KYBERNETIKA — VOLUME 37 (2001), NUMBER 2, PAGES 183 - 204

COMBINING FORECASTS USING THE LEAST TRIMMED SQUARES

JAN ÁMOS VÍŠEK[1]

Employing recently derived asymptotic representation of the least trimmed squares estimator, the combinations of the forecasts with constraints are studied. Under assumption of unbiasedness of individual forecasts it is shown that the combination without intercept and with constraint imposed on the estimate of regression coefficients that they sum to one, is better than others. A numerical example is included to support theoretical conclusions.

1. INTRODUCTION

It is more that thirty year ago when the paper by Bates and Granger [1] opened the question of possible improvement of forecast of process in question by combining k individual forecasts. Bates and Granger proposed to utilize the framework of linear regression model, namely to consider forecasted process as the response variable and the individual forecasts as explanatory ones.

Let us briefly discuss a background of this endeavour. One can trace out in the background the reasons of various types. What concerns heuristics they can be given as follows:

- Firstly, if k forecasters used efficiently information which they had at hand and if their information was at least partially disjunct, we may try by combining their forecasts to employ all the information simultaneously and hence (hopefully) to obtain better forecast.
- Secondly, if they used the information inefficiently and if the inefficiency did not "happened" in the same way by all of them, then, even in the case when all the forecasters had the same information, we may hope in an improvement of efficiency.

Of course, the real situation will be somewhere in-between these two possibilities.

But the task represents also very interesting theoretical challenge. Let us assume (or imagine, if you wish) that the forecasted process is "in fact" generated by a (dynamic) linear regression model with normally distributed random noise. The

[1] Research was supported by Grant 255/2000/A EK /FSV of the Grant Agency of Charles university.

converted commas around *in fact* indicate that usually we have no idea about a "mechanism" (if any[1]) which generates the forecasted process. Let us denote the design matrix of the respective model by X_t. We may then interpret the matrix X_t as the information accompanying the forecasted process, so that each single column represents one explanatory variable. Then, of course we may find the *best possible* prediction of the forecasted process by means of estimating corresponding regression coefficients (the *best possible* under given circumstances, i. e. e. g. in the case when we find that the data are contaminated – see e. g. Rubio and Víšek [10] and on the other hand the specification test indicates correlation between explanatory variables and disturbances – see Víšek [16], we can use robust version of instrumental variables – see Víšek [15] or [19] or in the case when the design matrix exhibits a collinearity we may use M-estimators or the least trimmed squares subject to some (linear) constraints – see Rubio et al [9] or Víšek [21], etc.). But the design matrix X_t is available neither to us nor to the individual forecasters. Nevertheless, let us assume that each forecaster has at hand some columns of this matrix, i. e. in other words, each of them has at hand some part of explanatory information. And we may assume that they have together the whole relevant information. Of course, sometimes this last assumption need not be realistic.

Assuming moreover that (typically) the common criterion of a quality of prediction for all the forecasters is the minimum of the sum of squared errors of prediction, the question appears:

Having at hand k individual forecasts, is it possible to reconstruct – not necessarily by linear combination - the best possible forecast?

Finally, the task of improving forecast has also its practical meaning. The fact that there is available several forecasts of process in question indicates that the forecasted process is of considerable importance at given time and place. But it hints:

Even a small improvement of quality of forecast may be appreciated a lot.

Since 1969, when the paper by Bates and Granger appeared, a large attention was devoted to the problems which of the many types of regression model should be used. In other words, the questions of the type:

- Should some constraints be imposed on the coefficients of the model or not?
- Should the intercept be included or not?
- Should the coefficients be considered stable or moving in time and how?

etc. were intensively studied. Less attention was paid to the problems of how the coefficients are to be estimated (in the cases when we decide for combining the forecasts by means of linear regression model). It is quite understandable since the *least squares principle*, sometimes with the maximum likelihood one, are still (unfortunately) nearly exclusive tools of econometrics and the advantages of robust methods did not yet attracted appropriate attention.

It is nowadays already well known that the least squares are extremely vulnerable to influential points, either outliers or leverage points. Moreover, even in the case

[1] Of course, it is a philosophical question, very interesting and in the interpretation of results also very important one, how far the idea that some mechanism generated data, is tenable.

that no influential point is present[2], but the residuals are not normally distributed, the least squares is optimal estimator only in the class of linear estimator. And unfortunately the restriction on the linear estimators is drastic, see e. g. again Hampel et al [4].

So, even today, more than thirty years after the pioneering paper by Bates and Granger the problem is still worthwhile to be studied. Of course, at first we may ask whether the idea to combine the forecasts in the framework of (linear) regression model is the most appropriate one in all situations. It is easy to see that in the case when the forecasted process (considered as vector, i. e. all past values of the forecasted process are taken as coordinates of one vector) is far away from the space generated by linear combinations of individual forecasts, no linear combination of individual forecasts may improve the situation too much. In opposite case the linear combination may give a satisfactory result. But then, admitting that among the values of forecasted process and/or of individual forecasts a portion of atypical points[3] mәy appear, one should consider an alternative method to the least squares.

In Rubio et al [9] and in Víšek [14] the first attempts were made to generalize result which is due to Clemen [3] and which holds for the ordinary least squares. It claims that in the case when the individual forecasts are unbiased, it is preferable to construct the combination of forecasts by means of the regression model without intercept subject to the constraint that the sum of coefficients and of course, also of estimates, is equal to one. It appeared that this result holds also for M-estimators and numerical illustration showed that the best results were obtained by M-estimators with redescending ψ-function. On the other hand, it is well known that firstly the breakdown point of the M-estimators is equal to k^{-1} where k is the dimension of regression model in question, i. e. the number of explanatory variables. In our cases, as it follows from the first lines of this paper, it is the number of individual forecasts we have taken into account. In other words, the M-estimators have the breakdown point limited by dimension of corresponding regression model (see Yohai and Maronna [23]). Secondly, the M-estimators are not scale- and regression equivariant. To reach scale- and regression-equivariance one needs to studentize the residuals by an appropriate scale-invariant and regression-equivariant scale estimator, see Bickel [2] or Jurečková and Sen [7]. Although to evaluate such an estimator is possible (see Víšek [17]), it is not very easy and quick. So it would be preferable to have a (theoretically supported) possibility to apply for combining forecasts such robust estimator which is scale- and regression-equivariant and, if possible, with adjustable breakdown point. One of such estimators is just the least trimmed squares. Although we were able already earlier to demonstrate that combining forecasts by the least trimmed squares can give good numerical results, we were not able to prove a theoretical result analogous to Clemen's one. Nowadays, following steps in Víšek [18] we can carry out corresponding proof.

First of all we shall introduce notations and simultaneously recall Clemen's result.

[2] There are however studies indicating that it is very rare (if not impossible) case, see Hampel et al [4] or Huber [6] and references given there.

[3] The word *atypical* means that such point may (but need not) belong to the "true" model, or in other words, the point does not necessarily represent contamination, nevertheless its value is such that it worsens the result of prediction anyway.

We shall keep notations which were used by Clemen and which we have also used in Víšek [14]. It will allow to follow easier the text to reader who is familiar with Clemen's one.

2. NOTATION

We shall denote by N the set of all positive integers, by R the real line and by R^+ its positive part. Moreover, by R^k we shall denote the k dimensional Euclidean space. Finally, we shall consider for any $t \in N$ the linear model

$$\theta^t = F_t \cdot \beta^0 + \varepsilon^t \tag{1}$$

where the forecasted process $\theta^t = (\theta_1, \theta_2, \ldots, \theta_t)^{\mathrm{T}}$ plays the role of response variable (the capital "T" indicates transposition). We shall assume that the first column of design matrix $F_t = (f_{ij})_{j=1,2,\ldots,k}^{i=1,2,\ldots,t}$ consists of ones, i. e. $f_{i1} = 1$ for $i = 1, 2, \ldots, t$, and the rest of it is created by $(k-1)$ individual forecasts (columns $2, 3, \ldots, k$). Regression coefficients are denoted by $\beta^0 = (\beta_1^0, \beta_2^0, \ldots, \beta_k^0)^{\mathrm{T}}$ and the vector of random disturbances in model by $\varepsilon^t = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_t)^{\mathrm{T}}$. Moreover, ith line of the matrix F_t (considered as column vector) will be denoted by f_i. (To avoid any misundeistanding what concerns vectors, we shall assume all of them to be column ones.) Now, the alternative form of the model (1) is

$$\theta_i = f_i^{\mathrm{T}} \beta^0 + \varepsilon_i, \quad i = 1, 2, \ldots, t.$$

Remark 1. Let us realize that the regression model given in (1) is not the regression mechanism we have mentioned at the introduction that generates the forecasted process. Notice that, to emphasize it, we have used for rows of corresponding design matrix of former model the letters X_t while for the latter model (i. e. (1) we utilized f_t.

