KYBERNETIKA — VOLUME *51* (2015), NUMBER 6, PAGES 973–993

EXPONENTIAL SMOOTHING BASED ON L-ESTIMATION

Přemysl Bejda and Tomáš Cipra

Robust methods similar to exponential smoothing are suggested in this paper. First previous results for exponential smoothing in L_1 are generalized using the regression quantiles, including a generalization to more parameters. Then a method based on the classical sign test is introduced that should deal not only with outliers but also with level shifts, including a detection of change points. Properties of various approaches are investigated by means of a simulation study. A real data example is used as an illustration.

Keywords: change point, exponential smoothing, quantiles, robust methods, sign test

Classification: 62M10, 62M20, 62N02

1. INTRODUCTION

Its simplicity and recursive computing scheme predetermine exponential smoothing to be widely used for time series, namely for smoothing and forecasting. It is an ad hoc procedure, but it is also relevant with respect to ARIMA models, see [1].

In [13] this method is shown to be still effective in practical problems.

However, the exponential smoothing, like many other statistical methods, is very sensitive to outliers. The literature dealing with robust exponential smoothing remains rather lacking. On the other hand, many other methods have been robustified, see, e.g., [11] Chapter 8. The first attempt to fulfill the gap was made in [2] (this method will be described later in our paper). Other authors have tried to apply robust versions of the Kalman filter to the state-space model associated with exponential smoothing (these methods employ M-estimation) – e. g., [3, 4, 14] and [5].

The approach based on the Kalman filter supposes that there can be a change in level in each step even if the change is rather small. On the other hand we suppose that level shifts appear in data only rarely, but the corresponding changes can then be quite significant.

Here we attempt to employ L-estimation and we also test for level shifts.

The general exponential smoothing supposes the model of the form

$$y_t = \boldsymbol{z}_t^\top \boldsymbol{a} + \varepsilon_t, \tag{1}$$

where $\{y_t\}$ is a given time series, $\boldsymbol{a}$ is a vector of parameters, $\boldsymbol{z}_t$ is a vector of fitting functions (both of these vectors are of dimension p), and ε_t is a white noise. The white

DOI: 10.14736/kyb-2015-6-0973

noise is usually supposed to be i.i.d. with normal distribution. Here we will admit that ε_t can be contaminated by distributions with heavy tails, but with a probability density symmetric around the origin.

The classical approach of exponential smoothing operates in the L_2 norm (see, e. g., [7]). Here one looks for adaptive estimates $\hat{\boldsymbol{a}}(t)$ at time t by minimizing

$$\sum_{i=1}^{t} \beta^{t-i}(y_i - \boldsymbol{z}_i^\top \boldsymbol{a})^2, \tag{2}$$

where β is a discount coefficient with a value between zero and one.

In this paper two robust approaches to the general exponential smoothing are considered:

1. the exponential smoothing using regression quantiles and implemented by means of a special algorithm in L_1 norm;
2. the approach combining the idea of the exponential smoothing with the classical sign test, which can also be applied to time series with level shifts.

2. EXPONENTIAL SMOOTHING BASED ON REGRESSION QUANTILES

The objective function to be minimized can, using the methodology of the regression quantiles with a robustifying effect (see [9]), be transformed to the form

$$\sum_{i=1}^{t} \beta^{t-i}\varrho_\alpha(y_i - \boldsymbol{z}_i^\top \boldsymbol{a}), \tag{3}$$

where $\alpha \in (0, 1)$ and

$$\varrho_\alpha(x) = |x|\{\alpha I[x > 0] + (1 - \alpha)I[x < 0]\}, \quad x \in \mathbb{R}.$$

If we minimize (3), denoting the solution as $\boldsymbol{a}^\alpha(t)$ for a given α, then the corresponding smoothed value of $\alpha-$quantile of y_t is $y_t^\alpha \equiv (\boldsymbol{z}_t)^\top \boldsymbol{a}^\alpha(t)$. It generalizes the L_1 approach of [2]. Indeed, in the case of the median, i. e., with $\alpha = 0.5$, instead of (3) we solve the minimization problem

$$\sum_{i=1}^{t} \beta^{t-i}|y_i - \boldsymbol{z}_i^\top \boldsymbol{a}| \tag{4}$$

(see, e. g., [2, 3] and [14]). The approach based on the regression quantiles can follow certain ideas used in the previous works with $\alpha = 0.5$.

First of all, we approximate (3) by

$$\sum_{i=t-T+1}^{t} \beta^{t-i}\varrho_\alpha(y_i - \boldsymbol{z}_i^\top \boldsymbol{a}) \tag{5}$$

where T should be sufficiently large such that the observations $y_{t-T}, y_{t-T-1}, \dots, y_1$ with exponential weights $\beta^T, \beta^{T+1}, \dots, \beta^{t-1}$ can be neglected.

The function (5) can be minimized using a suitable algorithm (see next Section).

2.1. Regression quantiles: Algorithm for generalized L_1 approach

Here an algorithm will be introduced for the approach from the previous Section looked upon as a generalized L_1 approach. First we will describe the case $p = 1$. Then we will deal with the general multi-dimensional case. However, the description of the algorithm for the multi-dimensional case is complex so that we will describe the idea behind it only very briefly.

2.2. Case $p = 1$

For the case of $p = 1$ we employ a simple algorithm. This algorithm has been introduced by [2] for $\alpha = 0.5$, and we will refer to it as C-algorithm.

We minimize

$$\widetilde{a_t^\alpha} = \operatorname*{argmin}_{a \in \mathbb{R}} \left\{ \sum_{i=t-T+1}^{t} \beta^{t-i} \varrho_\alpha (y_i - z_i a) \right\},$$

where z_i are known scalars. Without a loss of generality we can assume that $z_1, \ldots, z_t \neq 0$. In particular, for $z_i \equiv 1$ it includes the case of the classical constant trend in time series. For general z_i the algorithm has the following form:

1. Order the ratios $\frac{y_{t-T+1}}{z_{t-T+1}}, \ldots, \frac{y_t}{z_t}$ from the smallest to the largest one. Denote the ordered values by $v_{(1)} \leq v_{(2)} \leq \cdots \leq v_{(T)}$.

2. If $z_j > 0$, put $c_i^- \equiv \alpha \beta^{t-j} |z_j|$ and $c_i^+ \equiv (1-\alpha)\beta^{t-j}|z_j|$. If $z_j < 0$, put $c_i^- \equiv (1-\alpha)\beta^{t-j}|z_j|$ and $c_i^+ \equiv \alpha\beta^{t-j}|z_j|$. The index i is chosen in such a way that it corresponds to the order of the member with z_j in the sequence $v_{(1)}, \ldots, v_{(T)}$.

3. Find the index r ($r = \{1, \ldots, T\}$) which fulfills

$$\begin{aligned} \sum_{j=1}^{r-1} c_j^+ - \sum_{j=r}^{T} c_j^- &< 0 \\ \sum_{j=1}^{r} c_j^+ - \sum_{j=r+1}^{T} c_j^- &\geq 0 \end{aligned} \tag{6}$$

   and put

$$\widetilde{a_t^\alpha} = v_{(r)}.$$

In this way we construct the quantile estimates.

