Journal of Applied Mathematics, Statistics and Informatics (JAMSI), **5** (2009), No. 2

# LEAST WEIGHTED SQUARES IN ECONOMETRIC APPLICATIONS

JAN KALINA

**Abstract**

The least weighted squares (LWS) regression is a robust method for estimating parameters in linear regression models. It is asymptotically efficient for normal models while very robust for contaminated models. This paper broadens the scope of applications of the least weighted squares, mainly to econometrics. The idea of downweighting less reliable observations is applied to propose a modification for the instrumental variables estimator and to study a robust version of the correlation coefficient. Further we supply the least weighted squares regresssion by an asymptotic test of heteroscedasticity of the random errors. Finally computational aspects of the LWS estimator are studied.

**Mathematics Subject Classification 2000**: 62G35, 62J20, 62P20
**General Terms**: Robust statistics, Instrumental variables, Heteroscedasticity, Computational aspects
**Additional Key Words and Phrases**: robust regression, regression diagnostics, correlation analysis

## 1. LEAST WEIGHTED SQUARES REGRESSION

The least weighted squares (LWS) regression is a robust method with a high breakdown point proposed by Víšek (2001) to estimate the parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ in the linear regression model

$$Y_i = \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + e_i, \quad i = 1, \ldots, n, \tag{1}$$

or in the standard matrix notation $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$. The $i$-th row of $\boldsymbol{X}$ is the vector $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^T$ corresponding to the $i$-th observation for $i = 1, \ldots, n$.

The idea of the least weighted squares is to downweight less reliable data points. We choose sizes of the non-negative weights $w_1, w_2, \ldots, w_n$ at first, while these weights are assigned to particular observations after a permutation, which is determined automatically only during the computation based on the residuals. Let us consider (any) estimate $\boldsymbol{b} = (b_1, \ldots, b_p)^T \in \mathbb{R}^p$ of the parameter $\boldsymbol{\beta}$. By

$$u_i(\boldsymbol{b}) = y_i - b_1 X_{i1} - \cdots - b_p X_{ip}, \quad i = 1, \ldots, n, \tag{2}$$

we denote the residual corresponding to the $i$-th observation. Let us order squared residuals

$$u^2_{(1)}(\boldsymbol{b}) \leq u^2_{(2)}(\boldsymbol{b}) \leq \cdots \leq u^2_{(n)}(\boldsymbol{b}). \tag{3}$$

This research is supported by the grant 402/09/0557 (Robustification of selected econometric methods) of the Grant Agency of the Czech Republic.

The LWS estimator of $\boldsymbol{\beta}$ is defined as the argument of minimum of

$$\sum_{i=1}^{n} w_i u_{(i)}^2(\boldsymbol{b}) \tag{4}$$

over $\boldsymbol{b} \in \mathbb{R}^p$. The least trimmed squares (LTS) proposed by Rousseeuw and Leroy (1987) represent a special case of the least weighted squares with weights equal to zero or one only.

A different choice of weights was proposed by Čížek (2008). In his automatic data-adaptive procedure, the sizes of the weights are computed in two stages. The first stage uses the LWS estimator with fixed sizes of the weights as the initial estimator. In the second stage the empirical distribution function of squared residuals is compared with the theoretical distribution function under normality, which is accompanied by a complete trimming of data exceeding the 0.9999-quantile. Čížek (2008) proved the estimator to have high efficiency for normal errors and also a high breakdown point for contaminated data sets.

Further we broaden the scope of applications of the least weighted squares, namely study the idea of downweighting less reliable observations. Section 2 robustifies the instrumental variables estimator in this spirit. Section 3 studies a robust version of the correlation coefficient and its properties when the data-adaptive weights are used. Section 4 supplies the least weighted squares with fixed sizes of the weights with an asymptotic test of heteroscedasticity. Finally the applied Section 5 studies computational aspects of robust regression, with a focus on the least weighted squares with fixed sizes of the weights.