By $I\{\textit{property describing the set } A\}$ (instead of $I_{\{\textit{property describing the set } A\}}$) we shall denote the indicator of set A. The reason is the fact that in what follows we shall use for description of sets (somewhat) complicated expressions containing also indices.

Remark 2. There are well-known reasons for inclusion of the intercept into the model – except of a few situations when we are sure that the regression goes through the origin. Moreover, insisting on the absence of intercept implicitly indicates our belief in an absolute character of data which in turn means that we give up otherwise natural requirement of scale- and regression-equivariance of the estimator of the regression coefficients. Nevertheless, even with such situation we may meet as we shall see below. On the other hand, in the case of present paper we have started with the model with intercept to have a model as general as possible and we shall see later that another model may appear better (under some conditions).

We shall need the following assumptions.

Assumptions $\mathcal{A}$. The sequence $\{(f_i^{\mathrm{T}}, \varepsilon_i)^{\mathrm{T}}\}_{i=1}^{\infty}$ ($f_i \in R^p, \varepsilon_i \in R$) is the sequence of independent identically distributed random vectors with $f_{11} = 1$, $\mathbb{E}f_{1j} = 0$ and $\mathbb{E}f_{1j}^4 < \infty$ for $j = 2, 3, \ldots, k$ or $\mathbb{E}f_{1j} = 0$ and $\mathbb{E}f_{1j}^4 < \infty$ for $j = 1, 2, \ldots, k$. Moreover, $\mathbb{E}f_1 \cdot f_1^{\mathrm{T}} = Q$ is regular. The marginal distribution function of f_1, say H(x), is absolutely continuous and such that

$$t^{-\frac{1}{4}} \max_{1 \le i \le t,\ 1 \le j \le k} |f_{ij}| = \mathcal{O}_p(1). \tag{2}$$

The conditional distribution function $D(z|f)$ of random fluctuation ε_1 given f_1 is absolutely continuous with a bounded density $d(z|f)$ which is positive and has bounded derivative on the R. Denote $G(z|f)$ the conditional distribution function of e_1^2 given f_1. For some $\alpha \in [0, \frac{1}{2})$, u_α^2 will be the upper α-quantile of $G(z|f)$, i.e. $P(\varepsilon_1^2 > u_\alpha^2) = 1 - G(u_\alpha^2|f) = \alpha$ and $[(1-\alpha) - u_\alpha(d(u_\alpha|f) + d(-u_\alpha|f)] \neq 0$. Further

$$\mathbb{E}(\varepsilon_1 I\{\varepsilon_i^2 \le u_\alpha^2\}|f_1) = 0 \quad \text{and} \quad \mathbb{E}(\varepsilon_1^2 I\{\varepsilon_i^2 \le u_\alpha^2\}|f_1) = \sigma_{\varepsilon_1}^2 \tag{3}$$

with $\sigma_{\varepsilon_1}^2 \in (0, \infty)$. Finally, denote by $[a]$ the integer part of a and for any $t \in N$ put $h_t = [(1-\alpha)t]$.

Remark 3. Notice that assumptions in (3) are analogies of the orthogonality and sphericality conditions. Of course, when we shall recall Clemen's result for OLS, we will assume (for a moment) that "ordinary" orthogonality and sphericality conditions hold, i.e. that

$$\mathbb{E}(\varepsilon^t|F_t) = 0 \quad \text{and} \quad \mathbb{E}(\varepsilon^t \cdot [\varepsilon^t]^{\mathrm{T}}|F_t) = \sigma_0^2 \mathcal{I} \tag{4}$$

for any $t \in N$ (where "$\mathcal{I}$" denote the unit matrix). On the other hand, the assumption (3) is quite natural, since it corresponds to computational reality. As we shall see below, the evaluation of the least trimmed squares estimator is equivalent to the application of the ordinary least squares on a subset of data. The subset has size h and is given implicitly by the extremal problem – see (7) below. Nevertheless, since the ordinary least squares evaluates the estimate so that it corresponds to the assumption of centered random noise, (3) is "implicitly fulfilled" by numerical algorithm.

Remark 4. For any $1 \le j, \ell \le k$ we have

$$(F_t F_t^{\mathrm{T}})_{j\ell} = \sum_{i=1}^{t} f_{ij} f_{i\ell},$$

i.e.

$$F_t F_t^{\mathrm{T}} = \sum_{i=1}^{t} f_i f_i^{\mathrm{T}}.$$

Moreover, we have just assumed that $\mathbb{E}f_1 f_1^{\mathrm{T}} = Q$ exist and is regular (and hence the matrix Q has all elements finite). Together with the assumption that the explanatory vectors f_i's are i.i.d. it implies that

$$\lim_{t \to \infty} \frac{1}{t} F_t^{\mathrm{T}} F_t = Q \quad \text{a.e..} \tag{5}$$

Since the matrix Q has finite number of elements and its determinant is positive (remember that it is positive definite), there is $t_0 \in N$ such that for all $t > t_0$ determinant of $\frac{1}{t}F_t^{\mathrm{T}}F_t$ is also positive and the same is true about determinant of $F_t^{\mathrm{T}}F_t$. But it implies that $F_t^{\mathrm{T}}F_t$ is for $t > t_0$ also regular and hence we can evaluate an inversion matrix. Moreover, (5) then implies that

$$(F_t^{\mathrm{T}}F_t)^{-1} = O_p(t^{-1}). \tag{6}$$

3. THE LEAST TRIMMED SQUARES

Let us denote for any $\beta \in R^k$ by $r_i(\beta) = \theta_i - f_i^{\mathrm{T}}\beta$ the ith residual and by $r_{(i)}^2(\beta)$ the order statistics of squared residuals (for $i = 1, 2, \ldots, t$). In other words, it means that we have for any $\beta \in R^k$ (and any $\omega \in \Omega$)

$$0 \le r_{(1)}^2(\beta) \le r_{(2)}^2(\beta) \le \cdots \le r_{(t)}^2(\beta).$$

Finally, let us recall that the *least trimmed squares* estimator is given as

$$\hat{\beta}^{(LTS,t,h)} = \underset{\beta\in R^k}{\arg\min} \sum_{i=1}^{h} r_{(i)}^2(\beta) \tag{7}$$

where $\frac{t}{2} \le h \le t$. One can guess that the value of h implies the level of robustness of estimator, namely its breakdown point.

Let us recall that for $h = t$, $\hat{\beta}^{(LTS,t,h)}$ coincides with the least squares estimator $\hat{\beta}^{(LS,t)}$, given of course as

$$\hat{\beta}^{(LS,t)} = \underset{\beta\in R^k}{\arg\min} \sum_{i=1}^{t} r_i^2(\beta) = \underset{\beta\in R^k}{\arg\min} \sum_{i=1}^{t} (\theta_i - f_i^{\mathrm{T}}\beta)^2.$$

4. RECALLING CLEMEN'S RESULT

According to the well-known formula, the LS-estimate of β^0 is given as

$$\hat{\beta}^{(LS,t)} = (F_t^{\mathrm{T}}F_t)^{-1}F_t^{\mathrm{T}}\theta^t. \tag{8}$$

Naturally, one lag forward forecast is evaluated as

$$\hat{\theta}_{t+1}^{(LS,t)} = \sum_{j=1}^{k} f_{t+1,j}\, \hat{\beta}_j^{(LS,t)} \quad \text{where} \quad f_{t+1} = (1, f_{t+1,2}, \ldots, f_{t+1,k})^{\mathrm{T}}$$

or in an alternative (and more convenient) form can be written as

$$\hat{\theta}_{t+1}^{(LS,t)} = f_{t+1}^{\mathrm{T}}\hat{\beta}^{(LS,t)}.$$

A straightforward calculation gives the mean square error

$$\mathbb{E}\left\{(\hat{\theta}_{t+1}^{(LS,t)} - \theta_{t+1})^2 | F_t, f_{t+1}\right\} = \sigma_0^2(f_{t+1}^{\mathrm{T}}(F_t^{\mathrm{T}}F_t)^{-1}f_{t+1} + 1), \tag{9}$$

for σ_0 see (4).

In the case when the forecasts are unbiased (we can verify it by some test, see e. g. Holden and Peel [5]), we would expect that the model subject to the constraints

$$\beta_1^0 = 0, \qquad \beta_2^0 + \beta_3^0 + \ldots + \beta_k^0 = 1 \tag{10}$$

will be more suitable for the combination and we naturally try to estimate β^0 by an estimator which is subject to the same constraints. It is possible to impose other constraints e. g. that (10) holds and $\beta_i \in [0,1]$ for $i = 1, 2, \ldots, k$. Extensive literature discussing it and bringing arguments for and against can be found e. g. in the special issue of *Journal of Forecasting* devoted to twenty anniversary of the paper by Bates and Granger [1]. However, there are examples demonstrating that sometimes substantial improvement was achieved when some coefficients were either negative and/or some larger than 1. An explanation is simple. When the projection of the forecasted process into the linear space generated by the forecasts falls "outside" all forecasts (instead among them), it is clear that the best combination should contain also some negative coefficients. The heuristics for this case are straightforward. Simply all forecasters had interpreted some (common) information in a wrong way and hence only a possibility to "withdraw" this false step of all (or subtract this false interpretation, if you wish) can considerably improve the forecast. In such a case one may immediately object that then probably the individual forecasts are not unbiased and we should "recognize" it by previously mentioned test. However, it is sufficient to look into a one paper about testing unbiasedness of forecasts and one immediately learns that this topic is at least a bit controversial (see again Holden and Peel [5]).