The procedure can be performed recursively if one uses the ordering from the previous step and just adds the next observation. For this ordering we can also employ the heapsort or another sorting algorithm with a low computational complexity. If z_i is equal to 1 for all $i = 1, \ldots, t$, one obtains a robust version of simple exponential smoothing.

We will now show a result which relates to the robustness of the C-algorithm. The following definition can be found in [8].

Definition 2.1. Let $\boldsymbol{x}^0 = (x_1, \ldots, x_n)$. Consider the following value of an estimator $H_n(\boldsymbol{x}^0)$ of a functional H. From this initial sample $\boldsymbol{x}^0$ we replace m observations by any values; this new sample will be denoted by $\boldsymbol{x}^m$, and the pertaining value of the estimator by $H_n(\boldsymbol{x}^m)$.

Then the estimator H_n has the breakdown point

$$\varepsilon_n^*(H_n, \boldsymbol{x}^0) = \frac{m^*(\boldsymbol{x}^0)}{n},$$

where $m^*(\boldsymbol{x}^0)$ is the largest value of m for which

$$\sup_{\boldsymbol{x}^m} \|H_n(\boldsymbol{x}^m) - H_n(\boldsymbol{x}^0)\| < \infty.$$

Lemma 2.2. For each window of the length T the breakdown point is given by the ratio $\frac{j}{T}$ where j is given by

$$\operatorname*{argmax}_{j=1,\ldots,T} \left(\frac{1-\beta^{j-1}}{1-\beta^T} < \frac{1}{2} \right). \tag{7}$$

P r o o f . The observations closer to the current time have higher weights, therefore they influence the result of our algorithm the most. Let us send j most current observations into infinity (denote these numbers as outliers) such that the result of our procedure also converges to infinity. In our case, this sending of observations into infinity does not cause any loss of generality.

Now we are looking for the minimal number j. The sum of all weights is $S = \frac{1-\beta^T}{1-\beta}$. If the weights of the first j observations sum to more than one half of S, then the equations (6) have to be satisfied for a certain r in $j, \ldots, T$. Since j of the most current observations have the sum of weights $J = \frac{1-\beta^j}{1-\beta}$, it must hold that $\frac{J}{S} > \frac{1}{2}$ when we send our estimate to infinity. □

2.3. General case

The idea of the general case with $p > 1$ is the same as that of the previous algorithm. However, it is much more complicated since the principle of ordering is more complex here.

Let $\boldsymbol{z}_t \in \mathbb{R}^p$ be known vectors. We look for a suitable algorithm minimizing the objective function (5). The equation $y_i - \boldsymbol{z}_i^\top \boldsymbol{a} = 0$ represents a hyperplane in $\mathbb{R}^p$ for $\boldsymbol{a}$. It is straightforward that we have to look for the minimum of (5) in the intersection of such hyperplanes, i. e., at a point which is given by p different nonparallel hyperplanes, where p is the dimension of vectors $\boldsymbol{z}_t$. This point lies on lines which are also given by an intersection of appropriate hyperplanes. Another fact which must be employed is that, at the point minimizing our objective function, all signs of directional left-hand derivatives differ from signs of directional right-hand derivatives (we mean the left-hand derivative of an appropriate line). Such a minimization problem can be solved by applying a special table, where any row represents the line given by the intersection of the appropriate number of hyperplanes ($p-1$ if they are not parallel). In each column there must be a point which is given by the intersection of the line and another hyperplane. This table is updated in each step. After constructing this table we can employ a similar algorithm as in a one-dimensional case to find the minimum.

The problem of this approach is its high complexity. We will skip its detailed description.

3. SIGN TEST

An alternative approach to the robustification of the exponential smoothing described in this Section can provide even better results. It combines some ideas also employed for exponential smoothing with the classical sign test, see, e.g., [12] and [16]. It seems to be applicable also to data with level shifts, including detection of change points. Since the observations are not exponentially weighted, we cannot say that the method from this Section belongs among exponential smoothing methods; in fact, it is a recursive adaptive method.

3.1. Sign test: Constant trend

Let us assume the simplest version of the model (1)

$$y_t = a + \varepsilon_t, \tag{8}$$

where level shifts can occur.

A rough idea for a recursive estimate $\widehat{a}(t)$ of the level a can be described as follows:

1. Find the median of a segment of observations from the beginning.
2. If too many consequent observations lie under or above this median then there could be a level shift. The detected change point for the level shift occurs at the point where this pattern begins.
3. Estimate a by median up to this point and start this procedure again from the identified level shift.

The idea of a sign test can be exploited in the corresponding recursive algorithm. If we construct the median for a segment of observations then an observation will lie above or below this median with probability $\frac{1}{2}$ (for simplicity we neglect the possibility that the observation is exactly identical with the median). Let us apply an indicator I_i, where 1 means that the ith observation is above the median and 0 means that it is under the median. For a fixed i an arbitrary sum $S_i^k = \sum_{j=i+1}^{i+k} I_j$ for $k > 0$ has the binomial distribution. If it is too large or too low, one can claim a conjecture that there is a change point because too many observations lie above or below the median.

Similar to a sign test algorithm, we define the statistics

$$A_i^k = \frac{2S_i^k - k}{\sqrt{k}}. \tag{9}$$

We want to identify the potential level shifts in the series of statistics A_i^k (indexed with k). Therefore we employ a symmetrical interval $(-b, b)$, where b is determined empirically, and if A_i^k lies outside the interval $(-b, b)$ then we will indicate a level shift at time i and start from that point i as from the starting point of the new segment.

We will employ the following notation. The length of the time series $\{y_t\}$ is n. The median value found at time t since the beginning of the corresponding segment is M_t. Let t_j denote the time of the jth change point. The smoothed values $\hat{y}_t$ between two neighboring change points will be chosen equal; they are given by the last estimated value M_{t_j-1} before the next change point. The number of observations for the initial computation of the median in each segment is a fixed T. The symbol $\hat{y}_t$ denotes the estimate of a at time t and simultaneously the estimate of y_t due to the model (8).

Remark 3.1. One can look for more exact distributions. E. g., [10] has suggested statistics based on ordinary L_1-residuals.

One can also replace A_i^k by other statistics. E. g., the arcsine transformation should better approximate the normal distribution

$$2\sqrt{k}\left(\arcsin\sqrt{\frac{S_i^k}{k}} - \arcsin\sqrt{\frac{1}{2}}\right)$$

or even

$$\sqrt{4k+2}\left(\arcsin\sqrt{\frac{8S_i^k+3}{8k+6}} - \arcsin\sqrt{\frac{1}{2}}\right).$$

Instead of a sign test we can also employ some other nonparametric tests, e. g., Wilcoxon signed-rank test (see [15]). Then we proceed as follows. The median M_t is computed for all observations since the last change point, i. e., $y_{t_j}, \ldots, y_t$. The test investigates whether the location parameter of observations $y_{t-k}, \ldots, y_t$ is equal to M_t. Put $Y_i = y_i - M_t$. Order $|Y_{t-k}|, \ldots, |Y_t|$ and let R_i be the order of $|Y_i|$. Put $S^+ = \sum_{Y_i \geq 0} R_i$ and $S^- = \sum_{Y_i < 0} R_i$. If $\min(S^+, S^-)$ is too small within the framework of the Wilcoxon test, we have found a change point. This test has to be performed for all $k = 1, \ldots, t - t_j$. Obviously, the observations have to be ordered in each test of this type, which is more time-consuming than our procedure.