## 2. ROBUST INSTRUMENTAL VARIABLES ESTIMATION

The classical instrumental variables estimator has become a popular method in econometrics (Judge et al. 1985, Kmenta 1986). In the model (1) it is assumed that the errors $\boldsymbol{e}$ are not uncorrelated with the independent variables, while there is the total number $l$ ($l \geq p$) of instrumental variables available. Let the vector $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{il})^T$, $i = 1, \ldots, n$, of values of the instruments correspond to the $i$-th observation. Let the matrix $\mathbf{Z}$ contain these values using the notation $\mathbf{Z} = (Z_{ij})_{ij}$. The instrumental variables estimator of $\boldsymbol{\beta}$ is based on least squares estimation and therefore is very sensitive to outliers in the response as well in regressors. Before we study its robustification, we formulate a lemma, which can be proven easily.

Lemma 2.1. *For any diagonal matrix* $\mathbf{W}$*, denoting* $\boldsymbol{b} = (\mathbf{Z}^T\mathbf{W}\boldsymbol{X})^{-1}\mathbf{Z}\mathbf{W}\boldsymbol{Y}$*, it holds*

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})^T\mathbf{W}\mathbf{Z}\mathbf{Z}^T\mathbf{W}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}) = 0. \tag{5}$$

Víšek (2006) proposed an approximative algorithm for computing a robust version of the instrumental variables estimator for the special case with $l = p$ as an analogy of the FAST-LTS algorithm, assigning weights to the data based on squared residuals of an initial estimate. The new estimator is compared with the initial one in terms of the loss function

$$(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{b})^T \mathbf{W}\mathbf{Z}\mathbf{Z}^T\mathbf{W}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{b}) = \sum_{j=1}^{l}\left(\sum_{i=1}^{n} w_i Z_{ij} u_i(\boldsymbol{b})\right)^2. \tag{6}$$

However it follows from Lemma 2.1 that the loss (6) equals zero. Actually (6) is not a true loss function allowing to discriminate among different sets of weights and a modification of the FAST-LTS algorithm is not valid.

Therefore we present a completely different approach to robust instrumental variables estimation, which can be computed also for the general case with $l \geq p$. The estimator will be now proposed as a *two-stage procedure.* The first step computes the least weighted squares projection $\hat{\boldsymbol{X}}$ of the independent variables $\boldsymbol{X}$ to the linear space generated by the instruments. In the second step the regression parameters in the linear regression of the response $\boldsymbol{Y}$ against the projections $\hat{\boldsymbol{X}}$ are estimated as the least weighted squares squares estimator. This ensures the resulting estimator of $\boldsymbol{\beta}$ to be robust against outliers in the response, regressors and instruments. More schematically this two-stage estimator is defined in the following way:

*Definition* 2.2. We define the robust instrumental variables estimator as a two-stage method in the following way.

(1) Computing the least weighted squares in the linear regression $\boldsymbol{X} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{v}$ with some parameters $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_l)^T$ and random errors $\mathbf{v} = (v_1, \ldots, v_n)^T$, let the diagonal matrix containing the weights in the correct permutation (optimal permutation determined by the LWS) be denoted by $\mathbf{W}_1$. Let $\hat{\boldsymbol{X}}$ denote the projections obtained by

$$\hat{\boldsymbol{X}} = \mathbf{Z}\hat{\gamma} = \mathbf{Z}(\mathbf{Z}^T\mathbf{W}_1\mathbf{Z})^{-1}(\mathbf{Z}^T\mathbf{W}_1\boldsymbol{X}). \tag{7}$$

(2) Computing the least weighted squares in the linear regressin of $\boldsymbol{Y}$ against $\hat{\boldsymbol{X}}$, let the diagonal matrix containing the weights in the correct permutation (optimal permutation determined by the LWS) be denoted by $\mathbf{W}_2$. The robust IV estimator of $\boldsymbol{\beta}$ is then defined as

$$\boldsymbol{b} = \left(\hat{\boldsymbol{X}}^T\mathbf{W}_2\hat{\boldsymbol{X}}\right)^{-1}\hat{\boldsymbol{X}}^T\mathbf{W}_2\boldsymbol{Y}. \tag{8}$$

*Remark* 2.3. Only in the special case when both least weighted squares procedures assign the same weights to the data, this formula reduces to the explicit formula $(\mathbf{Z}^T\mathbf{W}\boldsymbol{X})^{-1}\mathbf{Z}\mathbf{W}\boldsymbol{Y}$, where $\mathbf{W}$ is the diagonal matrix with the weights on the main diagonal.

