Jan Ámos Víšek

Instrumental weighted variables under heteroscedasticity. Part II – Numerical study

Kybernetika, Vol. 53 (2017), No. 1, 26–58

Persistent URL: http://dml.cz/dmlcz/146707

KYBERNETIKA — VOLUME *53* (2017), NUMBER 1, PAGES 26–58

INSTRUMENTAL WEIGHTED VARIABLES UNDER HETEROSCEDASTICITY PART II – NUMERICAL STUDY

Jan Ámos Víšek

Results of a numerical study of the behavior of the *instrumental weighted variables* estimator – in a competition with two other estimators – are presented. The study was performed under various frameworks (homoscedsticity/heteroscedasticity, several level and types of contamination of data, fulfilled/broken orthogonality condition). At the beginning the optimal values of eligible parameters of estimatros in question were empirically established. It was done under the various sizes of data sets and various levels of the contamination of data. These values were then utilized in the numerical study. Its results indicate that *instrumental weighted variables* are as good as S- and W-*estimators* and under heteroscedasticity even better. The weight function of Tukey's type was used.

Keywords: heteroscedasticity of disturbances, numerical study of instrumental weighted variables.

Classification: 62J02, 62F35

1. DISCUSSING THE FRAMEWORK OF SIMULATIONS

The weak consistency and $\sqrt{n}$-consistency of the *instrumental weighted variables* (IWV) estimator, proved in Part I of this paper, are from the theoretical point of view presumably the property which is to be proved prior to any further research on any newly proposed estimator[1]. The further step is usually to establish the asymptotic representation, yielding typically asymptotic normality, see e. g. [18, 28] or [31]. Knowing the rate of convergence – typical $\sqrt{n}$-consistency, we need for a comparison of efficiency of the

DOI: 10.14736/kyb-2017-1-0026

[1]Although sometimes it is difficult. The history of point estimation recorded an illustrative example that waited for the proof of consistency of newly proposed estimator for rather long time. It was due to the fact that we looked as long as nearly 10 years for a solution of the problem of high breakdown point estimation of regression model. The problem was explicitly formulated in [6] and implicitly even earlier in Princeton study [1]. When the first feasible couple of proposals appeared (see [24, 25, 26, 29, 36]), people focused on checking that the required level of the breakdown point was reached and how the desired tools can be employed especially as diagnostic means. It caused that the proof of consistency under a general framework was delivered nearly 20 years later, see [38]. On the other hand, already at that time it was felt that one of important piece of information about any new estimator is information saying how it behaves for some small data sets and/or some data which became benchmarks, [26] and also [8].

estimators the results of the second order as in [16, 17, 19] or [32]. All these steps fulfill the requirements which became traditional in the classical statistics and when we try to keep them in the robust statistics we pay a sacrifice to tradition, the sacrifice which was nicely described in famous paper by Paul Halmos [15]. Nevertheless, it appeared that the knowledge of asymptotic normality is not sufficient and it inspired research on expansion (of a bit modified Edgeworth type, see [12] or [4]) on distribution function leading to results in the small sample asymptotics, see e. g. [13] or [23]. Although these results demonstrated that the asymptotics may start to work nicely from very small sample sizes (surprisingly from 3 or 4), the validity of approximation had to be checked by simulation studies anyway. So it seems that for any accountable application of an estimator we need some idea about the behaviour of the estimator for the finite samples of data under various circumstances delivered by well designed numerical study. However, even nowadays when computational means offer very fast computation it can be a bit problem. The structures of data and/or of their contamination can be so manifold that an "exhaustive" simulations would be endless (and consequently, the extent of resulting papers huge). Nevertheless, we hope that we offer results which demonstrate that simulation study can "discover" that our traditional ideas (making-up the statistical/econometric folklore) may be wrong.

What can be also fruitful is a comparison of new estimator with some alternative estimator(s) – all after it was also the recommendation of reviewer and we thank for it. The recommendation appeared more useful than it could be guessed at the first glance. The comparison with other estimators (representing high quality of estimators) revealed which kind of data and of contamination can cause really serious complications to the estimators of regression model. Our numerical study in the first draft of paper (which did not contain any comparison with other estimators) confirmed a good behaviour of IWV in the situation when the orthogonality condition is broken and a contamination is present. However, the framework was proposed in a way which seemed to be the most terrible for the estimator according to the traditional ideas, we can meet in the papers on robust estimation of regression model – see discussion below. May be that without comparing IWV with some other estimators we would never "discover" some gimmicks of estimating the underlying regression model. Thanks for it once again.

So, we looked for competitors for the *instrumental weighted variables.* The very first idea may be to employ M-instrumental variables as they were studied in [34]. Pros for the M-estimators is a quick algorithm by which we can compute (most of) them, see [2]. However on the other hand, they have only limited maximal breakdown point, see [21].

Recently many people started to employ not only OLS but also the *regression quantiles* (RQ), see [20]. It has probably various reasons. Firstly, the problems caused by contamination became more significant because of type and amount of data. Secondly, the RQ became available subroutine in STATA already a couple years ago, earlier than other possibilities, see below. Finally, RQ are easy to understand and to accept the underlying idea. However, the RQ are also M-estimators and the reasons given above hint that it would be useless.

Recently appeared several proposals, see e. g. [9] or [10], see also [14]. As the estimator proposed in [10] (which is based on S-estimator) was implemented and included into STATA too, see [11], it became more or less widely used and that is why we have selected them[2]. Of course, we should demonstrate that LWS and S-estimator are based on different ideas, i. e. that LWS is not a special case of S-estimator and vice versa – to have authentic competitors. We shall do it after recalling their definitions below.

In [11] the authors advocated to use also a robustification of the instrumental variables based on W-estimators (they will be recalled later on) which employ as an external rule for assigning the weights to residuals the Mahalanobis distance. As they claimed that this estimator exhibits a significant gain in efficiency with respect to the estimator proposed by Cohen Freue, Ortiz–Molina and Zamar [10], we have included also this estimator into our study[3].

1.1. Recalling the estimators used in numerical study

To facilitate understanding the results (collected in the tables below), we recall the basic notions. The estimators which are considered in the study from the region of the classical statistics and econometrics are:

The *ordinary least squares* (OLS)

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

where for $i = 1, 2, \ldots, n$ $\quad Y_i, X_i$ denotes response and explanatory variables, respectively.

Selecting the *instrumental variables* Z_i's the estimator by means of the method of *instrumental variables* (IV) $\hat{\beta}^{(IV,n)} = \hat{\beta}^{(IV,n)}(Y,X)$ is given as solution of the equations

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

Denoting for any $\beta \in R^p$ the residual $r_i(\beta) = Y_i - X_i'\beta$ and by $r^2_{(i)}(\beta)$ the ith order statistics among the squared residuals, the *least weighted squares* (LWS) are given by

$$\hat{\beta}^{(LWS,n,w)} = \hat{\beta}^{(LWS,n,w)}(Y,X) = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w_i r^2_{(i)}(\beta) \tag{1}$$

where $i = 1, 2, \ldots, n$ and $w_i \in [0,1]$ are non-increasing weights. It is only a technicality to show that, when denoting (for any $\beta \in R^p$) the empirical distribution function of the residuals by $F_\beta^{(n)}(r)$, $\hat{\beta}^{(LWS,n,w)}$ is one of the solutions of the *normal equations*

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

The *instrumental weighted variables* (IWV) $\hat{\beta}^{(IWV,n)} = \hat{\beta}^{(IWV,n)}(Y,X)$ are given as (any) solution of the equations

[2] All after, it was also the recommendation of the reviewer. On the other hand, for our study it was rather complicated to combine our procedures written in MATLAB – for LWS and IWV – which moreover include the results directly into TEX tables, with STATA and then to transfer the results into TEX tables. Therefore we followed [11] with their recommendation to consult [30] together with [7] and we wrote MATLAB code for the S-, W-, SIV- and WIV-estimators (the definitons are recalled below) – the code is available on the request.

[3] The MATLAB code is also available on the request.

$$\sum_{i=1}^{n} w\left(F_{\beta}^{(n)}(|r_i(\beta)|)\right) Z_i \left(Y_i - X_i^{'}\beta\right) = 0. \tag{2}$$

As we have already mentioned above, we selected two competitors to the *instrumental weighted variables*. They are modifications of S- and of W-estimators for the situation when the orthogonality condition is broken. Firstly, let us recall S- and W-estimators. Let $\rho : (-\infty, \infty) \rightarrow (0, \infty), \ \ \rho(x) = \rho(-x), \textit{nondecreasing on } (0, \infty)$. Then

$$\hat{\beta}^{(S,\rho,n)} = \underset{\beta \in R^p}{\arg\min} \left\{ \sigma \in R^+ : \sum_{i=1}^{n} \rho\left(\frac{r_i(\beta)}{\sigma}\right) = b \right\} \tag{3}$$

where $b = I\!\!E\rho(\frac{\varepsilon_1}{\sigma_0})$, is called the *S-estimator* (S), see [27]. We have used Tukey's ρ function – please, see details below in the paragraph on the optimality of all parameters of estimators included into simulation study.

S-estimator can be simply computed as

$$\hat{\beta}^{(S,\rho,n)} = \left[\widehat{\Sigma}_{XX}^{S}\right]^{-1} \cdot \widehat{\Sigma}_{XY}^{S}$$

where $\widehat{\Sigma}_{XX}^{S}$ and $\widehat{\Sigma}_{XY}^{S}$ is S-estimator of scatter matrix of explanatory variables and S-estimator of covariance between explanatory variables X and the response Y, for details see [11] (see also [30] for technicalities about estimating the covariance matrix and [7] for algorithm of estimating scatter and location by S-estimator).

Let us compute the Mahalanobis distances for individual observations

$$\hat{d}_i = \sqrt{(M_i - \hat{\mu}_M)' \left[\widehat{\Sigma}_{M}^{S}\right]^{-1} (M_i - \hat{\mu}_M)}$$

where M_i is the ith row of the matrix $M = (X, Y)$ (M_i are assumed to be column vector) and $\widehat{\Sigma}_{M}^{S}$ and $\hat{\mu}_{M}^{S}$ are the S-estimators of scatter matrix and location of data M, respectively. Further, delete those observations that are associated with $\hat{d}_i$'s larger than $\sqrt{\chi^2_{p+1,q}}$ where q is a confidence level, e. g. 99 %, and denote the reduced data as $\tilde{X}$ and $\tilde{Y}$. Then the *W-estimator* (W) is given as

$$\hat{\beta}^{(W,\rho,q_W,n)} = \left(\tilde{X}'\tilde{X}\right)^{-1} \tilde{X}\tilde{Y}, \tag{4}$$

see again [11].

1.2. Mutual relation of estimators

We have promised to show that LWS are not special case of S-estimators and vice versa. Showing that we indicate that the competitors are based on an alternative idea(s) than IWV and hence they can be used as serious competitors. The S-estimators were defined by the extremal problem (3). Putting $\tilde{\rho} = \rho(\sqrt{x})$ we can write

$$\hat{\beta}^{(S,\tilde{\rho},n)} = \underset{\beta \in R^p}{\arg\min} \left\{ \sigma \in R^+ : \sum_{i=1}^{n} \tilde{\rho}\left(\frac{r_i^2(\beta)}{\sigma^2}\right) = b \right\}. \tag{5}$$

On the other hand, denoting for any $\beta \in R^p$

$$\sigma = \sigma(\beta) = \sqrt{\sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_i^2(\beta)},$$

we obtain from (1)

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \left\{ \sigma \in R^+ \ : \ \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) \frac{r_i^2(\beta)}{\sigma^2} = 1 \right\}. \tag{6}$$

The comparison of (5) and (6) implies that if we have to prove that LWS can be represented as S-estimator, we have to show that for any weight function w there is a nondecreasing function ρ^* such that

$$\frac{1}{b}\rho^*\left(\frac{r_i^2(\beta)}{\sigma^2}\right) = w\left(\frac{i-1}{n}\right) \frac{r_i^2(\beta)}{\sigma^2}. \tag{7}$$

Now, let us realize that $r_{(i)}^2(\beta) \le r_{(i+1)}^2(\beta)$ implies – due to the assumption that ρ^* is nondecrasing on $(0,\infty)$

$$\rho^*\left(\frac{r_{(i)}^2(\beta)}{\sigma^2}\right) \le \rho^*\left(\frac{r_{(i+1)}^2(\beta)}{\sigma^2}\right)$$

while we can have

$$w\left(\frac{i-1}{n}\right) \frac{r_{(i)}^2(\beta)}{\sigma^2} \le w\left(\frac{i}{n}\right) \frac{r_{(i+1)}^2(\beta)}{\sigma^2}$$

as well as

$$w\left(\frac{i-1}{n}\right) \frac{r_{(i)}^2(\beta)}{\sigma^2} \ge w\left(\frac{i}{n}\right) \frac{r_{(i+1)}^2(\beta)}{\sigma^2}.$$

It indicates that we cannot find a nondecreasing function ρ^* fulfilling (7). So the conclusion is: The LWS is not a special case of S-estimator but – taking into account once again just performed considerations – we can also conclude that S-estimators cannot be represented as LWS by a special adjustment of the weights w_i's. Finally, turn to the modifications of $\hat{\beta}^{(LWS,n,w)}$, $\hat{\beta}^{(S,\rho,n)}$ and $\hat{\beta}^{(W,n)}$ for the situation when the orthogonality condition is broken. The IWV were already racalled in (2).

The *S-instrumental variables* (SIV) estimator, a modification of S-estimator is given as

$$\hat{\beta}^{(SIV,\rho,n)} = \left\{ \widehat{\Sigma}_{XZ}^{S} \left[\widehat{\Sigma}_{ZZ}^{S}\right]^{-1} \widehat{\Sigma}_{XZ}^{S} \right\}^{-1} \cdot \widehat{\Sigma}_{XZ}^{S} \left[\widehat{\Sigma}_{ZZ}^{S}\right]^{-1} \widehat{\Sigma}_{ZY}^{S}$$

where again $\widehat{\Sigma}_A^S$ is the S-estimator of the matrix A. Similarly, the *W-instrumental variables* (WIV) estimator is given as

$$\hat{\beta}^{(WIV,\rho,q_W,n)} = \left\{ \widehat{\Sigma}_{XZ}^{W} \left[\widehat{\Sigma}_{ZZ}^{W}\right]^{-1} \widehat{\Sigma}_{XZ}^{W} \right\}^{-1} \cdot \widehat{\Sigma}_{XZ}^{W} \left[\widehat{\Sigma}_{ZZ}^{W}\right]^{-1} \widehat{\Sigma}_{ZY}^{W},$$

i. e. $\hat{\beta}^{(WIV,n)}$ is computed along the same lines as $\hat{\beta}^{(SIV,\rho,n)}$ but all S-estimates are substituted by W-estimates, see again [11].