For theoretical considerations it will be convenient to have the constraints (10) in the matrix form

$$S \cdot \beta^0 = \gamma \tag{11}$$

where $S = (s_{ij})_{j=1,2,\ldots,k}^{i=1,2}$, $s_{11} = 1, s_{1j} = 0$ for $j = 2, 3, \ldots, k$, $s_{21} = 0$, $s_{2j} = 1$ for $j = 2, 3, \ldots, k$ and $\gamma = (0,1)^{\mathrm{T}}$. So, we shall look for the least squares estimator under the constraint (11), i. e. we shall consider

$$\tilde{\beta}^{(LS,t)} = \arg\min_{\beta \in R^k} \left\{\sum_{i=1}^{t}(\theta_i - f_i^{\mathrm{T}}\beta)^2,\ S\beta = \gamma\right\}$$

and we hope that it will work better than the OLS given in (8). Following Clemen [3] and using the Lagrangian technique we can find that

$$\tilde{\beta}^{(LS,t)} = \hat{\beta}^{(LS,t)} - (F_t^{\mathrm{T}}F_t)^{-1}S^{\mathrm{T}}\left[S(F_t^{\mathrm{T}}F_t)^{-1}S^{\mathrm{T}}\right]^{-1}(S\hat{\beta}^{(LS,t)} - \gamma). \tag{12}$$

Again a straightforward calculation yields the mean squared error of the prediction $\tilde{\theta}_{t+1}$ based on the "constrained" estimator $\tilde{\beta}^{(t)}$, namely

$$\mathbb{E}\left\{(\tilde{\theta}_{t+1}^{(LS,t)} - \theta_{t+1})^2 | F_t, f_{t+1}\right\} \tag{13}$$

$$= \sigma_0^2 \left\{ f_{t+1}^{\mathrm{T}} \left[(F_t^{\mathrm{T}} F_t)^{-1} - (F_t^{\mathrm{T}} F_t)^{-1} S^{\mathrm{T}} \left[S(F_t^{\mathrm{T}} F_t)^{-1} S^{\mathrm{T}} \right]^{-1} S(F_t^{\mathrm{T}} F_t)^{-1} \right] f_{t+1} + 1 \right\}.$$

Evidently the matrix

$$(F_t^{\mathrm{T}} F_t)^{-1} S^{\mathrm{T}} \cdot \left[S(F_t^{\mathrm{T}} F_t)^{-1} S^{\mathrm{T}} \right]^{-1} S(F_t^{\mathrm{T}} F_t)^{-1}$$

is positive semidefinite. It implies that (13) is not larger than (9).

We have already mentioned that the main goal of this paper is to confirm (or reject) validity of this Clemen result for the least trimmed squares. Since there is not a close formula for $\hat{\beta}^{(LTS,t,h)}$, we shall use, similarly as in the case of M-estimators (see Víšek [14]), its asymptotic representation. So first of all, we shall recall it and try to find also a representation for the least trimmed squares under constraint (11).

5. ASYMPTOTIC REPRESENTATION OF THE LEAST TRIMMED SQUARES UNDER A CONSTRAINT

Assertion 1. Let Assumptions $\mathcal{A}$ be fulfilled. Moreover write h instead of h_t. Then

$$\sqrt{t}\left(\hat{\beta}^{(LTS,t,h)} - \beta^0\right) = t^{-\frac{1}{2}} Q^{-1} \left[(1-\alpha) - u_\alpha(d(u_\alpha|f) + d(-u_\alpha|f))\right]^{-1} \times$$
$$\times \sum_{i=1}^{t} \left(\theta_i - f_i^{\mathrm{T}}\beta^0\right) f_i \cdot I\{\varepsilon_i^2 \le u_\alpha^2\} + o_p(1) \tag{14}$$
$$= t^{-\frac{1}{2}} Q^{-1} \left[(1-\alpha) - u_\alpha(d(u_\alpha|f) + d(-u_\alpha|f))\right]^{-1} \sum_{i=1}^{t} \varepsilon_i f_i \cdot I\{\varepsilon_i^2 \le u_\alpha^2\} + o_p(1)(15)$$

and $\hat{\beta}^{(LTS,t,h)}$ is asymptotically normal with mean value equal to β^0 and covariance matrix

$$V(\hat{\beta}^{(LTS,t,h)}, D) = Q^{-1} \left[(1-\alpha) - u_\alpha(d(u_\alpha|f) + d(-u_\alpha|f)) \right]^{-2} \int_{-u_\alpha}^{u_\alpha} z^2 \mathrm{d}D(z|f),$$

i. e.

$$\mathcal{L}\left(\sqrt{t}\left(\hat{\beta}^{(LTS,t,h)} - \beta^0\right)\right) \to \mathcal{N}(0, V(\hat{\beta}^{(LTS,t,h)}, D)) \quad \text{as } t \to \infty.$$

For the proof see Víšek [18].

Now, let us consider the least trimmed squares estimator of β^0 which is subject to the same constraints as given in (11), i. e. the estimator given by

$$\tilde{\beta}^{(LTS,t,h)} = \underset{\beta \in R^k}{\arg\min} \left\{ \sum_{i=1}^{h} r_{(i)}^2(\beta) \quad \text{together with} \quad S\beta = \gamma \right\}. \tag{16}$$

Employing Assertion 1, let us try to establish an asymptotic representation for $\tilde{\beta}^{(LTS,t,h)}$. In order to achieve it, we shall consider the linear regression model for the variables transformed in the following way. Put $\bar{\theta}_i = \theta_i - f_{i,2}, \bar{f}_{ij} = f_{i,j+2} - f_{i,2}$ for $i = 1, 2, \ldots, t$ and $j = 1, 2, \ldots k-2$ and define a mapping $\mathcal{T} : R^{k-2} \longrightarrow R^k$ which for any $\bar{\beta} \in R^{k-2}$ gives $\tilde{\beta} \in R^k$ so that

$$\tilde{\beta}_1 = 0, \quad \tilde{\beta}_2 = 1 - \sum_{j=1}^{k-2} \bar{\beta}_j, \quad \tilde{\beta}_\ell = \bar{\beta}_{\ell-2} \quad \text{for } \ell = 3, 4, \ldots k.$$

Let us notice that the image of the mapping $\mathcal{T}$ is the subset of R^k for which $S\beta = \gamma$. Keeping in mind (11), let us put $\bar{\beta}^0 = (\beta_3^0, \beta_4^0, \ldots, \beta_k^0)^{\mathrm{T}}$, i. e. $\bar{\beta}_\ell^0 = \beta_{\ell+2}^0$ for $\ell = 1, 2, \ldots, k-2$. Then $\mathcal{T}(\bar{\beta}^0) = \beta^0$ and we may write for the model (1) the following sequence of equations

$$\begin{aligned} \theta_i &= \sum_{j=1}^{k} f_{ij}\beta_j^0 + \varepsilon_i \qquad i = 1, 2, \ldots, t, && (17) \\ \theta_i &= \left(1 - \sum_{j=3}^{k} \beta_j^0\right) f_{i2} + \sum_{j=3}^{k} f_{ij}\beta_j^0 + \varepsilon_i, \\ \theta_i - f_{i2} &= \sum_{j=1}^{k-2} (f_{i,j+2} - f_{i,2})\bar{\beta}_j^0 + \varepsilon_i \end{aligned}$$

and finally

$$\bar{\theta}_i = \bar{f}_i^{\mathrm{T}} \bar{\beta}^0 + \varepsilon_i \tag{18}$$

which implies that the random disturbances in the regression model for the transformed variables $\bar{\theta}_i$'s and $\bar{f}_i$'s are the same as in the original model (1). Moreover, modifying a little the steps from (17) to (18) we obtain for any $\bar{\beta} \in R^{k-2}$ and $\tilde{\beta} = \mathcal{T}(\bar{\beta})$

$$\bar{\theta}_i - \sum_{j=1}^{k-2} \bar{f}_{ij}\bar{\beta}_j = \theta_i - \sum_{j=1}^{k} f_{ij}\tilde{\beta}_j$$

with $S\tilde{\beta} = \gamma$. But it implies that for any $\bar{\beta} \in R^{k-2}$ and $\tilde{\beta} = \mathcal{T}(\bar{\beta})$ we have for $i = 1, 2, \ldots, t$