Now we show two numerical illustrations. The first data set comes from Arrow et al. (1961), studied also by Kmenta (1986). The response

$$\log\frac{\text{wage}}{\text{price of product}} \tag{9}$$

and the independent variable

$$\log\frac{\text{value added}}{\text{price of product} \quad\cdot\quad \text{labor input}} \tag{10}$$

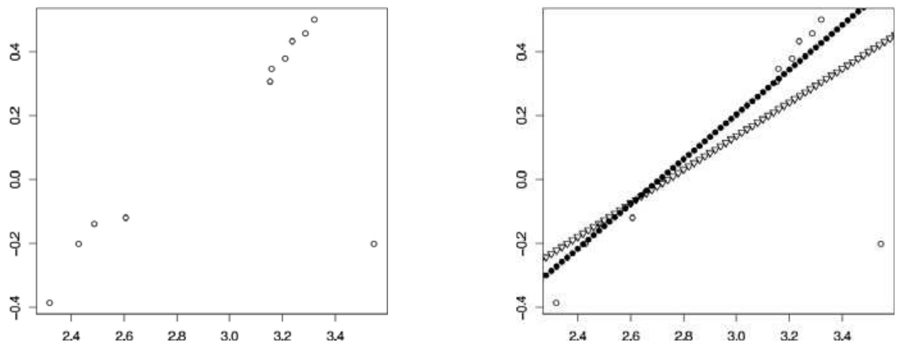


Fig. 1. Left: Furniture data. Right: Fitted values of furniture data by classical instrumental variables estimator (dark circles) and by robust instrumental variables estimator (grey triangles).

are both measured in furniture manufacturing industry in 11 different countries of the world. However to compare the classical and robust approaches, we introduce an outlier, namely we replace the response of the first observation from the value 0.768 to −0.2. The data after this modification are shown in Figure 1 (left), where the outlier can be seen in the right bottom corner of the plot. First the linear regression of the response against the regressor is considered. The least squares estimate of $\boldsymbol{\beta}$ is $(-1.48, 0.540)^T$, while the least weighted squares with data-adaptive weights is different with $(-1.91, 0.706)^T$. Kmenta (1986) warns that additional variables may contribute to the variability of the response, so the disturbances may not be uncorrelated with the regressor. Therefore he uses an instrumental variable, namely the variable (10) measured in knitting mill industry. There is a very nice linear relationship of the regressor against the instrument.

The estimator of $\boldsymbol{\beta}$ using instrumental variables estimation is similar to the least squares estimate computed without instruments, namely $(-1.44, 0.526)^T$. The robust instrumental variables estimator is computed as the two-stage estimator with the least weighted squares in each stage with data-adaptive choice of weights. The estimate of $\boldsymbol{\beta}$ equals $(-1.89, 0.699)^T$, similarly with the LWS estimate without instruments. The classical and robust estimates using instrumental variables are shown in Figure 1 (right), so the robust approach truly brings an improvement over the classical one, which is deceived by the outlier.

The next example shows that the robust instrumental variables estimation must be used with care and that it does not give always a reasonable result even in simple situations. We study the relationship between two macroeconomic time series of yearly data measured in the United States in $10^9$ USD between 1980 and 2001. These investment data shown in Figure 2 (left) come from `www.stls.frb.org/fred` and were studied also by Kalina (2008*b*). The linear regression of real gross private domestic investment (INVESTMENT) against the real gross domestic product (GDP) is considered in the form

$$\mathsf{INVESTMENT}_t = \beta_0 + \beta_1 \mathsf{GDP}_t, \quad t = 1, \dots, n. \tag{11}$$