We will use the statistic A_i^k because of its simplicity and possibility of recursive calculations. If we have A_{i+1}^{k-1}, then $S_{i+1}^{k-1} = \frac{\sqrt{k-1}A_{i+1}^{k-1}+k-1}{2}$. Then we add $S_i^k = S_{i+1}^{k-1} + I_{i+1}$. It simplifies the computation of A_i^k so that the computational complexity of the statistics in each step is at worst n.

Now we summarize our algorithm in particular steps:

1. Put $t = 1, j = 1, t_1 = 1$.

2. Order the observations $y_{t_j}, \ldots, y_{t_j+T}$ and denote this ordered vector by K. Find the median for K and denote it as M_{t_j+T}. For each $i = t_j, \ldots, t_j + T$ compute statistics A_i^k such that $i + k = T$. If each A_i^k lies in the chosen interval $(-b, b)$, then put $t = t_j + T$ and go to Step 3. Otherwise put $t = i$ for the first i for which A_i^k lies outside the interval, put $M_{i-1} = M_{t_j+T}$ and go to Step 4.

3. $t = t + 1$. Add the observation y_t to the vector K and reorder it. Find the median M_t in this vector. Increase k by 1. If $y_t > M_t$, increase the previous S_i^k by 1. Recompute A_i^k. If each A_i^k lies in the interval $(-b, b)$, repeat this step. Otherwise put $t = i$ for the first i for which A_i^k lies outside the interval and go to Step 4.

4. Put $j = j + 1$, $M_j^1 = M_{t-1}$ and $t_j = t$.

5. From the observations $y_{t_j}, \ldots, y_{t_j+T}$, compute median M_j^2.

6. Compute $d_{1,t} = |M_j^1 - y_t|$ and $d_{2,t} = |M_j^2 - y_t|$.

7. If $d_{1,t} < d_{2,t}$ and $t - t_j \le \frac{T}{2}$, then put $t = t + 1$. Add the observation y_t to vector K and reorder it. Return to Step 6. Otherwise go to Step 8.

8. Put $t_j = t$ and estimates $\hat{y}_{t_{j-1}}, \ldots, \hat{y}_{t_j-1}$ equal to M_{t-1}. Reset K. Go to Step 2.

Remark 3.2. In this Remark we describe why Steps $4-7$ were implemented into our algorithm. Consider the case with a real change point at time t and the median changing from the value M_1 to M_2. If $y_{t-1} > M_1$ and $M_1 < M_2$ then the change point is found earlier than in reality. The same situation occurs if $y_{t-1} < M_1$ and $M_1 > M_2$. The reason is that our procedure distinguishes only up and down movements of the time series.

If we suppose that the errors are smaller than the difference between M_1 and M_2 (which is quite natural, but it also depends on the number of outliers), then we should measure the distances $d_1 = |M_1 - y_{t-1}|$ and $d_2 = |M_2 - y_{t-1}|$. If $d_1 < d_2$, we conjecture that there is no change point at time $t - 1$.

This modification is useful if the errors of the time series do not exceed the difference between M_j^1 and M_j^2.

3.2. Sign test: General case

We return to the general model (1) and sketch a rough idea of an algorithm which can be employed in such a general case. Then we focus only on a special case of the linear trend.

By the series of **pre-estimates** we mean a series of preliminary estimates of parameters obtained in each time t. I. e., in each time t it includes a vector of p components. These are not the final estimates of parameters but they enable us to construct such final estimates.

Let us suppose that we have robust parameter estimates $\hat{\boldsymbol{a}}_{t-1}$ from the previous step and the series of pre-estimates till time $t - 1$. By means of this information we will find the next member of the series of pre-estimates for time t and for any component of the $\boldsymbol{a}$. Without the loss of generality, suppose that we will first deal with the first component a_1 of the vector $\boldsymbol{a}$. We denote by $\boldsymbol{a}_{2,\ldots,p}$ the remaining components of $\boldsymbol{a}$, and adjust the current observation according to (1). Here we use $\hat{\boldsymbol{a}}_{t-1}$ instead of $\boldsymbol{a}$ and we suppose that $\varepsilon_t = 0$. In other words, we solve the equation

$$y_t = z_{1,t}a_1 + \boldsymbol{z}_{2,\dots,p;t}^\top \hat{\boldsymbol{a}}_{2,\dots,p;t-1} \tag{10}$$

for the unknown variable a_1. The solution of equation (10) is the first component of the vector of pre-estimates at time t. We repeat this procedure for each component and derive the new estimate $\hat{\boldsymbol{a}}_t$ as a median or another kind of robust estimator from the series of pre-estimates. By means of the sign test from the previous Section, we test each component of the series of pre-estimates, or we can test the series y_t itself.

The details of the algorithm will be described for a special case of the linear trend.

3.3. Sign test: Linear trend

Consider the model with linear trend

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t. \tag{11}$$

We have only two parameters; so the pre-estimates form two series. The members of these series of pre-estimates for β_0 and β_1 at times $\tau, \dots, t$ will be denoted by $\beta_0^{\tau,t}$ and $\beta_1^{\tau,t}$, respectively. The estimates of parameters of β_0 and β_1 at time t will be denoted by $\hat{\beta}_{0,t}$ and $\hat{\beta}_{1,t}$, respectively. The estimated time of the jth change point will be τ_j. The estimate of the observation y_t will be $\hat{y}_t$. The estimates of β_0 and β_1 computed after the jth change point by a relevant algorithm (see later) will be denoted by $\hat{\beta}_0^j$ and $\hat{\beta}_1^j$. These estimates are obtained not recursively, like the estimates $\hat{\beta}_{0,t}$ and $\hat{\beta}_{1,t}$, but from the observations after the change point. They serve as the initial estimates for the segment after each change point.

The length of window employed by the algorithm will be denoted by T.

First, we describe the situation after finding a change point at time τ_j:

1. Take T observations $y_{\tau_j}, \dots, y_{\tau_j+T-1}$ and find estimates $\hat{\beta}_0^j$ and $\hat{\beta}_1^j$, by means of a robust algorithm; the latter exploits the observations $y_{\tau_j}, \dots, y_{\tau_j+T-1}$.

   Usually the minimization of the $\sum_{t=\tau_j}^{\tau_j+T-1} |y_t - \hat{\beta}_0^j - \hat{\beta}_1^j t|$ delivers sufficiently robust estimates but we can also employ other robust algorithms like WLS, M-estimation, and others.

2. Put $\beta_0^{t,t} = y_t - \hat{\beta}_1^j t$ and $\beta_1^{t,t} = \frac{y_t - \hat{\beta}_0^j}{t}$ for $t = \tau_j, \dots, \tau_j + T - 1$.

