Robust Akaike Information Criterion for ARMA models

Claudio Agostinelli
Dipartimento di Statistica
Università Cà Foscari di Venezia
30121 Venezia, Italy
claudio@unive.it

Abstract. A robust version of the Akaike Information Criterion (AIC) [5] is defined to the aim of selecting the order of an ARMA model. It is an extended version of the classical criterion based on weighted likelihood methodology [12]. To achieve robustness a weight is associated to each component of the conditional log-likelihood [3]. This criterion is asymptotically equivalent to the classical one when no outliers are present; its robustness are studied in presence of additive and innovative outliers [7] with symmetric and asymmetric contamination by Monte Carlo simulations.

Keywords. AIC, additive and innovative outliers, ARMA, model selection, robustness, weighted likelihood.

J.E.L. classification: C12, C15, C22, C49.

1 Introduction

We consider the problem of robust selection of the order \((p,q)\) in a stationary and invertible ARMA\((p,q)\) model

\[
\Phi(B)(Z_t - \theta_0) = \Theta(B)\alpha_t
\]

when a small fraction of the observations are generated by a different Data Generation Process (DGP). These observations are called Outliers. We assume that the errors \(\{\alpha_t\}\) corresponding to the remaining observations are distributed as a white noise process with density NID\((0,\sigma^2)\). The parameter \(\theta_0\) is the mean of the process and \(B\) is the backward shift operator such that \(B^kZ_t = Z_{t-k}\). The polynomials \(\Phi(B) = (1 - \phi_1B - \cdots - \phi_pB^p)\) and \(\Theta(B) = (1 - \theta_1B - \cdots - \theta_qB^q)\) are assumed causal and without common factors; further, let \(\{\beta,\sigma^2\} = \{\phi_1, \phi_2, \cdots, \phi_p, \theta_1, \theta_2, \cdots, \theta_q, \theta_0\}, \sigma^2\) be the vector of the unknown parameters.

Outliers are observations that depart from the model of the bulk of the data. This kind of observations can lead to a significant impact on the effectiveness

Hereafter, we will introduce a robust selection criterion for the order of an ARMA model based on weighted likelihood and the Akaike Information Criterion. The procedure could be easily used to selecting a sub-set autoregression too.

A similar approach is presented in [1] for the linear regression model.

In Section 2 we briefly review the weighted likelihood methodology for ARMA models and the outliers types considered, in Section 3 we introduce the Weighted Akaike Information Criterion; Section 4 presents some examples while Section 5 presents the results of a Monte Carlo simulation.

2 Weighted Likelihood for ARMA models

In this section we review the robust estimation procedure of the coefficients of an ARMA model \((p \text{ and } q \text{ fixed and known})\) introduced in [3]. This method based on weighted likelihood can deal with two kinds of outliers as introduced in [7]: Additive outliers (AO) which are values with no effect on the subsequent observations in the time series and Innovative outliers (IO) which modify the subsequent observations according to the specified ARMA model. A possible representation is as follows

\[
y_t = \sum_{j=1}^{n} h_j v_j(B) \xi_t(t^*_j) + z_t
\]

where \(y_t\) is the observed contaminated time series, \(z_t\) is the (unobservable) non-contaminated time series which follows an ARMA structure, \(h_j\) is the intensity of the outlier, \(n\) is the number of outliers and \(\xi_t(t^*_j)\) is an indicator function which indicates the position of the outliers. The function \(v_j(B)\) takes the following form according to the type of the outlier

\[
v_j(B) = \begin{cases} 
1 & \text{for an additive outlier} \\
\Phi(B)/\Theta(B) & \text{for an innovative outlier}
\end{cases}
\]

At a first look it may appear that the more dangerous type of outliers would be the IO since they change the values of several observations, on the other hand, it is easy to check that the (observed) residuals \(r_t\) have the following form with respect to the uncontaminated \(a_t\) errors

\[
r_t = \frac{\Phi(B)}{\Theta(B)} (y_t - \theta_0) = a_t + \sum_{j=1}^{n} h_j \tilde{v}_j(B) \xi_t(t^*_j)
\]

where

\[
\tilde{v}_j(B) = \begin{cases} 
\Phi(B)/\Theta(B) & \text{for an additive outlier} \\
1 & \text{for an innovative outlier}
\end{cases}
\]
hence, while an IO affects only one $r_t$, an AO affects all subsequent residuals according to $\hat{e}_j(B)$ structure. Since most estimation methods are based on residuals, rather than the series, we may need special attention on the treatment of AO.