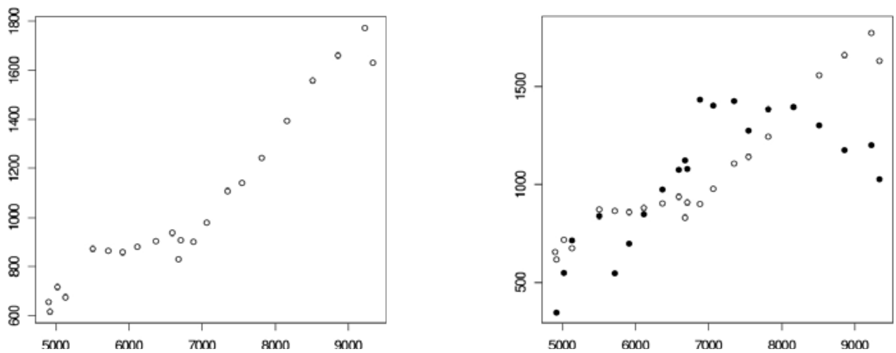


Fig. 2. Left: Investment data. Right: Fitted values of investment data using the robust instrumental variables estimator.

The least weighted squares estimate of $\boldsymbol{\beta}$ computed with linear weights equals $(-438.2, 0.2159)^T$ and is obtained for the weights

$$(16, 10, 19, 20, 6,\ \ 11, 18, 22, 17, 13,\ \ 8, 3, 5, 7, 15,\ \ 14, 21, 9, 4, 2,\ \ 1, 12)^T$$

to be transformed to $\sum_{i=1}^{n} w_i = 1$. Figure 3 (left) shows the scatter plot of the response against the independent variable; the data set apparently does not contain bad outliers.

Let us consider the model

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1}, \quad t = 1, \ldots, n. \tag{12}$$

The least squares estimate of $\boldsymbol{\beta}$ is $(-260, 0.102, 0.611)^T$ and the least weighted squares estimate with data-adaptive weights $(-37.2, 0.140, 0.437)^T$. Now because the disturbances in the model are autocorrelated, it may be desirable to use $X_{t-1}$ as instrument for $Y_{t-1}$ and to retain $X_t$ as an independent variable. In other words $Y_{t-1}$ is replaced by an instrument while $X_t$ is an instrument for itself. This estimator has a chance to be efficient, because the instrument is highly correlated with the regressor $Y_{t-1}$ as the sample correlation between them equals $r = 0.958$.

The robust instrumental variables estimate of $\boldsymbol{\beta}$ is $(-1438, 0.684, -2.211)^T$. This does not estimate the values of the response nicely as shown in Figure 2 (right), although the LWS fit in the model (12) corresponds well to the data. The fitted values are full circles and the pattern fails to follow the structure the data, although the relationship between the regressor and the instrument is nicely linear. The reason is the ill-conditioned design matrix of the linear fit of the second stage of Definition 4.2. Therefore the robust instrumental variables estimator must be handled with care.

## 3. ROBUST CORRELATION

This sections confronts the (classical) weighted correlation coefficient with known weights and the robust correlation coefficient based on least weighted squares. Firstly we evaluate the non-robustness of the weighted correlation. We study the sample weighted correlation coefficient between data vectors $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ with fixed weights $\boldsymbol{w} = (w_1, \ldots, w_n)^T$ satisfying $\sum_{i=1}^{n} w_i = 1$, which is computed as

$$r_W(\boldsymbol{X},\ \boldsymbol{Y};\ \boldsymbol{w}) = \frac{\sum_{i=1}^n w_i(X_i - \bar{X}_W)(Y_i - \bar{Y}_W)}{\sqrt{\sum_{i=1}^n [w_i(X_i - \bar{X}_W)^2] \sum_{j=1}^n [w_j(Y_j - \bar{Y}_W)^2]}}, \tag{13}$$

where $\bar{X}_W = \sum_{i=1}^n w_i X_i$ and $\bar{Y}_W = \sum_{i=1}^n w_i Y_i$ are weighted means. We will need the notation $S_W^2(\boldsymbol{X};\boldsymbol{w})$ for the weighted variance of $\boldsymbol{X}$ defined by

$$S_W^2(\boldsymbol{X};\boldsymbol{w}) = \sum_{i=1}^n w_i (X_i - \bar{X}_W)^2$$

and an analogous notation $S_W^2(\boldsymbol{Y};\boldsymbol{w})$ for the weighted variance of $\boldsymbol{Y}$.