$$\bar{r}_i(\bar{\beta}) = \bar{\theta}_i - \bar{f}_i^{\mathrm{T}} \bar{\beta} = \theta_i - f_i^{\mathrm{T}} \tilde{\beta} = r_i(\tilde{\beta})$$

and of course also

$$\bar{r}_{(i)}^2(\bar{\beta}) = r_{(i)}^2(\tilde{\beta}).$$

But then we have

$$\sum_{i=1}^{h} \bar{r}_{(i)}^2(\bar{\beta}) = \sum_{i=1}^{h} r_{(i)}^2(\tilde{\beta})$$

and $S\tilde{\beta} = \gamma$. However it means that if we find a solution of the problem

$$\bar{\beta}^{(LTS,t,h)} = \underset{\beta \in R^{k-2}}{\arg\min} \sum_{i=1}^{h} \bar{r}_{(i)}^2(\beta), \tag{19}$$

we immediately have the solution of the problem (16). It is evident that $\tilde{\beta}^{(LTS,t,h)} = \mathcal{T}(\bar{\beta}^{(LTS,t,h)})$. But then our plan is simple. We shall try to verify that Assumptions $\mathcal{A}$ hold also for transformed random variables $\bar{\theta}_i$'s and $\bar{f}_i$'s, so that we may write asymptotic representation (14) also for $\bar{\beta}^{(LTS,t,h)}$. Then we may try to modify this representation to obtain a representation of $\tilde{\beta}^{(LTS,t,h)}$. Finally, employing both representations, i. e. of $\hat{\beta}^{(LTS,t,h)}$ and of $\tilde{\beta}^{(LTS,t,h)}$, we may find which of these two estimators has smaller asymptotic variance. And that will be done in the rest of this paragraph.

Let us recall that we have denoted by $\mathcal{I}$ the identity (or if you wish, the unit) matrix. We shall prove

Lemma 1. Let $S \cdot \beta^0 = \gamma$ hold. Then under Assumptions $\mathcal{A}$ we have

$$\begin{aligned} &\sqrt{t}\ (\tilde{\beta}^{(LTS,t,h)} - \beta^0) \\ = \ & \frac{1}{\sqrt{t}}[(1-\alpha) - u_\alpha(d(u_\alpha|f) + d(-u_\alpha|f)]^{-1}\left\{\mathcal{I} - Q^{-1}S^{\mathrm{T}}(SQ^{-1}S^{\mathrm{T}})^{-1}S\right\} \times \\ & \times Q^{-1}\sum_{i=1}^{t} f_i\left(\theta_i - f_i^{\mathrm{T}}\beta^0\right) \cdot I\{\varepsilon_i^2 \le u_\alpha^2\} + o_p(1) \quad \text{as } t \to \infty. \end{aligned} \tag{20}$$

Proof. First of all, we shall show that for $t > t_0$ the matrix $SQ^{-1}S^{\mathrm{T}}$ is regular (for t_0 see Remark 2).

Due to the fact that the matrix S is created by two independent vectors, S may be "expanded" into a regular $(p \times p)$-matrix, say $\tilde{S}$, with first two lines equal just to S. $SQ^{-1}S^{\mathrm{T}}$ is then the main submatrix of the positive definite matrix $\tilde{S}Q^{-1}\tilde{S}^{\mathrm{T}}$, hence it is also positive definite and finally regular.

Now, we shall verify that for the transformed problem (19) the Assumptions $\mathcal{A}$ also hold, so that we shall be able to use Assertion 1. For any $\ell, r = 1, 2, \ldots, k-2$ we have

$$\begin{aligned} I\!\!E\bar{f}_{1,\ell}\bar{f}_{1,r} &= I\!\!E(f_{1,\ell+2} - f_{1,2})(f_{1,r+2} - f_{1,2}) \\ &= q_{\ell+2,r+2} - q_{\ell+2,2} - q_{2,r+2} + q_{2,2}. \end{aligned} \tag{21}$$

Denote the expression in (21) by $\bar{q}_{\ell,r}$ and the corresponding matrix by $\bar{Q}$, i. e. $\bar{Q} = (\bar{q}_{\ell,r})_{\ell,r=1,2,\ldots,k-2}$. Let us further consider the matrix $A = (a_{\ell j})_{\ell,j=1,2,\ldots,k}$ such that $a_{\ell\ell} = 1$, for $\ell = 1, 2, \ldots, k$, $a_{\ell 2} = -1$ for $\ell = 3, 4, \ldots, k$ and $a_{\ell j} = 0$ for all other indices. One easily finds that $(AQA^{\mathrm{T}})_{\ell+2,r+2} = \bar{q}_{\ell,r}$ for $\ell, r = 1, 2, \ldots, k-2$, i. e. that $\bar{Q}$ is one of the main submatrices of AQA^{T}. Since the matrix A is evidently regular, and the matrix Q is positive definite, AQA^{T} is also positive definite and regular. So the assumption about regularity of matrix $\bar{Q}$ holds. Employing similar arguments, we may easy verify (2). The validity of the rest of Assumptions $\mathcal{A}$ for the transformed variables (or if you wish, for the transformed problem (19)) follows from (17) and (18) and from properties of conditional moments. Now, recalling that we have denoted

$$(\bar{Q})_{\ell r} = I\!\!E(f_{1,\ell+2} - f_{12})(f_{1,r+2} - f_{1,2})$$

and putting

$$\vartheta = [(1-\alpha) - u_\alpha(d(u_\alpha|f) + d(-u_\alpha|f))],$$

we can apply Assertion 1 to the transformed setup and we obtain (from (15))

$$t^{\frac{1}{2}}\ (\bar{\beta}^{(LTS,t,h)} - \bar{\beta}^0) = t^{-\frac{1}{2}}\vartheta^{-1}\bar{Q}^{-1}\sum_{i=1}^{t}\bar{f}_i\varepsilon_i \cdot I\{\varepsilon_i^2 \le u_\alpha^2\} + o_p(1) \quad \text{as} \quad t \to \infty$$

or, denoting $\varepsilon_i \cdot I\{\varepsilon_i^2 \le u_\alpha^2\}$ by κ_i

$$\begin{aligned} & t^{\frac{1}{2}}\ \vartheta\ \sum_{j=1}^{k-2}\bar{q}_{\ell j}(\bar{\beta}_j^{(LTS,t,h)} - \bar{\beta}_j^0) \\ = \quad & t^{-\frac{1}{2}}\sum_{i=1}^{t}\bar{f}_{i,\ell}\kappa_i + o_p(1) \quad \text{as} \quad t \to \infty \quad \text{for} \quad \ell = 1, 2, \ldots, k-2. \end{aligned} \tag{22}$$

Similarly the representation (15) may be rewritten as

$$\begin{aligned} & t^{\frac{1}{2}}\ \vartheta\ \sum_{j=1}^{k} q_{\ell j}(\hat{\beta}_j^{(LTS,t,h)} - \beta_j^0) \\ = \quad & t^{-\frac{1}{2}}\sum_{i=1}^{t} f_{i,\ell}\kappa_i + o_p(1) \quad \text{as} \quad t \to \infty \quad \text{for} \quad \ell = 1, 2, \ldots, k. \end{aligned} \tag{23}$$

Using (21) we may modify (22) and we obtain

$$\begin{aligned} & t^{\frac{1}{2}}\ \vartheta \sum_{j=1}^{k-2}(q_{\ell+2,j+2} - q_{\ell+2,2} - q_{j+2,2} + q_{2,2})(\bar{\beta}_j^{(LTS,t,h)} - \beta_{j+2}^0) \\ = \quad & t^{-\frac{1}{2}}\sum_{i=1}^{t}(f_{i,\ell+2} - f_{i,2})\kappa_i + o_p(1) \quad \text{as} \quad t \to \infty \quad \text{for} \quad \ell = 1, 2, \ldots, k-2. \end{aligned}$$

Combining it with (23), we arrive at

$$\begin{aligned} & t^{\frac{1}{2}}\ \vartheta \left\{ \sum_{j=3}^{k}(q_{\ell j} - q_{j2})(\bar{\beta}_{j-2}^{(LTS,t,h)} - \beta_j^0) + (q_{2,2} - q_{\ell 2})\sum_{j=1}^{k-2}(\bar{\beta}_j^{(LTS,t,h)} - \beta_{j+2}^0) \right\} \\ = \quad & t^{-\frac{1}{2}}\ \vartheta \sum_{j=1}^{k}(q_{\ell j} - q_{2j})(\hat{\beta}_j^{(LTS,t,h)} - \beta_j^0) + o_p(1) \quad \text{as} \quad t \to \infty \quad \text{for} \quad \ell = 3, 4, \ldots, k. \end{aligned}$$