3. Find medians from $\beta_0^{\tau_j,\tau_j+T-1}$ and $\beta_1^{\tau_j,\tau_j+T-1}$ and denote these values by $\hat{\beta}_{0,\tau_j+T-1}$ and $\hat{\beta}_{1,\tau_j+T-1}$. Instead of medians we could use other robust estimates.

4. Compute the statistics (9) of the sign test and check whether the change point is indicated. These statistics will be discussed later.

Now let us consider a case in which a change point is not identified in the previous step. Suppose that the current time is $\tau_j + t$, where $t \geq T$. We know $\hat{\beta}_{0,\tau_j+t-1}$ and $\hat{\beta}_{1,\tau_j+t-1}$. We want to find $\hat{\beta}_{0,\tau_j+t}$ and $\hat{\beta}_{1,\tau_j+t}$.

1. Put $\beta_0^{\tau_j+t,\tau_j+t} = y_t - \hat{\beta}_{1,\tau_j+t-1} \cdot t$ and $\beta_1^{\tau_j+t,\tau_j+t} = \frac{y_t - \hat{\beta}_{0,\tau_j+t-1}}{t}$.
2. After computing the next member of the series of pre-estimates, find medians from $\beta_0^{\tau_j,\tau_j+t}$ and $\beta_1^{\tau_j,\tau_j+t}$, and denote these values by $\hat{\beta}_{0,\tau_j+t}$ and $\hat{\beta}_{1,\tau_j+t}$.
3. Compute the statistics of the sign test and check whether the change point is indicated. These statistics will be discussed later.

Similar to the constant trend, the interval $(-b, b)$ relevant for the test statistics will be found by simulation experiments presented later.

Remark 3.3. A similar idea for estimating the next member of the series of pre-estimates was introduced in [6] but there it is simply $\beta_1^{\tau_j+t,\tau_j+t} = y_{\tau_j+t} - y_{\tau_j+t-1}$. Such an approach has the disadvantage that it is not robust, since an outlier influences two members of the series of pre-estimates. This can be solved by replacing y_{τ_j+t-1} with its robust estimate $\hat{y}_{\tau_j+t-1} = \hat{\beta}_{0,\tau_j+t-1} + \hat{\beta}_{1,\tau_j+t-1} \cdot (t-1)$ (compare [4]). However, it is also inappropriate to put $\hat{\beta}_1^{\tau_j+t,\tau_j+t} = y_t - \hat{\beta}_{0,\tau_j+t-1} - \hat{\beta}_{1,\tau_j+t-1} \cdot (t-1)$, since in this case the residuals are summed up while in our algorithm they are divided by t. Denote $\delta_0 = \hat{\beta}_{0,t-1} - \beta_0$ and $\delta_1 = \hat{\beta}_{1,t-1} - \beta_1$. Suppose further that there is no change point for several observations around the time t so that the parameters are constant. Then

$$y_t - \hat{\beta}_{0,t-1} - \hat{\beta}_{1,t-1}(t-1) = \beta_1 + (\beta_0 - \hat{\beta}_{0,t-1}) + (\beta_1 - \hat{\beta}_{1,t-1})(t-1) + \varepsilon_t$$

$$= \beta_1 + \delta_0 + \delta_1(t-1) + \varepsilon_t$$

in comparison to

$$\frac{y_t - \hat{\beta}_{0,t-1}}{t} = \frac{\delta_0 + \beta_1 t + \varepsilon_t}{t} = \beta_1 + \frac{\delta_0 + \varepsilon_t}{t}.$$

Let us now return to the test statistics. The idea behind them is the same as for the constant trend. On the other hand there are other possible ways for these statistics to be computed. Here we consider two of them:

(a) The statistics for the series of pre-estimates $\beta_0^{\tau_j,\tau_j+t}$ and $\beta_1^{\tau_j,\tau_j+t}$: one uses the median statistics $\hat{\beta}_{0,\tau_j+t}$ and $\hat{\beta}_{1,\tau_j+t}$.

(b) The statistics for the original series $y_{\tau_j}, \ldots, y_{\tau_j+t}$: one uses the estimates $\hat{y}_t = \hat{\beta}_{0,\tau_j+t} + \hat{\beta}_{1,\tau_j+t} \cdot t$.

Remark 3.4. The following results for both statistics can be verified simply:

(a) It is not true that the series of pre-estimates always has to lie only above or below β_0^{old} and β_1^{old}. The pre-estimates $\beta_0^{T+j,T+j}$ can be biased more by an incorrect estimation of β_1 than by an incorrect estimation of β_0.

(b) A counterexample exists for this approach (see, e. g., Remark 3.5), but it can be easily solved. The advantage of approach (b) is also its simplicity.

Remark 3.5. Consider the following situation that can sometimes occur. A change point appears at time τ. But it can happen that the new trend is at first above the old one but after several observations falls below it. Then the change point can be indicated incorrectly. On the other hand, such a problem can be solved by computing a new trend and reviewing to see whether or not the change point could have occurred earlier.

Remark 3.6. Similar to Remark 3.2, some improvements are possible:

1. Put $\tau_j = t$.
2. Estimate $\hat{\beta}_0^j$ and $\hat{\beta}_1^j$ from observations $y_{\tau_1}, \ldots, y_{\tau_1+T-1}$ applying the initial algorithm.
3. Put $\hat{y}_t^1 = \hat{\beta}_{0,\tau_j} + \hat{\beta}_{1,\tau_j} \cdot t$ and $\hat{y}_t^2 = \hat{\beta}_0^j + \hat{\beta}_1^j t$ (see (12) below).
4. Compute $d_{1,t} = |\hat{y}_t^1 - y_t|$ and $d_{2,t} = |\hat{y}_t^2 - y_t|$.
5. If $d_{1,t} < d_{2,t}$ and $t - \tau_j \leq \frac{T}{2}$ then put $t = t + 1$ and return to Step 3.
6. Otherwise put $\tau_j = t$ as the final estimate of the change point.

After finding a change point, we smooth the series between the previous and current change points. First we obtain the initial estimates $\hat{\beta}_0^j$ and $\hat{\beta}_1^j$. Then we continue by searching $\hat{\beta}_{0,t}$ and $\hat{\beta}_{1,t}$ up to the time of the change point. We start this procedure in τ_{j-1} and stop exactly in τ_j. As soon as we have $\hat{\beta}_{0,\tau_j}$ and $\hat{\beta}_{1,\tau_j}$, we approximate all observations between τ_{j-1} and τ_j by

$$\hat{y}_t = \hat{\beta}_{0,\tau_j} + \hat{\beta}_{1,\tau_j} t. \tag{12}$$

It is possible not to repeat the whole algorithm for finding $\hat{\beta}_{0,\tau_j}$ and $\hat{\beta}_{1,\tau_j}$, employing the last estimates $\hat{\beta}_{0,t}$ and $\hat{\beta}_{1,t}$ in (12) instead, but such a simplistic approach gives slightly worse results. It is also possible to keep in memory all the previous $\hat{\beta}_{0,t}$ and $\hat{\beta}_{1,t}$.

The algorithms described in this Section will be referred to as *sign test algorithms*. In particular, we will distinguish the *constant sign test algorithm* and the *linear sign test algorithm*.