To make hard the construction of a robust estimation procedure, an AO (IO) could be always represented as a (possible infinite) sequence of IO (AO) see [3] for details. This new sequence generates the same observed time series $y_t$ and so we can not identify the structure of the outliers by the data. To avoid this problem, we seek for the smallest fraction of outliers that well describes the actual contamination pattern. The robust estimation procedure is based on (conditional) weighted likelihood and an algorithm to classify the outliers as AO or IO. Let $r_t(\beta) = r_t(\beta; z_*, a_*, z_T) = \theta_1 r_{t-1}(\beta) + \cdots + \theta_q r_{t-q}(\beta) + (z_t - \theta_0) - \phi_1(z_{t-1} - \theta_0) - \cdots - \phi_p(z_{t-p} - \theta_0)$ the residuals where $z_*$ and $a_*$ are the vectors of initial values, let $S(\beta) = \sum_{t=1}^T r_t^2(\beta)$ the square sum of the residuals then, under the normal assumption, we have the following expression for the (conditional) log–likelihood

$$l \left( r_t(\beta); \sigma^2 \right) = \log L \left( r_t(\beta); \sigma^2 \right) = -\frac{T}{2} \log 2\pi \sigma^2 - \frac{S(\beta)}{2\sigma^2}$$

and we define the score function as $u \left( r_t(\beta); \sigma^2 \right) = \frac{\partial}{\partial \beta} \log L \left( r_t(\beta); \sigma^2 \right)$. The weighted likelihood estimating equations has the form

$$\sum_{t=1}^T w \left( r_t(\beta); \sigma^2, \hat{F}(\beta) \right) u \left( r_t(\beta); \sigma^2 \right) = 0 \quad (2)$$

where $w \left( r_t(\beta); \sigma^2, \hat{F}(\beta) \right)$ are weights depending on the unknown parameters $\{\beta, \sigma^2\}$ as defined in the follows.

Let a non-parametric kernel density estimator of the residuals as

$$f^* \left( r_t(\beta), \hat{F}(\beta) \right) = \int k(\beta; r, g) \ d\hat{F}(\beta) \quad \forall t = 1, \cdots, T$$

and the smoothed model

$$m^* \left( r_t(\beta); \sigma^2 \right) = \int k(\beta; r, g) \ dM(r; \sigma^2) \quad \forall t = 1, \cdots, T$$

where $\hat{F}(\beta)$ is the empirical cumulative distribution function based on the residuals $r_t(\beta)$, $M(\sigma^2)$ is normal distribution function with mean 0 and variance $\sigma^2$, $k(\beta; r, g)$ is a kernel density and $g$ is its bandwidth.

By definition of the Pearson residuals and weight function [12] we have

$$\delta_t = \delta \left( r_t(\beta); \sigma^2, \hat{F}(\beta) \right) = \frac{f^* \left( r_t(\beta), \hat{F}(\beta) \right) - m^* \left( r_t(\beta); \sigma^2 \right)}{m^* \left( r_t(\beta); \sigma^2 \right)} \quad \forall t = 1, \cdots, T$$
and

\[ w \left( r_t(\beta); \sigma^2, \hat{F}(\beta) \right) = w(\delta_t) = \min \left\{ 1, \frac{[A(\delta_t) + 1]^+}{\delta_t + 1} \right\} \quad \forall t = 1, \ldots, T \]

where \([\cdot]^+\) indicates the positive part and \(A(\cdot)\) is the Residual Adjustment Function (RAF) of [10] that operates on Pearson residuals as the Huber \(\psi\)-function operates on the structural residuals. When \(A(\delta) = \delta\) the weights are equal to one and this corresponds to maximum likelihood. An example of RAF is \(A(\delta) = 2(\delta + 1)^{1/2} - 1\) namely the Hellinger RAF that we will use in our examples and simulations. For an extensive discussion on the concept of the RAF see [10].