2. DESCRIPTION OF THE FRAMEWORK OF NUMERICAL STUDIES

Data generating model was employed as follows

$$Y_{it} = \beta_1^0 \cdot X_{it1} + \beta_2^0 \cdot X_{it2} + \cdots + \beta_5^0 \cdot X_{it5} + \varepsilon_{it} = X_{it}'\beta^0 + \varepsilon_{it},\ i = 1, 2, \ldots, n,\ \ t = 1, 2, \ldots, T$$

where we denoted by $X_{it} = (X_{it1}, X_{it2}, \ldots, X_{it5})'$ and by $\beta^0 = (\beta_0^0, \beta_1^0, \ldots, \beta_4^0)'$. In other words, *we generated panel data containing n blocks, each block created by T observations.* Throughout the whole numerical study we kept $X_{it1} = 1$ for all $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$ and the values of coordinates of β^0 were

β^0	5.000	−4.000	3.000	−2.000	1.000

Since all the estimators (we have included into this numerical study) are regression equivariant, the values of coordinates of β^0 can be selected arbitrarily[4]. To make easier reading this paper we have selected such which is easy to remember and for the sake of space we will not repeat them. Explanatory variables were generated as follows.

2.1. Generating explanatory variables

. For each i ($i = 1, 2, \ldots, n$) we generated 4-dimensional random vector U_i with mutually independent coordinates, each of them distributed according to standard normal distribution. Moreover we generated an innovation sequence of independent identically distributed 4-dimensional vectors, $\{v_{i,t}\}_{i=1,t=1}^{n,T+1}$ (say), again with mutually independent coordinates, each of them distributed according to standard normal distribution. Then we put $V_{i,t,1} = 1$ for all $i = 1, 2, \ldots, n$ and $t = 0, 1, 2, \ldots T + 1$ and $V_{i,0,j} = U_{i,j-1}$ for all $i = 1, 2, \ldots, n$ and $j = 2, 3, 4$ and 5. We continue for $\theta \in [0, 1]$ with

$$V_{i,t,j} = (1 - \theta) \cdot V_{i,t-1,j} + \theta \cdot v_{i,t,j-1}, \quad i = 1, 2, \ldots, n, \quad t = 1, 2, \ldots, T + 1, \quad j = 2, 3, 4, 5.$$

Finally, we put $X_{i,t} = V_{i,t+1}$ and $Z_{i,t} = V_{i,t}$ for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$ and from some data we made good leverage points – see a discussion below..

2.2. Generating the disturbances

. Disturbances ε_i's were distributed as follows: We generated i. i. d. sequence

$$\{\varepsilon_{i,t}\}_{i=1,t=1}^{n,T+1}$$

each element of sequence being distributed according to the standard normal distribution. When *the orthogonality condition was assumed to be broken*[5] but the disturbances are still homoscedastic (of course, it is a case which is not very frequent that the disturbances are correlated with explanatory variables and they are still homoscedastic – however we should address all possibilities), we changed the explanatory variables on

$$X_{i,t} = V_{i,t+1} + \gamma \cdot \varepsilon_{i,t}, \quad i = 1, 2, \ldots, n \ \text{ and } \ t = 1, 2, \ldots, T \tag{8}$$

[4]We make couple of experiments whether the estimators which are theoretically regression-equivariant, are really also empirically regression-equivariant. It appeared that they are.

[5]One can learn that the most of results reported below in tables were found for the broken orthogonality condition.

If moreover the *disturbances were assumed to be heteroscedastic*, we put

$$\tilde{\varepsilon}_{i,t} = \tau_{i,t} \cdot \varepsilon_{i,t}, \quad i = 1, 2, \ldots, n \text{ and } t = 1, 2, \ldots, T \tag{9}$$

where $\tau_{i,t}$'s were generated as a sequence of i. i. d. r. v.'s distributed according to the uniform distribution on the interval $(0.5, a)$ where a is specified below[6].

2.3. How were the data contaminated?

Finally, we should describe how the data were contaminated. Naturally, when proposing a new estimator (with a hope that it will be good) we should try to examine it on data which can cause the estimators of given type (in our case, the estimator of regression coefficients) the most serious problems. There is an idea (we suppose well spread and which we shared too) that the most serious problem are caused for any estimator of regression coefficients by bad leverage points. On the over hand, usually the leverage points are easier detectable than outliers. Nevertheless even if we recognize leverage points, e. g. by hat matrix corresponding to the matrix (Y, X), see [8], it need not be easy to decide whether they are good or bad leverage points. Of course, at the first glance it may seem that when there is only several leverage points, we can try to compute the estimates with and without them and everything will be clear. However, let us realize that in fact some of these detected leverage points can be good and some may be bad. Hence we have to take into account data sets created from our original data by deleting all possible subgroups of the group of k leverage points we have detected. It means that the extent of computation will increase with $k!$. That is why we usually prefer estimators which are able to distinguish between good and bad leverage points themselves. All robust estimators included into our numerical study are able to do it. Nevertheless, the above discussion implies that the presence of good leverage points in data set can cause some problems to the estimators in question.

2.3.1. Generating leverage points

Therefore we selected always one block of data and multiplied the values of all X's and Z's in this block by 20 (of course without the first coordinate – please see the construction of X's and Z's). The corresponding values of response variable were computed correctly. In other words, the data contained a group of good leverage points.

2.3.2. Generating outliers

We also selected one block and contaminated all data in this block, creating from them either the outliers or the bad leverage points (it is specified below). It also implies that the extent of contamination was the same as the extent of group of good leverage points. As the recommendation of reviewer of the first version of paper was to make the

[6] It means that the disturbances depend on explanatory variables as they are in fact a part of explanatory variables, see (8). On the other hand, variances of disturbances were generated separately because otherwise they would not be (generally) bounded (9). Moreover, it is necessary to realize that if the variance of a disturbance is rather large, such observation would be assumed as contamination and the influence would be depressed anyway. That is why we have used rather limited values of a, namely, 1.5, 3.5 and 5.5.

study for various sizes of data sets, the simulations were performed for total numbers of observation in data sets from 100 to 500 with step 100 (it is given at the head of tables).

2.4. Reporting the results

We have generated 500 sets, each containing $n \cdot T$ observations. The values of n and T are specified at the head of tables. Let us recall that n is the number of blocks while T is number of observations in each block. As in each dataset one block was contaminated, i. e. T observation among $n \cdot T$ observations, the level of contamination was $n^{-1} \cdot 100\,\%$. That is why – when we want to find the results for different sample sizes $n \cdot T = 100, 200, \ldots, 500$ with the same level of contamination we changed $T = \frac{100}{n}, \frac{200}{n}, \ldots, \frac{500}{n}$. Then the estimates of regression coefficients were computed, say

$$\left\{ \hat{\beta}^{(index,k)} = (\hat{\beta}_0^{(index,k)}, \hat{\beta}_1^{(index,k)}, \hat{\beta}_2^{(index,k)}, \hat{\beta}_3^{(index,k)}, \hat{\beta}_4^{(index,k)})' \right\}_{k=1}^{500} \tag{10}$$

where the abbreviations $OLS, IV, LWS, S, W, IWV, SIV$ and WIV at the position of *"index"* indicate the method employed for the computation, namely OLS for the *Ordinary Least Squares*, IV – for the *Instrumental Variables*, LWS – for the *Least Weighted Squares*, S – for *S-estimator*, W – for *W-estimator*, IWV for the *Instrumental Weighted Variables*, SIV – for *S-instrumental variables estimator* and finally WIV – for *W-instrumental variables estimator.*

Further, the empirical means and empirical mean square error of estimates of coefficients (over these 500 repetitions) were computed, i. e. we report values (for $j = 1, 2, 3, 4$ and 5)

$$\hat{\beta}_j^{(index)} = \frac{1}{500} \sum_{k=1}^{500} \hat{\beta}_j^{(index,k)} \quad \text{and} \quad \widehat{\mathrm{MSE}}\left(\hat{\beta}_j^{(index)}\right) = \frac{1}{500} \sum_{k=1}^{500} \left[\hat{\beta}_j^{(index,k)} - \beta_j^0\right]^2 \tag{11}$$

where (let's recall it once again) the abbreviations OLS, IV, LWS, S, W, IWV, SIV and WIV at the position of *"index"* indicate the method employed for the computation. The results are given in tables (starting with Table 8) in the form: The first cell of each row indicates the method, e. g. $\beta^{(OLS)}$, the next 5 cells contain then just $\hat{\beta}^{(OLS)}_{\left(\widehat{MSE}(\hat{\beta}^{(OLS)})\right)}$ for the first, the second up to the fifth coordinate. The respective framework (homoscedastic/heteroscedastic disturbances, contaminated/noncontaminated data, broken/fulfilled orthogonal condition) is further specified above each table and in its head.

Although the study was performed on very fast computer (with the frequency of processor 3.8 GHz and 16 MB of working memory), the simulations when we looked for the optimal values of free parameters and when 500 sets containing 500 observations were considered, were rather time-consuming. That is why for estimators having two assignable parameters (LWS and W-estimator) we had to select some reasonable pattern of possibilities. A special attention was devoted to the weights for LWS because the very first results indicated that the conclusions can be different from those made in the first version of paper, see discussion below. Prior to answering a question how to select the free parameters of estimators (or constants, if you want) let us repeat: The data were generated as the panel data – n blocks, each containing T observations, the contamination affects just all observations of one block.

3. HOW TO ASSIGN THE FREE PARAMETERS OF ESTIMATORS?

All robust estimators, we have decided to include into the simulation study, have some assignable parameters and objective functions. We have performed some preliminary simulations to establish empirically approximately optimal values of these parameters. We made these simulation individually for "parent" estimators (i. e. for LWS, S- and W-estimators) and assumed that the same values hold approximately also for IWV, SIV and WIV. Let us recall that all simulations were 500 times repeated. The number of observations are given in tables. As the criterion of quality was used the *aggregated mean square error*, i. e.

$$\widehat{\text{AMSE}}^{(index)} = \sum_{j=1}^{p} \widehat{\text{MSE}} \left(\hat{\beta}_j^{(index)} \right). \tag{12}$$

Because in the most cases the values of $\widehat{\text{AMSE}}$ are very low, the values given in tables below are in fact $10 \cdot \widehat{\text{AMSE}}$.

The weight function $w(r) : [0,1] \to [0,1]$ for LWS is equal to 1 for $0 \le r \le h$, it is equal to 0 for $g \le r \le 1$ and it decreases from 1 to 0 for $h \le r \le g$, i. e. putting $c = g - h$ and $y = g - r$ we compute

$$w(r) = 3\frac{y^2}{c^2} - 3\frac{y^4}{c^4} + \frac{y^6}{c^6}, \tag{13}$$

i. e. between h and g the weight function borrowed the shape from Tukey's ρ which is recalled below.

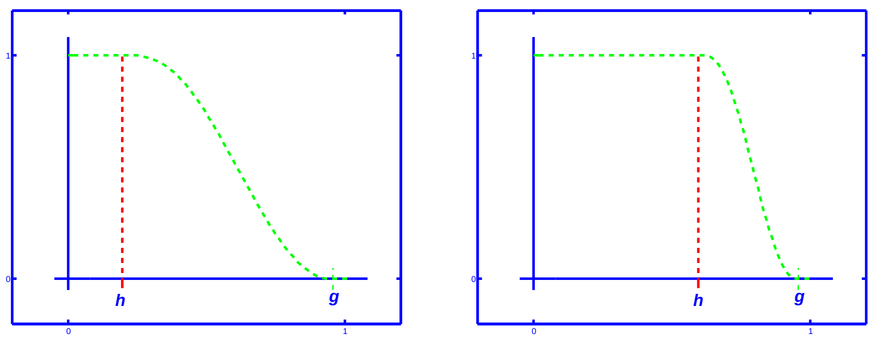

Fig. 1. The examples of possible shapes of weight function.

Then the weights w_i are given by $w_i = w(\frac{i-1}{n})$. The same weights were used for the *Least Weighted Squares* and for the *Instrumental Weighted Variables*. The values of h and g were selected according to numerical results presented below, for different sample sizes and various levels of contamination. For AMSE please see (12), for how the data were contaminated see subsection 2.3. Notice that AMSE is large for OLS even for low level of contamination.

$T = 1, n \cdot T = 100, AMSE^{(OLS)} = 78.45$									
$g - h = 0.06$									
h	0.840	0.860	0.880	0.900	0.920	**0.930**	0.940	0.950	0.960
g	0.900	0.920	0.940	0.960	0.980	**0.990**	1.000	1.000	1.000
AMSE	1.014	0.881	0.884	0.738	0.650	**0.548**	2.892	4.750	8.831
$g - h = 0.04$									
h	0.860	0.880	0.900	0.920	0.940	**0.950**	0.960	0.970	0.980
g	0.900	0.920	0.940	0.960	0.980	**0.990**	1.000	1.000	1.000
AMSE	0.942	0.896	0.798	0.721	0.624	**0.539**	9.528	20.577	38.529
$g - h = 0.02$									
h	0.880	0.900	0.920	0.940	0.960	**0.970**	0.980	0.990	0.999
g	0.900	0.920	0.940	0.960	0.980	**0.990**	1.000	1.000	1.000
AMSE	0.942	0.845	0.755	0.677	0.603	**0.584**	40.163	74.265	71.638
$g - h = 0.01$									
h	0.870	0.890	0.910	0.930	0.950	**0.970**	0.980	0.990	0.999
g	0.880	0.900	0.920	0.940	0.960	**0.980**	0.990	1.000	1.000
AMSE	0.967	0.981	0.842	0.762	0.670	**0.585**	0.595	73.181	69.653
$T = 2, n \cdot T = 200, AMSE^{(OLS)} = 65.29$									
$g - h = 0.06$									
h	0.840	0.860	0.880	0.900	0.920	**0.930**	0.940	0.950	0.960
g	0.900	0.920	0.940	0.960	0.980	**0.990**	1.000	1.000	1.000
AMSE	0.535	0.475	0.460	0.384	0.340	**0.318**	1.148	2.807	4.434
$g - h = 0.04$									
h	0.860	0.880	0.900	0.920	0.940	**0.950**	0.960	0.970	0.980
g	0.900	0.920	0.940	0.960	0.980	**0.990**	1.000	1.000	1.000
AMSE	0.534	0.481	0.420	0.361	0.326	**0.303**	3.726	9.495	24.030
$g - h = 0.02$									
h	0.880	0.900	0.920	0.940	0.960	**0.970**	0.980	0.990	0.999
g	0.900	0.920	0.940	0.960	0.980	**0.990**	1.000	1.000	1.000
AMSE	0.482	0.445	0.395	0.375	0.299	**0.295**	22.890	54.299	68.076

$g - h = 0.01$									
h	0.890	0.910	0.930	0.950	0.970	**0.980**	0.990	0.999	0.999
g	0.900	0.920	0.940	0.960	0.980	**0.990**	1.000	1.000	1.000
AMSE	0.462	0.448	0.397	0.346	0.291	**0.271**	51.941	82.688	67.959

$\boldsymbol{T = 3, n \cdot T = 300, AMSE^{(OLS)} = 72.35}$									
$g - h = 0.06$									
h	0.840	0.860	0.880	0.900	0.920	**0.930**	0.940	0.950	0.960
g	0.900	0.920	0.940	0.960	0.980	**0.990**	1.000	1.000	1.000
AMSE	0.371	0.326	0.298	0.265	0.242	**0.229**	0.851	1.332	2.563

$g - h = 0.04$									
h	0.860	0.880	0.900	0.920	0.940	**0.950**	0.960	0.970	0.980
g	0.900	0.920	0.940	0.960	0.980	**0.990**	1.000	1.000	1.000
AMSE	0.361	0.323	0.293	0.263	0.241	**0.206**	2.893	6.194	15.508

$g - h = 0.02$									
h	0.880	0.900	0.920	0.940	0.960	**0.970**	0.980	0.990	0.999
g	0.900	0.920	0.940	0.960	0.980	**0.990**	1.000	1.000	1.000
AMSE	0.360	0.331	0.273	0.261	0.200	**0.208**	15.502	47.702	75.658

$g - h = 0.01$									
h	0.890	0.910	0.930	0.950	0.970	**0.980**	0.990	0.999	0.999
g	0.900	0.920	0.940	0.960	0.980	**0.990**	1.000	1.000	1.000
AMSE	0.341	0.317	0.290	0.245	0.218	**0.216**	41.953	74.692	70.938

Tab. 1. The contamination by bad leverage pointson the level of 1 %, $n = 100$.