THEOREM 3.1. *Let $r_W(\boldsymbol{X},\boldsymbol{Y};\boldsymbol{w})$ denote the sample weighted correlation coefficient between data vectors $\boldsymbol{X} = (X_1,\ldots,X_n)^T$ and $\boldsymbol{Y} = (Y_1,\ldots,Y_n)^T$ with fixed weights $\boldsymbol{w} = (w_1,\ldots,w_n)^T$. Let a new observation $(X_{n+1},Y_{n+1})^T$ be added to the sample with a weight $w_{n+1}$. The weighted correlation coefficient for the data $\boldsymbol{X}^* = (X_1,\ldots,X_n,X_{n+1})^T$ and $\boldsymbol{Y}^* = (Y_1,\ldots,Y_n,Y_{n+1})^T$ with weights $\boldsymbol{w}^* = (w_1,\ldots,w_n,w_{n+1})^T$ transformed to $\sum_{i=1}^{n+1} w_i^* = 1$ is equal to*

$$r_W(\boldsymbol{X}^*,\boldsymbol{Y}^*;\boldsymbol{w}^*) = \frac{r_W(\boldsymbol{X},\boldsymbol{Y};\boldsymbol{w}) + uv}{\sqrt{1+u^2}\sqrt{1+v^2}},$$

*where*

$$u = \frac{X_{n+1} - \bar{X}_W}{S_W(\boldsymbol{X};\boldsymbol{w})}\sqrt{\frac{w_{n+1}}{1+w_{n+1}}} \quad \text{and} \quad v = \frac{Y_{n+1} - \bar{Y}_W}{S_W(\boldsymbol{Y};\boldsymbol{w})}\sqrt{\frac{w_{n+1}}{1+w_{n+1}}}.$$

PROOF. This is a weighted analogy of Mustonen (2005). □

*Remark* 3.2. The local robustness of the weighted correlation coefficient can be measured by means of the sample influence function defined as $n[r_W(\boldsymbol{X}^*,\boldsymbol{Y}^*;\boldsymbol{w}^*) - r_W(\boldsymbol{X},\boldsymbol{Y};\boldsymbol{w})]$. An observation with a very different value of the $x$ or $y$ coordinate has a large influence of the estimator and also a large weight $w_{n+1}$ influences the estimator very much. Therefore $r_W(\boldsymbol{X},\boldsymbol{Y};\boldsymbol{w})$ is non-robust with respect to outliers as well as non-robust with respect to the weights. In practice robust correlation measures should be used instead.

Kalina (2007) proposed two correlation measures based on the least weighted squares regression. Now we study properties of such robust correlation, which is based on the LWS regression with the data-adaptive choice of weights (Čížek 2008) computed in the model $Y_i = \beta_0 + \beta_1 X_i, \quad i = 1,\ldots,n$. Let the weights determined by the LWS regression be denoted by $\boldsymbol{w}_{LWS}$. The robust correlation coefficient $r_{LWS}$ based on the LWS regression is now defined as the weighted correlation $r_W$ in (13) with the weights $\boldsymbol{w}_{LWS}$. This is robust for contaminated models and efficient for normal models, which will be now formulated precisely.

THEOREM 3.3. *Let $(X_1,Y_1)^T,\ldots,(X_n,Y_n)^T$ be a sequence of independent identically distributed random vectors, which are almost surely in a general position for $n > p$. Let $\epsilon_n^0$ denote the finite-sample breakdown point of an initial estimator of $\boldsymbol{\beta}$ and $\sigma^2$ in the model (1). Then the finite-sample breakdown point of $r_{LWS}$ is larger than or equal to* $\min\{e_n^0,\ \{[n+1)/2] - (p+1)\}/n\}$.

Proof. This is a consequence of the result of Čížek (2008) for the least weighted squares regression. The breakdown point corresponds to the percentage of data needed to "break down" (Rousseeuw and Leroy 1987). The robust correlation inherits the breakdown of the robust regression fit. □

Theorem 3.4. *Let* $(X_1, Y_1)^T, \ldots, (X_n, Y_n)^T$ *be a sequence of independent identically distributed random vectors, let* $\rho$ *be the true correlation between* $X_1$ *and* $Y_1$ *and let* $e_i \sim \mathsf{N}(0, \sigma^2)$, $i = 1, \ldots, n$. *Let the initial estimator* $\hat{\beta}^0$ *and* $\hat{\sigma}^2$ *of* $\boldsymbol{\beta}$ *and* $\sigma^2$ *in (1) be consistent for* $n \to \infty$. *Then* $r_{LWS}$ *is a consistent estimator of the correlation coefficient* $\rho$ *as* $n \to \infty$.