4. SIMULATION STUDY

In this Section, simulations for models with constant and linear trends are presented with the aim to find the optimal arrangement of the corresponding procedures and compare different methods.

4.1. Simulation study: Constant trend

We have generated, in Matlab and C, time series of length $n = 100$ with a constant trend $y_t = a + \varepsilon_t$. The errors are i.i.d. $\mathrm{N}(0, 1)$ but they are, with probability p, contaminated by other distributions specified in Tables 1, 2 and 3 (e. g. $\mathrm{N}(0, 100)$ with probabilities $p = 5\%$ or $p = 10\%$). For particular situations we have always generated $N = 1000$ series of the same type.

Let a_t denote the actual value of a at time t and let $\widehat{y}_t$ be the estimate of a_t (i. e., the smoothed value) based on our algorithm. If we also want to stress the series number i, we add the index i to the notation, e. g., $y_{t,i}$.

Moreover, the level shift occurs at time $t = 50$. In particular, for each time series the values a_t for $t = 1, \ldots, 49$ are constant, generated by the uniform distribution on the interval (-10,10) for each trajectory and the same rule holds for $t = 50, \ldots, 100$ (all samples are independent).

There is a criterion MAE (Mean Absolute Error) to be minimized with respect to the technical coefficients b and T

$$\mathrm{MAE} = \frac{1}{Nn} \sum_{i=1}^{N} \sum_{t=1}^{n} |a_{t,i} - \widehat{y}_{t,i}|.$$

The MAE criterion's value will differ for the case of forecasting (see later).

In Table 1 we look for b and T such that they minimize the criterion. If a function or a coefficient is indexed by min, then the value has been obtained by minimizing MAE.

We compare the values of $\mathrm{MAE}_{\min}$ with the values obtained under the condition that the technical coefficients b and T are fixed (then the objective function is not indexed and the values of coefficients are given in the legend of the corresponding Table).

Due to the first column and other computations it turns out that the results do not depend too much on the technical coefficient T. More precisely, for a wide range of values of T, the results are almost the same. Therefore for the sign test algorithm with the constant trend, one can recommend the choice of this value between 35 and 50. But we should also take into account the approximate length between two change points. The boundary b can be recommended equal somewhere around 3 (on the other hand, the results are not too dependent on this value and we can choose any value between 2.5 and 3.5). The values for b are appropriate, especially for smoothing. For forecasting we should recommend lower values (between 1.5 and 2), because we could forecast incorrectly after a level shift. A recommendation for routine application of a sign test algorithm with constant trend follows: b equal to 3 and T higher than 20 or equal to average conjectural length between two change points.

Let us now compare the sign test algorithm with other algorithms. It is also interesting to see the average size of residuals. It can help us to decide whether the estimate by a specific algorithm really smoothes the values of the series y_t in the direction of a_t (in other words, whether the estimate is closer to a_t than the original series y_t). For this purpose we employ the average of absolute errors

$$\frac{1}{Nn} \sum_{i=1}^{N} \sum_{j=1}^{n} |a_{t,i} - y_{t,i}|.$$

Distribution	MAE	$\mathrm{MAE}_{\min}$	$b_{\min}$	$T_{\min}$
$p = 0\%$	0.164	0.157	2.760	49
$p = 5\%$				
$\mathrm{N}(0, 100)$	0.198	0.168	3.159	50
Cauchy	0.173	0.155	2.904	50
$\mathrm{U}(-10, 10)$	0.190	0.173	2.866	50
$p = 10\%$				
$\mathrm{N}(0, 100)$	0.217	0.191	2.935	50
Cauchy	0.181	0.164	3.007	49
$\mathrm{U}(-10, 10)$	0.215	0.187	3.135	45
$p = 40\%$				
$\mathrm{U}(0, 50)$	1.77	1.662	2.232	52

Tab. 1. (constant trend by the sign test algorithm): In the first column there are values of MAE for fixed $b = 2$ and $T = 50$. In the second column there are the minimal values of MAE. There are also minimal values of technical coefficients $b_{\min}$ and $T_{\min}$ in the third and fourth columns.

We will employ the following indices:

(i) average of absolute errors (Err);

(ii) simple exponential smoothing (index Exp);

(iii) C-algorithm (index C);

(iv) sign test algorithm (index S);

(v) M-estimation of simple exponential smoothing (index M).

E. g., MAE_{Exp} means that we substitute the estimates from the exponential smoothing algorithm to the MAE criterion. The same technical coefficients $(\beta_{Exp}, \beta_C, b, T)$ are applied to each kind of outliers so their choice is rough.

The M-estimation method is adopted from [5]. Under assumption that the level shifts appear quite rarely, the method [5] gives moderate results. On the other hand, when we generate the data according to [5] then the robust method of [5] gives better results than the method described here. Obviously it depends on the nature of the data:

1. if we suppose a small change in each step then M-estimation should be preferred.

2. in the case of significant rare jumps one should employ the method from this paper.

Table 2 shows that according to the MAE criterion, the sign test algorithm works best by a significant margin.

The sign test methods construct the smoothed value of the series based on its current, past and future values. However, the exponential smoothing based methods, employed for comparison in our simulation study, use just current and past values of the series. Thus, in the case of smoothing, the comparison is not "fair".

Distribution	Err	MAE_{Exp}	MAE_C	MAE_S	MAE_M
$p = 0\%$	0.797	0.487	0.603	0.164	0.583
$p = 5\%$					
$\text{N}(0, 100)$	1.144	0.752	0.654	0.198	0.600
Cauchy	1.020	0.685	0.613	0.173	0.579
$\text{U}(-10, 10)$	1.007	0.635	0.648	0.190	0.609
$p = 10\%$					
$\text{N}(0, 100)$	1.513	1.017	0.730	0.217	0.637
Cauchy	1.468	1.108	0.632	0.181	0.586
$\text{U}(-10, 10)$	1.217	0.771	0.708	0.215	0.639
$p = 40\%$					
$\text{U}(0, 50)$	10.444	10.006	6.982	1.677	4.655

Tab. 2. (constant trend smoothing: comparison of algorithms): (i) the average of absolute errors of simulated time series; the values of MAE for (ii) the simple exponential smoothing with $\beta_{Exp} = 0.6$; (iii) the C-algorithm with $\beta_C = 0.6$; (iv) the sign test with $b = 2$ and $T = 50$; (v) M-estimation of simple exponential smoothing with $\alpha_M = 0.8$ and $\nu_M = 0.5$.

Let us turn our attention to forecasting. If we want to forecast, then the crucial thing is to detect the level shift faster than in the case of smoothing, because after level shift it may happen that we predict too many observations in a wrong way (according to old observations). We also employ the parameter b equal to 2 and not higher in the case of smoothing, to keep comparable conditions for all algorithms.

In the case of forecasting, we employ the following function

$$\text{MAE}_f = \frac{1}{N(n-10)} \sum_{i=1}^{N} \sum_{t=10}^{n-1} |a_{t+1,i} - \widehat{y}_{t+1,i|t}|,$$

where $\widehat{y}_{t+1,i|t}$ stands for an estimate of $a_{t+1,i}$, if we know the observations $y_1, \ldots, y_t$.

