(OLS,n,C,\delta)}$.

The answer on the question is a bit more complicated:

Estimating the model under collinearity

Consider the constraints $C\beta + \kappa = 0$ (it implies that C has p columns).

What is (or what can be) the number of rows of C

(or, equivalently, the dimension of κ) ?

The rows of C are (linearly) independent,

otherwise we delete some of them.

Denote the number of rows of C, i. e. the $\mathrm{rank}(C)$ by ℓ.

If $\ell = p$, then the p-tuple of equations $C \cdot \beta + \kappa = 0$

identifies uniquely β and hence

the whole extremal problem has solution given by constraints.

It is a nonsence,hence

$\ell < p$.

We have of course also $\dim(\kappa) = \ell$.

Solving the "extremal problem with constraints"

So we have C of type $\ell \times p$. Then there is a matrix $\tilde{C}$ (of type $p - \ell \times p$) so that $\begin{bmatrix} C \\ \tilde{C} \end{bmatrix}$ is regular and $C' \cdot \tilde{C} = 0$.

Let's give an example.

Let

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

The matrix $\tilde{C}$ can be selected as

$$\tilde{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

Solving the "extremal problem with constraints"

Now, let β^* fulfills constraints, i. e. $C\beta^* + \kappa = 0$ and define the mapping $\beta(\lambda) : R^{p-\ell} \rightarrow \{\beta \in R^p : C\beta + \kappa = 0\}$ by

$$\beta(\lambda) = \tilde{C}' \cdot \lambda + \beta^*.$$

- Let's verify that it is mapping into $\{\beta \in R^p : C\beta + \kappa = 0\}$.
  We have for any $\lambda \in R^{p-\ell}$ $\quad C \cdot \beta(\lambda) = \underbrace{C\tilde{C}'}_{0} \cdot \lambda + C\beta^* = \kappa.$
- The mapping is one-to-one.
  Assume that for a pair $\lambda_1, \lambda_2 \in R^{p-\ell}$ $\quad \beta(\lambda_1) = \beta(\lambda_2)$
  which means that $\tilde{C}' \cdot (\lambda_1 - \lambda_2) = 0.$
  As the matrix $\tilde{C}$ is of full rank, we have $\lambda_1 - \lambda_2 = 0.$
- The mapping is on $\{\beta \in R^p : C\beta + \kappa = 0\}$.
  We have for any $\bar{\beta} \in \{\beta \in R^p : C\beta + \kappa = 0\}$ $\quad C\left(\bar{\beta} - \beta^*\right) = 0.$
  Hence $\bar{\beta} - \beta^* \perp C.$
  So, $\exists \lambda \in R^{p-\ell}$ such that $\bar{\beta} - \beta^* = \tilde{C}' \cdot \lambda$ (combination of columns of $\tilde{C}'$),
  i. e. $\bar{\beta} = \tilde{C}' \cdot \lambda + \beta^*.$

An example

Let $\beta^0 = (1, -2,\ 3, -4,\ 5)'$ and consider the constraint:

$$\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4 + \hat{\beta}_5 = 3.$$

Then

$$C = \left[\begin{array}{ccccc} 1 & 1 & 1 & 1 & 1 \end{array}\right] \quad \text{and} \quad \kappa = \left[\begin{array}{c} 3 \end{array}\right].$$

The matrix $\tilde{C}$ and β^* can be selected as

$$\tilde{C} = \left[\begin{array}{ccccc} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{array}\right] \quad \text{and} \quad \beta^* = [\,1 \quad -2 \quad 3 \quad -4 \quad 5\,]'.$$

Solving the “extremal problem with constraints”$_{(continued)}$

Let again β^* fulfills constraints, i. e. $C\beta^* + \kappa = 0$ and put

$$\tilde{Y} = Y - X\beta^*, \quad \tilde{X} = X \cdot \tilde{C}'.$$

Then for any $\lambda \in R^{n-\ell}$ and $\beta(\lambda) = \tilde{C}' \cdot \lambda + \beta^*$ we have:

$$\forall\ i \quad \tilde{Y}_i - \tilde{X}_i'\lambda = Y_i - X_i'\beta^* - X_i' \cdot \tilde{C}'\lambda = Y_i - X_i'\left(\tilde{C}'\lambda + \beta^*\right) = Y_i - X_i'\beta(\lambda).$$

It means that

$$\hat{\lambda}^{(OLS,n)} = \underset{\lambda \in R^{p-\ell}}{\arg\min} \left\{ \sum_{i=1}^{n} \left(\tilde{Y}_i - \tilde{X}_i'\lambda\right)^2 \right\} = \underset{\lambda \in R^{p-\ell}}{\arg\min} \left\{ \sum_{i=1}^{n} \left(Y_i - X_i'\beta(\lambda)\right)^2 \right\}$$

$$= \underset{\bar{\beta} \in \left\{\beta \in R^p : C\beta + \kappa = 0\right\}}{\arg\min} \left\{ \sum_{i=1}^{n} \left(Y_i - X_i'\bar{\beta}\right)^2 \right\}.$$

Robustifying the Ordinary Least squares with constraints

It means that for

$$\hat{\lambda}^{(OLS,n)} = \arg\min_{\lambda \in R^{p-\ell}} \left\{ \sum_{i=1}^{n} \left(\tilde{Y}_i - \tilde{X}_i' \lambda \right)^2 \right\},$$

$\hat{\beta}(\lambda) = \tilde{C}' \cdot \hat{\lambda}^{(OLS,n)} + \beta^*$ solves

$$\hat{\beta}(\lambda) = \arg\min_{\beta \in R^p} \left\{ \sum_{i=1}^{n} (Y_i - X_i'\beta)^2 ; \;\; C \cdot \beta + \kappa = 0 \right\}.$$

Estimating robustly the model under collinearity

Robustifying the least squares with constraints:

The least weighted squares with constraints

$$\hat{\beta}^{(LWS,n,C)} = \arg\min_{\beta \in R^p} \left\{ \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta); \quad C \cdot \beta + \kappa = 0 \right\}$$

They can be calculated as LWS without constraints for data $\left(\tilde{Y}, \tilde{X}\right)$ - that's all.

Numerical study

We have generated 1000 data-sets, as follows

$$\left\{ \left\{ X_i^{(k)},\ e_i^{(k)},\ \varepsilon_i^{(k)} \right\}_{i=1}^{100} \right\}_{k=1}^{1000} \quad \text{and} \quad \left\{ \left\{ \sigma_i^{(k)} \right\}_{i=1}^{100} \right\}_{k=1}^{1000}$$

with $X_i^{(k)} = \left(X_{i1}^{(k)}, X_{i2}^{(k)}, X_{i3}^{(k)} \right)'$'s, $e_i^{(k)}$'s and $\varepsilon_i^{(k)}$'s normally distributed,

and $\sigma_i^{(k)}$'s uniformly distributed over $[0.5, 3.5]$,

$$U_{ij}^{(k)} = X_{ij}^{(k)}, j = 1, 2, 3 \quad \text{and} \quad U_{i,3+\ell}^{(k)} = 0.5 * X_{(i,p-1-\ell)}^{(k)} + 0.5 * X_{(i,p-2-\ell)}^{(k)} + \eta \cdot \varepsilon_{i\ell}^{(k)},$$

for $\ell = 1, 2,\ \ i = 1, 2, ..., 100,\ \ k = 1, 2, ..., 1000$. Then we calculated for

$$\beta^0 = [\ 1,\ \ -2,\ \ 3,\ \ -4,\ \ 5\ ]'$$

$$W_i^{(k)} = \sum_{j=1}^{p} U_{ij}^{(k)} \cdot \beta_j^0 + e_i^{(k)} \cdot \sigma_i^{(k)} \quad \text{and employed data} \quad \left\{ \left\{ W_i^{(k)}, U_i^{(k)} \right\}_{i=1}^{100} \right\}_{k=1}^{1000}.$$

Numerical study (*continued*)

For each dataset we obtained, say

$$\left\{ \hat{\beta}^{(index,k)} = (\hat{\beta}_1^{(index,k)}, \hat{\beta}_2^{(index,k)}, \ldots, \hat{\beta}_5^{(index,k)})' \right\}_{k=1}^{1000}$$

for indeces *OLS, LWS, LTS, OLSC, LWSC* and *LTSC* an we refer

$$\hat{\beta}_j^{(index)} = \frac{1}{1000} \sum_{k=1}^{1000} \hat{\beta}_j^{(index,k)} \quad \text{and} \quad \widehat{\text{MSE}}\left(\hat{\beta}_j^{(index)}\right) = \frac{1}{1000} \sum_{k=1}^{1000} \left(\hat{\beta}_j^{(index,k)} - \beta_j^0\right)^2 .$$

Employing one constraint

We will consider one constraint:

$$\sum_{j=1}^{5} \hat{\beta}_j = 3.$$

Then

$$C = [\ 1 \quad 1 \quad 1 \quad 1 \quad 1\ ] \quad \text{and} \quad \kappa = [\ 3\ ].$$

The matrix $\tilde{C}$ and β^* can be selected as

$$\tilde{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix} \quad \text{and} \quad \beta^* = [\,1 \quad -2 \quad 3 \quad -4 \quad 5\,]'.$$

Notice that the rows of matrix $\tilde{C}$ are not orthogonal each to other.