Proof. This is a consequence of the result of Čížek (2008), who proves that $\lim_{n\to\infty} w_n(t) = 1$ for all $t \in (0,1)$ for the weights defined as values of the weight function $w(t)$ with $t \in (0,1)$. Therefore $r_{LWS}$ is asymptotically equivalent to the (non-weighted) sample correlation coefficient. □

*Remark* 3.5. For normal models without outliers, the robust estimator $r_{LWS}$ has asymptotically the same efficiency as the classical correlation coefficient (without weighting).

*Remark* 3.6. There is another possibility to define a correlation measure based on assigning weights to data in an implicit way. Let us assume the sizes of the weights to be fixed. Let the correlation measure be defined as the maximal weighted correlation coefficient over all possible permutations of the weights to the observations. Clearly this measure has a breakdown equal to zero and is not robust. To prove this, consider a majority of data fluctuating around a constant independently on the regressor, contaminated by a small percentage of data with a larger value of the regressor and a large value of the response.

The robust correlation coefficient $r_{LWS}$ has desirable properties under both normal and contaminated models. This justifies its using in the image analysis applications of Kalina (2007). Further applications can be suggested to robustify the autocorrelation function in the context of financial time series.

## 4. HETEROSCEDASTICITY TESTS FOR THE LEAST WEIGHTED SQUARES

Although robust regression was originally intended as a mere diagnostic tool for ordinary least squares regression to detect outliers in the response and in the independent variables, it requires to be accompanied by diagnostic tools as well. These include tests of heteroscedasticity of the errors $\boldsymbol{e}$ and Durbin-Watson test for independence of the regression errors $\boldsymbol{e}$ in a time series context. Here we present Goldfeld-Quandt test of heteroscedasticity for the least weighted squares with fixed sizes of the weights, which is asymptotically equivalent to the classical test for the ordinary least squares context. Other heteroscedasticity tests for robust regression are presented in Kalina (2008*a*). Durbin-Watson test for the least weighted squares is studied by Kalina (2004).

Golfeld-Quandt test for the model (1) is a test of the null hypothesis of homoscedasticity

$$H_0: \ \mathsf{var}\, \boldsymbol{e} = \sigma^2$$

against the alternative formulated in the specific form

$$H_1 : \ \mathsf{var}\, \boldsymbol{e} = \sigma^2 \cdot \mathsf{diag}\{k_1, \ldots, k_n\}. \tag{14}$$

Constants $k_1, \ldots, k_n$ must be specified before the computation and usually correspond to the values of one of the independent variables or fitted values of the response (Greene 2002 or Cipra 2008).

The test however does not depend on the values $k_1, \ldots, k_n$. It is required to divide the observations to three parts according to the values $k_1, \ldots, k_n$. The sizes of the subsets are however arbitrary and selected subjectively, for example $r \doteq n/3$. The estimates of the parameters in the linear regression models only for the first and third subset will be now computed using the least weighted squares. Let $\boldsymbol{u}^I$ denote the residuals of the regression computed for the first subset, let $\bar{u}^I$ denote their mean and let $\sum_I (u_i^I - \bar{u}^I)^2$ represent the sum of squared centered residuals in the first subset of the data. In an analogous way $\boldsymbol{u}^{III}$, $\bar{u}^{III}$ and $\sum_{III} (u_i^{III} - \bar{u}^{III})^2$ will be denoted using the residuals of the regression computed only for the third subset. Let the first subset contain $r_1$ observations and the third subset $r_3$ observations. The test is based on the test statistic

$$\frac{\sum_{III} (u_i^{III} - \bar{u}^{III})^2}{\sum_I (u_i^I - \bar{u}^I)^2}. \tag{15}$$

There is a difference from the classical least squares statistic, which is based on residual sum of squares comparing $\sum_I (u_i^I)^2$ and $\sum_{III} (u_i^{III})^2$, because the mean of least squares residuals equals zero. The test is based on the following theorem.