From Table 3 we see that our algorithm is still the best, but it is much closer to the others. The values of MAE_f are generally worse. This is given by the fact that for prediction it is harder to switch to another level after a level shift, and the prediction is always delayed.

In Figure 1 we present a time series with large residuals and also a level shift. In this picture we can visually compare how the algorithms are able to deal with these obstacles. For instance the M algorithm deals quite well with high residuals but is unable to react fast enough to a quite large level shift. The C-algorithm and the exponential smoothing algorithm behave in a similar way, nevertheless the exponential smoothing is more sensitive to violations in data.

Distribution	Err	$\text{MAE}_{f,Exp}$	$\text{MAE}_{f,C}$	$\text{MAE}_{f,S}$	$\text{MAE}_{f,M}$
$p = 0\%$	0.797	0.564	0.675	0.560	0.672
$p = 5\%$					
N(0, 100)	1.150	0.835	0.724	0.575	0.699
Cauchy	1.136	0.894	0.681	0.552	0.673
U(−10, 10)	1.010	0.710	0.716	0.575	0.701
$p = 10\%$					
N(0, 100)	1.514	1.094	0.794	0.596	0.738
Cauchy	1.268	0.958	0.699	0.562	0.686
U(−10, 10)	1.226	0.848	0.771	0.581	0.723
$p = 40\%$					
U(0, 50)	10.531	10.198	7.046	2.961	4.799

Tab. 3. (constant trend forecasting: comparison of algorithms): (i) the average of absolute errors of simulated time series; the values of MAE_f for (ii) the simple exponential smoothing with $\beta_{Exp} = 0.6$; (iii) the C-algorithm with $\beta_C = 0.6$; (iv) the sign test with $b = 2$ and $T = 50$; (v) M-estimation of simple exponential smoothing with $\alpha_M = 0.8$ and $\nu_M = 0.5$.

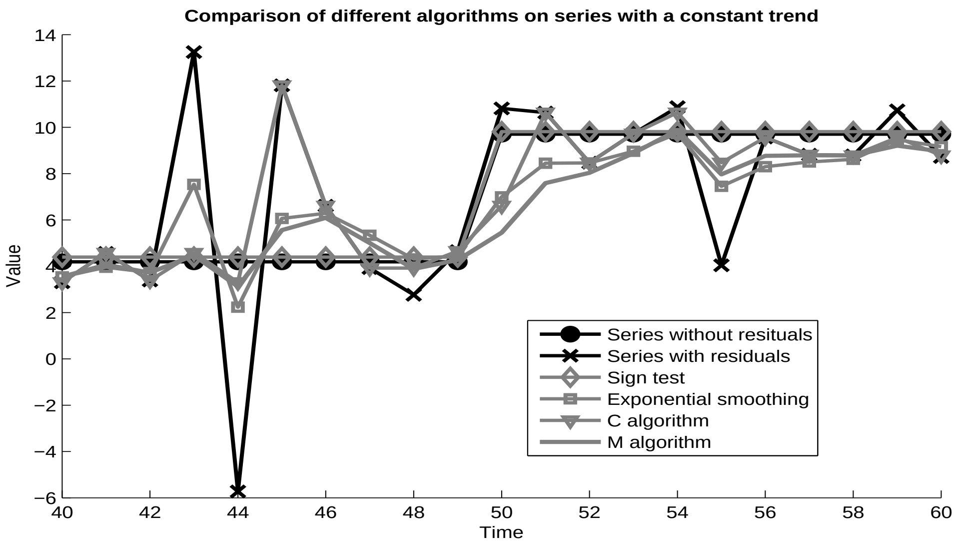

Fig. 1. (constant trend): Comparison of different estimates of time series ($\alpha = 0.5, \beta_C = 0.6, \beta_{Exp} = 0.6, T = 50, b = 2, \alpha_M = 0.8$ and $\nu_M = 0.5$).

4.2. Simulation study: Linear trend

In the linear case the situation differs from the case of a constant trend, since we also have to deal with a slope. It is possible to generate the series by many methods. We choose only one here. The notation is the same as above (e. g., $\mathrm{MAE} = \frac{1}{Nn}\sum_{i=1}^{N}\sum_{t=1}^{n}|\beta_{0,i} + \beta_{1,i}t - \widehat{y}_{t,i}|$).

Similar to the previous case, we generate the time series of length $n = 100$, and also the $N = 100$ series for each distribution. The number N differs from the case of constant trend, because the computation in the linear case can already be quite time-consuming if $N = 1000$, especially when we want to minimize and look for the most suitable technical coefficients.

The constant term is initialized by the uniform distribution on the interval $\langle -10; 10\rangle$ and the linear term similarly on the interval $\langle -5; 5\rangle$. The constant and slope coefficients are changed at time $t = 50$ for each time series. The new linear term also follows the uniform distribution on the interval $\langle -5; 5\rangle$. To avoid too large a gap between observations 49 and 50, we put the constant term β_0 for the new line such that $a_{50} = a_{49} + u$, where u has the uniform distribution on the interval $\langle -10; 10\rangle$. The residuals are constructed in the same way as for a constant trend. Now we have to deal with coefficients T and b. We will employ the same methods as in the case of a constant trend.

We employ the version of the sign test algorithm which follows Remark (b) (see paragraph 3.3), because the algorithm following Remark (a) does not work properly in some cases.

Let us look now at Table 4, where we display the optimal technical coefficients. For the linear trend by the sign test, algorithm one can recommend choice of the window in a length equal to the probable length between two neighboring level shifts. The coefficient b should be chosen between 2.2 and 2.5 for the improved version of algorithm (see Remark 3.6). Once more the choice of technical coefficients is not too important, because we get very similar results for quite a wide range of values (see Table 4).

We can compare the sign test algorithm with the double exponential smoothing. The coefficient was chosen to be optimal according to MAE and with respect to [1], where its author recommends an interval for values of the coefficient for the double exponential smoothing method. In the contaminated cases our algorithms give significantly better results.

In Table 5 the authors also employs M-estimation of double exponential smoothing and compares it with other algorithms. The detailed description of this method can be found in [5]. The results for M-estimation are better for a higher contamination in (relative) comparison to other methods (similarly as for the constant trend): this approach should be used in the case of small jumps occuring quite often since the M-estimation algorithm has been suggested originally just for such a type of data. Otherwise we should employ our algorithm.

Similar to the case of a constant trend we will study the ability of algorithms to forecast. From Table 6 we see that our algorithm is the best in almost all cases except for the non-contaminated case, but we have to employ the improved algorithm. In the case of forecasting it is more suitable to use lower b around 1.5. We get even better results then.

Distribution	MAE	$\text{MAE}_{\min}$	$b_{\min}$	$T_{\min}$
$p = 0\%$	0.231	0.219	2.293	53
$p = 5\%$				
$\text{N}(0, 100)$	0.289	0.250	2.520	52
Cauchy	0.271	0.221	2.286	53
$\text{U}(-10, 10)$	0.244	0.226	2.222	52
$p = 10\%$				
$\text{N}(0, 100)$	0.299	0.291	2.218	50
Cauchy	0.343	0.212	2.437	51
$\text{U}(-10, 10)$	0.265	0.241	2.220	53
$p = 40\%$				
$\text{U}(0, 50)$	2.305	2.113	2.01	49

Tab. 4. (linear trend by the sign test algorithm): This Table employs the improved version of the algorithm according to Remark 3.6. In the first column there are values of MAE for fixed $b = 2$ and $T = 50$. In the second column there are the minimal values of MAE. There are also minimal values of technical coefficients $b_{\min}$ and $T_{\min}$ in the third and fourth columns.