Numerical study (*continued*)

TABLE 1

1. The disturbances are heteroscedastic ($0.5 \leq \sigma_i^2 \leq 3.5$) and independent from explanatory variables.
2. Data are not contaminated but there is a collinearity, the mean value of condition numbers was equal to 41.25.
3. The level of robustness was fixed: the number of observations h taken into account by LTS was 95% of n, the weight function w had $h = 85\%$ and $g = 95\%$ of n.
4. The collinearity is depressed by one constraint condition (see previous slide).

$\hat{\beta}^{OLS}$ $(\mathrm{MSE}(\hat{\beta}^{OLS}))$	$1.00_{(0.808)}$	$-2.00_{(1.619)}$	$3.01_{(0.855)}$	$-4.01_{(3.413)}$	$5.00_{(3.228)}$
$\hat{\beta}^{LWS}$ $(\mathrm{MSE}(\hat{\beta}^{LWS}))$	$1.02_{(0.851)}$	$-1.94_{(1.745)}$	$3.04_{(0.966)}$	$-4.08_{(3.839)}$	$4.96_{(3.378)}$
$\hat{\beta}^{LTS}$ $(\mathrm{MSE}(\hat{\beta}^{LTS}))$	$1.03_{(0.868)}$	$-1.93_{(1.760)}$	$3.05_{(0.961)}$	$-4.09_{(3.813)}$	$4.94_{(3.455)}$
$\hat{\beta}^{OLSC}$ $(\mathrm{MSE}(\hat{\beta}^{OLSC}))$	$1.00_{(0.797)}$	$-1.99_{(1.601)}$	$3.01_{(0.839)}$	$-4.01_{(3.351)}$	$5.00_{(3.187)}$
$\hat{\beta}^{LWSC}$ $(\mathrm{MSE}(\hat{\beta}^{LWSC}))$	$1.03_{(0.829)}$	$-1.94_{(1.697)}$	$3.04_{(0.912)}$	$-4.07_{(3.633)}$	$4.94_{(3.303)}$
$\hat{\beta}^{LTSC}$ $(\mathrm{MSE}(\hat{\beta}^{LTSC}))$	$1.03_{(0.851)}$	$-1.94_{(1.703)}$	$3.03_{(0.946)}$	$-4.06_{(3.766)}$	$4.94_{(3.387)}$

The other tables for the mean values of the condition numbers 1.16, 9.95, 19.95, 31.01, 32.79, 48.97, 61.22, 69.78 and 98.47 are available on http://samba.fsv.cuni.cz/~visek/Constraints_Heteroscedasticity_Numerical_Study/visek_NumericalStudy.

What about to employing two constraints ?

Now we consider two constraints:

$$\sum_{j=1}^{5} \hat{\beta}_j = 3 \qquad \text{and} \qquad \hat{\beta}_1 + \hat{\beta}_2 = -1.$$

Then

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \kappa = \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

The matrix $\tilde{C}$ and β^* can be selected as

$$\tilde{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix} \quad \text{and} \quad \beta^* = [\,1 \quad -2 \quad 3 \quad -4 \quad 5\,]'.$$

Notice that the rows of matrix C are not orthogonal each to other.
The same is true about the rows of $\tilde{C}$.

Numerical study (*continued*)

TABLE 2

Nearly the same framework as in previous case - except of underlined.

1. The disturbances are heteroscedastic ($0.5 \leq \sigma_i^2 \leq 3.5$) and independent from explanatory variables.
2. Data are not contaminated but there is a collinearity, the mean value of condition numbers was equal to 32.79.
3. The level of robustness was fixed: the number of observations h taken into account by LTS was 95% of n, the weight function w had $h = 85\%$ and $g = 95\%$ of n.
4. The collinearity is depressed by two constraints (see previous slide).

$\hat{\beta}^{OLS}_{(\mathrm{MSE}(\hat{\beta}^{OLS}))}$	$1.00_{(0.559)}$	$-2.01_{(1.179)}$	$2.98_{(0.609)}$	$-3.97_{(3.402)}$	$5.00_{(2.224)}$
$\hat{\beta}^{LWS}_{(\mathrm{MSE}(\hat{\beta}^{LWS}))}$	$1.02_{(0.543)}$	$-2.00_{(1.086)}$	$2.99_{(0.546)}$	$-3.97_{(2.172)}$	$4.96_{(2.154)}$
$\hat{\beta}^{LTS}_{(\mathrm{MSE}(\hat{\beta}^{LTS}))}$	$1.01_{(0.557)}$	$-2.00_{(1.165)}$	$2.99_{(0.594)}$	$-3.98_{(2.360)}$	$4.97_{(2.219)}$
$\hat{\beta}^{OLSC}_{(\mathrm{MSE}(\hat{\beta}^{OLSC}))}$	$1.01_{(0.115)}$	$-2.01_{(0.115)}$	$2.99_{(0.455)}$	$-3.97_{(1.795)}$	$4.98_{(0.449)}$
$\hat{\beta}^{LWSC}_{(\mathrm{MSE}(\hat{\beta}^{LWSC}))}$	$1.01_{(0.108)}$	$-2.01_{(0.108)}$	$2.99_{(0.425)}$	$-3.97_{(1.686)}$	$4.98_{(0.425)}$
$\hat{\beta}^{LTSC}_{(\mathrm{MSE}(\hat{\beta}^{LTSC}))}$	$1.01_{(0.112)}$	$-2.01_{(0.112)}$	$2.99_{(0.444)}$	$-3.97_{(1.760)}$	$4.98_{(0.443)}$

The other tables for the mean values of the condition numbers 1.16, 9.95, 19.95, 31.01, 41.25, 48.97, 61.22, 69.78 and 98.47 are available again on http://samba.fsv.cuni.cz/~visek/Constraints_Heteroscedasticity_Numerical_Study/visek_NumericalStudy.

The results are much better than in Table 1 - compare MSE's.

What about to employ two orthogonal constraints ?

Let $\beta^0 = (1, -2,\ 3, -4,\ 5)'$ and consider two orthogonal constraints:

$$\hat\beta_1 + \hat\beta_2 = -1 \qquad \text{and} \qquad \hat\beta_3 + \hat\beta_4 = -1.$$

Then

$$C = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \kappa = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

The matrix $\tilde{C}$ and β^* can be selected as

$$\tilde{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \beta^* = [\,1 \ \ -2 \ \ 3 \ \ -4 \ \ 5\,]'.$$

How the transformation of variables looks like ? An example

We have

$$\tilde{Y}_i = Y_i - X_i' \beta^* \qquad \text{and} \qquad \tilde{X}_i = \tilde{C} X_i$$

i. e.

$$\tilde{X}_{i1} = X_{i1} - X_{i2}\,, \qquad \tilde{X}_{i2} = X_{i3} - X_{i4} \quad \text{and} \quad \tilde{X}_{i3} = X_{i5}.$$

Finally,

$$\hat{\beta} = \tilde{C}' \hat{\lambda} + \beta*$$

i. e.

$$\hat{\beta}_1 = \hat{\lambda}_1 + 1, \qquad \hat{\beta}_2 = -\hat{\lambda}_1 - 2, \qquad \hat{\beta}_3 = \hat{\lambda}_2 + 3$$

$$\hat{\beta}_4 = -\hat{\lambda}_2 - 4 \qquad \text{and} \qquad \hat{\beta}_5 = \hat{\lambda}_3 + 5.$$

Numerical study(*continued*)

TABLE 3

Nearly the same framework as in previous case - except of underlined.