Theorem 4.1. *Let the test statistic (15) be computed from the residuals of the least weighted squares regression fits as described above. Assuming normal errors* $\boldsymbol{e}$ *in the model (1), the test statistic (15) has asymptotically* $F_{r_3-p, r_1-p}$ *distribution under the null hypothesis* $H_0$.

Proof. The technical proof is based on the asymptotical behavior of the LWS estimator, its asymptotical linearity and asymptotical representation of the first order, as studied for example in Víšek (2009), using technical lemmas of Kalina (2004) and Kalina (2008*a*). □

## 5. COMPUTATIONAL ASPECTS OF ROBUST REGRESSION

The computation of both the LTS and LWS estimators is intensive and an approximative algorithm must be used already for moderate sample sizes. For the least trimmed squares, the FAST-LTS algorithm (Rousseeuw and van Driessen 1999) is well-established and implemented in the `R` software package, library `robustbase`. We point out that the function `ltsReg()` computes a two-stage reweighted version; Rousseeuw and Leroy (1987) call it reweighted least squares (RLS) and find it useful based on simulations, however without studying its theoretical properties.

The true LTS estimate with $n$ data points retaining $h$ of them and ignoring $n-h$ observations can be computed in this way:

```
library(robustbase)
fit=ltsReg(x,y,alpha=h/n);
h=fit$quan;
```

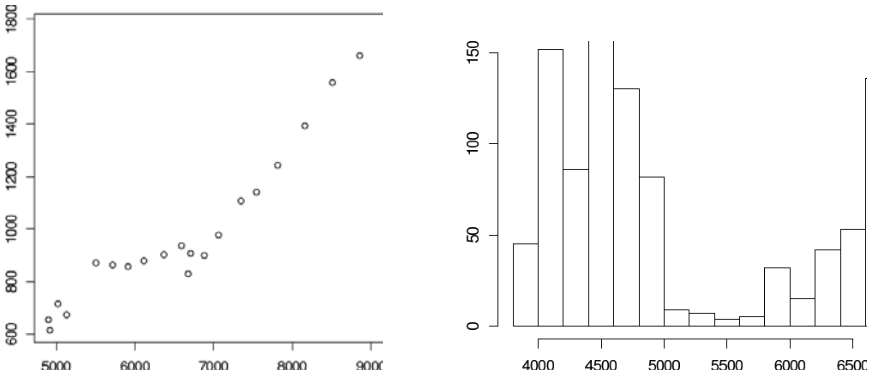


Fig. 3. Left: investment data. Right: loss computed in particular 1000 independent repetitions of the LWS estimator with linear weights.

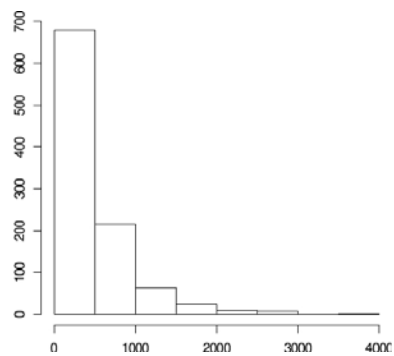


Fig. 4. Number of repetitions of the LWS estimator with linear weights needed to reach the optimal loss.

```
w=rep(0,n);
for (i in 1:h)
  {j=fit$best[i];
   w[j]=1;}
lts=lm(y~x, weights=w);
```

The FAST-LTS algorithm can be modified to compute the LWS estimator with fixed sizes of the weights. We have implemented the algorithm in C++ so the computation of the LTS and LWS estimators is very fast. The algorithm searches for such permutation of fixed weights $w_1, \ldots, w_n$, which minimizes (4). This was described by Kalina (2008*b*). The following argument gives reasons for this algorithm. The sum is minimized by such weights, which correspond to the weighted squared residuals, so the observation with the smallest squared residual obtains the largest weight etc. It follows clearly that the output of this algorithm corresponds to the LWS estimator, because this is defined by minimization of (4).

Just like the FAST-LTS algorithm, the algorithm for the least weighted squares is based on independent repetitions of an iterative procedure, which iterates as long as the loss function keeps improving. Further we make a difference between the repetitions and the iterations within each repetition. While the recommended number of repetitions for the LTS is 10 000, we now examine the number of needed repetitions for the least weighted squares for a numerical example with the investment data from Section 2.