Similarly as the simulations were performed for $T = 1, 2$ and 3, we prolong these study for $T = 4$ and $T = 5$ and the results were used in the simulations (however for the sake of space we do not present them here – they can be obtained on a request). It follows from previous table that the optimal selection is $= 0.98$ and $g = 0.99$ for $n \cdot T = 200, 400$ and 500. For $n \cdot T = 100$ the optimal values are $h = 0.97$ and $g = 0.98$ and for $n \cdot T = 300$, the values $h = 0.95$ and $g = 0.99$ seem to be the best. On the other hand the differences between values from left to the point of minima are of order of 10 %. It means that probably (and we have the same experience from similar other study) very important requirement is that g is selected so that all leverage points have chance to obtain zero weight. The results of such a study significantly depend on the topology of data and of contamination. As we already mentioned it in the previous text, we try to prepare the framework of simulation in such a way to scrutinize the estimator by data which can cause the most terrible problems. However our idea about the "Achilles heel" of the estimator can be nastily influenced by the ideas which are handed down by "statistical folklore". To make more complex idea about the optimal selection of the weight function for various variants of (more or less "sneak") contamination would require tens or hundreds of hours of simulations.

A relatively stable results (for different sample sizes and different values of $g - h$) was taken as the reason why we simulated for the 5 % contamination only following table:

$T = 5, n \cdot T = 100, AMSE^{(OLS)} = 244.65, g - h = 0.01$									
h	0.890	0.900	0.910	0.920	0.930	**0.940**	0.950	0.960	0.970
g	0.900	0.910	0.920	0.930	0.940	**0.950**	0.960	0.970	0.980
AMSE	1.037	1.118	2.210	0.942	1.009	**0.771**	80.323	154.661	235.378

$T = 10, n \cdot T = 200, AMSE^{(OLS)} = 222.48, g - h = 0.01$									
h	0.890	0.900	0.910	0.920	**0.930**	0.940	0.950	0.960	0.970
g	0.900	0.910	0.920	0.930	**0.940**	0.950	0.960	0.970	0.980
AMSE	0.585	0.530	0.503	0.774	**0.470**	0.850	29.118	94.434	154.646

$T = 15, n \cdot T = 300, AMSE^{(OLS)} = 192.81, g - h = 0.01$									
h	0.890	0.900	0.910	0.920	0.930	**0.940**	0.950	0.960	0.970
g	0.900	0.910	0.920	0.930	0.940	**0.950**	0.960	0.970	0.980
AMSE	0.374	0.365	0.352	0.301	0.320	**0.287**	9.372	47.487	111.402

$T = 20, n \cdot T = 400, AMSE^{(OLS)} = 199.40, g - h = 0.01$									
h	0.890	0.900	0.910	0.920	0.930	**0.940**	0.950	0.960	0.970
g	0.900	0.910	0.920	0.930	0.940	**0.950**	0.960	0.970	0.980
AMSE	0.277	0.266	0.254	0.242	0.217	**0.203**	2.954	27.287	75.459

$T = 25, n \cdot T = 500, AMSE^{(OLS)} = 136.35, g - h = 0.01$									
h	0.890	0.900	0.910	0.920	0.930	**0.940**	0.950	0.960	0.970
g	0.900	0.910	0.920	0.930	0.940	**0.950**	0.960	0.970	0.980
AMSE	0.216	0.206	0.195	0.178	0.177	**0.165**	1.191	11.966	46.933

Tab. 2. The contamination by bad leverage points on the level of 5 % (please see again subsection 2.3), $n = 20$.

The optimal selection of h and g is 0.94 and 0.95, respectively, except of $n \cdot T = 200$ for which we obtained as optimal the values $h = 0.93$ and $g = 0.94$. Again the situation is rather flexible what concerns value of h especially for larger values of n.

The situation for a bit higher contamination – 20 % – is a somewhat different from the situation for small and modest contamination (as given in Table 1 and 2). Let us give firstly the table:

$T = 20, n \cdot T = 100, AMSE^{(OLS)} = 306.86$, $g - h = 0.01$								
h	0.510	0.520	0.530	0.540	0.550	0.560	0.570	**0.580**
g	0.520	0.530	0.540	0.550	0.560	0.570	0.580	**0.590**
AMSE	20.429	19.197	12.602	16.535	20.375	8.268	10.373	**7.946**
h	0.590	0.600	0.610	0.620	0.630	0.640	0.650	0.660
g	0.600	0.610	0.620	0.630	0.640	0.650	0.660	0.670
AMSE	8.435	12.572	8.867	10.234	19.079	21.086	15.732	16.590

$T = 40, n \cdot T = 200, AMSE^{(OLS)} = 191.80$, $g - h = 0.01$									
h	0.600	0.610	0.620	0.630	0.640	0.650	0.660	0.670	0.680
g	0.610	0.620	0.630	0.640	0.650	0.660	0.670	0.680	0.690
AMSE	4.450	1.018	0.622	2.342	4.528	0.473	0.579	1.678	7.405
h	**0.690**	0.700	0.710	0.720	0.730	0.740	0.750	0.760	0.770
g	**0.700**	0.710	0.720	0.730	0.740	0.750	0.760	0.770	0.780
AMSE	**0.397**	1.026	1.078	2.911	7.317	6.472	8.529	8.236	7.960

$T = 60, n \cdot T = 300, AMSE^{(OLS)} = 142.41$, $g - h = 0.01$									
h	0.660	0.670	0.680	0.690	0.700	0.710	0.720	0.730	0.740
g	0.670	0.680	0.690	0.700	0.710	0.720	0.730	0.740	0.750
AMSE	0.261	0.478	0.241	0.468	0.211	0.199	4.530	0.181	1.532
h	0.750	0.760	0.770	0.775	0.780	0.785	0.790	**0.800**	0.810
g	0.760	0.770	0.780	0.785	0.790	0.795	0.800	**0.810**	0.820
AMSE	0.155	0.156	1.508	1.265	0.145	2.651	2.358	**0.130**	5.963

$T = 80, n \cdot T = 400, AMSE^{(OLS)} = 117.75$, $g - h = 0.01$									
h	0.680	0.690	0.700	0.710	0.720	0.730	0.740	0.750	0.760
g	0.690	0.700	0.710	0.720	0.730	0.740	0.750	0.760	0.770
AMSE	0.158	0.146	4.845	0.137	0.132	0.119	0.115	0.099	0.177
h	0.770	0.780	0.790	0.795	**0.800**	0.805	0.810	0.820	0.830
g	0.780	0.790	0.800	0.805	**0.810**	0.815	0.820	0.830	0.840
AMSE	0.105	0.095	4.681	6.151	**0.089**	0.194	0.155	0.152	0.259

$T = 100, n \cdot T = 500, AMSE^{(OLS)} = 94.87, g - h = 0.01$									
h	0.740	0.750	0.760	0.770	0.780	0.785	0.790	0.795	**0.800**
g	0.750	0.760	0.770	0.780	0.790	0.795	0.800	0.805	**0.810**
AMSE	0.085	0.087	0.080	0.079	0.066	0.070	0.071	0.065	**0.061**
h	0.810	0.815	0.820	0.825	0.830	0.835	0.845	0.855	0.865
g	0.820	0.825	0.830	0.835	0.840	0.845	0.855	0.865	0.875
AMSE	0.068	0.085	0.117	0.153	0.174	0.224	0.433	0.868	1.364

Tab. 3. The contamination by bad leverage points on the level of 20 % (please see again subsection 2.3), $n = 5$.

For the S-, W-, SIV- and WIV-estimators we have employed Tukey's ρ because it became to be one of the most popular one, see again [10] or [14]. It is given for some $c > 0$ as

$$\begin{aligned} \rho_c(x) &= \frac{x^2}{2} - \frac{x^4}{2 \cdot c^2} + \frac{x^6}{6 \cdot c^4} \quad \text{for} |x| \leq c, \\ &= \frac{c^2}{6} \quad \text{otherwise.} \end{aligned}$$

Let us recall that the value of $b = I\!\!E \rho \left(\frac{\varepsilon_1}{\sigma_0} \right)$ given as

$$b = p \frac{\chi^2_{p+2}(c^2)}{2} - p \cdot (p+2) \frac{\chi^2_{p+4}(c^2)}{2 \cdot c^2} + p \cdot (p+2) \cdot (p+4) \frac{\chi^2_{p+6}(c^2)}{6 \cdot c^4} + \frac{c^2}{6}(1 - \chi^2_p(c^2)),$$

($\chi^2_k(x)$ denotes the value of χ^2 distribution function with k degrees of freedom at the point x and let us recall that p is the dimension of regression model in question, including the intercept), see [7]. The approximately optimal values of c_S – the constants for Tukey's ρ for the S-estimator, again for different sample sizes and different levels of contamination were assigned respecting the results of preliminary numerical study result of which is collected in the next table.

c_S	6	7	8	9	10	11	12	13	14
$b(c,p)$	2.054	2.165	2.239	2.292	2.330	2.359	2.381	2.398	2.412
$T = 1, n \cdot T = 100$									
AMSE	0.671	0.660	**0.629**	0.643	0.641	0.649	0.648	0.686	0.679

$T = 2, n \cdot T = 200$									
AMSE	0.365	0.358	0.366	**0.340**	0.348	0.355	0.367	0.380	0.409

$T = 3, n \cdot T = 300$									
AMSE	0.264	0.259	**0.237**	0.259	0.249	0.238	0.271	0.269	0.301
$T = 4, n \cdot T = 400$									
AMSE	0.213	0.207	**0.201**	0.207	0.207	0.212	0.213	0.222	0.243
$T = 5, n \cdot T = 500$									
AMSE	0.186	0.179	0.177	**0.168**	0.173	0.169	0.174	0.184	0.208

Tab. 4. The contamination by bad leverage points on the level of 1 % (please see again subsection 2.3), $n = 100$.

c_S	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5
$b(c,p)$	1.764	1.884	1.979	2.054	2.115	2.165	2.205	2.239	2.268
$T = 5, n \cdot T = 100$									
AMSE	1.592	1.351	1.317	**1.250**	1.344	1.328	1.535	34.410	93.376
$T = 10, n \cdot T = 200$									
AMSE	1.034	0.916	0.907	0.895	0.894	**0.807**	0.910	1.746	10.923
$T = 15, n \cdot T = 300$									
AMSE	0.949	0.882	0.773	0.808	0.766	**0.763**	0.844	1.274	6.399
$T = 20, n \cdot T = 400$									
AMSE	0.948	0.782	0.775	0.713	**0.666**	0.724	0.807	0.922	2.843
$T = 25, n \cdot T = 500$									
AMSE	0.955	0.788	0.706	0.658	**0.656**	0.661	0.771	1.012	2.652

Tab. 5. The contamination by bad leverage points on the level of 5 %, $n = 20$.

Similarly as the simulations were performed for the contamination level of 1 % and 5 % in Tables 4 and 5, we prolong these study for further the contamination levels up to 20 % and the results were used in the simulations (however for the sake of space we do not present them here). Finally, approximately optimal values of c_W – the constants for Tukey's ρ and the confidence levels q_W for W-estimator were selected according to the results which are presented in the tables which follow. Notice that the approximately optimal values c_S and c_W are different. (In this case we present results for the contamination level 1 % and 20 % but the results for the other levels were simulated, too, and used in the simulations below, for q_W see (4).)

c_W	7.000	8.000	9.000	10.000	11.000	12.000	13.000	14.000	15.000
q_W	0.992	0.993	0.994	0.995	0.996	0.997	0.998	0.999	1.000
$\boldsymbol{T = 1, n \cdot T = 100}$									
AMSE	0.764	0.717	0.705	0.717	**0.671**	0.745	0.752	0.735	0.786
$\boldsymbol{T = 2, n \cdot T = 200}$									
AMSE	0.423	0.383	0.385	0.381	**0.373**	0.394	0.414	0.418	0.432
$\boldsymbol{T = 3, n \cdot T = 300}$									
AMSE	0.297	0.288	0.274	**0.264**	0.272	0.300	0.295	0.301	0.301
$\boldsymbol{T = 4, n \cdot T = 400}$									
AMSE	0.236	0.223	0.224	**0.208**	0.226	0.238	0.250	0.245	0.247
$\boldsymbol{T = 5, n \cdot T = 500}$									
AMSE	0.211	0.204	0.191	0.189	**0.185**	0.200	0.197	0.214	0.209

Tab. 6. The contamination by bad leverage points on the level of 1 %, $n = 100$.

c	3.500	4.000	4.500	5.000	5.500	6.000	6.500
q_W	0.995	0.996	0.997	0.998	0.999	0.999	1.000
$\boldsymbol{T = 20, n \cdot T = 100}$							
AMSE	1302.371	1031.556	**295.990**	472.084	524.253	649.898	936.679
$\boldsymbol{T = 40, n \cdot T = 200}$							
AMSE	200.431	158.875	**105.816**	147.129	206.742	323.318	854.863
$\boldsymbol{T = 60, n \cdot T = 300}$							
AMSE	136.927	68.122	**40.891**	69.065	113.084	185.740	930.297
$\boldsymbol{T = 80, n \cdot T = 400}$							
AMSE	55.491	59.487	**27.614**	46.116	85.877	136.653	890.311
$\boldsymbol{T = 100, n \cdot T = 500}$							
AMSE	49.50	36.091	**16.640**	31.927	57.246	105.820	898.305

Tab. 7. The contamination by bad leverage points on the level of 20 %, $n = 5$.