Since we have assumed that $\vartheta \neq 0$ we may omit it and taking into account that $\tilde{\beta}_j^{(LTS,t,h)} = \bar{\beta}_{j-2}^{(LTS,t,h)}$ for $j = 3, 4, \ldots, k$, $\tilde{\beta}_2^{(LTS,t,h)} = 1 - \sum_{j=1}^{k}\bar{\beta}_j^{(LTS,t,h)}$ and $\tilde{\beta}_1^{(LTS,t,h)} = 0$ (and also $\beta_1^0 = 0$), we obtain

$$t^{\frac{1}{2}}\ \sum_{j=1}^{k}(q_{\ell j} - q_{2j})(\tilde{\beta}_j^{(LTS,t,h)} - \beta_j^0)$$

$$= \quad t^{\frac{1}{2}} \sum_{j=1}^{k} (q_{\ell,j} - q_{2,j})(\hat{\beta}_j^{(LTS,t,h)} - \beta_j^0) + o_p(1)$$
$$\text{as } t \to \infty \qquad \text{for } \ell = 3, 4, \dots, k$$

and finally

$$t^{\frac{1}{2}} \sum_{j=1}^{k} q_{\ell j}(\tilde{\beta}_j^{(LTS,t,h)} - \beta_j^0)$$
$$= \quad t^{\frac{1}{2}} \sum_{j=1}^{k} \left\{ q_{\ell j}(\hat{\beta}_j^{(LTS,t,h)} - \beta_j^0) + q_{2j}(\tilde{\beta}_j^{(LTS,t,h)} - \hat{\beta}_j^{(LTS,t,h)}) \right\} + o_p(1)$$
$$\text{as } t \to \infty \quad \text{for } \ell = 3, 4, \dots, k.$$

Putting

$$\lambda_\ell = t^{\frac{1}{2}} \sum_{j=1}^{k} q_{\ell j}(\tilde{\beta}_j^{(LTS,t,h)} - \hat{\beta}_j^{(LTS,t,h)}) \quad \text{for} \quad \ell = 1, 2, \tag{24}$$

we obtain

$$t^{\frac{1}{2}} \sum_{j=1}^{k} q_{1,j}(\tilde{\beta}_j^{(LTS,t,h)} - \beta_j^0)$$
$$= \quad t^{\frac{1}{2}} \sum_{j=1}^{k} q_{1,j}(\hat{\beta}_j^{(LTS,t,h)} - \beta_j^0) + \lambda_1 \quad \text{as } t \to \infty, \tag{25}$$

and for $\ell = 2, \dots, k$ we have

$$t^{\frac{1}{2}} \sum_{j=1}^{k} q_{\ell,j}(\tilde{\beta}_j^{(LTS,t,h)} - \beta_j^0)$$
$$= \quad t^{\frac{1}{2}} \sum_{j=1}^{k} q_{\ell,j}(\hat{\beta}_j^{(LTS,t,h)} - \beta_j^0) + \lambda_2 + o_p(1) \quad \text{as } t \to \infty \tag{26}$$

(notice, please, that (25) is just an equality). Moreover

$$S\tilde{\beta}^{(LTS,t,h)} = \gamma \tag{27}$$

Putting $\lambda = (\lambda_1, \lambda_2)^{\mathrm{T}}$, we may rewrite (25) – (27) into the matrix form

$$t^{\frac{1}{2}} \, Q(\tilde{\beta}^{(LTS,t,h)} - \beta^0) = t^{\frac{1}{2}} \, Q(\hat{\beta}^{(LTS,t,h)} - \beta^0) + S^{\mathrm{T}}\lambda + o_p(1) \quad \text{and} \quad S\tilde{\beta}^{(LTS,t,h)} = \gamma.$$

We have thus obtained

$$t^{\frac{1}{2}} \, Q\tilde{\beta}^{(LTS,t,h)} = t^{\frac{1}{2}} \, Q\hat{\beta}^{(LTS,t,h)} + S^{\mathrm{T}}\lambda + o_p(1) \quad \text{and} \quad S\tilde{\beta}^{(LTS,t,h)} = \gamma.$$

Expressing $\tilde{\beta}^{(LTS,t,h)}$ as $\hat{\beta}^{(LTS,t,h)} + t^{-\frac{1}{2}}Q^{-1}S^{\mathrm{T}}\lambda + o_p(t^{-\frac{1}{2}})$ we may write

$$\gamma = S\tilde{\beta}^{(LTS,t,h)} = S\left[\hat{\beta}^{(LTS,t,h)} + t^{-\frac{1}{2}}Q^{-1}S^{\mathrm{T}}\lambda + o_p(t^{-\frac{1}{2}})\right]$$

and so

$$\lambda = t^{\frac{1}{2}}\ \left[SQ^{-1}S^{\mathrm{T}}\right]^{-1}(\gamma - S\hat{\beta}^{(LTS,t,h)}) + o_p(1)$$

(keep in mind that at the beginning of the proof we have shown that $SQ^{-1}S^{\mathrm{T}}$ is regular). This means that we have arrived at

$$\tilde{\beta}^{(LTS,t,h)} = \hat{\beta}^{(LTS,t,h)} + Q^{-1}S^{\mathrm{T}}\left[SQ^{-1}S^{\mathrm{T}}\right]^{-1}(\gamma - S\hat{\beta}^{(LTS,t,h)}) + o_p(t^{-\frac{1}{2}}).$$

We have assumed that $\gamma = S\beta^0$ so that

$$\begin{aligned} & t^{\frac{1}{2}}\ (\tilde{\beta}^{(LTS,t,h)} - \beta^0) \\ = \ & t^{\frac{1}{2}}\ \left\{(\hat{\beta}^{(LTS,t,h)} - \beta^0) + Q^{-1}S^{\mathrm{T}}\left[SQ^{-1}S^{\mathrm{T}}\right]^{-1}S(\beta^0 - \hat{\beta}^{(LTS,t,h)})\right\} + o_p(1). \end{aligned}$$

Now, employing asymptotic representation of $\sqrt{t}(\hat{\beta}^{(LTS,t,h)} - \beta^0)$ once again we obtain

$$\sqrt{t}\,(\tilde{\beta}^{(LTS,t,h)}-\beta^0) = \frac{1}{\sqrt{t}}\vartheta^{-1}\left\{\mathcal{I} - Q^{-1}S^{\mathrm{T}}(SQ^{-1}S^{\mathrm{T}})^{-1}S\right\}\times$$

$$\times Q^{-1}\sum_{i=1}^{t} f_i\ \left[\varepsilon_i \cdot I\{\varepsilon_i^2 \le u_\alpha^2\}\right] + o_p(1) \quad \text{as } t \to \infty$$

which concludes the proof. □

Remark 5. Since $S\cdot\beta^0 = \gamma$ implies that $\beta_1^0 = 0$ and the same holds for $\tilde{\beta}^{(LTS,n,h)}$, it may be of interest to verify that the expansion (20) is consistent with it.

First of all, notice that due to Assumptions $\mathcal{A}$ in the case when we assume model with intercept (and hence the constraint $S\cdot\beta^0 = \gamma$ has a sense), the matrix Q has the form

$$\begin{bmatrix} 1, & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0}, & H \end{bmatrix}$$

where $\mathbf{0} = (0, 0, \dots, 0)^{\mathrm{T}}$. Due to the assumption about regularity of Q, also H is regular. Hence

$$Q^{-1} = \begin{bmatrix} 1, & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0}, & H^{-1} \end{bmatrix}.$$

Now let us look on the structure of matrix $\mathcal{I} - Q^{-1}S^{\mathrm{T}}(SQ^{-1}S^{\mathrm{T}})^{-1}S$. Taking into account the structure of matrices Q^{-1} and S we easy verify that

$$(SQ^{-1}S^{\mathrm{T}})^{-1} = \begin{bmatrix} 1, & 0 \\ 0, & v \end{bmatrix}$$

where $v = \left[\sum_{i=1}^{k} \sum_{\ell=1}^{k} \tilde{q}_{i\ell}\right]^{-1}$ for $\tilde{q}_{i\ell} = \left[Q^{-1}\right]_{i\ell}$. But then

$$S^{\mathrm{T}}(SQ^{-1}S^{\mathrm{T}})^{-1}S = \begin{bmatrix} 1, & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0}, & \Upsilon \end{bmatrix}$$

where $(\Upsilon)_{i\ell} = v$ (for all i and ℓ) and finally

$$Q^{-1}S^{\mathrm{T}}(SQ^{-1}S^{\mathrm{T}})^{-1}S = \begin{bmatrix} 1, & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0}, & H^{-1}\Upsilon \end{bmatrix}$$

It means that

$$\mathcal{I} - Q^{-1}S^{\mathrm{T}}(SQ^{-1}S^{\mathrm{T}})^{-1}S = \begin{bmatrix} 0, & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0}, & \mathcal{I} - H^{-1}\Upsilon \end{bmatrix}.$$