5. EXAMPLE: GDP OF CHINA

The example deals with the annual values of Chinese GDP in the period 1952 - 2014 from National Bureau of Statistics of China (see Table 7). Since the GDP of China has a characteristic exponential growth, we apply a logarithmic transformation to the data. We employ the improved algorithm for a linear trend. We choose the technical coefficients $b = 2.2$ and $T = 10$ since we suppose that the change points can appear in periods of approximately ten years.

The points detected by the algorithm as change points (namely 1961, 1982, 1994 and 2002) should correspond to the significant economic changes (see also Figure 2). In 1958, Mao Tse-tung announced the Great Leap Forward. Our model indicates this event in 1961. In this year the constant term of our model decreases and the slope does not change so much. It is interesting that we get absolutely identical results for a wide range of values b. Until 1978, our model shows a low but stable growth. After 1978, when the crucial reforms of the Chinese economy began, the growth of GDP speeds up. However, these reforms were realized only in several economic zones along the coast. A more rapid growth started in the early 1980s when the reforms were introduced in further areas. In 1990s the growth was still quite fast, but it was accompanied by a high rate of inflation. In the period 2003-2006 other reforms (e. g., the protection of private property) were approved. In 2006 the 11th Five-Year Economic Program was accepted which aimed at education, medical care, etc. These changes of the Chinese economy are visible in GDP and the algorithm reflects them too.

Distribution	Err	MAE_{Exp}	MAE_{SL_b}	MAE_{SLT_b}	MAE_M
$p = 0\%$	0.803	0.729	0.556	0.280	2.999
$p = 5\%$					
$\text{N}(0, 100)$	1.183	1.091	0.577	0.286	2.360
Cauchy	1.052	0.985	0.530	0.232	3.025
$\text{U}(-10, 10)$	1.000	0.899	0.544	0.249	3.498
$p = 10\%$					
$\text{N}(0, 100)$	1.514	1.442	0.664	0.304	2.921
Cauchy	1.200	1.133	0.522	0.244	2.957
$\text{U}(-10, 10)$	1.196	1.081	0.609	0.307	3.611
$p = 40\%$					
$\text{U}(0, 50)$	10.488	10.448	5.791	2.237	11.435

Tab. 5. (linear trend smoothing: comparison of algorithms): In the first column one calculates the errors i. e., $Err = \frac{1}{Nn}\sum_{i=1}^{N}\sum_{j=1}^{n}|a_{t,i} - y_{t,i}|$. In the further columns the values of MAE for different algorithms are calculated. In the second column we present the double exponential smoothing with $\beta = 0.84$ (Exp). Next column contains the values for linear sign test algorithms without the improvement, connected with looking for change points from Remark 3.6. The coefficients for the unimproved version are $T = 10$, $b = 2$. The next column contains the values for the linear sign test algorithm ($T = 50$, $b = 2$), now considering the improvement. The last column contains values of double exponential smoothing employing M-estimation with parameters $\alpha_M = 0.8$ and $\nu_M = 0.7$.

6. CONCLUSIONS

Let us summarize the results and compare the algorithms as to complexity and robustness.

6.1. Conclusion: Complexity

First consider the constant sign test algorithm. We have n steps of our algorithm and in each step we have to put one observation in a proper place to find a median. We have to do n operations and calculate and check the statistics. There are in total $3n$ operations in each step. The complexity of our algorithm is then $O(n^2)$. In reality it depends on b (also on the number of real level shifts in the series); when b is small then we have to order only a few observations and so the complexity is close to n.

We proceed similarly if our model has the dimension p. Only we have to deal with p series in the same way as in the one-dimensional case so that the complexity is $O(pn^2)$. However, because of performing regression in L_1 norm and often recalculating estimates, the real time of computations is much higher.

The C-algorithm also has the complexity $O(n^2)$. In real time it is slightly faster than

Distribution	Err	$\mathrm{MAE}_{f,Exp}$	MAE_{f,SL_b}	MAE_{f,SLT_b}	$\mathrm{MAE}_{f,M}$
$p = 0\%$	0.796	0.979	1.589	1.036	3.058
$p = 5\%$					
N(0, 100)	1.133	1.302	1.582	1.042	3.178
Cauchy	0.939	1.155	1.559	1.038	4.119
U(−10, 10)	1.008	1.105	1.522	0.991	3.420
$p = 10\%$					
N(0, 100)	1.484	1.661	1.770	1.170	2.962
Cauchy	2.291	3.108	1.590	1.076	2.700
U(−10, 10)	1.196	1.292	1.632	1.072	3.908
$p = 40\%$					
U(0, 50)	10.576	11.064	6.226	4.884	13.564

Tab. 6. (linear trend forecasting: comparison of algorithms): In the first column one calculates the errors i. e., $Err = \frac{1}{Nn}\sum_{i=1}^{N}\sum_{j=1}^{n}|a_{t,i} - y_{t,i}|$. In the further columns, the values of MAE_f for different algorithms are calculated. In the second column we present the double exponential smoothing with $\beta = 0.84$ (Exp). Next column contains the values for linear sign test algorithms without the improvement connected with looking for change points from Remark 3.6. The coefficients for the unimproved version are $T = 10$, $b = 2$. The next column contains the values for the linear sign test algorithm ($T = 50$, $b = 2$), but now considering the improvement. The last column contains values of double exponential smoothing employing M-estimation with parameters $\alpha_M = 0.8$ and $\nu_M = 0.7$.

the one-dimensional sign test algorithm.

The exponential smoothing has the complexity n. In the general case it is $O(pn)$. It is also much simpler to implement and the idea behind it is easier to understand than for algorithms which were introduced in this paper. One can summarize that the idea behind the sign test algorithm is not too complicated and the speed of computation is satisfactory in comparison with other robust algorithms.

6.2. Conclusion: Robustness

The sign test algorithm is suggested to deal with both outliers and level shifts. The simulation study shows that the first objective is met fully and the second partly. In the case of many small level shifts, i. e., the case studied for instance in [5], the method from the present paper should be preferred. After applying the notes from Remark 3.2, the results seem to be better but they are still sensitive to the frequency of level shifts.

According to the simulation study the suggested algorithms give better results even in the case without outliers.