3.1. Assigning the free parametres in the case of real data ?

The above presented results of simulations will be used in what follows. Naturally, one can ask however how to select appropriately "free" parameters of estimators when processing real data. One possibility can be to use some estimator of contamination level, see e. g. [33]. It however assumes that we decide for some model of contamination and it need not be very reliable. Much better is probably to use *forward search*, see [3, 5, 35, 37], i. e. to start with with a rather conservative values of parameters (assuming high level of contamination) and in each step to decrease the level of robustness of the estimator in question up to the moment when a break of a "smooth" series of the estimates of models appears. It is true that to adjust in such a way carefully the weight functions takes some time. If we would be really careful, we start with conservative values of h and g (see FIGURE 2.) and then we increase (by small steps) g leaving h unchanged. When a break appears, we return g a bit back and start with the increase of h, again up to moment when either a break appears or we reach (nearly) the value of g. Then we can hope that we have adjusted the weight function so that the estimation will be efficient as much as possible under given (unknown) level of contamination and given (topology of) data. If we select the step rather large, our effort can be useless, if too small the process can be tiresome. Nevertheless, a potential way how to adjust a maximal efficiency of estimation exists and with the (fascinating) increase of speed of IT, we can hope that in (close) future it will be a routine step. The number inside tables have structure $\hat{\beta}^{(method)}_{\left(\widehat{MSE}(\hat{\beta}^{(method)})\right)}$, for more details see (11) and text below it.

3.2. Can we achieve efficiency by robust estimators?

In Table 8 we have collected results of study showing that in the case when there is no contamination the robust methods – if they employ *forward search* (see [3]) – can achieve the efficiency well comparable with the optimal method, i. e. the efficiency of OLS.

$h = 0.45, g = 0.55,\ c_S = 3,\ c_W = 5, q_W = 0.90$					
$\hat{\beta}^{(OLS)}$	$5.002_{(0.204)}$	$-4.001_{(0.163)}$	$3.002_{(0.139)}$	$-2.002_{(0.118)}$	$0.999_{(0.143)}$
$\hat{\beta}^{(LWS)}$	$5.012_{(1.721)}$	$-3.984_{(5.276)}$	$3.023_{(5.809)}$	$-2.025_{(5.741)}$	$0.999_{(6.079)}$
$\hat{\beta}^{(S)}$	$5.005_{(0.443)}$	$-3.999_{(5.071)}$	$3.008_{(4.599)}$	$-2.006_{(4.629)}$	$1.014_{(5.544)}$
$\hat{\beta}^{(W)}$	$5.006_{(0.702)}$	$-3.951_{(32.898)}$	$3.026_{(29.582)}$	$-2.014_{(31.191)}$	$1.013_{(31.668)}$
$h = 0.70, g = 0.85,\ c_S = 7,\ c_W = 10, q_W = 0.95$					
$\hat{\beta}^{(OLS)}$	$5.001_{(0.190)}$	$-4.001_{(0.135)}$	$2.999_{(0.130)}$	$-2.000_{(0.141)}$	$0.999_{(0.139)}$
$\hat{\beta}^{(LWS)}$	$4.997_{(0.539)}$	$-4.003_{(0.354)}$	$2.999_{(0.431)}$	$-2.008_{(0.477)}$	$0.998_{(0.464)}$
$\hat{\beta}^{(S)}$	$5.001_{(0.213)}$	$-4.012_{(1.374)}$	$2.995_{(1.450)}$	$-2.009_{(1.494)}$	$1.002_{(1.546)}$
$\hat{\beta}^{(W)}$	$5.000_{(0.217)}$	$-4.020_{(2.310)}$	$2.996_{(2.230)}$	$-2.009_{(2.197)}$	$1.009_{(2.375)}$

$h = 0.98, g = 0.995,\ c_S = 12,\ c_W = 16, q_W = 0.999$					
$\hat{\beta}^{(OLS)}$	$5.000_{(0.205)}$	$-4.000_{(0.135)}$	$3.000_{(0.129)}$	$-2.002_{(0.118)}$	$1.003_{(0.133)}$
$\hat{\beta}^{(LWS)}$	$5.000_{(0.219)}$	$-4.000_{(0.150)}$	$3.000_{(0.147)}$	$-2.002_{(0.137)}$	$1.002_{(0.149)}$
$\hat{\beta}^{(S)}$	$5.001_{(0.218)}$	$-4.003_{(1.319)}$	$2.997_{(1.282)}$	$-2.011_{(1.330)}$	$1.002_{(1.314)}$
$\hat{\beta}^{(W)}$	$5.000_{(0.225)}$	$-4.003_{(1.701)}$	$3.000_{(1.768)}$	$-2.010_{(1.808)}$	$0.998_{(1.728)}$
$h = 0.995, g = 1,\ c_S = 24,\ c_W = 35, q_W = 1$					
$\hat{\beta}^{(OLS)}$	$5.000_{(0.202)}$	$-3.996_{(0.119)}$	$3.001_{(0.169)}$	$-1.998_{(0.127)}$	$1.001_{(0.130)}$
$\hat{\beta}^{(LWS)}$	$5.000_{(0.201)}$	$-3.996_{(0.120)}$	$3.001_{(0.169)}$	$-1.998_{(0.129)}$	$1.001_{(0.130)}$
$\hat{\beta}^{(S)}$	$5.000_{(0.202)}$	$-3.997_{(0.134)}$	$3.002_{(0.179)}$	$-1.998_{(0.148)}$	$1.000_{(0.158)}$
$\hat{\beta}^{(W)}$	$5.000_{(0.201)}$	$-3.997_{(0.300)}$	$3.000_{(0.333)}$	$-1.998_{(0.292)}$	$1.001_{(0.348)}$

Tab. 8. Each data set contained 500 observations, 500 repetitions were performed. The data were not contaminated, disturbances were homoscedastic and the orthogonality condition held.

4. SIMULATIONS

After making the above presented preliminary simulations, we are prepared to carry out the simulations for all estimators simultaneously, keeping the approximately optimal values of parameters of estimators. The study should reveal the behavior of estimators which were proposed for the situations when the orthogonality is broken. All below given (patterns of) results of simulation study were obtained under this assumption – the orthogonality condition was broken. From the very merit of study it is clear that there are nearly endless number of combinations of "features" of situations (homo/heteroscedasticity with various level of heteroscedasticity, contamination by outliers and/or leverage points with various levels of contamination, etc.). We have selected some of them with the hope that they allow to create some idea about our possibilities to estimate (at least approximately) correctly the underlying model. We shall start with the simplest situation: the contamination by outliers and the disturbances are homoscedastic.

Prior to giving the tables of results of simulations, let us recall that numbers in the individual cells of tables reports values specified in (11), i. e.

$$\hat{\beta}^{(method)}_{\left(\widehat{MSE}(\hat{\beta}^{(method)})\right)}$$

for the first, the second up to the fifth regression coefficient.

In the first part of Table 9 the results for LWS, S- and W-estimator were included to give an idea how much the break of orthogonality condition influenced these estimators. In the rest of Table 9 we shall omit the results for LWS, S- and W-estimator (partially for the sake of space, partially because these results are not in the first interest of our simulations).

$\boldsymbol{T = 1, n \cdot T = 100, h = 0.98, g = 0.99, c_S = 8, c_W = 11, q_W = 0.996}$					
$\hat{\beta}^{(OLS)}$	$4.698_{(0.188)}$	$-3.590_{(0.395)}$	$3.116_{(0.177)}$	$-1.675_{(0.282)}$	$1.182_{(0.202)}$
$\hat{\beta}^{(IV)}$	$4.688_{(0.209)}$	$-3.818_{(0.288)}$	$2.892_{(0.189)}$	$-1.917_{(0.236)}$	$0.957_{(0.205)}$
$\hat{\beta}^{(LWS)}$	$5.002_{(0.006)}$	$-3.733_{(0.079)}$	$3.275_{(0.083)}$	$-1.729_{(0.081)}$	$1.270_{(0.080)}$
$\hat{\beta}^{(S)}$	$5.002_{(0.005)}$	$-3.731_{(0.077)}$	$3.275_{(0.081)}$	$-1.727_{(0.080)}$	$1.271_{(0.079)}$
$\hat{\beta}^{(W)}$	$5.002_{(0.005)}$	$-3.731_{(0.079)}$	$3.275_{(0.082)}$	$-1.729_{(0.080)}$	$1.272_{(0.080)}$
$\hat{\beta}^{(IWV)}$	$4.994_{(0.014)}$	$-3.996_{(0.015)}$	$3.025_{(0.015)}$	$-1.996_{(0.017)}$	$1.010_{(0.014)}$
$\hat{\beta}^{(SIV)}$	$4.993_{(0.012)}$	$-4.011_{(0.015)}$	$3.008_{(0.014)}$	$-2.008_{(0.015)}$	$0.996_{(0.014)}$
$\hat{\beta}^{(WIV)}$	$4.994_{(0.012)}$	$-4.012_{(0.015)}$	$3.011_{(0.015)}$	$-2.010_{(0.015)}$	$0.997_{(0.015)}$
$\boldsymbol{T = 2, n \cdot T = 200, h = 0.98, g = 0.99, c_S = 9, c_W = 11, q_W = 0.996}$					
$\hat{\beta}^{(IV)}$	$4.704_{(0.184)}$	$-3.818_{(0.286)}$	$2.855_{(0.226)}$	$-1.902_{(0.221)}$	$0.939_{(0.236)}$
$\hat{\beta}^{(IWV)}$	$5.004_{(0.006)}$	$-3.988_{(0.008)}$	$3.007_{(0.008)}$	$-1.997_{(0.008)}$	$1.008_{(0.008)}$
$\hat{\beta}^{(SIV)}$	$5.003_{(0.005)}$	$-3.999_{(0.008)}$	$2.992_{(0.008)}$	$-2.009_{(0.008)}$	$0.998_{(0.008)}$
$\hat{\beta}^{(WIV)}$	$5.003_{(0.005)}$	$-3.999_{(0.009)}$	$2.991_{(0.008)}$	$-2.009_{(0.008)}$	$0.998_{(0.008)}$
$\boldsymbol{T = 3, n \cdot T = 300, h = 0.98, g = 0.99, c_S = 8, c_W = 10, q_W = 0.995}$					
$\hat{\beta}^{(IV)}$	$4.705_{(0.170)}$	$-3.845_{(0.222)}$	$2.853_{(0.277)}$	$-1.877_{(0.217)}$	$0.951_{(0.176)}$
$\hat{\beta}^{(IWV)}$	$5.003_{(0.004)}$	$-3.992_{(0.005)}$	$3.006_{(0.006)}$	$-1.984_{(0.006)}$	$1.004_{(0.006)}$
$\hat{\beta}^{(SIV)}$	$5.005_{(0.004)}$	$-4.001_{(0.005)}$	$2.998_{(0.005)}$	$-1.993_{(0.005)}$	$0.999_{(0.006)}$
$\hat{\beta}^{(WIV)}$	$5.005_{(0.004)}$	$-4.001_{(0.006)}$	$2.998_{(0.006)}$	$-1.993_{(0.006)}$	$0.998_{(0.006)}$
$\boldsymbol{T = 4, n \cdot T = 400, h = 0.98, g = 0.99, c_S = 8, c_W = 10, q_W = 0.995}$					
$\hat{\beta}^{(IV)}$	$4.686_{(0.178)}$	$-3.802_{(0.340)}$	$2.852_{(0.242)}$	$-1.882_{(0.247)}$	$0.967_{(0.223)}$
$\hat{\beta}^{(IWV)}$	$5.002_{(0.003)}$	$-3.996_{(0.004)}$	$3.003_{(0.005)}$	$-1.996_{(0.005)}$	$1.009_{(0.005)}$
$\hat{\beta}^{(SIV)}$	$5.003_{(0.003)}$	$-4.003_{(0.004)}$	$2.994_{(0.005)}$	$-2.002_{(0.005)}$	$1.004_{(0.005)}$
$\hat{\beta}^{(WIV)}$	$5.003_{(0.003)}$	$-4.003_{(0.005)}$	$2.993_{(0.005)}$	$-2.002_{(0.005)}$	$1.004_{(0.005)}$
$\boldsymbol{T = 5, n \cdot T = 500, h = 0.98, g = 0.99, c_S = 9, c_W = 11, q_W = 0.996}$					
$\hat{\beta}^{(IV)}$	$4.693_{(0.160)}$	$-3.821_{(0.262)}$	$2.848_{(0.254)}$	$-1.932_{(0.235)}$	$0.964_{(0.264)}$
$\hat{\beta}^{(IWV)}$	$5.002_{(0.003)}$	$-4.003_{(0.004)}$	$3.001_{(0.004)}$	$-1.996_{(0.004)}$	$1.000_{(0.004)}$
$\hat{\beta}^{(SIV)}$	$5.002_{(0.002)}$	$-4.004_{(0.004)}$	$2.995_{(0.004)}$	$-2.000_{(0.004)}$	$0.995_{(0.004)}$
$\hat{\beta}^{(WIV)}$	$5.002_{(0.002)}$	$-4.004_{(0.004)}$	$2.995_{(0.004)}$	$-1.999_{(0.005)}$	$0.996_{(0.004)}$

Tab. 9. The contamination by outliers on the level of 1 %, $n = 100$. Please see section 2.1 up to 2.4. The disturbances were homoscedastic and the orthogonality condition was broken.