The classification–estimation algorithm works as follows

1. fix an initial value \(\{\beta_0, \sigma^2_0\}\) of the unknown parameters \(\{\beta, \sigma^2\}\);
2. fix a threshold \(0 \leq w_l \leq 1\);
3. let \(\hat{w}(t) = w \left( r_t(\beta_0); \sigma^2_0, \hat{F}(\beta_0) \right)\);
4. let \(\mathcal{O}\) be the set of observations such that \(\hat{w}(t) \leq w_l\);
5. let \(\mathcal{O}(i), i = 1, \ldots, 2^{d_O}\) be all possible subsets of \(\mathcal{O}\);
6. let \(\tilde{y}_t^{(i)}\) the time series associated to the set \(\mathcal{O}(i)\) build recursively as

\[ \tilde{y}_t^{(i)} = \begin{cases} y_t & t \notin \mathcal{O}(i) \\ \tilde{y}_t^{(i)} & t \in \mathcal{O}(i) \end{cases} \]

where

\[ \tilde{y}_t^{(i)} = \beta_0 \left( \tilde{y}_{t-1}^{(i)}, \tilde{y}_{t-2}^{(i)}, \ldots, \tilde{y}_{t-p}, \tilde{r}_{t-1}^{(i)}, \tilde{r}_{t-2}^{(i)}, \ldots, \tilde{r}_{t-q}^{(i)} \right) \]

and

\[ \tilde{r}_t^{(i)} = \tilde{y}_t^{(i)} - \hat{y}_t^{(i)} \]

7. For each residual \(\tilde{r}_t^{(i)}\) with \(t \notin \mathcal{O}(i)\) recalculate the corresponding weight

\[ w_{\mathcal{O}(i)}(t) = w \left( \tilde{r}_t^{(i)}(\beta_0); \sigma^2_0, \hat{F}(\beta_0) \right) \]

and then evaluate:

\[ \tilde{w}(i) = \frac{\sum_{t \notin \mathcal{O}(i)} w_{\mathcal{O}(i)}(t)}{T} \quad i = 1, \ldots, 2^{d_O} \]

which is an estimate of \(1 - \varepsilon\) where \(\varepsilon\) is the level of contamination (i.e. \(n/T\));
8. find the set \(\mathcal{O}(i^*)\) such that \(\tilde{w}(i^*) = \max_i \tilde{w}(i)\);
9. label the observations in \(\mathcal{O}(i^*)\) as Additive Outliers;
10. use \(\tilde{y}_t^{(i^*)}, \tilde{r}_t^{(i^*)}\) as your series and solve the weighted likelihood estimating equations (2);
11. use the new values of \(\{\beta_1, \sigma^2_1\}\) obtain from the previous step;
12. iterate from step 1 to 11 until a convergence criterion is reached.

The convergence criterion we use is a tolerance between the estimated parameters in two different steps and the fact that the observations classified as AO are equal between the two steps.
The observations in the set $O$ of the last step which are not classified as AOs are classified as IOs. Let $\hat{\beta}$ and $\hat{\sigma}^2$ the final values of the algorithm namely the weighted likelihood estimates.

Initial values $\{\beta_0, \sigma_0^2\}$ play an important role in the convergence of the algorithm. Good starting points could be based on resampling techniques as in [12]. Further details on this estimation method are available in [3].

3 Weighted Akaike Information Criterion

The Akaike Information Criterion [5] is defined as follows

$$\text{AIC} = -2 \sum_{t=1}^{T} l(r_t(\hat{\beta}; \hat{\sigma}^2)) + \alpha$$

where $\hat{\beta}$ is the maximum likelihood estimate of $\beta$, $\hat{\sigma}^2$ is an estimate of $\sigma^2$ and $\alpha = 2(p + q)$ (The value of $\alpha$ could assume different form, as for instance the Bayesian Information Criterion. Our method is easily extended to these cases too). There is an interesting relation between the AIC and the log–likelihood ratio test $\lambda$ when the set of hypothesis is

$$\begin{align*}
H_0 & : \text{the order of the model is } (p_0, q_0) \ (\beta_0) \\
H_1 & : \text{the order of the model is } (p_1, q_1) \ (\beta_1)
\end{align*}$$

where $p_1 \geq p_0, q_1 \geq q_0$. In this setting we have

$$\lambda = \alpha_1 - \alpha_0 - \{\text{AIC}(1) - \text{AIC}(0)\}$$

where $\text{AIC}(i) = -2 \sum_{t=1}^{T} l(r_t(\hat{\beta}_i; \hat{\sigma}_i^2)) + \alpha_i$ and $\alpha_i = 2(p_i + q_i), i = 0, 1$