1. The disturbances are heteroscedastic ($0.5 \leq \sigma_i^2 \leq 3.5$) and independent from explanatory variables.
2. Data are not contaminated but there is a collinearity, the mean value of condition numbers was equal to 32.95.
3. The level of robustness was fixed: the number of observations *h* taken into account by LTS was 95% of *n*, the weight function *w* had $h = 85\%$ and $g = 95\%$ of *n*.
4. We have employed two orthogonal constraints (see previous slide).

$\hat{\beta}^{OLS}_{(\text{MSE}(\hat{\beta}^{OLS}))}$	$1.05_{(0.593)}$	$-1.91_{(1.167)}$	$3.04_{(0.579)}$	$-4.08_{(2.297)}$	$4.91_{(2.359)}$
$\hat{\beta}^{LWS}_{(\text{MSE}(\hat{\beta}^{LWS}))}$	$1.04_{(0.573)}$	$-1.93_{(1.065)}$	$3.03_{(0.525)}$	$-4.06_{(2.076)}$	$4.92_{(2.279)}$
$\hat{\beta}^{LTS}_{(\text{MSE}(\hat{\beta}^{LTS}))}$	$1.04_{(0.615)}$	$-1.91_{(1.174)}$	$3.05_{(0.572)}$	$-4.11_{(2.275)}$	$4.93_{(3.435)}$
$\hat{\beta}^{OLSC}_{(\text{MSE}(\hat{\beta}^{OLSC}))}$	$1.00_{(0.005)}$	$-2.00_{(0.005)}$	$3.00_{(0.021)}$	$-4.00_{(0.021)}$	$5.01_{(0.015)}$
$\hat{\beta}^{LWSC}_{(\text{MSE}(\hat{\beta}^{LWSC}))}$	$1.00_{(0.004)}$	$-2.00_{(0.004)}$	$3.00_{(0.019)}$	$-4.00_{(0.019)}$	$5.01_{(0.014)}$
$\hat{\beta}^{LTSC}_{(\text{MSE}(\hat{\beta}^{LTSC}))}$	$1.00_{(0.004)}$	$-2.00_{(0.004)}$	$3.00_{(0.021)}$	$-4.00_{(0.021)}$	$5.01_{(0.015)}$

The other tables for some other mean values of the condition numbers are available again on http://samba.fsv.cuni.cz/˜visek/Constraints_Heteroscedasticity_Numerical_Study/visek_NumericalStudy.
The results are again even much better than in Table 2 - compare MSE's.

Hence we will keep these constraints in the rest of study.

We are going to make an idea what a contamination can cause.

TABLE 4

Nearly the same framework as in previous case - except of underlined.

1. The disturbances are heteroscedastic ($0.5 \leq \sigma_i^2 \leq 3.5$) and independent from explanatory variables.
2. There is a collinearity, the mean value of condition numbers was equal to <u>16.55</u>.
3. The level of robustness was fixed: the number of observations h taken into account by LTS was <u>90% of n</u>, the weight function w had <u>$h = 65\%$ and $g = 90\%$ of n</u>.
4. We have employed two orthogonal constraints.
5. <u>The 5% contamination by outliers $Y^{(contaminated)} = -2 * Y^{(original)}$</u>.

$\hat{\beta}^{OLS}_{(\mathrm{MSE}(\hat{\beta}^{OLS}))}$	$0.02_{(5.943)}$	$-1.44_{(8.821)}$	$3.41_{(4.669)}$	$-3.45_{(16.254)}$	$4.10_{(19.377)}$
$\hat{\beta}^{LWS}_{(\mathrm{MSE}(\hat{\beta}^{LWS}))}$	$1.00_{(0.160)}$	$-2.02_{(0.314)}$	$2.99_{(0.164)}$	$-3.97_{(0.620)}$	$5.01_{(0.612)}$
$\hat{\beta}^{LTS}_{(\mathrm{MSE}(\hat{\beta}^{LTS}))}$	$0.99_{(0.170)}$	$-2.02_{(0.336)}$	$2.99_{(0.182)}$	$-3.97_{(0.692)}$	$5.02_{(0.652)}$
$\hat{\beta}^{OLSC}_{(\mathrm{MSE}(\hat{\beta}^{OLSC}))}$	$0.25_{(0.958)}$	$-1.25_{(0.958)}$	$2.38_{(1.639)}$	$-3.38_{(1.639)}$	$3.66_{(3.231)}$
$\hat{\beta}^{LWSC}_{(\mathrm{MSE}(\hat{\beta}^{LWSC}))}$	$1.00_{(0.005)}$	$-2.00_{(0.005)}$	$3.00_{(0.023)}$	$-4.00_{(0.023)}$	$5.01_{(0.019)}$
$\hat{\beta}^{LTSC}_{(\mathrm{MSE}(\hat{\beta}^{LTSC}))}$	$1.00_{(0.006)}$	$-2.00_{(0.006)}$	$3.00_{(0.028)}$	$-4.00_{(0.028)}$	$5.01_{(0.023)}$

The other tables for some other mean values of the condition numbers are available again on http://samba.fsv.cuni.cz/˜visek/Constraints_Heteroscedasticity_Numerical_Study/visek_NumericalStudy.

The results are again even much better than in Table 2 - compare MSE's.

We are going to make an idea what a contamination can cause.

TABLE 4

Nearly the same framework as in previous cases - except of underlined.

1. There is a collinearity,
   the mean value of condition numbers was equal to <u>32.60</u>.

$\hat{\beta}^{OLS}$ $(\mathrm{MSE}(\hat{\beta}^{OLS}))$	$0.22_{(16.606)}$	$-1.54_{(31.568)}$	$2.12_{(18.144)}$	$-2.90_{(69.279)}$	$3.72_{(64.889)}$
$\hat{\beta}^{LWS}$ $(\mathrm{MSE}(\hat{\beta}^{LWS}))$	$1.06_{(0.636)}$	$-1.92_{(1.279)}$	$3.01_{(0.622)}$	$-4.03_{(3.485)}$	$4.88_{(2.541)}$
$\hat{\beta}^{LTS}$ $(\mathrm{MSE}(\hat{\beta}^{LTS}))$	$1.04_{(0.686)}$	$-1.95_{(1.286)}$	$3.00_{(0.653)}$	$-3.99_{(2.607)}$	$4.91_{(2.700)}$
$\hat{\beta}^{OLSC}$ $(\mathrm{MSE}(\hat{\beta}^{OLSC}))$	$0.27_{(0.915)}$	$-1.27_{(0.915)}$	$2.34_{(1.848)}$	$-3.34_{(1.848)}$	$3.62_{(3.488)}$
$\hat{\beta}^{LWSC}$ $(\mathrm{MSE}(\hat{\beta}^{LWSC}))$	$1.00_{(0.005)}$	$-2.00_{(0.005)}$	$3.00_{(0.026)}$	$-4.00_{(0.026)}$	$5.00_{(0.021)}$
$\hat{\beta}^{LTSC}$ $(\mathrm{MSE}(\hat{\beta}^{LTSC}))$	$1.00_{(0.006)}$	$-2.00_{(0.006)}$	$3.01_{(0.030)}$	$-4.01_{(0.030)}$	$5.00_{(0.024)}$

The other tables for some other level of contamination are available again on http://samba.fsv.cuni.cz/˜visek/Constraints_Heteroscedasticity_Numerical_Study/visek_NumericalStudy.

What about leverage points.

TABLE 5

Nearly the same framework as in previous cases - except of underlined.