$T = 5, n \cdot T = 100, h = 0.94, g = 0.95, c_S = 6, c_W = 7, q_W = 0.996$					
$\hat{\beta}^{(IV)}$	$3.363_{(5.230)}$	$-2.961_{(10.393)}$	$2.329_{(7.036)}$	$-1.541_{(9.936)}$	$0.672_{(6.559)}$
$\hat{\beta}^{(IWV)}$	$4.999_{(0.014)}$	$-4.006_{(0.033)}$	$3.000_{(0.025)}$	$-2.012_{(0.035)}$	$1.006_{(0.029)}$
$\hat{\beta}^{(SIV)}$	$4.441_{(1.480)}$	$-3.498_{(3.917)}$	$2.707_{(2.474)}$	$-1.747_{(4.473)}$	$0.730_{(3.128)}$
$\hat{\beta}^{(WIV)}$	$4.155_{(1.687)}$	$-4.011_{(0.052)}$	$2.976_{(0.055)}$	$-2.029_{(0.069)}$	$0.976_{(0.053)}$
$T = 10, n \cdot T = 200, h = 0.94, g = 0.95, c_S = 7, c_W = 7, q_W = 0.996$					
$\hat{\beta}^{(IV)}$	$3.449_{(3.795)}$	$-3.219_{(6.909)}$	$2.500_{(4.831)}$	$-1.640_{(5.246)}$	$0.911_{(5.894)}$
$\hat{\beta}^{(IWV)}$	$4.995_{(0.006)}$	$-3.999_{(0.011)}$	$2.994_{(0.014)}$	$-2.001_{(0.012)}$	$0.996_{(0.014)}$
$\hat{\beta}^{(SIV)}$	$4.647_{(0.767)}$	$-3.603_{(3.057)}$	$2.709_{(2.116)}$	$-1.775_{(2.637)}$	$0.942_{(2.309)}$
$\hat{\beta}^{(WIV)}$	$4.496_{(0.852)}$	$-4.001_{(0.026)}$	$2.984_{(0.029)}$	$-2.007_{(0.026)}$	$0.985_{(0.026)}$

$T = 15, n \cdot T = 300, h = 0.94, g = 0.95, c_S = 7, c_W = 7, q_W = 0.996$					
$\hat{\beta}^{(IV)}$	$3.412_{(3.342)}$	$-3.445_{(4.282)}$	$2.605_{(3.866)}$	$-1.629_{(4.034)}$	$0.747_{(3.620)}$
$\hat{\beta}^{(IWV)}$	$5.003_{(0.005)}$	$-4.002_{(0.009)}$	$2.999_{(0.010)}$	$-1.998_{(0.010)}$	$0.995_{(0.010)}$
$\hat{\beta}^{(SIV)}$	$4.831_{(0.309)}$	$-3.785_{(1.794)}$	$2.899_{(0.801)}$	$-1.889_{(1.488)}$	$0.895_{(1.513)}$
$\hat{\beta}^{(WIV)}$	$4.695_{(0.431)}$	$-4.016_{(0.025)}$	$2.985_{(0.023)}$	$-2.008_{(0.025)}$	$0.997_{(0.022)}$
$T = 20, n \cdot T = 400, h = 0.94, g = 0.95, c_S = 6.5, c_W = 6, q_W = 0.995$					
$\hat{\beta}^{(IV)}$	$3.347_{(3.372)}$	$-3.448_{(3.900)}$	$2.541_{(3.960)}$	$-1.651_{(3.352)}$	$0.807_{(2.882)}$
$\hat{\beta}^{(IWV)}$	$4.997_{(0.003)}$	$-3.997_{(0.007)}$	$3.000_{(0.008)}$	$-2.001_{(0.006)}$	$1.003_{(0.008)}$
$\hat{\beta}^{(SIV)}$	$4.851_{(0.227)}$	$-3.833_{(1.168)}$	$2.794_{(1.697)}$	$-1.843_{(1.393)}$	$0.854_{(1.625)}$
$\hat{\beta}^{(WIV)}$	$4.774_{(0.303)}$	$-3.997_{(0.028)}$	$2.969_{(0.029)}$	$-2.009_{(0.025)}$	$0.982_{(0.030)}$

$T = 25, n \cdot T = 500, h = 0.94, g = 0.95, c_S = 6.5, c_W = 5, q_W = 0.994$					
$\hat{\beta}^{(IV)}$	$3.464_{(2.761)}$	$-3.705_{(1.991)}$	$2.721_{(2.352)}$	$-1.734_{(2.864)}$	$0.870_{(2.297)}$
$\hat{\beta}^{(IWV)}$	$5.003_{(0.003)}$	$-4.002_{(0.006)}$	$3.003_{(0.006)}$	$-2.003_{(0.005)}$	$0.997_{(0.005)}$
$\hat{\beta}^{(SIV)}$	$4.914_{(0.113)}$	$-3.901_{(0.754)}$	$2.928_{(0.747)}$	$-1.995_{(0.625)}$	$1.000_{(0.620)}$
$\hat{\beta}^{(WIV)}$	$4.827_{(0.223)}$	$-4.006_{(0.023)}$	$3.000_{(0.024)}$	$-2.016_{(0.025)}$	$0.997_{(0.031)}$

Tab. 10. The contamination by outliers on the level of 5 %, $n = 20$. Please see section 2.1 up to 2.4. The disturbances were homoscedastic and the orthogonality condition was broken.

We can conclude from previous tables that the all estimators are able to cope with the contamination on level 1 % but already the contamination on level of 5 % causes some problems – see MSE of SIV estimator in Table 10. The results for WIV estimator is a bit

$T = 20, n \cdot T = 100, h = 0.58, g = 0.59, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.389_{(9.330)}$	$-3.346_{(8.974)}$	$2.881_{(10.165)}$	$-1.646_{(15.230)}$	$0.893_{(7.263)}$
$\hat{\beta}^{(IWV)}$	$4.956_{(0.089)}$	$-3.929_{(0.252)}$	$3.072_{(0.277)}$	$-1.950_{(0.366)}$	$1.005_{(0.593)}$
$\hat{\beta}^{(SIV)}$	$0.941_{(1429.524)}$	$-1.061_{(9096.120)}$	$3.474_{(3076.015)}$	$-2.894_{(3306.974)}$	$0.733_{(746.432)}$
$\hat{\beta}^{(WIV)}$	$1.127_{(1029.077)}$	$1.710_{(11087.383)}$	$-6.695_{(23453.720)}$	$-6.185_{(8304.501)}$	$-2.301_{(5891.612)}$

$T = 40, n \cdot T = 200, h = 0.69, g = 0.70, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.614_{(6.156)}$	$-3.856_{(0.346)}$	$2.929_{(0.363)}$	$-1.912_{(0.330)}$	$0.950_{(0.370)}$
$\hat{\beta}^{(IWV)}$	$4.986_{(0.011)}$	$-3.999_{(0.009)}$	$3.014_{(0.010)}$	$-1.992_{(0.012)}$	$1.009_{(0.013)}$
$\hat{\beta}^{(SIV)}$	$3.364_{(32.797)}$	$-1.444_{(1014.471)}$	$3.032_{(666.919)}$	$-1.679_{(179.352)}$	$0.096_{(695.465)}$
$\hat{\beta}^{(WIV)}$	$2.438_{(27.455)}$	$-3.611_{(365.097)}$	$0.710_{(539.432)}$	$-1.720_{(181.528)}$	$0.255_{(133.987)}$

$T = 60, n \cdot T = 300, h = 0.80, g = 0.81, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.646_{(5.790)}$	$-3.929_{(0.125)}$	$2.922_{(0.151)}$	$-1.983_{(0.137)}$	$0.993_{(0.135)}$
$\hat{\beta}^{(IWV)}$	$4.991_{(0.006)}$	$-4.004_{(0.003)}$	$3.004_{(0.003)}$	$-2.000_{(0.003)}$	$1.004_{(0.003)}$
$\hat{\beta}^{(SIV)}$	$3.178_{(9.167)}$	$-0.698_{(1085.072)}$	$4.801_{(3996.203)}$	$-1.105_{(223.362)}$	$1.442_{(252.287)}$
$\hat{\beta}^{(WIV)}$	$2.555_{(7.974)}$	$-2.507_{(83.470)}$	$0.933_{(81.321)}$	$-1.523_{(52.334)}$	$-0.011_{(100.188)}$

$T = 80, n \cdot T = 400, h = 0.80, g = 0.81, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.736_{(5.298)}$	$-3.959_{(0.077)}$	$2.944_{(0.061)}$	$-1.985_{(0.068)}$	$0.995_{(0.059)}$
$\hat{\beta}^{(IWV)}$	$4.992_{(0.005)}$	$-3.999_{(0.002)}$	$3.005_{(0.002)}$	$-1.999_{(0.002)}$	$1.004_{(0.002)}$
$\hat{\beta}^{(SIV)}$	$3.392_{(24.990)}$	$-3.256_{(153.722)}$	$-2.900_{(12166.945)}$	$-1.772_{(122.549)}$	$-1.832_{(3078.206)}$
$\hat{\beta}^{(WIV)}$	$2.089_{(100.889)}$	$-0.970_{(1198.133)}$	$8.832_{(27261.453)}$	$-1.977_{(354.271)}$	$3.185_{(4121.882)}$

$T = 100, n \cdot T = 500, h = 0.80, g = 0.81, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.756_{(5.158)}$	$-3.950_{(0.053)}$	$2.956_{(0.049)}$	$-1.980_{(0.042)}$	$1.001_{(0.041)}$
$\hat{\beta}^{(IWV)}$	$4.998_{(0.003)}$	$-4.001_{(0.001)}$	$2.997_{(0.001)}$	$-2.002_{(0.001)}$	$1.002_{(0.001)}$
$\hat{\beta}^{(SIV)}$	$3.303_{(3.767)}$	$-3.103_{(22.649)}$	$2.433_{(16.521)}$	$-1.414_{(19.176)}$	$1.188_{(16.369)}$
$\hat{\beta}^{(WIV)}$	$2.824_{(42.293)}$	$-0.517_{(1697.478)}$	$5.029_{(4815.141)}$	$-0.953_{(69.444)}$	$0.644_{(72.691)}$

Tab. 11. The contamination by outliers on the level of 20 %, $n = 5$. Please see section 2.1 up to 2.4. The disturbances were homoscedastic and the orthogonality condition was broken.

better, nevertheless MSE of WIV is also larger than of IWV. The 20 % contamination then destroys both SIV as well as WIV. In this situation even the empirical mean values of the estimates of regression coefficients are rather biased.

Let's continue with the situation when data are contaminated still by outliers but the disturbances are heteroscedastic. As it was already said above (see (9)) we generated uniformly distributed random variable τ_{it} for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$ on the interval $[0.5, a]$ and put $\tilde{\varepsilon}_{i,t} = \tau_{i,t} \cdot \varepsilon_{i,t}$ with $Y = X\beta^0 + \tilde{\varepsilon}$. In the next two tables $a = 5.5$.

$\boldsymbol{T = 1, n \cdot T = 100, h = 0.98, g = 0.99, c_S = 8, c_W = 11, q_W = 0.996}$					
$\hat{\beta}^{(OLS)}$	$4.785_{(0.166)}$	$-2.923_{(1.624)}$	$3.860_{(1.144)}$	$-0.962_{(1.506)}$	$1.945_{(1.340)}$
$\hat{\beta}^{(IV)}$	$4.752_{(0.383)}$	$-3.923_{(0.420)}$	$2.897_{(0.429)}$	$-1.980_{(0.458)}$	$0.926_{(0.454)}$
$\hat{\beta}^{(LWS)}$	$4.979_{(0.073)}$	$-3.012_{(1.202)}$	$3.992_{(1.219)}$	$-0.990_{(1.275)}$	$2.002_{(1.252)}$
$\hat{\beta}^{(S)}$	$4.987_{(0.061)}$	$-2.984_{(1.100)}$	$4.034_{(1.136)}$	$-0.973_{(1.123)}$	$2.032_{(1.125)}$
$\hat{\beta}^{(W)}$	$4.966_{(0.065)}$	$-2.956_{(1.182)}$	$4.060_{(1.208)}$	$-0.956_{(1.177)}$	$2.058_{(1.198)}$
$\hat{\beta}^{(IWV)}$	$4.943_{(0.299)}$	$-4.008_{(0.333)}$	$3.023_{(0.343)}$	$-1.988_{(0.367)}$	$0.980_{(0.350)}$
$\hat{\beta}^{(SIV)}$	$4.965_{(0.301)}$	$-4.068_{(0.464)}$	$2.964_{(0.463)}$	$-2.028_{(0.433)}$	$0.955_{(0.456)}$
$\hat{\beta}^{(WIV)}$	$4.903_{(1.574)}$	$-4.183_{(4.325)}$	$2.877_{(2.923)}$	$-2.093_{(1.337)}$	$0.895_{(1.018)}$

$\boldsymbol{T = 2, n \cdot T = 200, h = 0.98, g = 0.99, c_S = 9, c_W = 11, q_W = 0.996}$					
$\hat{\beta}^{(IV)}$	$4.795_{(0.254)}$	$-3.914_{(0.316)}$	$2.844_{(0.329)}$	$-1.973_{(0.266)}$	$0.893_{(0.271)}$
$\hat{\beta}^{(IWV)}$	$4.982_{(0.153)}$	$-4.023_{(0.183)}$	$2.946_{(0.175)}$	$-2.012_{(0.179)}$	$0.961_{(0.159)}$
$\hat{\beta}^{(SIV)}$	$4.997_{(0.157)}$	$-4.050_{(0.231)}$	$2.917_{(0.245)}$	$-2.043_{(0.237)}$	$0.941_{(0.204)}$
$\hat{\beta}^{(WIV)}$	$4.988_{(0.178)}$	$-4.073_{(0.348)}$	$2.906_{(0.392)}$	$-2.050_{(0.333)}$	$0.932_{(0.301)}$

$\boldsymbol{T = 3, n \cdot T = 300, h = 0.98, g = 0.99, c_S = 8, c_W = 0, q_W = 0.995}$					
$\hat{\beta}^{(IV)}$	$4.695_{(0.289)}$	$-3.859_{(0.382)}$	$2.872_{(0.344)}$	$-1.960_{(0.340)}$	$0.916_{(0.296)}$
$\hat{\beta}^{(IWV)}$	$4.993_{(0.099)}$	$-4.020_{(0.143)}$	$2.997_{(0.132)}$	$-2.034_{(0.129)}$	$0.983_{(0.110)}$
$\hat{\beta}^{(SIV)}$	$5.007_{(0.093)}$	$-4.026_{(0.170)}$	$2.970_{(0.172)}$	$-2.044_{(0.156)}$	$0.973_{(0.143)}$
$\hat{\beta}^{(WIV)}$	$5.003_{(0.100)}$	$-4.035_{(0.225)}$	$2.973_{(0.221)}$	$-2.059_{(0.227)}$	$0.970_{(0.202)}$

$\boldsymbol{T = 4, n \cdot T = 400, h = 0.98, g = 0.99, c_S = 8, c_W = 10, q_W = 0.995}$					
$\hat{\beta}^{(IV)}$	$4.680_{(0.251)}$	$-3.826_{(0.344)}$	$2.822_{(0.377)}$	$-1.918_{(0.304)}$	$0.942_{(0.381)}$
$\hat{\beta}^{(IWV)}$	$4.988_{(0.078)}$	$-3.991_{(0.090)}$	$2.970_{(0.083)}$	$-2.002_{(0.087)}$	$0.991_{(0.092)}$
$\hat{\beta}^{(SIV)}$	$4.999_{(0.070)}$	$-4.004_{(0.117)}$	$2.959_{(0.105)}$	$-1.999_{(0.114)}$	$0.992_{(0.113)}$
$\hat{\beta}^{(WIV)}$	$4.992_{(0.077)}$	$-4.016_{(0.168)}$	$2.946_{(0.149)}$	$-2.013_{(0.167)}$	$0.980_{(0.158)}$

$\boldsymbol{T = 5, n \cdot T = 500, h = 0.98, g = 0.99, c_S = 9, c_W = 11, q_W = 0.996}$					
$\hat{\beta}^{(IV)}$	$4.709_{(0.218)}$	$-3.775_{(0.372)}$	$2.839_{(0.349)}$	$-1.941_{(0.326)}$	$0.939_{(0.374)}$
$\hat{\beta}^{(IWV)}$	$4.991_{(0.057)}$	$-3.998_{(0.082)}$	$2.967_{(0.099)}$	$-2.004_{(0.083)}$	$0.987_{(0.103)}$

$\hat{\beta}^{(SIV)}$	$5.005_{(0.056)}$	$-4.008_{(0.098)}$	$2.951_{(0.120)}$	$-2.017_{(0.100)}$	$0.968_{(0.119)}$
$\hat{\beta}^{(WIV)}$	$4.999_{(0.060)}$	$-4.011_{(0.143)}$	$2.938_{(0.172)}$	$-2.022_{(0.138)}$	$0.958_{(0.161)}$

Tab. 12. The contamination by outliers on the level of 1 %, $n = 100$. The values of variance of the disturbances randomly selected from the interval $[0.5, 5.5]$. The orthogonality condition was broken (see again section 2.1 up to 2.4.).