The Weighted Akaike Information Criterion (WAIC) is defined using the weighted log–likelihood ratio test ([4], [2])

$$\lambda_w = -2 \sum_{t=1}^{T} w \left( r_t(\hat{\beta}_1; \hat{\sigma}_1^2, \hat{F}_n(\hat{\beta}_1)) \right) \times \left\{ l(r_t(\hat{\beta}_0; \hat{\sigma}^2) - l(r_t(\hat{\beta}_1; \hat{\sigma}_1^2) \right\}$$

where $\hat{\beta}_0$ e $\hat{\beta}_1$, are the weighted likelihood estimates of $\beta$ under the null hypotheses ($H_0$) and under the alternative ($H_1$) while $\hat{\sigma}^2$ is an estimate of $\sigma^2$, in general we have $\hat{\sigma}^2 = \hat{\sigma}_1^2$. Notice that the weights $w \left( r_t(\hat{\beta}_1; \hat{\sigma}_1^2, \hat{F}_n(\hat{\beta}_1) \right)$ are evaluated under the alternative hypothesis. This ensure the consistency of the test when no contamination is involved and very good robust properties under contamination [2] for details.
From the definition of $\lambda_w$ we have

$$\text{WAIC}(i) = -2 \sum_{t=1}^{T} w \left( r_t(w); \sigma_1^2, \hat{F}_n(w) \right) \times l(r_t(w); \sigma^2) + \alpha_i$$

for $i = 0, 1$. It is possible to use an estimate $\hat{\beta}_0$, instead of $w \beta_0$, evaluate as a solution of a weighted likelihood estimating equations where the weights are fixed to $w \left( r_t(w); \sigma_1^2, \hat{F}_n(w) \right)$; that is a solution of the equations

$$\sum_{t=1}^{T} w \left( r_t(w); \sigma_1^2, \hat{F}_n(w) \right) u \left( r_t(w); \sigma_0^2 \right) = 0$$

The use of $\hat{\beta}_0$ instead of $w \beta_0$ has two main advantages: i) it speeds up the calculations, ii) it stabilize the procedure. This is the procedure we recommend and that we use in the examples and simulations.

4 Examples

In this section we introduce the analysis of two time series, one simulated in order to illustrate the stability of the method with respect to AIC in presence of additive outliers and one real dataset which exibites the presence of one innovative outlier.

A time series of length 300 from an $AR(2)$ with $\phi_1 = 0.9$, $\phi_2 = -0.4$ and $\sigma = 0.5$ is simulated (see figure 1). We run the AIC and WAIC criteria and both identified as a DGP an $AR(4)$. We start to add additive outliers one by one and rerun for each series the two criteria. The outliers are generated with probability 0.5 from a $N(\mu = \max, \sigma = 0.5)$ otherwise from a $N(\mu = \min, \sigma = 0.5)$ where max and min indicate the maximum and minimum value of the actual series (the dataset are available from the author upon request). The last series is presented in figure 2 where the outliers are highlighted by circles. The percentage of contamination goes from 0 to 10%. The maximum order we allow is $p_1 = 6$ and $q_1 = 0$. The results are reported in table 1, where the position of the new introduced outlier at each step is reported in the second line and in the remaining rows the order selected by the two methods are reported. Both methods decrease the selected order as the contamination level increase. After 1% the AIC select order 3 and after 3% it always select order 1 until around 8% where a sample mean is suggested as the best model. The WAIC is stable around order 4 until 6% (remember that both methods suggest this order for the series free of contamination) and then it suggest order 2 for almost all the remaining series.
Fig. 1. Time series of length 300 from the $AR(2)$.

Fig. 2. Time series of length 300 from an $AR(2)$ of figure 1 contaminated with AOs (in red and circled the outliers).
The next example is based on a real dataset of 100 observations (see figure 3) made at a Chicago Steel Mill. on the Rockwell hardness (measured on rockwell "B" scale) of 100 coils produced in sequence [9].