1. There is a collinearity,
 the mean value of condition numbers was equal to <u>16.55</u>.
2. The 5% contamination by leverage points

$$X^{(contaminated)} = 3 * X^{(original)}, \quad Y^{(contaminated)} = -2 * Y^{(original)}.$$

$\hat{\beta}^{OLS}_{(\mathrm{MSE}(\hat{\beta}^{OLS}))}$	$-1.59_{(45.600)}$	$0.68_{(81.603)}$	$0.68_{(45.803)}$	$-1.53_{(155.949)}$	$0.50_{(169.327)}$
$\hat{\beta}^{LWS}_{(\mathrm{MSE}(\hat{\beta}^{LWS}))}$	$0.99_{(0.162)}$	$-2.01_{(0.312)}$	$3.01_{(0.163)}$	$-4.01_{(0.634)}$	$5.01_{(0.622)}$
$\hat{\beta}^{LTS}_{(\mathrm{MSE}(\hat{\beta}^{LTS}))}$	$1.00_{(0.190)}$	$-2.00_{(0.323)}$	$3.01_{(0.201)}$	$-4.01_{(0.745)}$	$4.99_{(0.717)}$
$\hat{\beta}^{OLSC}_{(\mathrm{MSE}(\hat{\beta}^{OLSC}))}$	$-1.86_{(10.474)}$	$0.86_{(10.474)}$	$0.70_{(13.424)}$	$-1.70_{(13.424)}$	$0.52_{(27.070)}$
$\hat{\beta}^{LWSC}_{(\mathrm{MSE}(\hat{\beta}^{LWSC}))}$	$1.00_{(0.005)}$	$-2.00_{(0.005)}$	$3.00_{(0.026)}$	$-4.00_{(0.026)}$	$5.00_{(0.021)}$
$\hat{\beta}^{LTSC}_{(\mathrm{MSE}(\hat{\beta}^{LTSC}))}$	$1.00_{(0.007)}$	$-2.00_{(0.007)}$	$3.01_{(0.037)}$	$-4.01_{(0.037)}$	$4.99_{(0.028)}$

The other tables for some other level of contamination are available again on http://samba.fsv.cuni.cz/~visek/Constraints_Heteroscedasticity_Numerical_Study/visek_NumericalStudy.

What about leverage points.

TABLE 6

The same framework as in previous cases
- only the mean value of condition numbers was equal to 32.60.

$\hat\beta^{OLS}_{(\mathrm{MSE}(\hat\beta^{OLS}))}$	$-1.67_{(145.901)}$	$0.47_{(268.296)}$	$0.98_{(139.653)}$	$-2.16_{(536.245)}$	$1.23_{(570.985)}$
$\hat\beta^{LWS}_{(\mathrm{MSE}(\hat\beta^{LWS}))}$	$0.94_{(0.625)}$	$-2.06_{(1.216)}$	$3.00_{(0.593)}$	$-4.00_{(2.335)}$	$5.12_{(2.471)}$
$\hat\beta^{LTS}_{(\mathrm{MSE}(\hat\beta^{LTS}))}$	$0.94_{(0.686)}$	$-2.07_{(1.260)}$	$2.99_{(0.687)}$	$-3.98_{(2.658)}$	$5.11_{(2.672)}$
$\hat\beta^{OLSC}_{(\mathrm{MSE}(\hat\beta^{OLSC}))}$	$-1.61_{(9.285)}$	$0.61_{(9.285)}$	$0.69_{(13.963)}$	$-1.69_{(13.963)}$	$0.65_{(26.999)}$
$\hat\beta^{LWSC}_{(\mathrm{MSE}(\hat\beta^{LWSC}))}$	$1.00_{(0.005)}$	$-2.00_{(0.005)}$	$3.01_{(0.025)}$	$-4.01_{(0.025)}$	$5.00_{(0.019)}$
$\hat\beta^{LTSC}_{(\mathrm{MSE}(\hat\beta^{LTSC}))}$	$1.00_{(0.009)}$	$-2.00_{(0.009)}$	$3.00_{(0.045)}$	$-4.00_{(0.045)}$	$5.00_{(0.043)}$

The other tables for some other level of contamination are available again on http://samba.fsv.cuni.cz/˜visek/Constraints_Heteroscedasticity_Numerical_Study/visek_NumericalStudy.

Significance of explanatory variable - for the Least Weighted Squares

We are going to give an idea of deriving
the significance of *individual regressor* - two steps:

The first one (which we have already seen in the seventh lecture
- hence only a brief repetition):

The *Least Weighted Squares* $\hat{\beta}^{(LWS,n,W)}$ can be
- at any point of a basic probabily space $(\Omega, \mathcal{A}, P)$ -
written as *Weighted Least Squares* $\hat{\beta}^{(WLS,n,W,\pi)}$.

The second one:

The classical derivation for significance of
individual regressor for OLS
can be generalised for the *classical WLS* $\hat{\beta}^{(WLS,n,W,\pi)}$.

Showing that $\hat{\beta}^{(LWS,n)}$ is $\hat{\beta}^{(LWS,n,\hat{\pi})}$ ...

We have seen in the seventh lecture:

$\underline{\forall \ \ (\omega \in \Omega) \ \ \exists \ \ (\pi = \pi(\omega) = \{\pi_1(\omega), \pi_2(\omega), ..., \pi_n(\omega)\})}$ so that

$$\hat{\beta}^{(LWS,n,W)}(\omega) = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w_i \left(Y_{\pi_i} - X'_{\pi_i}\beta\right)^2 = \hat{\beta}^{(WLS,n,W,\pi)}(\omega)$$

$$= \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} \left(w_i^{\frac{1}{2}} Y_{\pi_i} - w_i^{\frac{1}{2}} X'_{\pi_i}\beta\right)^2.$$

Notice the dependence of π on ω.

Let's recall how we have found it.

Recalling several facts

Let $\mathcal{P}$ be the set of all permutations of integers $\{1, 2, ..., n\}$.

For any $\pi \in \mathcal{P}, \pi = \{\pi_1, \pi_2, ..., \pi_n\}$ let

$$Y_\pi = (Y_{\pi_1}, Y_{\pi_2}, ..., Y_{\pi_n})', \ X_\pi = (X_{\pi_1}, X_{\pi_2}, ..., X_{\pi_n})' \text{ and } \varepsilon_\pi = (\varepsilon_{\pi_1}, \varepsilon_{\pi_2}, ..., \varepsilon_{\pi_n})'.$$

Put

$$\hat{\beta}^{(WLS,n,\pi)} = (X_\pi' W X_\pi)^{-1} X_\pi' W Y_\pi \text{ and } S_\pi^2 = \sum_{j=1}^{n} w_i \left(Y_{\pi_j} - X_{\pi_j}' \hat{\beta}^{(WLS,n,\pi)} \right)^2.$$

Deriving existence of $\hat{\beta}^{(LWS,n)}$

Then for any $\pi \in \mathcal{P}$

$$S_\pi^2 = \sum_{j=1}^{n} w_j \left(Y_{\pi_j} - X'_{\pi_j}\hat{\beta}^{(WLS,n,\pi)}\right)^2 \leq \min_{\beta \in R^p} \sum_{j=1}^{n} w_j \left(Y_{\pi_j} - X'_{\pi_j}\beta\right)^2 . \qquad (3)$$

Let $(\Omega, \mathcal{A}, P)$ be the basic *Probability Space*

and denote for any r. v. Z by $Z(\omega)$ its value at point $\omega \in \Omega$.

Finally, fix ω_0 and put

$$\hat{\pi}(\omega_0) = \arg\min_{\pi \in \mathcal{P}} S_\pi^2(\omega_0).$$

Then - due to (3)

$$S^2_{\hat{\pi}(\omega_0)}(\omega_0) \leq \min_{\pi \in \mathcal{P}} \min_{\beta \in R^p} \sum_{j=1}^{n} w_j \left(Y_{\pi_j} - X'_{\pi_j}\beta\right)^2 = \min_{\beta \in R^p} \min_{\pi \in \mathcal{P}} \sum_{j=1}^{n} w_j \left(Y_{\pi_j} - X'_{\pi_j}\beta\right)^2 .$$

Deriving existence and form of $\hat{\beta}^{(LWS,n)}$

It means

$$\hat{\beta}^{(WLS,n,\hat{\pi}(\omega_0))}(\omega_0) = \hat{\beta}^{(WLS,n)}\left(Y_{\pi(\omega_0)}(\omega_0), X_{\pi(\omega_0)}(\omega_0)\right) = \hat{\beta}^{(LWS,n,w)}(\omega_0)$$

Repeating it for all $\omega \in \Omega$, we prove the existence of $\hat{\beta}^{(LWS,n,w)}$.

Fix $\pi \in \mathcal{P}$ and put

$$B(\pi) = \{\omega \in \Omega \ : \ \pi = \hat{\pi}(\omega)\}.$$

Then

$$\omega \in B(\pi) \ \Rightarrow \ \hat{\beta}^{(LWS,n,w)}(\omega) = \hat{\beta}^{(WLS,n,\pi)}(\omega)$$

i. e.

$$\hat{\beta}^{(LWS,n,w)}(\omega) = \hat{\beta}^{(WLS,n,\pi)}(\omega) = \arg\min_{\beta \in R^p} \sum_{j=1}^{n} w\left(\frac{j-1}{n}\right)\left(Y_{\pi_j} - X'_{\pi_j}\beta\right)^2.$$

Due to i.i.d. framework, $\forall \pi \in \mathcal{P}$

$$P(B(\pi)) = (n!)^{-1}.$$

Deriving form of $\hat{\beta}^{(LWS,n)}$

Let's repeat:
We have shown that fixing $\pi \in \mathcal{P}$ and putting $B(\pi) = \{\omega \in \Omega : \pi = \pi(\omega)\}$,
we have

$$P(B(\pi)) = (n!)^{-1}$$

and one can easy verify that

$$\cup_{\pi \in \mathcal{P}} B(\pi) = \Omega.$$

All further considerations can be done conditionally,
on π (or on $B(\pi)$, if You want) for $\hat{\beta}^{(WLS,n,\pi)}$.