In the linear regression model (11), the minimal loss (4) obtained for the LWS regression with linear weights equals 3910.3. Figure 3 (right) shows the histogram of results of 1000 independent repetitions of the iterative computation; the LWS estimate is then the result corresponding to the minimal value of the loss. About 25 % of the resulting estimates have a loss similar to the least squares and are influenced also by less reliable data. The remaining about 75 % of the data have a remarkably smaller loss. Typically the observations in the years 1991, 1998, 1999 and 2000 have smaller weights; these are the less reliable observations determined by the true solution of the LWS. Therefore 75 % of repetitions are already remarkably more robust than the least squares solution.

Finally we studied the number of repetitions needed to obtain the minimal loss 3910.3. The result of this study is shown in Figure 4 for 1000 independent repetitions. In the mean there are 458 repetitions needed. The number of repetitions is below 1000 for 87.9 % of cases. Therefore the number of 10 000 recommended repetitions is extremely safe and 1000 repetitions would certainly give at least a good approximation to the true solution of the LWS regression in this numerical example.

The modification of the FAST-LTS algorithm for the data-adaptive weights of Čížek (2008) is straightforward, each iteration begins with an initial estimate of $\boldsymbol{\beta}$, the residuals are computed from it, the sizes of the weights are then computed, and these are assigned to particular observations based on the squared residuals. Other computational aspects of the least weighted squares regression are studied by Kalina (2008*b*).

## REFERENCES


Arrow K.J., Chenery H.B., Minhas B.S., Solow R.M. (1961): Capital-labor substitution and economic efficiency. *Rev. Econ. Stat.* **43**, No. 3, 225 – 250.

Cipra T. (2008): Financial econometrics. Ekopress, Prague. In Czech.

Čížek P. (2008): Efficient robust estimation of time-series regression models. *Appl. Math.* **53**, No. 3, 267 – 279.

Greene W.H. (2002): Econometric analysis. Fifth edition. Macmillan, New York.

Judge G.G., Griffiths W.E., Hill R.C., Lutkepohl H., Lee T.C. (1985): The theory and practice of econometrics. Wiley, New York.

Kalina J. (2004): Durbin-Watson test for least weighted squares. In Antoch J. (ed.): *COMPSTAT 2004, Proceedings in computational statistics*, Physica-Verlag, Heidelberg, 1287 – 1294.

Kalina J. (2007): Locating the mouth using weighted templates. *Journal of Applied Mathematics, Statistics and Informatics* **3**, No. 1, 111 – 125.

Kalina J. (2008*a*): Robust regression and diagnostic tools. In Kupka K. (ed.): *Data analysis 2007/II, Progressive methods of statistical data analysis and modelling for research and technical practice*, Trilobyte, Pardubice, 31 – 41.

Kalina J. (2008*b*): Robust GMM estimation. *Forum Statisticum Slovacum* **4**, No. 6, 78 – 83.

Kmenta J. (1986): Elements of econometrics. Macmillan, New York.

Mustonen S. (2005): Influence curves for the correlation coefficient. `http://www.survo.fi/papers/corrcurves.pdf`.

Rousseeuw P.J., Leroy A.M. (1987): Robust regression and outlier detection. Wiley, New York.

Rousseeuw P.J., van Driessen K. (1999): A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41**, 212 – 223.

Víšek J.Á. (2001): Regression with high breakdown point. In Antoch J., Dohnal G. (eds.): *Proceedings of ROBUST 2000, Summer School of JČMF*, JČMF and Czech statistical society, 324 – 356.

Víšek J.Á. (2006): Instrumental weighted variables – algorithm. In Rizzi A., Vichi M. (eds.): *COMPSTAT 2006, Proceedings in computational statistics.* Physica-Verlag, Heidelberg, 777 – 786.

Víšek J.Á. (2009): Consistency of the least weighted squares. Submitted to *Kybernetika.*

Jan Kalina
KPMS MFF UK
Sokolovská 83
CZ-186 75 Praha 8
Czech Republic
e-mail: kalina@karlin.mff.cuni.cz

Received July 2009