The autocorrelation function and the partial autocorrelation function (figure 4) together with the p–values of the Box–Pierce test [6] for different lags (figure 5, circles) indicate the absence of linear dependence structure or perhaps an autoregressive of order 1 where the corresponding p–value is 0.050. The first spike in the series is at position 8. The weighted likelihood estimator is stable over the autoregressive orders to indicate this observation as an innovative outlier (we check it until order 10). After substituting this observation with the mean of the remain cases the behavior of the Box–Pierce test is quite different and strongly indicating the absence of dependence structure in this dataset, in particular the p–value at lag 1 is equal to 0.133. We run the AIC and our method where we allow $p_1 = 4$ and $q_1 = 0$ as maximum orders.

In table 2 we report the results for this dataset. The introduced method is insensible to the presence of the innovative outlier and it chooses a model with only the intercept, while the AIC suggests an autoregressive model of order 1.

## 5 Monte Carlo Simulation

In this Monte Carlo simulation the reference model is $z_t = 1.2z_{t-1} - 0.47z_{t-2} + 0.06z_{t-3} + a_t$ where $a_t$ is distributed as a N$(0, 1)$. We did two different experiments. In the first experiment we generated 500 series of length 100. The contaminated series were generated from these (uncontaminated) series by adding a fraction $(\varepsilon = 5\%, 10\%, 15\%)$ of additive outliers following: i) a N$(0, 25)$ for the symmetric contamination distribution and ii) a N$(8, 1)$ for the asymmetric case.

In the second experiment the series were contaminated by innovative outliers with the same structure as before.
Fig. 3. Rockwell hardness of 100 coils produced in sequence. Observation in position 8 is circled.

<table>
<thead>
<tr>
<th>AIC</th>
<th>WAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>order intercept value</td>
<td>order intercept value</td>
</tr>
<tr>
<td>1 1 569.0      0      1 535.1</td>
<td></td>
</tr>
<tr>
<td>2 1 570.7      4      1 538.9</td>
<td></td>
</tr>
<tr>
<td>0 1 570.9      3      0 856.5</td>
<td></td>
</tr>
<tr>
<td>3 1 571.9      4      0 858.3</td>
<td></td>
</tr>
<tr>
<td>4 1 573.3      3      1 858.4</td>
<td></td>
</tr>
<tr>
<td>3 0 606.5      2      0 859.8</td>
<td></td>
</tr>
<tr>
<td>4 0 608.0      2      1 861.8</td>
<td></td>
</tr>
<tr>
<td>2 0 610.1      1      0 870.2</td>
<td></td>
</tr>
<tr>
<td>1 0 623.2      1      1 872.2</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. AIC and WAIC for the Rockwell hardness of 100 coils produced in sequence dataset.
Fig. 4. Rockwell hardness of 100 coils produced in sequence. Autocorrelation and Partial autocorrelation functions.
Fig. 5. Rockwell hardness of 100 coils produced in sequence. Box–Pierce test for different lags. In circle the p–values for the original dataset and in triangle the p–values of the modified series.

The maximum order allowed is $p_1 = 6$ and $q_1 = 0$. For all the series we run the two methods (AIC and WAIC) and we report the results in table 3 and 4 for the additive case and in table 5 and 6 for the innovative case. In each row we report the fraction of the series with the corresponding selected order. In the last two columns we report the mean and standard deviation (sd) of the value of the two criteria (AIC and WAIC) over the series. From the uncontaminated case we conclude that the performance of the WAIC is very comparable with that of the AIC in term of efficiency. In particular, since the small value of the coefficient of order 3 more than 65% of the series are classified as order 2 and a small fraction (around 16 – 19%) as order 3 (true model). Hence order 2 and 3 are both reasonable answer in the contamination case.

In presence of additive outliers the performance of AIC decrease rapidly as the contamination level increase: the mode of the distribution is on order 1 and a moderate fraction at order 2 for both symmetric and asymmetric contamination. The introduced methods (WAIC) performs well for 5% of contamination and reasonable until 10%, its results are somewhat better to AIC for higher contamination.