Then we take mean value over all conditions, i. e. over all $\pi \in \mathcal{P}$
but the situation for all π is the same, with the same probabilities,
hence an unconditional result is the same as conditional.

Deriving existence and form of $\hat{\beta}^{(LWS,n)}$

Let's rewrite the line (the third from bottom) from the last but one slide, for $\omega \in B(\pi)$

$$\hat{\beta}^{(LWS,n,w)}(\omega) = \hat{\beta}^{(WLS,n,\pi)}(\omega) = \underset{\beta \in R^p}{\arg\min} \sum_{j=1}^{n} w\left(\frac{j-1}{n}\right)\left(Y_{\pi_j} - X'_{\pi_j}\beta\right)^2. \quad (4)$$

Putting $\tilde{W} = diag\left\{w^{-\frac{1}{2}}(0), w^{-\frac{1}{2}}\left(\frac{1}{n}\right), ..., w^{-\frac{1}{2}}\left(\frac{n-1}{n}\right)\right\}$

and $\tilde{Y} = \tilde{W}Y$, $\tilde{X} = \tilde{W}X$, then (4) reads

$$\hat{\beta}^{(LWS,n,w)}(\omega) = \underset{\beta \in R^p}{\arg\min} \left\{\left(\tilde{Y} - \tilde{X}\beta\right)'\left(\tilde{Y} - \tilde{X}\beta\right)\right\}$$
$$= \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} = \hat{\beta}^{(OLS,n)}(\tilde{Y}, \tilde{X}).$$

Consider, for a while, the model

$$\tilde{Y} = \tilde{X}\beta^0 + \tilde{\varepsilon} \quad \text{with} \quad \mathcal{L}(\tilde{\varepsilon}) = \mathcal{N}\left(0, \sigma^2\tilde{W}^2\right)$$ (notice the heteroscedascity of $\tilde{\varepsilon}$).

Significance of explanatory variable - classical OLS case

Let's recall the simplest classical framework for finite-sample diagnostics:

Regression model

$$Y_i = X_i'\beta^0 + \varepsilon_i, \quad i = 1, 2, ..., n \quad \text{or} \quad Y = X\beta^0 + \varepsilon$$

Conditions :

$$\{(X_i', \varepsilon_i)'\}_{i=1}^{\infty} \text{ i.i.d.}, \; F_{X,\varepsilon}(x, v) = F_X(x) \cdot F_\varepsilon(v), \; F_\varepsilon(v) = \mathcal{N}(0, \sigma^2),$$
$$Q = \boldsymbol{E}[X_1 \cdot X_1'] \text{ is regular.}$$

Significance of ℓ-th explanatory variable $X_{i\ell}$ $\Leftrightarrow \neg H_0 : \; \hat{\beta}_\ell^{(OLS,n)} = 0$

Denote $c_{\ell,\ell}^2 = \left[(X'X)^{-1}\right]_{\ell,\ell}$ and $s_n^2 = \frac{1}{n-p}\sum_{i=1}^n \left(Y_i - X_i'\hat{\beta}^{(OLS,n)}\right)^2$.

Test is based on

$$\mathcal{L}\left(\frac{\hat{\beta}_\ell^{(OLS,n)} - \beta_\ell^0}{s_n \cdot c_{\ell,\ell}}\right) = \mathcal{L}(t_\ell) = t_{n-p}$$

(Fisher-Cochran theorem)

Recalling the derivation of t-statistics in the classical regression

First of all, let's recall that in any regression model
and hence also in our model for $\left(\tilde{Y},\tilde{X}\right)$

$$\tilde{r}\left(\hat{\beta}^{(OLS,n)}\left(\tilde{Y},\tilde{X}\right)\right)=\tilde{Y}-\tilde{X}\hat{\beta}^{(OLS,n)}\left(\tilde{Y},\tilde{X}\right)\perp\mathcal{M}\left(\tilde{X}\right)$$

($\mathcal{M}\left(\tilde{X}\right)$ is the set of all linear combinations of the columns of $\tilde{X}$).

Secondly,

$$\hat{\tilde{Y}}=\tilde{X}\hat{\beta}^{(OLS,n)}=\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y}$$

$$=\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\left(\tilde{X}\beta^0+\tilde{\varepsilon}\right)=\tilde{X}\beta^0+\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{\varepsilon},$$

i. e.

$$\mathcal{L}\left(\hat{\tilde{Y}}\right)=\mathcal{N}\left(\tilde{X}\beta^0,\sigma^2\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{W}^2\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right).$$

Recalling the derivation of t-statistics in the classical regression

As $\widehat{\tilde{Y}} = \tilde{X}\hat{\beta}^{(OLS,n)}$, i. e. $\widehat{\tilde{Y}}$ is a linear combination of the columns of $\tilde{X}$

$$\Rightarrow \quad \widehat{\tilde{Y}} \perp \tilde{r}\left(\hat{\beta}^{(OLS,n)}\left(\tilde{Y},\tilde{X}\right)\right) = \tilde{Y} - \tilde{X}\hat{\beta}^{(OLS,n)}\left(\tilde{Y},\tilde{X}\right)$$

and due to the normality of $\widehat{\tilde{Y}}$, it is independent with $r\left(\hat{\beta}^{(OLS,n)}\left(\tilde{Y},\tilde{X}\right)\right)$.

By standard way

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

$$= \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\left[\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}\tilde{Y}\right] = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\widehat{\tilde{Y}} \tag{5}$$

Then due to (5) $\hat{\beta}^{(OLS,n)}\left(\tilde{Y},\tilde{X}\right)$ is independent with $\tilde{r}\left(\hat{\beta}^{(OLS,n)}\left(\tilde{Y},\tilde{X}\right)\right)$

(we shall need it in a few minutes).

Recalling the derivation of t-statistics in the classical regression

Let's recall

$$\hat{\beta}^{(OLS,n)}\left(\tilde{Y},\tilde{X}\right)-\beta^0=\hat{\beta}^{(WLS,n,W,\pi)}-\beta^0=\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{\varepsilon},$$

i. e. due to normality of disturbances, also $\hat{\beta}^{(WLS,n,W,\pi)}-\beta^0$
is normally distributed (with heteroscadasticity).

Let's recall (as it is well-known from the classical regression)

$$\tilde{r}\left(\hat{\beta}^{(OLS,n)}\left(\tilde{Y},\tilde{X}\right)\right)=\tilde{r}\left(\hat{\beta}^{(WLS,n,W,\pi)}\right)=\left(I-\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right)\tilde{\varepsilon},$$

i. e. due to normality of disturbances, $\tilde{r}\left(\hat{\beta}^{(WLS,n,W,\pi)}\right)$
is normally distributed (with heteroscadasticity).