$T = 5, n \cdot T = 100, h = 0.94, g = 0.95, c_S = 6, c_W = 7, q_W = 0.996$					
$\hat{\beta}^{(IV)}$	$3.511_{(5.808)}$	$-3.502_{(9.829)}$	$2.116_{(7.547)}$	$-1.608_{(7.652)}$	$0.558_{(8.999)}$
$\hat{\beta}^{(IWV)}$	$4.880_{(0.799)}$	$-4.050_{(1.375)}$	$2.918_{(0.960)}$	$-2.006_{(1.153)}$	$0.914_{(1.717)}$
$\hat{\beta}^{(SIV)}$	$4.709_{(4.102)}$	$-4.278_{(54.866)}$	$2.522_{(21.485)}$	$-2.118_{(11.796)}$	$0.325_{(47.474)}$
$\hat{\beta}^{(WIV)}$	$4.343_{(8.304)}$	$-4.248_{(7.215)}$	$2.544_{(26.494)}$	$-2.304_{(7.009)}$	$0.255_{(62.928)}$

$T = 10, n \cdot T = 200, h = 0.94, g = 0.95, c_S = 7, c_W = 7, q_W = 0.996$					
$\hat{\beta}^{(IV)}$	$3.521_{(3.831)}$	$-3.338_{(6.166)}$	$2.525_{(5.436)}$	$-1.525_{(5.860)}$	$0.758_{(5.309)}$
$\hat{\beta}^{(IWV)}$	$4.950_{(0.233)}$	$-4.001_{(0.424)}$	$2.936_{(0.549)}$	$-1.974_{(0.654)}$	$0.961_{(0.386)}$
$\hat{\beta}^{(SIV)}$	$4.981_{(0.423)}$	$-4.024_{(5.608)}$	$2.838_{(3.459)}$	$-1.939_{(16.416)}$	$0.908_{(2.368)}$
$\hat{\beta}^{(WIV)}$	$4.622_{(0.778)}$	$-4.183_{(1.441)}$	$2.818_{(1.371)}$	$-2.085_{(2.913)}$	$0.857_{(1.134)}$

$T = 15, n \cdot T = 300, h = 0.94, g = 0.95, c_S = 7, c_W = 7, q_W = 0.996$					
$\hat{\beta}^{(IV)}$	$3.449_{(3.673)}$	$-3.606_{(4.660)}$	$2.550_{(5.346)}$	$-1.786_{(4.846)}$	$0.985_{(3.901)}$
$\hat{\beta}^{(IWV)}$	$4.950_{(0.158)}$	$-4.099_{(0.179)}$	$2.991_{(0.508)}$	$-2.068_{(0.467)}$	$1.008_{(0.392)}$
$\hat{\beta}^{(SIV)}$	$4.999_{(0.197)}$	$-4.136_{(0.695)}$	$2.897_{(0.833)}$	$-2.111_{(0.778)}$	$0.909_{(0.810)}$
$\hat{\beta}^{(WIV)}$	$4.769_{(0.568)}$	$-4.152_{(1.087)}$	$2.875_{(0.972)}$	$-2.133_{(1.340)}$	$0.873_{(0.830)}$

$T = 20, n \cdot T = 400, h = 0.94, g = 0.95, c_S = 6.5, c_W = 6, q_W = 0.995$					
$\hat{\beta}^{(IV)}$	$3.361_{(3.522)}$	$-3.472_{(4.207)}$	$2.525_{(3.479)}$	$-1.904_{(3.935)}$	$0.955_{(3.817)}$
$\hat{\beta}^{(IWV)}$	$4.924_{(0.101)}$	$-3.987_{(0.232)}$	$3.000_{(0.230)}$	$-1.972_{(0.209)}$	$0.988_{(0.214)}$
$\hat{\beta}^{(SIV)}$	$4.979_{(0.213)}$	$-4.086_{(1.085)}$	$2.847_{(2.208)}$	$-2.095_{(0.872)}$	$0.857_{(1.188)}$
$\hat{\beta}^{(WIV)}$	$4.773_{(0.400)}$	$-4.094_{(0.914)}$	$2.814_{(1.534)}$	$-2.073_{(0.839)}$	$0.826_{(1.003)}$

$T = 25, n \cdot T = 500, h = 0.94, g = 0.95, c_S = 6.5, c_W = 5, q_W = 0.994$					
$\hat{\beta}^{(OLS)}$	$3.455_{(2.813)}$	$-2.615_{(3.556)}$	$3.884_{(2.462)}$	$-0.739_{(3.314)}$	$2.025_{(2.714)}$
$\hat{\beta}^{(IV)}$	$3.441_{(2.904)}$	$-3.610_{(2.617)}$	$2.726_{(2.358)}$	$-1.773_{(2.594)}$	$0.844_{(2.508)}$
$\hat{\beta}^{(LWS)}$	$4.978_{(0.069)}$	$-2.940_{(1.781)}$	$4.120_{(2.015)}$	$-0.881_{(2.019)}$	$2.095_{(1.931)}$

$\hat{\beta}^{(S)}$	$4.984_{(0.043)}$	$-2.840_{(1.557)}$	$4.233_{(1.756)}$	$-0.839_{(1.563)}$	$2.207_{(1.660)}$
$\hat{\beta}^{(W)}$	$4.976_{(0.046)}$	$-2.832_{(1.639)}$	$4.224_{(1.794)}$	$-0.818_{(1.681)}$	$2.215_{(1.745)}$
$\hat{\beta}^{(IWV)}$	$4.956_{(0.068)}$	$-4.011_{(0.174)}$	$3.005_{(0.158)}$	$-2.020_{(0.183)}$	$0.999_{(0.187)}$
$\hat{\beta}^{(SIV)}$	$5.017_{(0.080)}$	$-4.092_{(0.614)}$	$2.913_{(0.621)}$	$-2.107_{(0.595)}$	$0.938_{(0.708)}$
$\hat{\beta}^{(WIV)}$	$5.077_{(11.854)}$	$-4.655_{(135.010)}$	$2.198_{(165.602)}$	$-2.034_{(15.789)}$	$-0.263_{(631.443)}$

Tab. 13. The contamination by outliers on the level of 5 %, $n = 20$. The values of variance of the disturbances randomly selected from the interval $[0.5, 5.5]$. The orthogonality condition was broken (see section 2.1 up to 2.4.).

The results collected in Tables 12 and 13 show that on contamination level 1 % and 5 % all estimators can give rather reliable information about the underlying model, although SIV and WIV have a bit larger MSE than IWV. For higher contamination the combination with heteroscedasticity can represent for SIV and WIV serious problem. When we learnt it, we became curious which level of heteroscedasticity starts to be "uncomfortable" for estimators in question. That is why we offer for this situation a bit more results.

$T = 20, n \cdot T = 100, h = 0.58, g = 0.59, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.501_{(7.860)}$	$-3.492_{(4.292)}$	$2.745_{(3.223)}$	$-1.770_{(3.881)}$	$0.835_{(4.077)}$
$\hat{\beta}^{(IWV)}$	$4.957_{(0.171)}$	$-3.941_{(0.400)}$	$3.023_{(0.106)}$	$-2.005_{(0.232)}$	$1.041_{(0.187)}$
$\hat{\beta}^{(SIV)}$	$1.956_{(754.757)}$	$-3.507_{(992.983)}$	$5.074_{(11118.891)}$	$5.326_{(16017.535)}$	$-1.504_{(2784.881)}$
$\hat{\beta}^{(WIV)}$	$2.550_{(47.141)}$	$-1.920_{(307.195)}$	$1.004_{(287.430)}$	$-1.728_{(321.087)}$	$0.932_{(443.731)}$

$T = 40, n \cdot T = 200, h = 0.69, g = 0.70, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.560_{(6.416)}$	$-3.887_{(0.516)}$	$2.851_{(0.545)}$	$-1.940_{(0.415)}$	$0.921_{(0.434)}$
$\hat{\beta}^{(IWV)}$	$4.990_{(0.010)}$	$-4.005_{(0.012)}$	$2.999_{(0.013)}$	$-1.994_{(0.007)}$	$1.002_{(0.012)}$
$\hat{\beta}^{(SIV)}$	$2.652_{(26.988)}$	$-1.107_{(701.649)}$	$1.391_{(113.934)}$	$-0.891_{(152.063)}$	$1.099_{(236.764)}$
$\hat{\beta}^{(WIV)}$	$2.448_{(27.360)}$	$0.048_{(893.196)}$	$1.389_{(841.463)}$	$-4.996_{(8666.455)}$	$-2.090_{(3797.268)}$

$T = 60, n \cdot T = 300, h = 0.80, g = 0.81, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.693_{(5.581)}$	$-3.926_{(0.115)}$	$2.911_{(0.144)}$	$-1.949_{(0.118)}$	$0.976_{(0.110)}$
$\hat{\beta}^{(IWV)}$	$4.986_{(0.006)}$	$-4.000_{(0.003)}$	$2.998_{(0.003)}$	$-2.001_{(0.003)}$	$0.999_{(0.003)}$
$\hat{\beta}^{(SIV)}$	$3.318_{(7.147)}$	$-2.052_{(174.964)}$	$2.134_{(81.094)}$	$-1.150_{(155.421)}$	$0.860_{(147.354)}$
$\hat{\beta}^{(WIV)}$	$2.572_{(7.812)}$	$-2.182_{(55.034)}$	$1.785_{(53.896)}$	$-1.076_{(119.606)}$	$0.373_{(161.460)}$

$T=80, n\cdot T=400, h=0.80, g=0.81, c_S=3, c_W=4.5, q_W=0.997$					
$\hat\beta^{(IV)}$	$2.708_{(5.414)}$	$-3.948_{(0.072)}$	$2.943_{(0.073)}$	$-1.965_{(0.063)}$	$0.956_{(0.082)}$
$\hat\beta^{(IWV)}$	$4.991_{(0.004)}$	$-4.003_{(0.002)}$	$2.999_{(0.001)}$	$-1.996_{(0.001)}$	$1.001_{(0.002)}$
$\hat\beta^{(SIV)}$	$3.197_{(4.481)}$	$-2.571_{(112.382)}$	$1.984_{(31.052)}$	$-1.171_{(74.374)}$	$0.606_{(29.182)}$
$\hat\beta^{(WIV)}$	$2.465_{(7.565)}$	$-6.464_{(6980.416)}$	$1.197_{(45.941)}$	$-1.841_{(140.763)}$	$-0.904_{(533.786)}$

$T=100, n\cdot T=500, h=0.80, g=0.81, c_S=3, c_W=4.5, q_W=0.997$					
$\hat\beta^{(IV)}$	$2.778_{(5.057)}$	$-3.942_{(0.053)}$	$2.952_{(0.050)}$	$-1.983_{(0.046)}$	$0.994_{(0.050)}$
$\hat\beta^{(IWV)}$	$4.987_{(0.003)}$	$-4.002_{(0.001)}$	$3.000_{(0.001)}$	$-2.002_{(0.001)}$	$1.001_{(0.001)}$
$\hat\beta^{(SIV)}$	$3.676_{(77.130)}$	$-4.222_{(910.660)}$	$1.401_{(358.280)}$	$-2.561_{(553.206)}$	$1.374_{(119.761)}$
$\hat\beta^{(WIV)}$	$2.587_{(6.335)}$	$-2.482_{(24.803)}$	$1.668_{(23.899)}$	$-1.330_{(25.399)}$	$0.539_{(23.336)}$

Tab. 14. The contamination by outliers on the level of 20 %, $n = 5$. The values of variance of the disturbances randomly selected from the interval $[0.5, 1.5]$. The orthogonality condition was broken (see section 2.1 up to 2.4).

$T=20, n\cdot T=100, h=0.58, g=0.59, c_S=3, c_W=4.5, q_W=0.997$					
$\hat\beta^{(IV)}$	$2.552_{(8.207)}$	$-3.539_{(18.056)}$	$2.922_{(36.527)}$	$-1.588_{(34.725)}$	$0.746_{(49.470)}$
$\hat\beta^{(IWV)}$	$4.727_{(0.593)}$	$-3.836_{(1.518)}$	$3.097_{(1.862)}$	$-1.779_{(1.961)}$	$1.155_{(1.202)}$
$\hat\beta^{(SIV)}$	$2.520_{(48.683)}$	$-2.522_{(422.614)}$	$0.526_{(339.076)}$	$-1.663_{(366.201)}$	$0.285_{(1003.163)}$
$\hat\beta^{(WIV)}$	$2.665_{(187.390)}$	$-3.508_{(2385.093)}$	$3.112_{(2780.185)}$	$1.030_{(1870.871)}$	$1.285_{(702.110)}$

$T=40, n\cdot T=200, h=0.69, g=0.70, c_S=3, c_W=4.5, q_W=0.997$					
$\hat\beta^{(IV)}$	$2.618_{(6.166)}$	$-3.863_{(0.583)}$	$2.842_{(0.522)}$	$-1.981_{(0.495)}$	$0.918_{(0.506)}$
$\hat\beta^{(IWV)}$	$4.861_{(0.085)}$	$-3.991_{(0.080)}$	$3.045_{(0.138)}$	$-1.972_{(0.069)}$	$1.037_{(0.089)}$
$\hat\beta^{(SIV)}$	$2.877_{(11.133)}$	$-3.565_{(1165.820)}$	$-2.807_{(4169.279)}$	$-3.432_{(1043.672)}$	$-0.177_{(287.683)}$
$\hat\beta^{(WIV)}$	$3.190_{(407.369)}$	$2.740_{(13753.909)}$	$4.543_{(4064.493)}$	$4.125_{(7558.794)}$	$-0.739_{(697.744)}$

$T=60, n\cdot T=300, h=0.80, g=0.81, c_S=3, c_W=4.5, q_W=0.997$					
$\hat\beta^{(IV)}$	$2.664_{(5.713)}$	$-3.925_{(0.158)}$	$2.913_{(0.150)}$	$-1.971_{(0.136)}$	$0.969_{(0.116)}$
$\hat\beta^{(IWV)}$	$4.890_{(0.043)}$	$-4.009_{(0.015)}$	$3.001_{(0.014)}$	$-1.985_{(0.015)}$	$1.009_{(0.018)}$
$\hat\beta^{(SIV)}$	$3.010_{(5.741)}$	$-3.284_{(119.051)}$	$1.498_{(73.685)}$	$-1.600_{(97.037)}$	$0.202_{(56.082)}$
$\hat\beta^{(WIV)}$	$2.436_{(9.441)}$	$-2.698_{(82.746)}$	$0.592_{(84.643)}$	$-1.446_{(89.175)}$	$-0.046_{(74.180)}$