6. COMBINING FORECASTS USING LEAST TRIMMED SQUARES ESTIMATOR WITH CONSTRAINTS

Prior to a comparison of the combined forecasts based on the least trimmed squares estimators with and without constraints, let us return for a moment to (9) and (13), and let us find how large is the difference of the corresponding mean squared errors. We easy find that the variance of the prediction $\hat{\theta}_{t+1}^{(LS,t)}$ is equal to

$$\operatorname{var}\left(\hat{\theta}_{t+1}^{(LS,t)} | F_t, f_{t+1}\right) = I\!E\left\{(\hat{\theta}_{t+1}^{(LS,t)} - I\!E\hat{\theta}_{t+1}^{(LS,t)})^2 | F_t, f_{t+1}\right\} = \sigma_0^2 f_{t+1}^{\mathrm{T}}(F_t^{\mathrm{T}} F_t)^{-1} f_{t+1}.$$

Taking into account (6), we conclude that $\operatorname{var}\left(\hat{\theta}_{t+1}^{(LS,t)} | F_t, f_{t+1}\right)$ is also of order $O_p(t^{-1})$. On the other side, the mean squared error is close to the conditional variance of θ_t which is equal to σ_0^2. Similarly, in (13), the term

$$f_{t+1}^{\mathrm{T}}\left[(F_t^{\mathrm{T}} F_t)^{-1} - (F_t^{\mathrm{T}} F_t)^{-1} S^{\mathrm{T}}\left[S(F_t^{\mathrm{T}} F_t)^{-1} S^{\mathrm{T}}\right]^{-1} S(F_t^{\mathrm{T}} F_t)^{-1}\right] f_{t+1} = O_p(t^{-1})$$

represents the conditional variance of the prediction $\tilde{\theta}_{t+1}^{(LS,t)}$. Now comparing (9) and (13) we conclude that the gain obtained by using the constrained least squares estimator instead of the unconstrained one, decreases in the rate t^{-1} when t increases. Nevertheless, having left aside how large the gain is, the expressions (9) and (13) imply that there are part of the mean squared error which may be influenced by the selection of our approach while remainders are given by circumstances which are beyond our control. In other words, the mean square error of the prediction consists of variance of process in question and of variance of prediction. The former is given, the later is under our control. So keeping in mind that in both (9) and (13) the terms $\sigma_0^2 f_{t+1}^{\mathrm{T}}(F_t^{\mathrm{T}} F_t)^{-1} f_{t+1}$ and

$$\sigma_0^2 f_{t+1}^{\mathrm{T}}\left[(F_t^{\mathrm{T}} F_t)^{-1} - (F_t^{\mathrm{T}} F_t)^{-} S^{\mathrm{T}}\left[S(F_t^{\mathrm{T}} F_t)^{-} S^{\mathrm{T}}\right]^{-1} S(F_t^{\mathrm{T}} F_t)^{-}\right] f_{t+1}$$

represent the conditional variances of $\hat{\theta}_{t+1}^{(LS,t)}$ and of $\tilde{\theta}_{t+1}^{(LS,t)}$, respectively, it was sufficient to compare the conditional variances of combined forecasts to find the respective gain (or loss). So in what follows, we will compare $\operatorname{var}(\hat{\theta}_{t+1}^{(LTS,t,h)} | F_t, f_{t+1})$ with

$\text{var}(\tilde{\theta}_{t+1}^{(LTS,t,h)}|F_t, f_{t+1})$ rather than their respective mean squared deviations from θ_{t+1}. Denoting

$$\theta_{t+1}^0 = f_{t+1}^{\mathrm{T}}\beta^0$$

and then using Assertion 1 we obtain

$$\begin{aligned}\hat{\theta}_{t+1}^{(LTS,t,h)} &= f_{t+1}^{\mathrm{T}}\hat{\beta}^{(LTS,t,h)} \\ &= \theta_{t+1}^0 + \frac{1}{t}\vartheta^{-1}f_{t+1}^{\mathrm{T}}Q^{-1}\sum_{i=1}^{t} f_i\ \varepsilon_i I\{\varepsilon_i^2 \le u_\alpha^2\} + o_p(t^{-\frac{1}{2}}) \quad \text{as} \quad t \to \infty,\end{aligned}$$

i. e.

$$\begin{aligned}&\sqrt{t}(\hat{\theta}_{t+1}^{(LTS,t,h)} - \theta_{t+1}^0) \\ &= \vartheta^{-1}f_{t+1}^{\mathrm{T}}Q^{-1}\frac{1}{\sqrt{t}}\sum_{i=1}^{t} f_i\ \varepsilon_i I\{\varepsilon_i^2 \le u_\alpha^2\} + o_p(1) \quad \text{as} \quad t \to \infty. \qquad (28)\end{aligned}$$

Unfortunately, this relation does not permit us to obtain either $I\!\!E\left(\hat{\theta}_{t+1}^{(LTS,t,h)}|F_t, f_{t+1}\right)$ or $\text{var}\left(\hat{\theta}_{t+1}^{(LTS,t,h)}|F_t, f_{t+1}\right)$ because of presence of $o_p(t^{-\frac{1}{2}})$ within it. Due to this well-known problem (see e. g. Huber [6]), in such cases we usually consider the asymptotic mean and the asymptotic variance (more precisely, the mean and the variance of the asymptotic distribution of the given statistic). It has even advantage against the precise (mean and) variance of the respective statistic because it depresses the influence of large and rarely appearing values of the statistic in question (these values are "hidden" in the term $o_p(1)$). In other words, it eliminates the influence of values of the statistic which the statistic attains for ω's from the sets of very small probability, or, still in other words, it avoids misleading effect of atypical values of the statistic in question, see Víšek [12] and compare also Huber [6], page 74_4. Taking this into account we will be able to give a generalization of Clemen's result in the following theorem and corollary. Earlier however we shall prepare a lemma.

Lemma 2. Let Assumptions $\mathcal{A}$ be fulfilled. Then the random vectors

$$\xi^{(t)} = t^{-\frac{1}{2}}\vartheta^{-1}Q^{-1}\sum_{i=1}^{t} f_i\ \varepsilon_i I\{\varepsilon_i^2 \le u_\alpha^2\}$$

and

$$\zeta^{(t)} = t^{-\frac{1}{2}}\vartheta^{-1}\left[Q^{-1} - Q^{-1}S^{\mathrm{T}}(SQ^{-1}S^{\mathrm{T}})^{-1}SQ^{-1}\right]\sum_{i=1}^{t} f_i\ \varepsilon_i I\{\varepsilon_i^2 \le u_\alpha^2\}$$

are asymptotically distributed as k-dimensional normal vectors with zero means and covariance matrices given by

$$\vartheta^{-2}\sigma_{\varepsilon_1}Q^{-1} \qquad (29)$$

and

$$\vartheta^{-2}\sigma_{\varepsilon_1} Q^{-1}\left[\mathcal{I} - S^{\mathrm{T}}(SQ^{-1}S^{\mathrm{T}})^{-1}SQ^{-1}\right],$$

respectively.

Proof. We shall use Varadarajan theorem (see Assertion A.1 of Appendix). Let $b \in R^k$ be nonzero otherwise arbitrary vector. We shall verify of course the assumptions of the central limit theorem for $b^{\mathrm{T}}\xi^{(t)}$ and $b^{\mathrm{T}}\zeta^{(t)}$. Denoting for $j, \ell = 1, 2, \ldots, k$ $\tilde{q}_{j\ell} = (Q^{-1})_{j\ell}$, we have

$$\begin{aligned} b^{\mathrm{T}}\xi^{(t)} &= t^{-\frac{1}{2}}\vartheta^{-1}b^{\mathrm{T}}Q^{-1}\sum_{i=1}^{t} f_i\ \varepsilon_i I\{\varepsilon_i^2 < u_\alpha^2\} \\ &= t^{-\frac{1}{2}}\vartheta^{-1}\sum_{i=1}^{t}\varepsilon_i I\{\varepsilon_i^2 < u_\alpha^2\})\sum_{j=1}^{k} b_j \sum_{\ell=1}^{k}\tilde{q}_{j\ell}f_{i\ell}. \end{aligned}$$

Put

$$W_i = t^{-\frac{1}{2}}\vartheta^{-1}\varepsilon_i I\{\varepsilon_i^2 < u_\alpha^2\})\sum_{j=1}^{k} b_j \sum_{\ell=1}^{k}\tilde{q}_{j\ell}f_{i\ell}.$$

For $i = 1, 2, \ldots, t$ and $j, \ell = 1, 2, \ldots, k$ we have $\mathbb{E}\left\{\mathbb{E}\left[\varepsilon_i I\{\varepsilon_i^2 < u_\alpha^2\})f_{i\ell}|f_i\right]\right\} = 0$, so that

$$\mathbb{E}W_i = 0. \tag{30}$$

Similarly

$$\begin{aligned} \mathbb{E}W_i^2 &= t^{-1}\vartheta^{-2}\sigma_{\varepsilon_1}^2\mathbb{E}\left[\sum_{j=1}^{k} b_j \sum_{\ell=1}^{k}\tilde{q}_{j\ell}f_{i\ell}\right]^2 \\ &= t^{-1}\vartheta^{-2}\sigma_{\varepsilon_1}^2\mathbb{E}\left\{\left[\sum_{j=1}^{k} b_j \sum_{\ell=1}^{k}\tilde{q}_{j\ell}f_{i\ell}\right]\left[\sum_{r=1}^{k} b_r \sum_{s=1}^{k}\tilde{q}_{rs}f_{is}\right]\right\} \\ &= t^{-1}\vartheta^{-2}\sigma_{\varepsilon_1}^2 b^{\mathrm{T}}Q^{-1}\mathbb{E}\left[F_t^{\mathrm{T}}F_t\right]Q^{-1}b \\ &= \vartheta^{-2}\sigma_{\varepsilon_1}^2 b^{\mathrm{T}}Q^{-1}b < \infty. \end{aligned} \tag{31}$$

Since the sequence $\{W_i\}_{i=1}^{\infty}$ is the sequence of independent and identically distributed random variables, and b was arbitrarily selected vector from R^k, taking into account (30) and (31) and employing Lindeberg–Lévy and Varadarajan theorems, we conclude the proof of the first assertion of lemma.