In presence of innovative outliers the AIC suggests order 2 with increasing frequency as the contamination level increase, while the WAIC suggests order 3 (true order) as the contamination level increase rather then order 2 specially in the asymmetric contamination case.
Claudio Agostinelli

<table>
<thead>
<tr>
<th></th>
<th>0%</th>
<th>1%</th>
<th>2%</th>
<th>3%</th>
<th>4%</th>
<th>5%</th>
<th>6%</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0% W</strong></td>
<td>0.000</td>
<td>0.000</td>
<td>0.648</td>
<td>0.183</td>
<td>0.067</td>
<td>0.042</td>
<td>0.060</td>
<td>265.4</td>
<td>(14.9)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.000</td>
<td>0.000</td>
<td>0.697</td>
<td>0.161</td>
<td>0.063</td>
<td>0.036</td>
<td>0.043</td>
<td>285.1</td>
</tr>
<tr>
<td><strong>5% W</strong></td>
<td>0.000</td>
<td>0.073</td>
<td>0.573</td>
<td>0.146</td>
<td>0.084</td>
<td>0.062</td>
<td>0.062</td>
<td>271.8</td>
<td>(20.1)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.002</td>
<td>0.616</td>
<td>0.198</td>
<td>0.084</td>
<td>0.038</td>
<td>0.024</td>
<td>0.038</td>
<td>388.3</td>
</tr>
<tr>
<td><strong>10% W</strong></td>
<td>0.000</td>
<td>0.235</td>
<td>0.369</td>
<td>0.138</td>
<td>0.096</td>
<td>0.068</td>
<td>0.103</td>
<td>289.9</td>
<td>(31.4)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.004</td>
<td>0.556</td>
<td>0.268</td>
<td>0.058</td>
<td>0.040</td>
<td>0.032</td>
<td>0.042</td>
<td>433.2</td>
</tr>
<tr>
<td><strong>15% W</strong></td>
<td>0.000</td>
<td>0.383</td>
<td>0.245</td>
<td>0.133</td>
<td>0.054</td>
<td>0.103</td>
<td>0.338</td>
<td>461.2</td>
<td>(28.4)</td>
</tr>
</tbody>
</table>

**Table 3.** Symmetric contamination – Additive outliers.

<table>
<thead>
<tr>
<th></th>
<th>0%</th>
<th>1%</th>
<th>2%</th>
<th>3%</th>
<th>4%</th>
<th>5%</th>
<th>6%</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0% W</strong></td>
<td>0.000</td>
<td>0.000</td>
<td>0.689</td>
<td>0.145</td>
<td>0.073</td>
<td>0.059</td>
<td>0.028</td>
<td>278.6</td>
<td>(15.5)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.006</td>
<td>0.672</td>
<td>0.164</td>
<td>0.070</td>
<td>0.054</td>
<td>0.034</td>
<td>286.6</td>
<td>(14.9)</td>
</tr>
<tr>
<td><strong>5% W</strong></td>
<td>0.000</td>
<td>0.030</td>
<td>0.560</td>
<td>0.175</td>
<td>0.091</td>
<td>0.055</td>
<td>0.089</td>
<td>257.0</td>
<td>(16.4)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.000</td>
<td>0.540</td>
<td>0.286</td>
<td>0.058</td>
<td>0.058</td>
<td>0.032</td>
<td>0.032</td>
<td>452.9</td>
</tr>
<tr>
<td><strong>10% W</strong></td>
<td>0.000</td>
<td>0.120</td>
<td>0.400</td>
<td>0.149</td>
<td>0.146</td>
<td>0.094</td>
<td>0.091</td>
<td>290.7</td>
<td>(61.6)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.052</td>
<td>0.478</td>
<td>0.288</td>
<td>0.096</td>
<td>0.032</td>
<td>0.024</td>
<td>0.030</td>
<td>500.1</td>
</tr>
<tr>
<td><strong>15% W</strong></td>
<td>0.000</td>
<td>0.209</td>
<td>0.283</td>
<td>0.206</td>
<td>0.108</td>
<td>0.094</td>
<td>0.101</td>
<td>407.6</td>
<td>(88.7)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.126</td>
<td>0.434</td>
<td>0.232</td>
<td>0.110</td>
<td>0.052</td>
<td>0.032</td>
<td>0.014</td>
<td>526.8</td>
</tr>
</tbody>
</table>