Recalling the derivation of t-statistics in the classical regression

Recalling that $\tilde{W} = diag\left\{w_1^{\frac{1}{2}}, w_2^{\frac{1}{2}}, \ldots, w_n^{\frac{1}{2}}\right\}$ and

$$\mathcal{L}\left(\tilde{\varepsilon}\right) = \mathcal{N}\left(0, \sigma^2 \tilde{W}^2\right),$$

we have from

$$\hat{\beta}^{(OLS,n)}\left(\tilde{Y}, \tilde{X}\right) - \beta^0 = \hat{\beta}^{(WLS,n,W,\pi)} - \beta^0 = \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{\varepsilon},$$

$$\mathbb{E}\left\{\hat{\beta}^{(OLS,n)}\left(\tilde{Y}, \tilde{X}\right) - \beta^0\right\} = \mathbb{E}\left\{\hat{\beta}^{(WLS,n,W,\pi)} - \beta^0\right\} = 0$$

and

$$\operatorname{cov}\left\{\hat{\beta}^{(OLS,n)}\left(\tilde{Y}, \tilde{X}\right) - \beta^0\right\} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{W}^2\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}.$$

$$= \left(X'\tilde{W}^2X\right)^{-1} \cdot \sum_{i=1}^{n} w_i^2 \cdot X_i \cdot X_i' \left(X'\tilde{W}^2X\right)^{-1}.$$

Recalling the derivation of t-statistics in the classical regression

Denote

$$\left[\mathrm{cov}\left\{\hat{\beta}^{(OLS,n)}\left(\tilde{Y},\tilde{X}\right)-\beta^0\right\}\right]_{\ell\ell}=\left[\mathrm{cov}\left\{\hat{\beta}^{(WLS,n,W,\pi)}\left(Y,X\right)-\beta^0\right\}\right]_{\ell\ell}$$

$$=\left[\left(X'\tilde{W}^2X\right)^{-1}\cdot\sum_{i=1}^{n}w_i^2\cdot X_i\cdot X_i'\left(X'\tilde{W}^2X\right)^{-1}\right]_{\ell\ell}\underbrace{=}_{(denote)}d_{n,\ell}(W,X).$$

Then

$$\mathcal{L}\left(\frac{\hat{\beta}_\ell^{(OLS,n)}\left(\tilde{Y},\tilde{X}\right)-\beta^0}{\sigma d_{n,\ell}(W,X)}\right)=\mathcal{L}\left(\frac{\hat{\beta}_\ell^{(LWS,n,W)}-\beta^0}{\sigma d_{n,\ell}(W,X)}\right)=\mathcal{N}(0,1).$$

Establishing the result

We can show (similarly as in the OLS-regression),

$$\mathcal{L}\left(\sigma^{-2}\tilde{r}'\left(\hat{\beta}^{(WLS,n,W)}\right)\cdot\tilde{r}\left(\hat{\beta}^{(WLS,n,W)}\right)\right)=\mathcal{L}\left(\sigma^{-2}\cdot RSS\right)=\chi^2_{generalized}\left(n-p\right)$$

in the sense that $\chi^2_{generalized}\left(n-p\right)$ is distribution of the sum of squares of $n-p$ independent r. v.'s normally distributed with zero mean but variance not equal one, but w_i.

Let us recall that for the "classical $\chi^2(n-p)$ we have (written symbolically)

$$\mathbb{E}\chi^2(n-p)=n-p$$

and hence in the denominatot of t-statistics we put

$$\left\{\frac{\tilde{r}'\left(\hat{\beta}^{(WLS,n,W)}\right)\cdot\tilde{r}\left(\hat{\beta}^{(WLS,n,W)}\right)}{(n-p)\cdot\sigma^2}\right\}^{\frac{1}{2}}.$$

So, to conclude the derivation we need to calculate $\mathbb{E}\chi^2_{generalized}\left(n-p\right)$.

Establishing the result

Let us recall that $W^{\frac{1}{2}} = \tilde{W} = \mathrm{diag}\left\{w_1^{\frac{1}{2}}, w_2^{\frac{1}{2}}, \ldots, w_n^{\frac{1}{2}}\right\}$ and $\tilde{\varepsilon} = \tilde{W}\varepsilon$

and that (as it is well-known from the classical regression)

$$\tilde{r}\left(\hat{\beta}^{(WLS,n)}\right) = \underbrace{\left(I - \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right)}_{\textit{denote it by } \tilde{M}}\tilde{\varepsilon} = \tilde{M}\tilde{W}\varepsilon.$$

$$\mathbb{E}\left[\tilde{r}'\left(\hat{\beta}^{(WLS,n)}\right) \cdot \tilde{r}\left(\hat{\beta}^{(WLS,n)}\right)\right] = \sigma^2 \cdot tr\left[W\left(I - \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right)\right]$$

$$= \sigma^2 \cdot \left(\sum_{i=1}^{n} w_i(1 - d_{ii})\right)$$

where $d_{ii} = \left[\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right]_{ii}$. So, $\frac{\tilde{r}'\left(\hat{\beta}^{(WLS,n)}\right)\cdot\tilde{r}\left(\hat{\beta}^{(WLS,n)}\right)}{\sigma^2\sum_{i=1}^{n} w_i(1-d_{ii})} \to 1.$

Recalling the classical regression for $\tilde{Y}, \tilde{X}$

We conclude

THEOREM

$$\mathcal{L}\left(\frac{\hat{\beta}_\ell^{(LWS,n,W)} - \beta^0}{d_{n,\ell}(W,X)} \cdot \left[\frac{\sum_{i=1}^n w_i(1-d_{ii})}{RSS} \right]^{\frac{1}{2}} \right) = t_{generalized}\,(n-p)$$

in the sense that $t_{generalized}\,(n-p)$ is a ratio of standard normal r. v. and $\chi^2_{generalized}\,(n-p)$ r. v. and numerator and denominator are independent

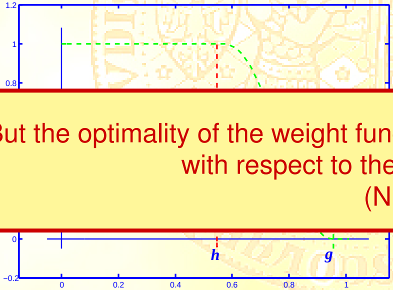
(in the appendix which follows after the end of lecture, an alternative derivation with more details is given, see also ICORS2011).

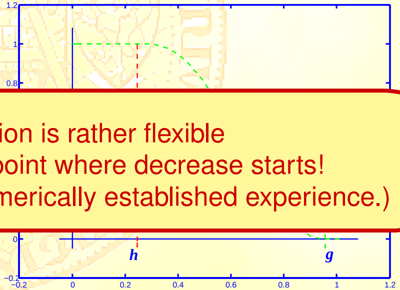
Function generating the weights

Let me recall that we already know
that under low contamination, the intuitively optimal (left)
and really optimal (right) weight functions are
(in the *sense of mean square error* of the estimates of regression coefficients).

$w_\ell = w\left(\frac{\ell-1}{n}\right)$

Contamination : 4% outliers





But the optimality of the weight function is rather flexible
with respect to the point where decrease starts!
(Numerically established experience.)

Framework of simulations

1. For each value of $n = 20, 30, ..., 190$ we generated 5000 times

$$\frac{\hat{\beta}_\ell^{(LWS,n,W)} - \beta^0}{d_{n,\ell}(W,X)} \cdot \left[\frac{\sum_{i=1}^n w_i(1-d_{ii})}{RSS}\right]^{\frac{1}{2}}.$$

2. The 4875 and 4975 order statistics among these 5000 values were found.

3. We repeated it 100 times and have found empirical means and the roots of mean square errors over these 100 repetitions (these roots of mean square errors are in parentheses).

TABLE 1

The simulated quantiles for 5%.

n	20	30	40	50	60	70
$\hat{t}^{LWS}_{0.975}(n)_{(\mathrm{var})}$	2.148 $_{(0.047)}$	2.087 $_{(0.040)}$	2.056 $_{(0.046)}$	2.027 $_{(0.045)}$	2.017 $_{(0.046)}$	2.012 $_{(0.045)}$
$t_{0.975}(n)$	2.085	2.043	2.022	2.009	2.000	1.995

n	80	90	100	110	120	130
$\hat{t}^{LWS}_{0.975}(n)$	2.008 $_{(0.040)}$	1.999 $_{(0.041)}$	1.992 $_{(0.040)}$	1.991 $_{(0.041)}$	1.990 $_{(0.040)}$	1.988 $_{(0.040)}$
$t_{0.975}(n)$	1.990	1.987	1.984	1.982	1.980	1.978

n	140	150	160	170	180	190
$\hat{t}^{LWS}_{0.975}(n)$	1.986 $_{(0.043)}$	1.989 $_{(0.041)}$	1.975 $_{(0.035)}$	1.974 $_{(0.035)}$	1.973 $_{(0.035)}$	1.973 $_{(0.035)}$
$t_{0.975}(n)$	1.977	1.976	1.975	1.974	1.974	1.973

By the way, the 0.975-upper quantile of the standard normal distribution is equal to 1.959964, i. e. $\Phi(1.959964) = 0.975$.