The second assertion of lemma can be proved along similar lines. □

Theorem 1. Let Assumptions $\mathcal{A}$ be fulfilled. Then the conditional asymptotic variances

$$\operatorname{var}\left(\sqrt{t}\ (\hat{\theta}_{t+1}^{(LTS,t,h)} - \theta_{t+1}^0)|f_{t+1}\right) \quad \text{and} \quad \operatorname{var}\left(\sqrt{t}\ (\tilde{\theta}_{t+1}^{(LTS,t,h)} - \theta_{t+1}^0)|f_{t+1}\right)$$

are given by

$$\vartheta^{-2}\sigma_{\varepsilon_1} f_{t+1}^{\mathrm{T}} Q^{-1} f_{t+1} \tag{32}$$

and

$$\vartheta^{-2}\sigma_{\varepsilon_1} f_{t+1}^{\mathrm{T}} \left[Q^{-1} - Q^{-1}S^{\mathrm{T}}(SQ^{-1}S^{\mathrm{T}})^{-1}SQ^{-1}\right] f_{t+1}. \tag{33}$$

Proof. Let us fix an arbitrary $\delta \in (0,1)$ and find $K > 0$ so that for any $t \in N$ for

$$B_t = \{\omega \in \Omega : \|f_t(\omega)\| > K\}$$

we have $P(B_t) < \delta$ (due to the fact that all f_i's are identically distributed, it is possible). Moreover, denoting by $D_{\xi^{(t)}}(y)$ the distribution function of $\xi^{(t)}$ and by $D_{\mathcal{N}(0,V)}(y)$ k-dimensional normal distribution with zero mean and covariance matrix V given by (29), due to Lemma 2 we may find $t_0 \in N$ so that for any $t > t_0$

$$\sup_{y \in R^k} |D_{\xi^{(t)}}(y) - D_{\mathcal{N}(0,V)}(y)| < \frac{\delta}{K}.$$

Now for any $t > t_0$ and any $\omega \in B_{t+1}$ we have

$$\sup_{y \in R^k} |D_{f_{t+1}^{\mathrm{T}}\xi^{(t)}}(y) - D_{\mathcal{N}(0,f_{t+1}^{\mathrm{T}}Vf_{t+1})}(y)| < \delta$$

where we have denoted by $D_{f_{t+1}^{\mathrm{T}}\xi^{(t)}}(y)$ and $D_{\mathcal{N}(0,f_{t+1}^{\mathrm{T}}Vf_{t+1})}(y)$ the distribution functions of $f_{t+1}^{\mathrm{T}}\xi^{(t)}$ and of normal random variable with zero mean and variance $f_{t+1}^{\mathrm{T}}Vf_{t+1}$. Since δ was arbitrary, the proof of the first assertion of the theorem follows.

The second assertion can be proved in a similar way. □

Corollary 1. Let the Assumptions $\mathcal{A}$ be fulfilled. Then the approximate confidence interval (on any significance level) for $\tilde{\theta}_{t+1}^{(LTS,t,h)}$ is not wider than that one for $\hat{\theta}_{t+1}^{(LTS,t,h)}$.

Proof follows from the definite positivity of the matrices at (32) and (33).

7. NUMERICAL ILLUSTRATION

As we have promised we shall give now a numerical example demonstrating how the theoretical result works. To offer comparison with the previous results for *LS*- and for *M*-estimators, we shall use the same data as in Víšek [14]. They were originally given in Holden and Peel [5] and they describe the economic growth in United Kingdom since 1977/1 to 1985/2. The abbreviation in the next table means that the forecasts were prepared by the Henley Centre for Forecasting, by the London Business School, by the National Institute of Economic and Social Research, by the Organization for Economic Co-operation and Development and, finally, by Phillips and Drew. The data are presented in the following table.

Table 1. Economic growth in U. K.

Case	Year	HCF	LBS	NI	OECD	PD	Growth
1	1977/1	2.5875	2.650	1.270	1.125	-0.400	1.76899
2	1977/2	3.0375	2.360	3.310	1.000	1.000	3.62319
3	1977/3	3.4500	2.240	3.150	1.875	1.500	3.40205
4	1977/4	3.0750	2.050	2.570	1.500	-0.400	2.76075
5	1978/1	3.1000	3.470	3.460	2.875	-3.000	2.04499
6	1978/2	2.9125	3.340	1.470	2.000	2.200	3.39661
7	1978/3	3.2125	1.660	0.830	2.125	3.000	2.79163
8	1978/4	3.1375	2.820	2.620	1.750	4.500	2.58706
9	1979/1	2.7000	3.160	2.960	1.875	3.500	2.30461
10	1979/2	1.9250	3.100	1.980	1.500	0.900	–2.70532
11	1979/3	0.3375	-0.930	1.100	2.625	-0.400	–3.68575
12	1979/4	–0.1375	-0.100	0.820	1.000	0.800	–5.04364
13	1980/1	–1.9000	-0.980	1.850	–1.625	1.500	–3.91773
14	1980/2	–1.0125	-0.040	0.470	–0.500	–3.700	–2.58193
15	1980/3	–0.6375	-0.200	1.600	2.750	–2.600	-0.50352
16	1980/4	–0.5500	1.980	1.130	–1.000	–5.000	2.04290
17	1981/1	1.4000	2.270	–0.050	–1.000	–5.600	1.63099
18	1981/2	–0.4500	2.480	–0.230	–1.250	–4.500	2.34455
19	1981/3	0.5500	2.560	0.150	–0.250	–2.400	1.31579
20	1981/4	1.4500	2.470	0.530	0.750	–0.500	1.10111
21	1982/1	–1.7500	2.790	0.310	1.000	1.000	3.10932
22	1982/2	1.6375	3.020	1.090	1.750	1.800	2.49004
23	1982/3	1.9375	2.910	0.860	1.750	1.200	4.09591
24	1982/4	2.2875	2.180	1.850	1.625	0.400	4.05940
25	1983/1	1.6250	2.210	1.780	1.500	1.300	3.11285
26	1983/2	2.1375	2.120	1.250	1.625	2.400	2.62390
27	1983/3	2.5125	2.920	1.200	2.375	3.000	2.69714
28	1983/4	2.0875	2.430	1.100	2.250	3.400	2.66413
29	1984/1	2.5000	2.360	1.980	2.250	1.700	3.30189
30	1984/2	2.2500	4.050	3.050	1.750	3.900	4.92424
31	1984/3	2.1000	2.220	3.740	2.750	2.710	3.45794
32	1984/4	2.3500	2.180	2.950	2.000	2.980	2.78035
33	1985/1	2.8300	3.400	1.360	3.630	2.810	2.37442
34	1985/2	2.4500	2.600	1.350	2.880	2.740	1.35379

The next table (Table 2) was taken from Víšek [14] to give the reader a possibility to compare directly the predictions prepared by means of the least squares and by means of the least trimmed squares. Both tables (Table 2 and 3) gather the successive sums of squared differences $(\hat{\theta}_{t+1}^{(LTS,t,h)} - \theta_{t+1})^2$ and $(\tilde{\theta}_{t+1}^{(LTS,t,h)} - \theta_{t+1})^2$ for the period since 1982/2 to 1985/2, i. e. for the same period as in Víšek [14]. (We have started from 1982/2 and not from 1982/1 because the combined forecast prepared by means of $\hat{\beta}^{(LS,t)}$ had very large error just when predicting on 1982/1.) As it is indicated at the head of tables we have considered all possible models, i. e. models with or without intercept, with or without constraints and the sums

$$\sum_{t=22}^{\ell} \left(\hat{\theta}_{t+1}^{(LTS,t,h)} - \theta_{t+1}\right)^2 \quad \text{and} \quad \sum_{t=22}^{\ell} \left(\tilde{\theta}_{t+1}^{(LTS,t,h)} - \theta_{t+1}\right)^2 \tag{34}$$

for $\ell = 22, 23, \ldots, 34$ were collected in the tables.