**Table 4.** Asymmetric contamination – Additive outliers.

<table>
<thead>
<tr>
<th></th>
<th>0%</th>
<th>1%</th>
<th>2%</th>
<th>3%</th>
<th>4%</th>
<th>5%</th>
<th>6%</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0% W</strong></td>
<td>0.000</td>
<td>0.006</td>
<td>0.685</td>
<td>0.145</td>
<td>0.073</td>
<td>0.059</td>
<td>0.028</td>
<td>278.6</td>
<td>(15.5)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.006</td>
<td>0.672</td>
<td>0.164</td>
<td>0.070</td>
<td>0.054</td>
<td>0.034</td>
<td>286.6</td>
<td>(14.9)</td>
</tr>
<tr>
<td><strong>5% W</strong></td>
<td>0.000</td>
<td>0.030</td>
<td>0.560</td>
<td>0.175</td>
<td>0.091</td>
<td>0.055</td>
<td>0.089</td>
<td>257.0</td>
<td>(16.4)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.000</td>
<td>0.540</td>
<td>0.286</td>
<td>0.058</td>
<td>0.058</td>
<td>0.032</td>
<td>0.032</td>
<td>452.9</td>
</tr>
<tr>
<td><strong>10% W</strong></td>
<td>0.000</td>
<td>0.120</td>
<td>0.400</td>
<td>0.149</td>
<td>0.146</td>
<td>0.094</td>
<td>0.091</td>
<td>290.7</td>
<td>(61.6)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.052</td>
<td>0.478</td>
<td>0.288</td>
<td>0.096</td>
<td>0.032</td>
<td>0.024</td>
<td>0.030</td>
<td>500.1</td>
</tr>
<tr>
<td><strong>15% W</strong></td>
<td>0.000</td>
<td>0.209</td>
<td>0.283</td>
<td>0.206</td>
<td>0.108</td>
<td>0.094</td>
<td>0.101</td>
<td>407.6</td>
<td>(88.7)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.126</td>
<td>0.434</td>
<td>0.232</td>
<td>0.110</td>
<td>0.052</td>
<td>0.032</td>
<td>0.014</td>
<td>526.8</td>
</tr>
</tbody>
</table>

**Table 5.** Symmetric contamination – Innovative outliers.

<table>
<thead>
<tr>
<th></th>
<th>0%</th>
<th>1%</th>
<th>2%</th>
<th>3%</th>
<th>4%</th>
<th>5%</th>
<th>6%</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0% W</strong></td>
<td>0.000</td>
<td>0.006</td>
<td>0.689</td>
<td>0.145</td>
<td>0.073</td>
<td>0.059</td>
<td>0.028</td>
<td>278.6</td>
<td>(15.5)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.006</td>
<td>0.672</td>
<td>0.164</td>
<td>0.070</td>
<td>0.054</td>
<td>0.034</td>
<td>286.6</td>
<td>(14.9)</td>
</tr>
<tr>
<td><strong>5% W</strong></td>
<td>0.000</td>
<td>0.030</td>
<td>0.560</td>
<td>0.175</td>
<td>0.091</td>
<td>0.055</td>
<td>0.089</td>
<td>257.0</td>
<td>(16.4)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.000</td>
<td>0.540</td>
<td>0.286</td>
<td>0.058</td>
<td>0.058</td>
<td>0.032</td>
<td>0.032</td>
<td>452.9</td>
</tr>
<tr>
<td><strong>10% W</strong></td>
<td>0.000</td>
<td>0.120</td>
<td>0.400</td>
<td>0.149</td>
<td>0.146</td>
<td>0.094</td>
<td>0.091</td>
<td>290.7</td>
<td>(61.6)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.052</td>
<td>0.478</td>
<td>0.288</td>
<td>0.096</td>
<td>0.032</td>
<td>0.024</td>
<td>0.030</td>
<td>500.1</td>
</tr>
<tr>
<td><strong>15% W</strong></td>
<td>0.000</td>
<td>0.209</td>
<td>0.283</td>
<td>0.206</td>
<td>0.108</td>
<td>0.094</td>
<td>0.101</td>
<td>407.6</td>
<td>(88.7)</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>0.126</td>
<td>0.434</td>
<td>0.232</td>
<td>0.110</td>
<td>0.052</td>
<td>0.032</td>
<td>0.014</td>
<td>526.8</td>
</tr>
</tbody>
</table>

**Table 6.** Asymmetric contamination – Innovative outliers.
References