$T = 80, n \cdot T = 400, h = 0.80, g = 0.81, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.721_{(5.373)}$	$-3.949_{(0.081)}$	$2.960_{(0.071)}$	$-1.983_{(0.074)}$	$0.976_{(0.068)}$
$\hat{\beta}^{(IWV)}$	$4.889_{(0.033)}$	$-4.000_{(0.007)}$	$3.007_{(0.008)}$	$-1.998_{(0.007)}$	$1.007_{(0.007)}$
$\hat{\beta}^{(SIV)}$	$3.038_{(4.949)}$	$-2.998_{(27.380)}$	$1.908_{(28.086)}$	$-1.626_{(23.583)}$	$0.495_{(31.963)}$
$\hat{\beta}^{(WIV)}$	$2.487_{(7.324)}$	$-2.208_{(60.841)}$	$1.410_{(34.009)}$	$-1.150_{(64.874)}$	$0.482_{(50.356)}$

$T = 100, n \cdot T = 500, h = 0.80, g = 0.81, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.765_{(5.111)}$	$-3.937_{(0.061)}$	$2.966_{(0.055)}$	$-1.973_{(0.050)}$	$0.994_{(0.056)}$
$\hat{\beta}^{(IWV)}$	$4.895_{(0.026)}$	$-3.998_{(0.005)}$	$3.002_{(0.005)}$	$-1.994_{(0.005)}$	$1.006_{(0.005)}$
$\hat{\beta}^{(SIV)}$	$3.171_{(4.081)}$	$-2.752_{(23.800)}$	$2.564_{(21.325)}$	$-1.667_{(20.350)}$	$0.937_{(21.342)}$
$\hat{\beta}^{(WIV)}$	$2.540_{(6.525)}$	$-2.303_{(31.289)}$	$1.975_{(27.224)}$	$-1.483_{(26.282)}$	$0.725_{(26.367)}$

Tab. 15. The contamination by outliers on the level of 20 %, the orthogonality condition was broken, $n = 5$ (see section 2.1 up to 2.4.). The values of variance of the disturbances randomly selected from the interval $[0.5, 3.5]$.

Similar results as in Table 15 were achieved for the same framework but with the variance of disturbances uniformly randomly selected from the interval $[0.5, 5.5]$. For the low level of heteroscedasticity ($\sigma \in [0.5, 1.5]$) IWV are able – more or less – to cope even with high level of contamination of 20 % but the other estimators have serious problems. With increasing level of heteroscedasticity even IWV started to have some problems, nevertheless for large sample size ($n \geq 300$) IWV still gives some relatively good results with acceptable MSE. For high level of heteroscedasticity ($\sigma \in [0.5, 5.5]$) all estimators give a bit biased estimates. Such situations will require some additional numerical study which e. g. we can try to rid of the heteroscedasticity by estimating it, see [41].

Now, let us turn to the results of simulations when contamination was caused by leverage points. For the sake of space let us focus directly on the case when disturbances were heteroscedastic.

$T = 5, n \cdot T = 100, h = 0.94, g = 0.95, c_S = 6, c_W = 7, q_W = 0.996$					
$\hat{\beta}^{(IV)}$	$5.018_{(1.169)}$	$-2.897_{(3.809)}$	$2.895_{(1.627)}$	$-1.622_{(1.893)}$	$0.851_{(2.068)}$
$\hat{\beta}^{(IWV)}$	$5.007_{(0.149)}$	$-4.005_{(0.321)}$	$2.937_{(0.406)}$	$-2.043_{(0.346)}$	$0.922_{(0.454)}$
$\hat{\beta}^{(SIV)}$	$5.020_{(0.155)}$	$-4.016_{(0.297)}$	$2.922_{(0.350)}$	$-2.063_{(0.286)}$	$0.896_{(0.344)}$
$\hat{\beta}^{(WIV)}$	$5.017_{(0.160)}$	$-4.013_{(0.327)}$	$2.925_{(0.390)}$	$-2.062_{(0.337)}$	$0.896_{(0.374)}$

$T = 10, n \cdot T = 200, h = 0.94, g = 0.95, c_S = 7, c_W = 7, q_W = 0.996$					
$\hat{\beta}^{(IV)}$	$4.972_{(0.675)}$	$-3.029_{(2.935)}$	$2.942_{(1.516)}$	$-1.573_{(1.680)}$	$1.016_{(1.574)}$
$\hat{\beta}^{(IWV)}$	$4.998_{(0.073)}$	$-4.032_{(0.219)}$	$2.936_{(0.203)}$	$-2.018_{(0.207)}$	$0.949_{(0.191)}$

$\hat{\beta}^{(SIV)}$	$4.991_{(0.070)}$	$-4.038_{(0.226)}$	$2.922_{(0.217)}$	$-2.019_{(0.221)}$	$0.957_{(0.180)}$
$\hat{\beta}^{(WIV)}$	$4.990_{(0.075)}$	$-4.043_{(0.268)}$	$2.927_{(0.256)}$	$-2.035_{(0.321)}$	$0.944_{(0.222)}$

$\boldsymbol{T = 15, n \cdot T = 300, h = 0.94, g = 0.95, c_S = 7, c_W = 7, q_W = 0.996}$					
$\hat{\beta}^{(IV)}$	$5.024_{(0.542)}$	$-2.896_{(3.230)}$	$2.971_{(1.504)}$	$-1.567_{(1.916)}$	$0.999_{(1.608)}$
$\hat{\beta}^{(IWV)}$	$4.993_{(0.042)}$	$-3.987_{(0.172)}$	$2.977_{(0.160)}$	$-2.007_{(0.167)}$	$0.981_{(0.149)}$
$\hat{\beta}^{(SIV)}$	$4.990_{(0.044)}$	$-3.991_{(0.164)}$	$2.963_{(0.156)}$	$-2.016_{(0.160)}$	$0.974_{(0.171)}$
$\hat{\beta}^{(WIV)}$	$4.990_{(0.047)}$	$-4.019_{(0.229)}$	$2.952_{(0.224)}$	$-2.024_{(0.199)}$	$0.959_{(0.230)}$

$\boldsymbol{T = 20, n \cdot T = 400, h = 0.94, g = 0.95, c_S = 6.5, c_W = 6, q_W = 0.995}$					
$\hat{\beta}^{(IV)}$	$4.992_{(0.407)}$	$-2.865_{(3.313)}$	$3.003_{(1.475)}$	$-1.432_{(1.695)}$	$0.876_{(1.550)}$
$\hat{\beta}^{(IWV)}$	$4.997_{(0.034)}$	$-4.023_{(0.138)}$	$2.967_{(0.153)}$	$-2.008_{(0.147)}$	$0.977_{(0.139)}$
$\hat{\beta}^{(SIV)}$	$5.000_{(0.031)}$	$-4.043_{(0.157)}$	$2.960_{(0.143)}$	$-2.012_{(0.147)}$	$0.972_{(0.136)}$
$\hat{\beta}^{(WIV)}$	$4.999_{(0.034)}$	$-4.037_{(0.215)}$	$2.959_{(0.210)}$	$-2.015_{(0.207)}$	$0.961_{(0.199)}$

$\boldsymbol{T = 25, n \cdot T = 500, h = 0.94, g = 0.95, c_S = 6.5, c_W = 5, q_W = 0.994}$					
$\hat{\beta}^{(IV)}$	$4.999_{(0.284)}$	$-2.830_{(3.054)}$	$2.958_{(1.293)}$	$-1.474_{(1.607)}$	$0.990_{(1.319)}$
$\hat{\beta}^{(IWV)}$	$4.998_{(0.023)}$	$-4.014_{(0.141)}$	$2.989_{(0.134)}$	$-1.991_{(0.125)}$	$0.982_{(0.121)}$
$\hat{\beta}^{(SIV)}$	$5.001_{(0.023)}$	$-4.023_{(0.151)}$	$2.993_{(0.124)}$	$-1.985_{(0.131)}$	$0.990_{(0.109)}$
$\hat{\beta}^{(WIV)}$	$5.002_{(0.025)}$	$-4.043_{(0.226)}$	$2.970_{(0.193)}$	$-1.989_{(0.196)}$	$1.003_{(0.165)}$

Tab. 16. The contamination by 5 % of leverage points, the orthogonality condition was broken, $n = 20$ (see section 2.1 up to 2.4.). The values of variance of the disturbances randomly selected from the interval $[0.5, 5.5]$.

Similarly as in the case when the contamination was caused by outliers our estimators are able to cope with the contamination by leverage points for the levels of 5 %, respectively, quite well. However the contamination on level 20 % starts to be a problem. That is why we give these results separately for three levels of heteroscedasticity.

$T = 20, n \cdot T = 100, h = 0.58, g = 0.59, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.718_{(44.623)}$	$-3.267_{(20.595)}$	$1.995_{(58.747)}$	$-1.734_{(18.512)}$	$1.787_{(230.342)}$
$\hat{\beta}^{(IWV)}$	$4.995_{(0.050)}$	$-3.955_{(0.104)}$	$3.033_{(0.184)}$	$-1.962_{(0.159)}$	$1.036_{(0.156)}$
$\hat{\beta}^{(SIV)}$	$5.700_{(78.064)}$	$-0.390_{(6546.254)}$	$5.057_{(2128.281)}$	$-5.040_{(11608.136)}$	$3.045_{(2285.545)}$
$\hat{\beta}^{(WIV)}$	$1.147_{(2459.371)}$	$-2.043_{(125.595)}$	$-1.004_{(2999.820)}$	$-2.401_{(500.419)}$	$-2.878_{(11236.357)}$

$T = 40, n \cdot T = 200, h = 0.69, g = 0.70, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.749_{(12.503)}$	$-3.538_{(1.214)}$	$2.600_{(1.397)}$	$-1.690_{(1.136)}$	$0.828_{(1.002)}$
$\hat{\beta}^{(IWV)}$	$4.995_{(0.010)}$	$-3.996_{(0.009)}$	$3.007_{(0.008)}$	$-1.995_{(0.010)}$	$1.015_{(0.009)}$
$\hat{\beta}^{(SIV)}$	$5.022_{(2.068)}$	$-3.884_{(5.831)}$	$2.994_{(4.365)}$	$-1.982_{(3.615)}$	$1.201_{(28.764)}$
$\hat{\beta}^{(WIV)}$	$3.183_{(14.375)}$	$-3.984_{(54.870)}$	$1.869_{(149.714)}$	$-1.839_{(8.457)}$	$0.631_{(18.723)}$

$T = 60, n \cdot T = 300, h = 0.80, g = 0.81, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.727_{(9.674)}$	$-3.637_{(0.473)}$	$2.742_{(0.426)}$	$-1.863_{(0.341)}$	$0.894_{(0.347)}$
$\hat{\beta}^{(IWV)}$	$4.992_{(0.007)}$	$-4.001_{(0.004)}$	$2.999_{(0.003)}$	$-1.997_{(0.003)}$	$1.001_{(0.003)}$
$\hat{\beta}^{(SIV)}$	$4.969_{(0.238)}$	$-3.876_{(15.808)}$	$3.075_{(35.115)}$	$-2.084_{(30.430)}$	$0.961_{(1.528)}$
$\hat{\beta}^{(WIV)}$	$3.085_{(17.244)}$	$-3.422_{(74.521)}$	$2.965_{(79.514)}$	$-1.197_{(396.247)}$	$2.140_{(730.823)}$

$T = 80, n \cdot T = 400, h = 0.80, g = 0.81, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.705_{(8.226)}$	$-3.713_{(0.290)}$	$2.815_{(0.292)}$	$-1.868_{(0.225)}$	$0.916_{(0.197)}$
$\hat{\beta}^{(IWV)}$	$4.996_{(0.004)}$	$-4.001_{(0.002)}$	$3.000_{(0.002)}$	$-2.001_{(0.002)}$	$1.000_{(0.002)}$
$\hat{\beta}^{(SIV)}$	$4.922_{(3.138)}$	$-4.132_{(3.583)}$	$2.638_{(63.278)}$	$-2.490_{(115.262)}$	$0.606_{(79.288)}$
$\hat{\beta}^{(WIV)}$	$2.929_{(6.441)}$	$-3.819_{(1.468)}$	$2.755_{(2.121)}$	$-2.054_{(1.516)}$	$0.843_{(1.717)}$

$T = 100, n \cdot T = 500, h = 0.80, g = 0.81, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.807_{(6.898)}$	$-3.735_{(0.224)}$	$2.824_{(0.151)}$	$-1.877_{(0.166)}$	$0.899_{(0.148)}$
$\hat{\beta}^{(IWV)}$	$5.000_{(0.004)}$	$-4.001_{(0.001)}$	$3.001_{(0.001)}$	$-2.001_{(0.001)}$	$1.002_{(0.001)}$
$\hat{\beta}^{(SIV)}$	$5.003_{(0.008)}$	$-4.049_{(0.197)}$	$2.968_{(0.209)}$	$-2.020_{(0.176)}$	$0.949_{(0.218)}$
$\hat{\beta}^{(WIV)}$	$2.976_{(5.717)}$	$-3.881_{(2.252)}$	$2.946_{(1.833)}$	$-1.923_{(4.867)}$	$0.923_{(2.226)}$

Tab. 17. Contamination – 20 % of leverage points, the orthogonality condition – broken, $n = 5$. The values of variance of the disturbances randomly selected from the interval $[0.5, 1.5]$.

Similarly as above we made for this framework the simulations for some other levels of contaminations which confirmed that the results are plausible in the same way as in Table 18.