TABLE 2

The simulated quantiles for 1%.

n	20	30	40	50	60	70
$\hat{t}^{LWS}_{0.995}(n)$	2.999 $_{(0.100)}$	2.825 $_{(0.082)}$	2.766 $_{(0.080)}$	2.702 $_{(0.085)}$	2.688 $_{(0.077)}$	2.678 $_{(0.079)}$
$t_{0.995}(n)$	2.845	2.748	2.705	2.678	2.661	2.651

n	80	90	100	110	120	130
$\hat{t}^{LWS}_{0.995}(n)$	2.659 $_{(0.067)}$	2.644 $_{(0.075)}$	2.633 $_{(0.077)}$	2.627 $_{(0.063)}$	2.629 $_{(0.070)}$	2.626 $_{(0.071)}$
$t_{0.995}(n)$	2.640	2.632	2.625	2.619	2.614	2.612

n	140	150	160	170	180	190
$\hat{t}^{LWS}_{0.995}(n)$	2.619 $_{(0.072)}$	2.621 $_{(0.073)}$	2.609 $_{(0.079)}$	2.609 $_{(0.070)}$	2.620 $_{(0.078)}$	2.602 $_{(0.078)}$
$t_{0.995}(n)$	2.611	2.610	2.609	2.608	2.606	2.605

Again, the 0.995-upper quantile of the standard normal distribution is equal to 2.575, i. e. $\Phi(2.575) = 0.995$.

THANKS FOR ATTENTION

An appendix

Significance of explanatory variable

Let's recall the simplest classical framework for finite-sample diagnostics:

Regression model

$$Y_i = X_i'\beta^0 + \varepsilon_i, \quad i = 1, 2, ..., n \qquad \text{or} \qquad Y = X\beta^0 + \varepsilon$$

Conditions :

$$\{(X_i', \varepsilon_i)'\}_{i=1}^{\infty} \text{ i.i.d., } F_{X,\varepsilon}(x, v) = F_X(x) \cdot F_\varepsilon(v), \ F_\varepsilon(v) = \mathcal{N}(0, \sigma^2),$$
$$Q = \boldsymbol{E}[X_1 \cdot X_1'] \text{ is regular,}$$

Significance of ℓ-th explanatory variable $X_{i\ell}$ $\Leftrightarrow$ $H_0: \ \hat{\beta}_\ell^{(LS,n)} = 0$

Denote $c_{\ell,\ell}^2 = \left[(X'X)^{-1}\right]_{\ell,\ell}$ and $s_n^2 = \frac{1}{n-p}\sum_{i=1}^n \left(Y_i - X_i'\hat{\beta}^{(LS,n)}\right)^2.$

Test is based on

$$\mathcal{L}\left(\frac{\hat{\beta}_\ell^{(LS,n)} - \beta_\ell^0}{s_n \cdot c_{\ell,\ell}}\right) = t_{n-p}$$

(Fisher-Cochran theorem)

Significance of explanatory variable

Let's recall the classical "treatment" under heteroscedaticity:(*we shall need it*)

Conditions :

$$\{(X_i', \varepsilon_i)'\}_{i=1}^{\infty} \text{ i.d.}, \; F_{X,\varepsilon_i}(x,v) = F_X(x) \cdot F_{\varepsilon_i}(v), \underline{F_{\varepsilon_i}(v) = \mathcal{N}\left(0, \sigma_i^2\right)},$$

$$\mathbf{QE}\left[X_1 \cdot X_1'\right] \text{ is regular,}$$

Significance of ℓ-th explanatory variable $X_{i\ell}$ $\Leftrightarrow$ $H_0 : \; \hat{\beta}_\ell^{(LS,n)} = 0$

Denote

$$d_{\ell,\ell}^2 = \left[\left(X'X\right)^{-1} \sum_{i=1}^{n} r_i^2 \left(\hat{\beta}^{(LS,n)}\right) \cdot X_i \cdot X_i' \left(X'X\right)^{-1} \right]_{\ell,\ell} .$$

(Halbert White estimator - 1980)

Test is based on

$$\mathcal{L}\left(\frac{\hat{\beta}_\ell^{(LS,n)} - \beta_\ell^0}{d_{\ell,\ell}} \right) \approx t_{n-p} \quad (?!?!)$$

$\left(\mathcal{L}\left((n-p) \cdot d_{\ell,\ell}^2\right) \approx \chi_{n-p}^2 \;\; ?!?!\right)$

Recalling several facts

Let's $W = diag\{w_1, w_2, ..., w_n\}$ be a weight matrix.

Then the classical *Weighted Least Squares* is given as

$$\hat{\beta}^{(WLS,n)} = (X'WX)^{-1} X'WY.$$

Let $\mathcal{P}$ be the set of all permutations of integers $\{1, 2, ..., n\}$.

Fix $\pi \in \mathcal{P}, \pi = \{i_1, i_2, ..., i_n\}$ and put

$$Y_\pi = (Y_{i_1}, Y_{i_2}, ..., Y_{i_n})', \quad X_\pi = (X_{i_1}, X_{i_2}, ..., X_{i_n})' \quad (\text{and} \quad \varepsilon_\pi = (\varepsilon_{i_1}, \varepsilon_{i_2}, ..., \varepsilon_{i_n})')$$

and consider model

$$Y_\pi = X_\pi \beta^0 + \varepsilon_\pi.$$

Then denote

$$\hat{\beta}^{(WLS,n,\pi)} = (X_\pi' W X_\pi)^{-1} X_\pi' W Y_\pi \ \text{ and } \ S_\pi^2 = \sum_{j=1}^{n} w_i \left(Y_{i_j} - X_{i_j}' \hat{\beta}^{(WLS,n,\pi)} \right)^2.$$

Deriving existence of $\hat{\beta}^{(LWS,n)}$

Then for any $\pi \in \mathcal{P}$

$$S_\pi^2 = \sum_{j=1}^{n} w_i \left(Y_{i_j} - X_{i_j}' \hat{\beta}^{(WLS,n,\pi)}\right)^2 \leq \min_{\beta \in R^p} \sum_{j=1}^{n} w_i \left(Y_{i_j} - X_{i_j}'\beta\right)^2 . \qquad (6)$$

Let $(\Omega, \mathcal{A}, P)$ be the basic *Probability Space*

and write $Z(\omega)$ for the value of r.v. at point $\omega \in \Omega$.

Finally, assume $w_1 \geq w_2 \geq ... \geq w_n$, fix ω_0 and put

$$\hat{\pi}(\omega_0) = \arg\min_{\pi \in \mathcal{P}} S_\pi^2(\omega_0).$$

Then - due to (6)

$$S^2_{\hat{\pi}(\omega_0)}(\omega_0) \leq \min_{\beta \in R^p} \min_{\pi \in \mathcal{P}} \sum_{j=1}^{n} w_i \left(Y_{i_j} - X_{i_j}'\beta\right)^2 .$$

Deriving form of $\hat{\beta}^{(LWS,n)}$

It means

$$\hat{\beta}^{(WLS,n,\hat{\pi}(\omega_0))}(\omega_0) = \hat{\beta}^{(LWS,n,w)}(\omega_0)$$

Repeat it for all $\omega \in \Omega$.

Fix $\pi \in \mathcal{P}$ and put

$$B(\pi) = \{\omega \in \Omega \;:\; \pi = \hat{\pi}(\omega)\}.$$

Then

$$\omega \in B(\pi) \;\Rightarrow\; \hat{\beta}^{(LWS,n,w)}(\omega) = \hat{\beta}^{(WLS,n,\pi)}(\omega)$$

i. e.

$$\hat{\beta}^{(LWS,n,w)}(\omega) = \hat{\beta}^{(WLS,n,\pi)}(\omega) = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w_i \left(Y_{\pi_1} - X'_{\pi}\beta\right)^2.$$

Due to i.i.d. framework, $\forall \pi \in \mathcal{P}$

$$P(B(\pi)) = (n!)^{-1}$$

and <u>all B's are the "same"</u>.

Deriving form of $\hat{\beta}^{(LWS,n)}$

Let's rewrite one line of previuos slide, for $\omega_0 \in B(\pi)$

$$\hat{\beta}^{(LWS,n,w)}(\omega_0) = \hat{\beta}^{(WLS,n,\pi)}(\omega_0) = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w_i \left(Y_\pi - X'_\pi \beta \right)^2 . \qquad (7)$$

and drop for a while ω_0 and π.