Year	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964
ln(GDP)	11.13	11.32	11.36	11.42	11.54	11.58	11.78	11.88	11.89	11.71	11.65	11.73	11.89
Estimate	11.13	11.23	11.33	11.44	11.54	11.65	11.75	11.85	11.96	11.66	11.73	11.80	11.87
Year	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977
ln(GDP)	12.05	12.14	12.09	12.06	12.18	12.33	12.40	12.44	12.52	12.54	12.62	12.60	12.68
Estimate	11.94	12.00	12.07	12.14	12.20	12.27	12.34	12.41	12.47	12.54	12.61	12.67	12.74
Year	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
ln(GDP)	12.81	12.92	13.03	13.10	13.19	13.30	13.49	13.71	13.85	14.01	14.23	14.35	14.45
Estimate	12.81	12.88	12.94	13.01	13.18	13.35	13.51	13.67	13.84	14.00	14.16	14.33	14.49
Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
ln(GDP)	14.60	14.81	15.08	15.39	15.63	15.78	15.89	15.95	16.01	16.12	16.22	16.31	16.43
Estimate	14.65	14.82	14.98	15.53	15.63	15.72	15.82	15.92	16.02	16.12	16.22	16.27	16.42
Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014		
ln(GDP)	16.59	16.74	16.90	17.10	17.27	17.36	17.53	17.70	17.79	17.89	17.97		
Estimate	16.58	16.73	16.89	17.04	17.20	17.36	17.51	17.67	17.82	17.98	18.13		

Tab. 7. Natural logarithm of Chinese GDP in 100 billions of Chinese yuan estimated by the improved linear sign test algorithm ($b = 2.2$, $T = 10$).

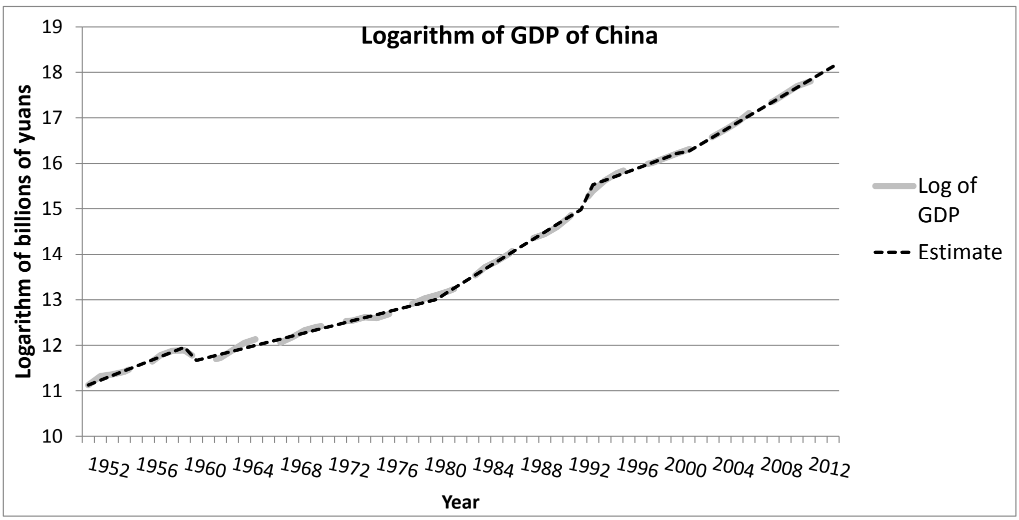

Fig. 2. Natural logarithm of Chinese GDP in billions of Chinese yuan estimated by the improved linear sign test algorithm ($b = 2.2$, $T = 10$).

6.3. Conclusions: Possible generalizations

The algorithms introduced in this paper are not easy to be generalized to ARMA processes. The sign test algorithms can be generalized, similarly as the C-algorithm, to look for quantiles. Another possible generalization includes seasonality. We can also employ estimator of location and other suitable tests, not only median and sign tests.

7. ACKNOWLEDGEMENT

This paper was written with the support of the Czech Science Foundation Project 402/12/G097 Dynamic Models in Economics.

The authors thank the anonymous reviewer for valuable comments and suggestions improving the original text.

(Received November 9, 2014)

REFERENCES

[1] R. G. Brown: Smoothing, Forecasting and Prediction of Descrete Time Series. Prentice-Hall, Englewood Cliffs, NJ 1962.

[2] T. Cipra: Robust exponential smoothing. J. Forecasting *11* (1992), 1, 57–69. DOI:10.1002/for.3980110106

[3] T. Cipra and R. Romera: Recursive time series methods in L_1-norm. In: L_1-Statistical Analysis and Related Methods (Y. Dodge ed.), pp. 233–243, 1992. DOI:10.2307/2289888

[4] S. Gelper, R. Fried, and Ch. Croux: Robust forecasting with exponential and Holt–Winters smoothing. J. Forecasting *29* (2010), 3, 285–300. DOI:10.2139/ssrn.1089403

[5] T. Hanzák and T. Cipra: Exponential smoothing for time series with outliers. Kybernetika *47* (2011), 2, 165–178.

[6] Ch. C. Holt: Forecasting seasonals and trends by exponentially weighted moving averages. Int. J. Forecasting *20* (2004), 5–10. DOI:10.1016/j.ijforecast.2003.09.015

[7] R. J. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder: Forecasting with Exponential Smoothing. Springer, Berlin – Heidelberg 2008. DOI:10.1007/978-3-540-71918-2

[8] J. Jurečková and J. Picek: Robust Statistical Methods with R. Chapman and Hall/CRC, 2005. DOI:10.1201/9781420035131

[9] R. Koenker and G. Bassett: Regression quantiles. Econometrica *46* (1978), 1, 33–50. DOI:10.2307/1913643

[10] A. Koubková: Forecasting seasonals and trends by exponentially weighted moving averages. In: COMPSTAT 2004 Proceedings (J. Antoch, ed.), Springer Verlag 2004, pp. 1345–1352. DOI:10.1007/978-3-7908-2656-2

[11] R. A. Maronna, D. R. Martin, and V. J. Yohai: Robust Statistics: Theory and Methods. John Wiley and Sons Ltd, Chichester 2006. DOI:10.1002/0470010940

[12] G. H. Moore and W. A. Wallis: Time series significance tests based on signs of differences. J. Amer. Statist. Assoc. *38* (1943), 222, 153–164. DOI:10.1080/01621459.1943.10501791

[13] M. Papageorgiou, A. Kotsialos, and A. Poulimenos: Long-term sales forecasting using holt-winters and neural network methods. J. Forecasting *25* (2005), 5, 353–368. DOI:10.1002/for.943

[14] R. Romera and T. Cipra: On practical implementation of robust Kalman filtering. Commun. Statist. – Simulation and Computation *24* (1995), 2, 461–488. DOI:10.1080/03610919508813252

[15] F. Wilcoxon: Individual comparisons by ranking methods. Biometrics Bull. *1* (1945), 6, 80–83. DOI:10.2307/3001968

[16] J. Wolfowitz: Asymptotic distribution of runs up and down. Ann. Math. Statist. *15* (1944), 2, 163–172. DOI:10.1214/aoms/1177731281

Přemysl Bejda, Department of Statistics, Mathematical — Physical Faculty, Charles University in Prague, Sokolovská 83, 186 75 Praha 8. Czech Republic.
e-mail: premyslbejda@gmail.com

Tomáš Cipra, Department of Statistics, Mathematical — Physical Faculty, Charles University in Prague, Sokolovská 83, 186 75 Praha 8. Czech Republic.
e-mail: cipra@karlin.mff.cuni.cz