$T = 20, n \cdot T = 100, h = 0.58, g = 0.59, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.944_{(38.147)}$	$-2.682_{(39.809)}$	$2.150_{(22.508)}$	$-1.705_{(23.816)}$	$0.887_{(16.748)}$
$\hat{\beta}^{(IWV)}$	$4.962_{(0.185)}$	$-3.702_{(1.978)}$	$3.049_{(0.699)}$	$-1.856_{(0.668)}$	$1.131_{(0.925)}$
$\hat{\beta}^{(SIV)}$	$4.841_{(6.915)}$	$-2.619_{(460.734)}$	$3.032_{(102.980)}$	$-3.074_{(511.159)}$	$-0.441_{(1317.492)}$
$\hat{\beta}^{(WIV)}$	$2.332_{(297.376)}$	$-0.625_{(1421.285)}$	$1.739_{(668.655)}$	$1.063_{(5434.978)}$	$0.606_{(1981.772)}$
$T = 40, n \cdot T = 200, h = 0.69, g = 0.70, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.856_{(12.802)}$	$-3.451_{(1.552)}$	$2.630_{(1.185)}$	$-1.721_{(1.053)}$	$0.840_{(1.125)}$
$\hat{\beta}^{(IWV)}$	$4.973_{(0.048)}$	$-3.945_{(0.093)}$	$3.034_{(0.074)}$	$-1.946_{(0.147)}$	$1.041_{(0.174)}$
$\hat{\beta}^{(SIV)}$	$4.984_{(1.708)}$	$-3.931_{(21.026)}$	$2.887_{(46.724)}$	$-2.059_{(29.152)}$	$0.890_{(13.403)}$
$\hat{\beta}^{(WIV)}$	$3.118_{(30.149)}$	$-3.589_{(18.656)}$	$1.885_{(175.413)}$	$-2.040_{(12.851)}$	$0.859_{(14.252)}$
$T = 60, n \cdot T = 300, h = 0.80, g = 0.81, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.822_{(8.650)}$	$-3.660_{(0.546)}$	$2.735_{(0.427)}$	$-1.853_{(0.327)}$	$0.902_{(0.382)}$
$\hat{\beta}^{(IWV)}$	$4.995_{(0.027)}$	$-3.997_{(0.011)}$	$3.004_{(0.012)}$	$-1.998_{(0.012)}$	$1.011_{(0.012)}$
$\hat{\beta}^{(SIV)}$	$5.048_{(0.556)}$	$-4.198_{(9.033)}$	$2.785_{(8.295)}$	$-2.441_{(16.643)}$	$0.443_{(20.833)}$
$\hat{\beta}^{(WIV)}$	$3.110_{(9.403)}$	$-4.181_{(45.182)}$	$2.437_{(42.881)}$	$-2.881_{(291.920)}$	$0.522_{(71.068)}$
$T = 80, n \cdot T = 400, h = 0.80, g = 0.81, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.731_{(8.279)}$	$-3.736_{(0.347)}$	$2.795_{(0.282)}$	$-1.860_{(0.245)}$	$0.877_{(0.217)}$
$\hat{\beta}^{(IWV)}$	$4.979_{(0.018)}$	$-4.000_{(0.007)}$	$3.001_{(0.008)}$	$-1.993_{(0.008)}$	$0.995_{(0.007)}$
$\hat{\beta}^{(SIV)}$	$5.012_{(0.261)}$	$-4.112_{(1.236)}$	$2.888_{(2.719)}$	$-2.103_{(2.628)}$	$0.852_{(1.419)}$
$\hat{\beta}^{(WIV)}$	$2.949_{(6.615)}$	$-3.789_{(10.501)}$	$2.703_{(2.540)}$	$-1.999_{(2.796)}$	$1.011_{(11.101)}$
$T = 100, n \cdot T = 500, h = 0.80, g = 0.81, c_S = 3, c_W = 4.5, q_W = 0.997$					
$\hat{\beta}^{(IV)}$	$2.901_{(7.015)}$	$-3.767_{(0.187)}$	$2.812_{(0.164)}$	$-1.881_{(0.143)}$	$0.961_{(0.134)}$
$\hat{\beta}^{(IWV)}$	$4.982_{(0.017)}$	$-3.994_{(0.005)}$	$2.997_{(0.005)}$	$-1.997_{(0.005)}$	$0.996_{(0.005)}$
$\hat{\beta}^{(SIV)}$	$4.993_{(0.039)}$	$-4.139_{(0.802)}$	$2.890_{(0.755)}$	$-2.149_{(0.735)}$	$0.917_{(0.550)}$
$\hat{\beta}^{(WIV)}$	$3.044_{(5.943)}$	$-4.110_{(1.721)}$	$2.777_{(2.034)}$	$-2.093_{(1.678)}$	$0.876_{(1.426)}$

Tab. 18. Contamination – 20 % of leverage points, the orthogonality condition – broken, $n = 5$. The values of variance of the disturbances randomly selected from the interval $[0.5, 3.5]$.

5. CONCLUSIONS OF SIMULATION STUDIES

Let us recall once again that $\hat{\beta}^{(IWV,n,w)}$, $\hat{\beta}^{(SIV,\rho,n)}$ and $\hat{\beta}^{(WIV,\rho,q_W,n)}$ do not need any studentization of residuals. It is of course an advantage with respect to M-estimators because the studentization need not be simple task. Of course, we pay for it by the complexity of proofs (see Part I of this paper or [34]). But it is a disadvantage which is not important for the potential user. The complexity of proofs has its roots in the fact that all these estimators estimate implicitly scale of disturbances. However they do it by different ways and moreover S- and W-estimators estimate also the covariance matrix of data which causes that they may have problems in some situation when data contain except of contamination also good leverage points. For all these estimators we have codes in different languages and the speed of their computation is very high. It allows to accommodate the weight function (for LWS and IWV) as well as c_S, c_W and q_W (for S- and W-estimators, respectively) for given data. We start with conservatively selected values and we decrease robustness of estimator in repeated computation of estimator so long when some break of estimated regression coefficients appears. In other words, we pursuit *forward search*, just as in [3], and we so attain maximal possible efficiency of estimator for given contamination of data (for an example with economic data see [35]). In the case when data are not contaminated we reach (or nearly reach) the efficiency of the *ordinary least squares*, see [40]. In [22] one can find an attempt to solve the problem of optimality of estimator also by theoretical tools.

The results demonstrate that *instrumental weighted variables* are in the sense of MSE very well comparable with SIV- and WIV-estimators and for some kind of data – when there are not only outliers and/or bad leverage points but also good leverage points (see Figure 2), they give (much) better results than their competitors (see Table 10 and 11). In other words, the numerical study "discovered" the fact that in the presence of good leverage points some robust estimators which can have larger problems with the outliers than with the leverage points. The explanation is as follows:

Any estimator which estimates (implicitly or explicitly) standard deviation (or more generally, covariance matrix of data, including response and explanatory variables) will consider good leverage points in the upper right corner dangerous while the influence of group of outliers need not be depressed. If we increase the robustness of estimator – to depress the fateful influence of outliers, the influence of good leverage points is also depressed. In other words, we loose the information brought by the good leverage points and consequently MSE of the estimator increases.

6. ACKNOWLEDGEMENT

We would like to thank two anonymous referees for valuable comments and suggestions how to improve the quality of paper and for recommendations od suitable estimators for the comparison with our proposal. The paper was written with the support of the Czech Science Foundation project No. P402/12/G097 "DYME – Dynamic Models in Economics".

(Received July 28, 2014)

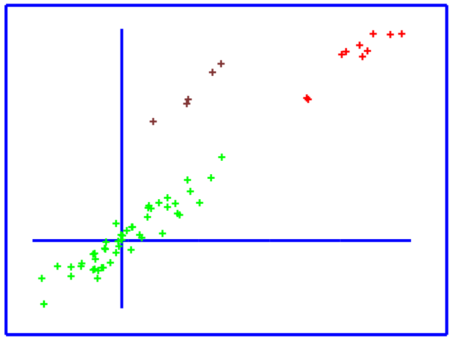

Fig. 2. The data which can (not necessarily do) cause problems to the estimator which – explicitly or implicitly – estimate covariances of data.

REFERENCES

[1] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey: Robust Estimates of Location: Survey and Advances. Princeton University Press, Princeton 1972.

[2] J. Antoch and J. Á. Víšek: Robust estimation in linear models and its computational aspects. In: Contributions to Statistics: Computational Aspects of Model Choice (J. Antoch, ed.), Springer Verlag 1992, pp. 39–104. DOI:10.1007/978-3-642-99766-2_4

[3] A. C. Atkinson and M. Riani: Robust Diagnostic Regresion Analysis. Springer Verlag, New York 2000. DOI:10.1007/978-1-4612-1160-0

[4] O. Barndorff-Nielsen and D. R. Cox: Edgeworth and saddle-point approximations with statistical applications. J. Royal Statist. Soc. *B 41* (1979), 3, 279–312.

[5] V. Benáček and J. Á. Víšek: Determining factors of trade specialization and growth of a small economy in transition. Impact of the EU opening-up on Czech exports and imports. IIASA, IR series, no. IR-03-001, 2002, pp. 1–41.

[6] P. J. Bickel: One-step Huber estimates in the linear model. JASA *70* (1975), 428–433. DOI:10.2307/2285834

[7] N. A. Campbell, H. P. Lopuhaa, and P. J. Rousseeuw: On calculation of a robust *S*-estimator if a covariance matrix. Statist. Medcine *17* (1998), 2685–2695. DOI:10.1002/(sici)1097-0258(19981215)17:23¡2685::aid-sim35¿3.3.co;2-n

[8] S. Chatterjee and A. S. Hadi: Sensitivity Analysis in Linear Regression. J. Wiley and Sons, New York 1988. DOI:10.1002/9780470316764

[9] P. Čížek: Generalized Method of Trimmed Moments. Center Discussion Paper 2009-25, Tilburg University 2009. DOI:10.2139/ssrn.1382807

[10] G. V. Cohen-Freue, H. Ortiz-Molina, and R. H. Zamar: Natural robustification of the ordinary instrumental variables estimator. Biometrics *69* (2013), 641–650. DOI:10.1111/biom.12043

[11] R. Desborges and V. Verardi: A robust instrumental-variable estimator. The Stata J. *12* (2012), 169–181.

[12] F. Y. Edgeworth: On observations relating to several quantities. Phil. Mag. (5th Series) *24* (1887), 222.

[13] C. A. Field and E. M. Ronchetti: Small Sample Asymptotics. Institute of Mathematical Statistics Monograph Series, Hayward 1990.

[14] D. Gervini V. J. and Yohai: A class of robust and fully efficient regression estimators. Annal Statist. *30* (2002), 583–616. DOI:10.1214/aos/1021379866

[15] P. R. Halmos: Applied mathematics is a bad mathematics. In: Mathematics Tomorrow (L. Steen, ed.), Springer Verlag, New York 1981, pp. 9–20. DOI:10.1007/978-1-4613-8127-3_2

[16] J. Jurečková: Robust estimation of location and regression parameters and their second order asymptotic relations. In: 9th Prague Conf. on Inform. Theory, Reidel 1983, pp. 19–32. DOI:10.1007/978-94-009-7013-7_2

[17] J. Jurečková and P. K. Sen: Uniform second order asymptotic linearity of M-statistics in linear models. Statistics and Decisions *7* (1989), 263–276. DOI:10.1524/strm.1989.7.3.263

[18] J. Jurečková and M. Malý: The asymptotics for studentized k-step M-estimators of location. Sequential Analysis *14* (1995), 229–245. DOI:10.1080/07474949508836334

[19] J. Jurečková and A. H. Welsh: Asymptotic relations between L- and M-estimators in the linear model. Ann. Inst. Statist. Math. *42* (1990), 671–698. DOI:10.1007/bf02481144

[20] R. Koenker and G. Bassett: Regression quantiles. Econometrica *46* (1978), 33–50. DOI:10.2307/1913643

[21] R. A. Maronna and V. J. Yohai: Asymptotic behaviour of general M-estimates for regression and scale with random carriers. Z. Wahrscheinlichkeitstheorie verw. Gebiete *58* (1981), 7–20. DOI:10.1007/bf00536192

[22] L. Mašíček: Optimality of the least weighted squares estimator. Kybernetika *40* (2004), 715–734.

[23] J. Robinson: Saddle point approximations for permutation tests and confidence intervals. J. Roy. Statist. Soc. *B 44* (1982), 1, 91–101.

[24] P. J. Rousseeuw: Multivariate estimation with high breakdown point. In: Proc. Fourth Pannonian Symposium on Mathematical Statistics, Bad Tatzmannsdorf 1983, pp. 283–297. DOI:10.1007/978-94-009-5438-0_20

[25] P. J. Rousseeuw: Least median of square regression. J. Amer. Statist. Association *79* (1984), 871–880. DOI:10.1080/01621459.1984.10477105

[26] P. J. Rousseeuw and A. M. Leroy: Robust Regression and Outlier Detection. J. Wiley and Sons, New York 1987. DOI:10.1002/0471725382

[27] P. J. Rousseeuw and V. Yohai: Robust regressiom by means of S-estimators. In: Robust and Nonlinear Time Series Analysis (J. Franke, W. Hărdle, and R. D. Martin, eds.), Lecture Notes in Statistics No. 26, Springer Verlag, New York 1984, pp. 256–272. DOI:10.1007/978-1-4615-7821-5_15

[28] A. M. Rubio and J. Á. Víšek: A note on asymptotic linearity of M-statistics in non-linear models. Kybernetika *32* (1996), 353–374. DOI:10.1016/s0020-7462(96)00064-9

[29] A. F. Siegel: Robust regression using repeated medians. Biometrica *69* (1982), 242–244. DOI:10.1093/biomet/69.1.242

[30] V. Verardi and A. McCathie: The S-estimator of multivariate location and scatter in Stata. The Stata J. *12* (2012), 299–307.

[31] A. H. Welsh: Bahadur representation for robust scale estimators based on regression residuals. Ann. Statist. *14* (1986), 1246–1251. DOI:10.1214/aos/1176350064

[32] J. Á. Víšek: On second order efficiency of a robust test and approximations of its error probabilities. Kybernetika *8* (1983), 3, 387–407.

[33] J. Á. Víšek: Estimating contamination level. In: Proc. Fifth Pannonian Symposium on Mathematical Statistics, Visegrád 1985, pp. 401–414.

[34] J. Á. Víšek: Robust instruments. In: Robust'98 (J. Antoch and G. Dohnal, eds.), Union of Czechoslovak Mathematicians and Physicists 1998, pp. 195–224.

[35] J. Á. Víšek: The robust regression and the experiences from its application on estimation of parameters in a dual economy. In: Proc. Conference Macromodels'99, Wroclaw University 1999, pp. 424–445.

[36] J. Á. Víšek: Regression with high breakdown point. In: ROBUST 2000 (J. Antoch and G. Dohnal, eds.), The Union of the Czech Mathematicians and Physicists and the Czech Statistical Society 2001, pp. 324–356.

[37] J. Á. Víšek: Character of the Czech economy in transition. In: Proc. The Czech society on the break of the third millennium, Karolinum, Publishing House of the Charles University 2000, pp. 181–205.

[38] J. Á. Víšek: The least trimmed squares. Part I – Consistency. Kybernetika *42* (2006), 1–36.

[39] J. Á. Víšek: Robust error-term-scale estimate. IMS Coll., Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: Festschrift for Jana Jurečková *7* (2010), 254–267.

[40] J. Á. Víšek: Least weighted squares with constraints and under heteroscedasticity. Bull. Czech Econometr. Soc. *20* (2012), 31, 21–54.

[41] J. M. Wooldridge: Introductory Econometrics. A Modern Approach. MIT Press, Cambridge 2006. Second edition 2009.

Jan Ámos Víšek, Department of Macroeconomics and Econometrics, Faculty of Social Sciences, Charles University, Opletalova 26, Praha 1, the Czech Republic.
e-mail: visek@fsv.cuni.cz