Table 2. Cumulative losses of forecasts – Least squares.

Period	Forecasted value	Cumulative losses			
		With intercept		Without intercept	
		Without constraints	With constraints	Without constraints	With constraints
1982/2	2.490	0.073	0.310	0.000	0.457
1982/3	4.096	3.527	1.304	2.076	1.269
1982/4	4.059	9.739	5.012	7.250	4.697
1983/1	3.113	12.356	6.236	9.406	5.825
1983/2	2.624	13.328	6.236	9.515	5.825
1983/3	2.697	13.370	7.012	9.748	6.620
1983/4	2.664	13.691	7.077	9.763	6.704
1984/1	3.302	15.099	7.741	11.121	7.282
1984/2	4.924	15.708	9.312	14.081	8.752
1984/3	3.458	18.479	12.556	19.705	12.553
1984/4	2.780	18.997	12.710	20.253	12.978
1985/1	2.374	21.240	15.418	21.803	14.653
1985/2	1.354	22.444	17.910	23.574	16.709

Table 3. Cumulative losses of forecasts – the Least Trimmed squares.

Period	Forecasted value	Cumulative losses			
		With intercept		Without intercept	
		Without constraints	With constraints	Without constraints	With constraints
1982/2	2.490	0.1811	0.1822	0.7018	0.1257
1982/3	4.096	4.1263	4.8516	3.7062	4.6367
1982/4	4.059	8.5138	7.9669	7.8661	8.1897
1983/1	3.113	9.5638	8.5913	8.5033	9.1743
1983/2	2.624	11.7590	8.9688	8.6232	9.4879
1983/3	2.697	12.1565	9.7919	8.9712	10.2122
1983/4	2.664	14.0899	9.9348	9.0334	10.4225
1984/1	3.302	14.3538	10.0405	9.5536	11.3630
1984/2	4.924	17.0939	12.2915	14.0979	12.3501
1984/3	3.458	18.4300	13.7886	14.4764	12.7249
1984/4	2.780	18.4305	14.0241	14.4825	12.8664
1985/1	2.374	22.4658	22.6739	15.2322	12.9524
1985/2	1.354	23.6412	23.5476	16.0211	13.2033

We may see that the cumulative sums given in the last row of Table 2 and 3 are smaller for models without intercept. It may seem to be in a contradiction with the assertion that it is usually recommended not to delete intercept from the regression model, even in the case when it is indicated (by the corresponding t-statistics and p-value) that the intercept is not significant, see e. g. Víšek [13].

First of all, we have to distinguish between the situations when we look for an explanation of data and when we look for a forecast (in the latter one, the quality of the forecast is typically measured by the mean square error). In situation when we look for a forecast, as after all Clemen's result showed, under assumption that the "true" model does not include the intercept, the results are better for model without intercept.

One can compare the desired results, namely sum of squared errors of the forecasts prepared by means of the ordinary least squares (Table 2) and by means of the least trimmed squares (Table 3). Relatively small values of sums of squared errors of forecasts prepared by means of the least squares (see the last value in the last row of Table 2) indicated that the data are not too much contaminated. Nevertheless the least trimmed squares have succeeded to improve a bit (about 21 %) the final results.

8. CONCLUSIONS

It suffices a brief look into e. g. *Journal of Forecasting* and one cannot longer hesitate that combining the forecasts occupies a considerable part of theoretical research and plays an important role in applications. After all, we have already in the introduction reminded the reason for it. We have also recalled why employment of robust methods, especially with high breakdown point, may be effective in combining the forecasts. Hence it is plausible that the theoretical considerations which had been made in present paper have confirmed that Clemen's result could be generalized also on the least trimmed squares. The short numerical study presented above than gave a very first idea how useful the employment of such method may be. On the other hand, as the rows at the middle of the Tables indicate, the asymptotics will work better for larger data. To give a more complete picture of possibilities of robust procedures in combining the forecasts a large study is to be carry out with more contaminated data. That is why the implementation of method is offered to be sent on request.

APPENDIX

Assertion A.1. Let D_t be the distribution function of k-dimensional vector $(f_{i1}, f_{i2}, \ldots, f_{ik})^T$, $i = 1, 2, \ldots$ and $D_{b,t}$ the distribution function of the linear combination $b_1 f_{i1} + b_2 f_{i2} + \ldots + b_k f_{ik}$. Necessary and sufficient condition for the convergence of the distribution function D_t to a k-dimensional d. f. D is that $D_{b,t}$ converges to a d. f. for any b.

For the proof see Rao [8] (also Varadarajan [11] or Wald and Wolfowitz [22]).

Specification of the assertion for normal distribution (which also shows that respective moment correspond) can be found also in Rao [8] (such assertion is not isolated there, however it is simple consequence of Assertion A.1).

(Received September 10, 1999.)

REFERENCES

[1] J. M. Bates and C. W. J. Granger: The combination of forecasts. Oper. Res. Quarterly *20* (1969), 451–468.

[2] P. J. Bickel: One-step Huber estimates in the linear model. J. Amer. Statist. Assoc. *70* (1975), 428–433.

[3] R. T. Clemen: Linear constraints and efficiency of combined forecasts. J. of Forecasting *6* (1986), 31–38.

[4] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel: Robust Statistics – The Approach Based on Influence Functions. Wiley, New York 1986.

[5] K. Holden and D. A. Peel: Unbiasedness, efficiency and the combination of economic forecasts. J. of Forecasting *8* (1989), 175–188.

[6] P. J. Huber: Robust Statistics. Wiley, New York 1981.

[7] J. Jurečková and P. K. Sen: Regression rank scores scale statistics and studentization in linear models. In: Proceedings of the Fifth Prague Symposium on Asymptotic Statistics, Physica Verlag, Heidelberg 1993, pp. 111–121.

[8] R. C. Rao: Linear Statistical Inference and Its Applications. Wiley, New York 1973.

[9] A. M. Rubio, L. Z. Aguilar and J. Á. Víšek: Combining the forecasts using constrained M-estimators. Bull. Czech Econometric Society *4* (1996), 61–72.

[10] A. M. Rubio and J. Á. Víšek: Estimating the contamination level of data in the framework of linear regression analysis. Qűestiió *21* (1997), 9-36.

[11] V. S. Varadarajan: A useful convergence theorem. Sankhyã *20* (1958), 221–222.

[12] J. Á. Víšek: Stability of regression model estimates with respect to subsamples. Computational Statistics *7* (1992), 183-203.

[13] J. Á. Víšek: Statistická analýza dat. (Statistical Data Analysis – a textbook in Czech.) Publishing House of the Czech Technical University Prague 1997.

[14] J. Á. Víšek: Robust constrained combination of forecasts. Bull. Czech Econometric Society *5* (1998), 8, 53–80.

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

[16] J. Á. Víšek: Robust specification test. In: Proceedings of Prague Stochastics'98 (M. Hušková, P. Lachout and J. Á. Víšek, eds.), Union of Czech Mathematicians and Physicists 1998, pp. 581–586.

[17] J. Á. Víšek: Robust estimation of regression model. Bull. Czech Econometric Society *9* (1999), 57–79.

[18] J. Á. Víšek: The least trimmed squares – random carriers. Bull. Czech Econometric Society *10* (1999), 1–30.

[19] J. Á. Víšek: Robust instrumental variables and specification test. In: PRASTAN 2000, Proceedings of the conference "Mathematical Statistics and Numerical Mathematics and Their Applications", (M. Kalina, J. Kalická, O. Nanásiová and A. Handlovičová, eds.), Comenius University, pp. 133–164.

[20] J. Á. Víšek: Regression with high breakdown point. In: Proceedings of ROBUST 2000, Nečtiny, Union of the Czech Mathematicians and Physicists and The Czech Statistical Society. Submitted.

[21] J. Á. Víšek: A new paradigm of point estimation. In: Proceedings of seminar "Data Processing", TRYLOBITE, Pardubice 2000. Submitted.

[22] A. Wald and J. Wolfowitz: Statistical tests based on permutations of the observations. Ann. Math. Statist. *15* (1944), 358–372.

[23] V. J. Yohai and R. A. Maronna: Asymptotic behaviour of M-estimators for the linear model. Ann. Statist. *7* (1979), 258–268.

Doc. RNDr. Jan Ámos Víšek, CSc., Department of Macroeconomics and Econometrics, Institute of Economic Studies, Faculty of Social Sciences, Charles University, Opletalova 26, 110 01 Praha 1, and

Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Praha 8. Czech Republic.

e-mail: visek@mbox.fsv.cuni.cz