Then putting $\tilde{W} = diag\left\{ w_1^{-\frac{1}{2}}, w_2^{-\frac{1}{2}}, ..., w_n^{-\frac{1}{2}} \right\}$ and $\tilde{Y} = \tilde{W}Y$, $\tilde{X} = \tilde{W}X$, (7) implies

$$\hat{\beta}^{(LWS,n,w)}(Y,X) = \left(\tilde{X}'\tilde{X} \right)^{-1} \tilde{X}'\tilde{Y} = \hat{\beta}^{(OLS,n)}(\tilde{Y},\tilde{X}).$$

Consider now the model

$$\tilde{Y} = \tilde{X}\beta^0 + \tilde{\varepsilon} \quad \text{with} \quad \mathcal{L}(\tilde{\varepsilon}) = \mathcal{N}\left(0, \sigma^2 \tilde{W}^2 \right)$$

($\tilde{W}^2$ is known $\Rightarrow$ heteroscedasticity is governed by one unknown parameter σ^2)

Recalling the classical regression for $\tilde{Y}, \tilde{X}$

By standard way (please read only last but one line)

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

$$= \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\left[\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}\tilde{Y}\right] = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\hat{\tilde{Y}} \qquad (8)$$

As $\hat{\tilde{Y}} = \tilde{X}\hat{\beta}^{(OLS,n)}$, i. e. $\hat{\tilde{Y}}$ is a linear combimation of the columns of $\tilde{X}$

$$\Rightarrow \quad \hat{\tilde{Y}} \perp \tilde{r}\left(\hat{\beta}^{(WLS,n)}\right) = \tilde{Y} - \tilde{X}\hat{\beta}^{(OLS,n)}$$

and due to the normality of disturbances $\hat{\tilde{Y}}$ is independent with $\tilde{r}$.

Then due to (8) $\hat{\beta}^{(LWS,n,w)}$ is independent with $\tilde{r}\left(\hat{\beta}^{(LWS,n,w)}\right)$

(we shall need it in a few minutes).

Recalling the classical regression for $\tilde{Y}, \tilde{X}$

Please read again only the last line

Regress ℓ-th column of matrix $\tilde{X}$ on all other columns of this matrix
and denote the residuals by $u^{(\ell)}$.

Recall that we consider $\tilde{Y} = \tilde{X}\beta^0 + \tilde{\varepsilon}$ with $\tilde{Y} = \tilde{W}Y$, etc. Then
(see e. g. Wooldridge (2003))

$$\hat{\beta}_\ell^{(LWS,n,w)}(Y,X) - \beta_\ell^0 = \hat{\beta}_\ell^{(OLS,n)}(\tilde{Y},\tilde{X}) - \beta_\ell^0 = \left[u^{(\ell)}\right]' \tilde{\varepsilon} \cdot \left\| u^{(\ell)} \right\|^{-2}$$

with

$$\operatorname{var}\left(\hat{\beta}_\ell^{(LWS,n,w)}\right) = \sigma^2 \sum_{i=1}^n w_i \left[u_i^{(\ell)}\right]^2 \left\| u^{(\ell)} \right\|^{-4} \underbrace{=}_{denote} \sigma^2 D_n^2(W,X).$$

Then

$$\mathcal{L}\left(\frac{\hat{\beta}_\ell^{(LWS,n,w)} - \beta_\ell^0}{\sigma D_n(W,X)}\right) = \mathcal{N}(0,1). \tag{9}$$

Recalling the classical regression for $\tilde{Y}, \tilde{X}$

Let us recall that $W^{\frac{1}{2}} = \tilde{W} = \operatorname{diag}\left\{ w_1^{\frac{1}{2}}, w_2^{\frac{1}{2}}, \ldots, w_n^{\frac{1}{2}} \right\}$ and that it is well-known from the classical regression

$$\tilde{r}\left(\hat{\beta}^{(WLS,n)}\right) = \underbrace{\left(I - \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}' \right)}_{\text{denote it by } \tilde{M}} \tilde{\varepsilon} = \tilde{M}\tilde{W}\varepsilon.$$

$$\mathbb{E}\left[\tilde{r}'\left(\hat{\beta}^{(WLS,n)}\right) \cdot \tilde{r}\left(\hat{\beta}^{(WLS,n)}\right)\right] = \sigma^2 \cdot tr\left[W\left(I - \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}' \right)\right]$$
$$= \sigma^2 \cdot \left(\sum_{i=1}^{n} w_i(1 - d_{ii}) \right)$$

where $d_{ii} = \left[\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right]_{ii}$. So, $s_n^2 = \frac{\tilde{r}'\left(\hat{\beta}^{(WLS,n)}\right)\cdot\tilde{r}\left(\hat{\beta}^{(WLS,n)}\right)}{\sum_{i=1}^{n} w_i(1-d_{ii})} \quad \to \quad \mathbb{E}s_n^2 = \sigma^2$.

Recalling the classical regression for $\tilde{Y}, \tilde{X}$

Let us recall that $W^{\frac{1}{2}} = \tilde{W} = \mathrm{diag}\left\{ w_1^{\frac{1}{2}}, w_2^{\frac{1}{2}}, ..., w_n^{\frac{1}{2}} \right\}$ and that it is well-known from the classical regression

$$\tilde{r}\left(\hat{\beta}^{(WLS,n)}\right) = \underbrace{\left(I - \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right)}_{\textit{denote it by } \tilde{M}} \tilde{\varepsilon} = \tilde{M}\tilde{W}\varepsilon.$$

Then (recall also that $\boldsymbol{E}\tilde{\varepsilon} = 0$ and $\mathrm{cov}(\tilde{\varepsilon}) = \sigma^2 W$ where $W = \mathrm{diag}\{w_1, w_2, ..., w_n\}$)

$$\tilde{r}'\left(\hat{\beta}^{(WLS,n)}\right) \cdot \tilde{r}\left(\hat{\beta}^{(WLS,n)}\right) = [\tilde{\varepsilon}]'\,\tilde{M}\tilde{\varepsilon} = [\tilde{\varepsilon}]'\,W^{-\frac{1}{2}}\underbrace{W^{\frac{1}{2}}\tilde{M}W^{\frac{1}{2}}}_{M}\underbrace{W^{-\frac{1}{2}}\tilde{\varepsilon}}_{\eta} = \eta' M \eta$$

with $\mathcal{L}(\eta) = \mathcal{N}\left(0, \sigma^2 I\right)$.

Recalling the classical regression for $\tilde{Y}, \tilde{X}$

Let $Q'MQ = \Lambda$ where $Q = [q_1, q, ..., q_n]$ and $\Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, ..., \lambda_n\}$ with $M \cdot q_i = \lambda_i \cdot q_i, \quad \text{i. e.} \ \ \lambda_i > 0, \ \ i = 1, 2, ..., n - p, \quad \lambda_i = 0 \ \text{ otherwise } \ \text{and} \ \ \mathrm{Q'Q = QQ'} = \mathbf{I}.$

Then $M = Q\Lambda Q'$ and

$$\tilde{r}'\left(\hat{\beta}^{(WLS,n)}\right) \cdot \tilde{r}\left(\hat{\beta}^{(WLS,n)}\right) = [\tilde{\varepsilon}]' \, Q\Lambda Q' \tilde{\varepsilon} = \xi'\xi$$

with $\xi = \Lambda^{\frac{1}{2}} Q\eta$ (notice that the last p coordinates of $\xi = 0$), i.e. $\mathbb{E}\xi = 0$ and $\mathrm{cov}(\xi) = \sigma\Lambda$.

Recalling the classical regression for $\tilde{Y}, \tilde{X}$

Then

$$\mathcal{L}(\xi) = \mathcal{N}\left(0, \sigma^2 \cdot \Lambda\right).$$

Finally,

$$\mathcal{L}\left(\sigma^{-2}\tilde{r}'\left(\hat{\beta}^{(WLS,n)}\right) \cdot \tilde{r}\left(\hat{\beta}^{(WLS,n)}\right)\right) = \chi^2_{generalized}(n-p)$$

in the sense that
$\chi^2_{generalized}(n-p)$ is distribution
of the sum of squares of $n-p$ independent r. v.'s normally distributed
with zero mean but variance not equal one.

We conclude,

$$\mathcal{L}\left(\frac{\hat{\beta}^{(LWS,n,w)} - \beta^0}{D_n(W,X)} \cdot \left[\frac{\sum_{i=1}^n w_i(1-d_{ii})}{\tilde{r}'\left(\hat{\beta}^{(WLS,n)}\right) \cdot \tilde{r}\left(\hat{\beta}^{(WLS,n)}\right)}\right]^{\frac{1}{2}}\right) = t_{generalized}(n-p).$$

THANKS FOR ATTENTION